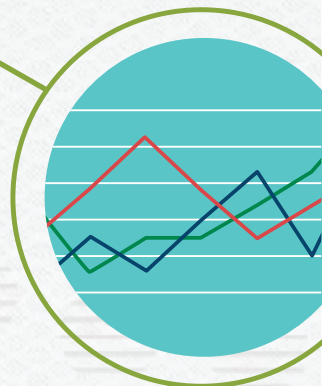




FINANCE AND DEVELOPMENT

BACK TO BASICS

Economic concepts explained





FINANCE AND DEVELOPMENT

BACK TO BASICS

Economic concepts explained

© 2017 International Monetary Fund

Commissioning editor: James L. Rowe
Managing editor: Marina Primorac
Production editor: Lijun Li
Copy editor: Lucy Morales
Cover design: IMF Multimedia Services

Back to **Basics**: Economic concepts explained

v Foreword

Maurice Obstfeld

I. THE BIG PICTURE

2 What Is Capitalism?

Free markets may not be perfect but they are probably the best way to organize an economy

Sarwat Jahan and Ahmed Saber Mahmud

4 What Is Keynesian Economics?

The central tenet of this school of thought is that government intervention can stabilize the economy

Sarwat Jahan, Ahmed Saber Mahmud, and Chris Papageorgiou

6 Micro and Macro: The Economic Divide

Economics is split into two realms: the overall economy and individual markets

G. Chris Rodrigo

8 Economic Models: Simulations of Reality

Economists build simplified descriptions to enhance their understanding of how things work

Sam Ouliaris

10 Econometrics: Making Theory Count

For economic theory to be a useful tool for policymaking, it must be quantifiable

Sam Ouliaris

12 Supply and Demand: Why Markets Tick

Buyers and sellers meet and at the right price all products are sold

Irena Asmundson

14 Gross Domestic Product: An Economy's All

When it is growing, especially if inflation is not a problem, workers and businesses are generally better off than when it is not.

Tim Callen

16 Monetarism: Money Is Where It's At

Its emphasis on money's importance gained sway in the 1970s

Sarwat Jahan and Chris Papageorgiou

II. HOW ECONOMIES FUNCTION

20 What Is Direct Investment?

Investors often seek profits from a long-term stake in a foreign operation

Tadeusz Galeza and James Chan

22 The Output Gap: Veering from Potential

Economists look for the difference between what an economy is producing and what it could produce

Sarwat Jahan and Ahmed Saber Mahmud

24 Structural Policies: Fixing the Fabric of the Economy

Monetary and fiscal policies deal with short-term economic fluctuations, but an economy's problems often go deeper

Khaled Abdel-Kader

26 Money: At the Center of Transactions

Without it, modern economies could not function

Irena Asmundson and Ceyda Oner

28 Price: The Language of Exchange

A price is the amount of money a buyer gives a seller in exchange for a good or a service. But it can be more than that

Irena Asmundson

30 Inflation: Prices on the Rise

Inflation measures how much more expensive a set of goods and services has become over a certain period, usually a year

Ceyda Oner

32 Unemployment: The Curse of Joblessness

The number of people at work is generally closely related to whether an economy is growing at a reasonable rate

Ceyda Oner

34 Recession: When Bad Times Prevail
It is a sustained period when economic output falls and unemployment rises
Stijn Claessens and M. Ayhan Kose

36 Fiscal Policy: Taking and Giving Away
Governments use spending and taxing powers to promote stable and sustainable growth
Mark Horton and Asmaa El-Ganainy

38 Externalities: Prices Do Not Capture All Costs
There are differences between private returns or costs and the costs or returns to society as a whole
Thomas Helbling

40 International Trade: Commerce among Nations
Nations are almost always better off when they buy and sell from one another
Brad McDonald

42 What Are Real Exchange Rates?
What is the value of a country's goods against those of another country, a group of countries, or the rest of the world, at the prevailing exchange rate?
Luis A. V. Catão

44 Purchasing Power Parity: Weights Matter
At what rate would the currency of one country have to be converted into that of another to buy the same goods and services in each country?
Tim Callen

46 Capital Accounts: Liberalize or Not?
There are both benefits and costs to easing restrictions on capital that flows across a country's borders
M. Ayhan Kose and Eswar Prasad

48 Current Account Deficits: Is There a Problem?
There can be consequences when the amount a country spends abroad is wildly different from what it receives from the outside world
Atish Ghosh and Uma Ramakrishnan

III. FINANCE

52 Shadow Banks: Out of the Eyes of Regulators
Many financial institutions that act like banks are not supervised like banks.
Laura E. Kodres

54 LIBOR: World Reference Point
The London interbank rate is used widely as a benchmark but has come under fire
John Kiff

56 Banks: At the Heart of the Matter
Institutions that match up savers and borrowers help ensure that economies function smoothly
Jeanne Gobat

58 What Are Money Markets?
They provide a means for lenders and borrowers to satisfy their short-term financial needs
Randall Dodd

60 Markets: Exchange or Over the Counter
How securities are traded plays a critical role in price determination and stability
Randall Dodd

IV. ECONOMICS IN ACTION

64 Financial Services: Getting the Goods
How consumers and businesses acquire financial products such as loans and insurance
Irena Asmundson

66 Strategic Thinking
Game theory analyzes behavior when decisions must take into account the potential actions of opponents
Sarwat Jahan and Ahmed Saber Mahmud

68 Taxes in Practice
It is hard to design a fair and efficient revenue system
Ruud De Mooij and Michael Keen

70 Taxing Principles
Making the best of a necessary evil
Ruud De Mooij and Michael Keen

72 Inflation Targeting: Holding the Line
Central banks use interest rates to steer price increases toward a publicly announced goal
Sarwat Jahan

74 Regressions: An Economist Obsession
A basic statistical tool for distinguishing between correlation and causality
Rodney Ramcharan

76 What Are Remittances?
For many countries, money transfers from citizens working abroad are a lifeline for development
Dilip Ratha

FOREWORD

PITY THE ECONOMIST. Financial stock prices have recovered from the great financial crisis of 2007–09—and, in some countries, are touching all-time highs. But the stock of economists remains in the tank.

Not only did we fail to see the financial crisis coming—as Queen Elizabeth II memorably observed during a 2008 visit to the London School of Economics—but our economic policies seem to most people not to have restored the global economy quite to its bloom of earlier decades.

Thus, politicians these days scorn “experts” and encourage voters to ignore them. They feel free to disavow even their own economic analysts in favor of convenient alternative views. Economic forecasting is widely derided as useless—or worse. How often have I been asked by a journalist, “Why should we believe anything you say, when you were wrong about ___?”

There are sadly many ways to fill in that blank, and some criticism of economics is well justified. Yet, just as it is important not to overstate what economics can do, it is critical not to understate it. Economics is far from a precise science—who would expect to predict with any accuracy global outcomes that depend on the individual actions of about 5 billion working-age individuals, not to mention the intervention of natural and man-made disasters? At the same time, however, economics provides an essential tool for understanding, and to some degree shaping, those events.

The articles in this collection, all from the International Monetary Fund’s quarterly magazine *Finance & Development* and updated in 2017, illustrate the rich diversity of questions that economics can illuminate. The best economic analysis clarifies thought: it is a mental discipline that helps make sense of complex events, ranging from famines, to bank runs, to housing shortages. It can proceed from the bottom up—focusing on the decisions of individuals and how they hang together at the economy level—or from the top down—reminding us how an economy’s resources, its technology, and its trading opportunities limit what its people and government can consume over time.

As you will see reading these pages, economics is less than a science—which is what gets economists into trouble—but, looked at in another way, it is more. One of the architects of the International Monetary Fund, John Maynard Keynes, called economics “an easy subject at which few excel.... The paradox finds its explanation, perhaps, in that the master-economist must possess a rare combination of gifts. He must be mathematician, historian, statesman, philosopher—in some degree.” It may be the curse of economists that their subject cannot be reduced to a routine technical exercise—like dentistry, to use another example of Keynes’s. But that is also what makes economics so fascinating. **FD**

— *Maurice Obstfeld*
Economic Counsellor and
Director of Research Department
International Monetary Fund

I THE BIG PICTURE

What Is Capitalism?

Free markets may not be perfect but they are probably the best way to organize an economy

Sarwat Jahan and Ahmed Saber Mahmud

CAPITALISM is often thought of as an economic system in which private actors own and control property in accord with their interests, and demand and supply freely set prices in markets in a way that can serve the best interests of society.

The essential feature of capitalism is the motive to make a profit. As Adam Smith, the 18th century philosopher and father of modern economics, said: “It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest.” Both parties to a voluntary exchange transaction have their own interest in the outcome, but neither can obtain what he or she wants without addressing what the other wants. It is this rational self-interest that can lead to economic prosperity.

In a capitalist economy, capital assets—such as factories, mines, and railroads—can be privately owned and controlled, labor is purchased for money wages, capital gains accrue to private owners, and prices allocate capital and labor between competing uses (see “Supply and Demand,” p. 12).

Although some form of capitalism is the basis for nearly all economies today, for much of the past century it was but one of two major approaches to economic organization. In the other, *socialism*, the state owns the means of production, and state-owned enterprises seek to maximize social good rather than profits.

Pillars of capitalism

Capitalism is founded on the following pillars:

- private property, which allows people to own tangible assets such as land and houses and intangible assets such as stocks and bonds;
- self-interest, through which people act in pursuit of their own good, without regard for sociopolitical pressure. Nonetheless, these uncoordinated individuals end up benefiting society as if, in the words of Smith’s 1776 *Wealth of Nations*, they were guided by an invisible hand;
- competition, through firms’ freedom to enter and exit markets, maximizes social welfare, that is, the joint welfare of both producers and consumers;
- a market mechanism that determines prices in a decentralized manner through interactions between buyers and sellers—prices, in return, allocate resources, which naturally

seek the highest reward, not only for goods and services but for wages as well;

- freedom to choose with respect to consumption, production, and investment—dissatisfied customers can buy different products, investors can pursue more lucrative ventures, workers can leave their jobs for better pay; and
- limited role of government, to protect the rights of private citizens and maintain an orderly environment that facilitates proper functioning of markets.

The extent to which these pillars operate distinguishes various forms of capitalism. In free markets, also called *laissez-faire* economies, markets operate with little or no regulation. In *mixed economies*, so called because of the blend of markets and government, markets play a dominant role, but are regulated to a greater extent by government to correct market failures, such as pollution and traffic congestion; promote social welfare; and for other reasons, such as defense and public safety. Mixed capitalist economies predominate today.

The many shades of capitalism

Economists classify capitalism into different groups using various criteria. Capitalism, for example, can be simply sliced into two types, based on how production is organized. In *liberal market economies*, the competitive market is prevalent and the bulk of the production process takes place in a decentralized manner akin to the free-market capitalism seen in the United States and the United Kingdom. *Coordinated market economies*, on the other hand, exchange private information through non-market institutions such as unions and business associations—as in Germany and Japan (Hall and Soskice 2001).

More recently, economists have identified four types of capitalism distinguished according to the role of entrepreneurship (the process of starting businesses) in driving innovation and the institutional setting in which new ideas are put into place to spur economic growth (Baumol, Litan, and Schramm 2007).

In *state-guided capitalism*, the government decides which sectors will grow. Initially motivated by a desire to foster growth, this type of capitalism has several pitfalls: excessive investment, picking the wrong winners, susceptibility to corruption, and difficulty withdrawing support when it is no longer appropriate. *Oligarchic capitalism* is oriented toward protecting and

enriching a very narrow fraction of the population. Economic growth is not a central objective, and countries with this variety have a great deal of inequality and corruption.

Big-firm capitalism takes advantage of economies of scale. This type is important for mass production of products. Entrepreneurial capitalism produces breakthroughs like the automobile, telephone, and computer. These innovations are usually the product of individuals and new firms. However, it takes big firms to mass-produce and market new products, so a mix of big-firm and entrepreneurial capitalism seems best. This is the kind that characterizes the United States more than any other country.

The Keynesian critique

During the Great Depression of the 1930s, the advanced capitalist economies suffered widespread unemployment. In his 1936 *General Theory of Employment, Interest, and Money*, British economist John Maynard Keynes argued that capitalism struggles to recover from slowdowns in investment because a capitalist economy can remain indefinitely in equilibrium with high unemployment and no growth. Keynesian economics challenged the notion that laissez-faire capitalist economies could operate well on their own without state intervention to promote aggregate demand and fight high unemployment and deflation of the sort seen during the 1930s. He postulated that government intervention (by cutting taxes and increasing government spending) was needed to pull the economy out of the recession (see “What Is Keynesian Economics?” p. 4). These actions sought to temper the boom and bust of the business cycle and to help capitalism recover following the Great Depression. Keynes never intended to replace the market-based economy with a different one; he asserted only that periodic government intervention was necessary.

The forces that generally lead to the success of capitalism can also usher in its failure. Free markets can flourish only when governments set the rules that govern them—such as laws that ensure property rights—and support markets with proper infrastructure, such as roads and highways to move goods and people. Governments, however, may be influenced by organized private interests that try to leverage the power of regulations to protect their economic position at the expense of the public interest—for example, by repressing the same free market that bred their success.

Thus, according to Rajan and Zingales (2003), society must “save capitalism from the capitalists”—that is, take appropriate steps to protect the free market from powerful private interests that seek to impede its efficient functioning. When political interest and the capitalist class combine, “crony capitalism” may emerge, and nepotism will be more rewarding than efficiency. The concentration of ownership of productive assets must be limited to ensure competition. And, because competition

begets winners and losers, losers must be compensated. Free trade and strong competitive pressure on incumbent firms will also keep powerful interests at bay. The public needs to see the virtues of free markets and oppose government intervention in the market to protect powerful incumbents at the expense of overall economic prosperity.

The essential feature of capitalism is the motive to make a profit.

Economic growth under capitalism may have far surpassed that of other economic systems, but *inequality* remains one of its most controversial attributes. Do the dynamics of private capital accumulation inevitably lead to the concentration of wealth in fewer hands, or do the balancing forces of growth, competition, and technological progress reduce inequality? Economists have taken various approaches to finding the driver of economic inequality. The most recent study analyzes a unique collection of data going back to the 18th century to uncover key economic and social patterns (Piketty 2014). It finds that in contemporary market economies, the rate of return on investment frequently outstrips overall growth. With compounding, if that discrepancy persists, the wealth held by owners of capital will increase far more rapidly than other kinds of earnings (wages, for example), eventually outstripping them by a wide margin. Although this study has as many critics as admirers, it has added to the debate on wealth distribution in capitalism and reinforced the belief among many that a capitalist economy must be steered in the right direction by government policies and the general public to ensure that Smith’s invisible hand continues to work in society’s favor. **FD**

SARWAT JAHAN is a senior economist in the IMF’s Asia and Pacific Department, and **AHMED SABER MAHMUD** is associate director in the Applied Economics Program at Johns Hopkins University.

References:

- Baumol, William J., Robert E. Litan, and Carl J. Schramm. 2007. *Good Capitalism, Bad Capitalism, and the Economics of Growth and Prosperity*. New Haven, Connecticut: Yale University Press.
- Hall, Peter A., and David Soskice, eds. 2001. *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. New York: Oxford University Press.
- Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, Massachusetts: Belknap Press.
- Rajan, Raghuram, and Luigi Zingales. 2003. *Saving Capitalism from the Capitalists: Unleashing the Power of Financial Markets to Create Wealth and Spread Opportunity*. New York: Crown Publishing Group.

What Is Keynesian Economics?

The central tenet of this school of thought is that government intervention can stabilize the economy

Sarwat Jahan, Ahmed Saber Mahmud, and Chris Papageorgiou

DURING THE GREAT DEPRESSION of the 1930s, existing economic theory was unable either to explain the causes of the severe worldwide economic collapse or to provide an adequate public policy solution to jump-start production and employment.

British economist John Maynard Keynes spearheaded a revolution in economic thinking that overturned the then-prevailing idea that free markets would automatically provide full employment—that is, that everyone who wanted a job would have one as long as workers were flexible in their wage demands (see box). The main plank of Keynes's theory, which has come to bear his name, is the assertion that aggregate demand—measured as the sum of spending by households, businesses, and the government—is the most important driving force in an economy. Keynes further asserted that free markets have no self-balancing mechanisms that lead to full employment. Keynesian economists justify government intervention through public policies that aim to achieve full employment and price stability.

The revolutionary idea

Keynes argued that inadequate overall demand could lead to prolonged periods of high unemployment. An economy's output of goods and services is the sum of four components: consumption, investment, government purchases, and net exports (the difference between what a country sells to and buys from foreign countries). Any increase in demand has to come from one of these four components. But during a recession, strong forces often dampen demand as spending goes down. For example, during economic downturns uncertainty often erodes consumer confidence, causing them to reduce their spending, especially on discretionary purchases like a house or a car. This reduction in spending by consumers can result in less investment spending by businesses, as firms respond to weakened demand for their products. This puts the task of increasing output on the shoulders of the government. According to Keynesian economics, state intervention is necessary to moderate the booms and busts in economic activity, otherwise known as the business cycle.

There are three principal tenets in the Keynesian description of how the economy works:

- Aggregate demand is influenced by many economic decisions—public and private. Private sector decisions can sometimes lead to adverse macroeconomic outcomes, such as reduction in

consumer spending during a recession. These market failures sometimes call for active policies by the government, such as a fiscal stimulus package (explained below). Therefore, Keynesian economics supports a mixed economy guided mainly by the private sector but partly operated by the government.

- Prices, and especially wages, respond slowly to changes in supply and demand, resulting in periodic shortages and surpluses, especially of labor.

- Changes in aggregate demand, whether anticipated or unanticipated, have their greatest short-run effect on real output and employment, not on prices. Keynesians believe that, because prices are somewhat rigid, fluctuations in any component of spending—consumption, investment, or government expenditures—cause output to change. If government spending increases, for example, and all other spending components remain constant, then output will increase. Keynesian models of economic activity also include a multiplier effect; that is, output changes by some multiple of the increase or decrease in spending that caused the change. If the fiscal multiplier is greater than one, then a one dollar increase in government spending would result in an increase in output greater than one dollar.

Stabilizing the economy

No policy prescriptions follow from these three tenets alone. What distinguishes Keynesians from other economists is their belief in activist policies to reduce the amplitude of the business cycle, which they rank among the most important of all economic problems.

Rather than seeing unbalanced government budgets as wrong, Keynes advocated so-called that act against the direction of the business cycle. For example, Keynesian economists would advocate deficit spending on labor-intensive infrastructure projects to stimulate employment and stabilize wages during economic downturns. They would raise taxes to cool the economy and prevent inflation when there is abundant demand-side growth. Monetary policy could also be used to stimulate the economy—for example, by reducing interest rates to encourage investment. The exception occurs during a liquidity trap, when increases in the money stock fail to lower interest rates and, therefore, do not boost output and employment.

Keynes argued that governments should solve problems in the short run rather than wait for market forces to fix things

over the long run, because, as he wrote, “In the long run, we are all dead.” This does not mean that Keynesians advocate adjusting policies every few months to keep the economy at full employment. In fact, they believe that governments cannot know enough to fine-tune successfully.

Keynesianism evolves

Even though his ideas were widely accepted while Keynes was alive, they were also scrutinized and contested by several contemporary thinkers. Particularly noteworthy were his arguments with the Austrian School of Economics, whose adherents believed that recessions and booms are a part of the natural order and that government intervention only worsens the recovery process.

Keynesian economics dominated economic theory and policy after World War II until the 1970s, when many advanced economies suffered both inflation and slow growth, a condition dubbed “stagflation.” Keynesian theory’s popularity waned then because it had no appropriate policy response for stagflation. Monetarist economists doubted the ability of governments to regulate the business cycle with fiscal policy and argued that judicious use of monetary policy (essentially controlling the supply of money to affect interest rates) could alleviate the crisis (see “Monetarism,” p. 16). Members of the monetarist school also maintained that money can have an effect on output in

the short run but believed that in the long run, expansionary monetary policy leads to inflation only. Keynesian economists largely adopted these critiques, adding to the original theory a better integration of the short and the long run and an understanding of the long-run neutrality of money—the idea that a change in the stock of money affects only nominal variables in the economy, such as prices and wages, and has no effect on real variables, like employment and output.

The main plank of Keynes’s theory is that aggregate demand is the most important driving force in an economy.

Both Keynesians and monetarists came under scrutiny with the rise of the new classical school during the mid-1970s. The new classical school asserted that policymakers are ineffective because individual market participants can anticipate the changes from a policy and act in advance to counteract them. A new generation of Keynesians that arose in the 1970s and 1980s argued that even though individuals can anticipate correctly, aggregate markets may not clear instantaneously; therefore, fiscal policy can still be effective in the short run.

The global financial crisis of 2007–08 caused a resurgence in Keynesian thought. It was the theoretical underpinnings of economic policies in response to the crisis by many governments, including in the United States and the United Kingdom. As the global recession was unfurling in late 2008, Harvard professor N. Gregory Mankiw wrote in the *New York Times*, “If you were going to turn to only one economist to understand the problems facing the economy, there is little doubt that the economist would be John Maynard Keynes. Although Keynes died more than a half-century ago, his diagnosis of recessions and depressions remains the foundation of modern macroeconomics. Keynes wrote, ‘Practical men, who believe themselves to be quite exempt from any intellectual influence, are usually the slave of some defunct economist.’ In 2008, no defunct economist is more prominent than Keynes himself.”

But the 2007–08 crisis also showed that Keynesian theory had to better include the role of the financial system. Keynesian economists are rectifying that omission by integrating the real and financial sectors of the economy. [FD](#)

KEYNES THE MASTER

Keynesian economics gets its name, theories, and principles from British economist John Maynard Keynes (1883–1946), who is regarded as the founder of modern macroeconomics. His most famous work, *The General Theory of Employment, Interest and Money*, was published in 1936. But its 1930 precursor, *A Treatise on Money*, is often regarded as more important to economic thought. Until then economics analyzed only static conditions—essentially doing detailed examination of a snapshot of a rapidly moving process. Keynes, in *Treatise*, created a dynamic approach that converted economics into a study of the flow of incomes and expenditures. He opened up new vistas for economic analysis.

In *The Economic Consequences of the Peace* in 1919, Keynes predicted that the crushing conditions the Versailles peace treaty placed on Germany to end World War I would lead to another European war.

He remembered the lessons from Versailles and from the Great Depression when he led the British delegation at the 1944 Bretton Woods conference—which set down rules to ensure the stability of the international financial system and facilitated the rebuilding of nations devastated by World War II. Along with US Treasury official Harry Dexter White, Keynes is considered the intellectual founding father of the International Monetary Fund and the World Bank, which were created at Bretton Woods.

SARWAT JAHAN is a senior economist in the IMF’s Asia and Pacific Department, **AHMED SABER MAHMUD** is the associate director of applied Economics at Johns Hopkins University, and **CHRIS PAPAGEORGIOU** is a deputy division chief in the IMF’s Research Department.

Micro and Macro: The Economic Divide

Economics is split into two realms: the overall economy and individual markets

G. Chris Rodrigo

PHYSICISTS LOOK AT the big world of planets, stars, galaxies, and gravity. But they also study the minute world of atoms and the tiny particles that comprise those atoms.

Economists also look at two realms. There is big-picture **macroeconomics**, which is concerned with how the overall economy works. It studies such things as employment, gross domestic product, and inflation—the stuff of news stories and government policy debates. Little-picture **microeconomics** is concerned with how supply and demand interact in individual markets for goods and services.

In macroeconomics, the subject is typically a nation—how all markets interact to generate big phenomena that economists call *aggregate variables*. In the realm of microeconomics, the object of analysis is a single market—for example, whether price rises in the automobile or oil industries are driven by supply or demand changes. The government is a major object of analysis in macroeconomics—for example, studying the role it plays in contributing to overall economic growth or fighting inflation. Macroeconomics often extends to the international sphere because domestic markets are linked to foreign markets through trade, investment, and capital flows. But microeconomics can have an international component as well. Single markets often are not confined to single countries; the global market for petroleum is an obvious example.

The macro/micro split is institutionalized in economics, from beginning courses in “principles of economics” through to postgraduate studies. Economists commonly consider themselves microeconomists or macroeconomists. The American Economic Association publishes several academic journals including two called *Microeconomics* and *Macroeconomics*.

Why the divide?

It was not always this way. In fact, from the late 18th century until the Great Depression of the 1930s, economics was economics—the study of how human societies organize the production, distribution, and consumption of goods and services. The field began with the observations of the earliest economists, such as Adam Smith, the Scottish philosopher popularly credited with

being the father of economics—although scholars were making economic observations long before Smith authored *The Wealth of Nations* in 1776. Smith’s notion of an invisible hand that guides someone seeking to maximize his or her own well-being to provide the best overall result for society as a whole is one of the most compelling notions in the social sciences. Smith and other early economic thinkers such as David Hume gave birth to the field at the onset of the Industrial Revolution.

Economic theory developed considerably between the appearance of Smith’s *The Wealth of Nations* and the Great Depression, but there was no separation into microeconomics and macroeconomics. Economists implicitly assumed that either markets were in equilibrium—such that prices would adjust to equalize supply and demand—or that in the event of a transient shock, such as a financial crisis or a famine, markets would quickly return to equilibrium. In other words, economists believed that the study of individual markets would adequately explain the behavior of what we now call aggregate variables, such as unemployment and output.

The severe and prolonged global collapse in economic activity that occurred during the Great Depression changed that. It was not that economists were unaware that aggregate variables could be unstable. They studied business cycles—as economies regularly changed from a condition of rising output and employment to reduced or falling growth and rising unemployment, frequently punctuated by severe changes or economic crises. Economists also studied money and its role in the economy. But the economics of the time could not explain the Great Depression. Economists operating within the classical paradigm of markets always being in equilibrium had no plausible explanation for the extreme “market failure” of the 1930s.

If Adam Smith is the father of economics, John Maynard Keynes is the founding father of macroeconomics. Although some of the notions of modern macroeconomics are rooted in the work of scholars such as Irving Fisher and Knut Wicksell in the late 19th and early 20th centuries, macroeconomics as a distinct discipline began with Keynes’s masterpiece, *The General Theory of Employment, Interest and Money*, in 1936. Its main concern is

the instability of aggregate variables. Whereas early economics concentrated on equilibrium in individual markets, Keynes introduced the simultaneous consideration of equilibrium in three interrelated sets of markets—for goods, labor, and finance. He also introduced “disequilibrium economics,” which is the explicit study of departures from general equilibrium. His approach was taken up by other leading economists and developed rapidly into what is now known as macroeconomics.

Coexistence and complementarity

Microeconomics is based on models of consumers or firms (which economists call agents) that make decisions about what to buy, sell, or produce—with the assumption that those decisions result in perfect market clearing (demand equals supply) and other ideal conditions. Macroeconomics, on the other hand, began from observed divergences from what would have been anticipated results under the classical tradition.

Today the two fields coexist and complement each other.

Microeconomics, in its examination of the behavior of individual consumers and firms, is divided into consumer demand theory, production theory (also called the theory of the firm), and related topics such as the nature of market competition, economic welfare, the role of imperfect information in economic outcomes, and at the most abstract, general equilibrium, which deals simultaneously with many markets. Much economic analysis is microeconomic in nature. It concerns such issues as the effects of minimum wages, taxes, price supports, or monopoly on individual markets and is filled with concepts that are recognizable in the real world. It has applications in trade, industrial organization and market structure, labor economics, public finance, and welfare economics. Microeconomic analysis offers insights into such disparate efforts as making business decisions or formulating public policies.

Macroeconomics is more abstruse. It describes relationships among aggregates so big as to be hard to apprehend—such as national income, savings, and the overall price level. The field is conventionally divided into the study of national economic growth in the long run, the analysis of short-run departures from equilibrium, and the formulation of policies to stabilize the national economy—that is, to minimize fluctuations in growth and prices. Those policies can include spending and taxing actions by the government or monetary policy actions by the central bank.

Bridging the micro/macro divide

Like physical scientists, economists develop theory to organize and simplify knowledge about a field and to develop a conceptual framework for adding new knowledge. Science begins with the accretion of informal insights, particularly with observed regular relationships between variables that are so stable they can be codified into “laws.” Theory is developed by pinning down

those invariant relationships through both experimentation and formal logical deductions—called models (see “Economic Models,” p. 8).

Since the Keynesian revolution, the economics profession has had essentially two theoretical systems, one to explain the small picture, the other to explain the big picture (micro and macro are the Greek words, respectively, for “small” and “big”). Following the approach of physics, for the past quarter century or so, a number of economists have made sustained efforts to merge microeconomics and macroeconomics. They have tried to develop microeconomic foundations for macroeconomic models on the grounds that valid economic analysis must begin with the behavior of the elements of microeconomic analysis: individual households and firms that seek to optimize their conditions.

There have also been attempts to use very fast computers to simulate the behavior of economic aggregates by summing the behavior of large numbers of households and firms. It is too early to say anything about the likely outcome of this effort. But within the field of macroeconomics there is continuing progress in improving models, whose deficiencies were exposed by the instabilities that occurred in world markets during the global financial crisis that began in 2008.

How they differ

Contemporary microeconomic theory evolved steadily without fanfare from the earliest theories of how prices were determined. Macroeconomics, on the other hand, is rooted in empirical observations that existing theory could not explain. How to interpret those anomalies has always been controversial. There are no competing schools of thought in microeconomics—which is unified and has a common core among all economists. The same cannot be said of macroeconomics—where there are, and have been, competing schools of thought about how to explain the behavior of economic aggregates. The best-known schools are the New Keynesians and the New Classicals. But these divisions have been narrowing over the past few decades (Blanchard, Dell’Ariccia, and Mauro 2010).

Microeconomics and macroeconomics are not the only distinct subfields in economics. Econometrics, which seeks to apply statistical and mathematical methods to economic analysis, is widely considered the third core area of economics. Without the major advances in econometrics made over the past century or so, much of the sophisticated analysis achieved in microeconomics and macroeconomics would not have been possible. **FD**

G. CHRIS RODRIGO, a writer on economics issues, was previously a visiting scholar in the IMF’s Research Department.

Reference:

Blanchard, Olivier, Giovanni Dell’Ariccia, and Paolo Mauro. 2010. “Rethinking Macroeconomic Policy,” IMF Staff Position Note 10/03. Washington: International Monetary Fund.

Economic Models: Simulations of Reality

Economists build simplified descriptions to enhance their understanding of how things work

Sam Ouliaris

THE MODERN ECONOMY is a complex machine. Its job is to allocate limited resources and distribute the resulting output among a large number of agents—mainly individuals, firms, and governments—allowing for the possibility that each agent’s action can directly (or indirectly) affect other agents’ actions.

Adam Smith labeled the machine the “invisible hand.” In *The Wealth of Nations*, published in 1776, Smith, widely considered the father of economics, emphasized the economy’s self-regulating nature—that agents independently seeking their own gain may produce the best overall result for society as well. Today’s economists build models—road maps of reality, if you will—to enhance our understanding of the invisible hand.

As economies allocate goods and services, they emit measurable signals that suggest there is order driving the complexity. For example, the annual output of advanced economies oscillates around an upward trend. There also seems to be a negative relationship between inflation and the rate of unemployment in the short term. At the other extreme, equity prices seem to be stubbornly unpredictable.

Economists call such empirical regularities “stylized facts.” Given the complexity of the economy, each stylized fact is a pleasant surprise that invites a formal explanation. Learning more about the process that generates these stylized facts should help economists and policymakers understand the inner workings of the economy. They may then be able to use this knowledge to nudge the economy toward a more desired outcome (for example, avoiding a global financial crisis).

Interpreting reality

An economic model is a simplified description of reality, designed to yield hypotheses about economic behavior that can be tested. An important feature of an economic model is that it is necessarily subjective in design because there are no objective measures of economic outcomes. Different economists will make different judgments about what is needed to explain their interpretations of reality.

There are two broad classes of economic models—theoretical and empirical. Theoretical models seek to derive verifiable implications about economic behavior under the assumption that agents maximize specific objectives subject to constraints that are well

defined in the model (for example, an agent’s budget). They provide qualitative answers to specific questions—such as the implications of asymmetric information (when one side to a transaction knows more than the other) or how best to handle market failures.

In contrast, empirical models aim to verify the qualitative predictions of theoretical models and convert these predictions to precise, numerical outcomes. For example, a theoretical model of an agent’s consumption behavior would generally suggest a positive relationship between expenditure and income. The empirical adaptation of the theoretical model would attempt to assign a numerical value to the average amount expenditure increases when income increases.

Economic models generally consist of a set of mathematical equations that describe a theory of economic behavior. The aim of model builders is to include enough equations to provide useful clues about how rational agents behave or how an economy works (see box). The structure of the equations reflects the model builder’s attempt to simplify reality—for example, by assuming an infinite number of competitors and market participants with perfect foresight. Economic models can be quite simple in practice: the demand for apples, for example, is inversely related to price if all other influences remain constant. The less expensive the apples,

A USEFUL MODEL

The standard model of supply and demand taught in introductory economics is a good example of a useful economic model. Its basic purpose is to explain and analyze prices and quantities traded in a competitive market. The model’s equations determine the level of supply and demand as a function of price and other variables (for example, income). The market-clearing price is determined by the requirement that supply equal demand at that price. Demand is usually set to decline and supply to increase with price, yielding a system that moves toward the market-clearing price—that is, equilibrium—without intervention. The supply-demand model can explain changes, for example, in the global equilibrium price of gold. Did the gold price change because demand changed or because of a one-time increase in supply, such as an exceptional sale of central bank gold stockpiles?

the more are demanded. Or models can be rather complex: some models that seek to predict the real level of output of an economy use thousands of complex formulations that go by such names as “nonlinear, interconnected differential equations.”

Economic models can also be classified in terms of the regularities they are designed to explain or the questions they seek to answer. For example, some models explain the economy’s ups and downs around an evolving long-run path, focusing on the demand for goods and services without being too exact about the sources of growth in the long run. Other models are designed to focus on structural issues, such as the impact of trade reforms on long-term production levels, ignoring short-term oscillations. Economists also build models to study “what-if” scenarios, such as the impact on the overall economy of introducing a value-added tax.

How economists build empirical models

Despite their diversity, empirical economic models have features in common. Each will allow for inputs, or exogenous variables, which do not need to be explained by the model. These include policy variables, such as government spending and tax rates, or non-policy variables, like the weather. Then there are the outputs, called dependent variables (for example, the inflation rate), which the model will seek to explain when some or all of the exogenous variables come into play.

Every empirical model will also have coefficients that determine how a dependent variable changes when an input changes (for example, the responsiveness of household consumption to a \$100 decrease in income tax). Such coefficients are usually estimated (assigned numbers) based on historical data. Last, empirical model builders add a catchall variable to each behavioral equation to account for idiosyncrasies of economic behavior at the individual level. (In the example above, agents will not respond identically to a \$100 tax rebate.)

There are, however, fundamental differences among economists regarding how an empirical model’s equations should be derived. Some economists insist that the equations must assume maximizing behavior (for example, an agent chooses its future consumption to maximize its level of satisfaction subject to its budget), efficient markets, and forward-looking behavior. Agents’ expectations and how they react to policy changes play a vital role in the resulting equations. Consequently, users of the model should be able to track the effect of specific policy changes without having to worry about whether the change itself alters agents’ behavior.

Other economists favor a more nuanced approach. Their preferred equations reflect, in part, what their own experience has taught them about observed data. Economists that build models this way are, in essence, questioning the realism of the behavioral constructs in the more formally derived models. Incorporating experience, however, often means it’s impossible to untangle the effect of specific shocks or predict the impact of a policy change because the underlying equations do not

explicitly account for changes in agent behavior. The gain, these same economists would argue, is that they do a better job of prediction (especially for the near term).

What makes a good economic model?

Irrespective of the approach, the scientific method (lots of sciences, such as physics and meteorology, create models) requires that every model yield precise and verifiable implications about the economic phenomena it is trying to explain. Formal evaluation involves testing the model’s key implications and assessing its ability to reproduce stylized facts. Economists use many tools to test their models, including case studies, lab-based experimental studies, and statistics.

Still, the randomness of economic data often gets in the way, so economists must be precise when saying that a model “successfully explains” something. From a forecasting perspective that means errors are unpredictable and irrelevant (zero) on average. When two or more models satisfy this condition, economists generally use the volatility of the forecast errors to break the tie—smaller volatility is generally preferred.

An objective signal that an empirical model needs to be revised is if it produces systematic forecasting errors. Systematic errors imply that one or more equations of the model are incorrect. Understanding why such errors arise is an important part of the regular assessment economists make of models.

Why models fail

All economic models, no matter how complicated, are subjective approximations of reality designed to explain observed phenomena. It follows that the model’s predictions must be tempered by the randomness of the underlying data it seeks to explain and by the validity of the theories used to derive its equations.

A good example is the ongoing debate over existing models’ failure to predict or untangle the reasons for the recent global financial crisis. Insufficient attention to the links between overall demand, wealth, and—in particular—excessive financial risk taking has been blamed. In the next few years there will be considerable research into uncovering and understanding the lessons from the crisis. This research will add new behavioral equations to current economic models. It will also entail modifying existing equations (for example, those that deal with household saving behavior) to link them to the new equations modeling the financial sector. The true test of the enhanced model will be its ability to consistently flag levels of financial risk that require a preemptive policy response.

No economic model can be a perfect description of reality. But the very process of constructing, testing, and revising models forces economists and policymakers to tighten their views about how an economy works. This in turn promotes scientific debate over what drives economic behavior and what should (or should not) be done to deal with market failures. Adam Smith would probably approve. **FD**

SAM OULIARIS is a former IMF staff member.

Econometrics: Making Theory Count

For economic theory to be a useful tool for policymaking, it must be quantifiable

Sam Ouliaris

ECONOMISTS DEVELOP ECONOMIC MODELS to explain consistently recurring relationships. Their models link one or more economic variables to other economic variables (see “Economic Models,” p. 8). For example, economists connect the amount individuals spend on consumer goods to disposable income and wealth, and expect consumption to increase as disposable income or wealth increases (that is, the relationship is positive).

There are often competing models for explaining the same recurring relationship, but few if any economic models provide useful clues as to the magnitude of the association. Yet this is what matters most to policymakers. When setting monetary policy, for example, central bankers need to know the likely impact of changes in official interest rates on inflation and on the growth rate of GDP. It is in cases like this that economists turn to econometrics.

Econometrics uses economic theory, mathematics, and statistical inference to quantify economic phenomena. In other words, it converts theoretical economic models into useful tools for economic policymaking. A key objective of econometrics is to convert qualitative statements (such as “the relationship between two or more variables is positive”) into quantitative statements (such as “consumption expenditure increases by 95 cents for every one dollar increase in disposable income”). Econometricians—practitioners of econometrics—transform models developed by economic theorists into versions that can be estimated and then used to inform policymaking. As Stock and Watson (2007) put it, “econometric methods are used in many branches of economics, including finance, labor economics, macroeconomics, microeconomics, and economic policy.” Economic policy decisions are rarely made without econometric analysis to assess their impact.

A daunting task

Certain features of economic data make it challenging for economists to quantify economic models. Unlike researchers in the physical sciences, econometricians are rarely able to conduct controlled experiments in which only one economic variable is changed. Instead, econometricians estimate economic relationships using data generated by a complex system of related equations, in which all variables may change at the same time.

That raises the question of whether there is even enough information in the data to identify the unknowns in the model.

The discipline of econometrics can be divided into theoretical and applied components.

Theoretical econometricians investigate the properties of existing statistical tests and procedures for estimating unknowns in the model. They also seek to develop new statistical procedures that are valid (or robust) despite the peculiarities of economic data—such as their tendency to change

Econometrics converts theoretical economic models into useful tools for economic policymaking.

simultaneously in response to the same shock(s). Theoretical econometrics relies heavily on mathematics, theoretical statistics, and numerical methods to prove that the new procedures have the ability to draw correct inferences.

Applied econometricians, by contrast, use econometric techniques developed by the theorists to translate qualitative economic statements into quantitative ones. Because applied econometricians are closer to the data, they often run into—and alert their theoretical counterparts to—data attributes that lead to problems with existing estimation techniques. For example, the econometrician might discover that the variance of the data (how much individual values in a series differ from the overall average) is changing over time.

The main tool of econometrics is the linear multiple regression model, which provides a formal approach to estimating how a change in one economic variable, the explanatory variable, affects the variable being explained, the dependent variable—taking into account the impact of all the other determinants of the dependent variable. This qualification is important because a regression seeks to estimate the marginal impact of a particular explanatory variable after taking into account the impact of the other explanatory variables in the model (see “Regressions: An Economist Obsession,” p. 74). For example, the model may try

to isolate the effect of a 1 percentage point increase in taxes on average household consumption expenditure, holding constant other determinants of consumption, such as pretax income, wealth, and interest rates.

Stages of development

The methodology of applied econometrics is well established.

The first step is to suggest a theory or hypothesis to explain the data being examined. The explanatory variables in the model are specified, and the expected sign and/or magnitude of the relationship between each explanatory variable and the dependent variable are clearly stated. At this stage of the analysis, applied econometricians rely heavily on economic theory to formulate the hypothesis. For example, a tenet of international economics is that prices across open borders move together after allowing for nominal exchange rate movements (that is, purchasing power parity). The empirical relationship between domestic prices and foreign prices (adjusted for nominal exchange rate movements) should be positive, and they should move together approximately one for one.

The second step is the specification of a statistical model that captures the essence of the theory the economist is testing. The model proposes a specific mathematical relationship between the dependent variable and the explanatory variables—on which, unfortunately, economic theory is usually silent. By far the most common approach is to assume log linearity—meaning that any change in an explanatory variable will always produce the same percentage change in the dependent variable (that is, a straight-line relationship).

Because it is impossible to account for every influence on the dependent variable, a catchall variable is added to the statistical model to complete its specification. The role of the catchall is to represent all the determinants of the dependent variable that cannot be accounted for explicitly—because of either the complexity of the data or its absence. Economists usually assume that this “error” term averages to zero and is unpredictable, simply to be consistent with the premise that the statistical model accounts for all the important explanatory variables.

The third step involves using an appropriate statistical procedure and an econometric software package to estimate the unknown parameters (coefficients) of the model using economic data. This is often the easiest part of the analysis thanks to readily available economic data and excellent econometric software. Still, the famous GIGO (garbage in, garbage out) principle of computing also applies to econometrics. Just because something can be computed doesn’t mean it makes economic sense to do so.

The fourth step is by far the most important: administering the smell test. Does the estimated model make economic sense—that is, yield meaningful economic predictions? For example, are the signs of the estimated parameters that connect the dependent variable to the explanatory variables consistent

with the predictions of the underlying economic theory? (In the household consumption example, for instance, the validity of the statistical model would be in question if it predicted a decline in consumer spending when income increased). If the estimated parameters do not make sense, how should the econometrician change the statistical model to yield sensible estimates? And does a more sensible estimate imply an economically significant effect? This step, in particular, calls on and tests the applied econometrician’s skill and experience.

Testing the hypothesis

The main tool of the fourth stage is hypothesis testing, a formal statistical procedure during which the researcher makes a specific statement about the true value of an economic parameter, and a statistical test determines whether the estimated parameter is consistent with that hypothesis. If it is not, the researcher must either reject the hypothesis or make new specifications to the underlying economic model and start over.

If all four stages proceed well, the result is a tool that can be used to assess the empirical validity of an abstract economic model. The empirical model may also be used to construct a forecast of the dependent variable, potentially helping policymakers make decisions about changes in monetary and/or fiscal policy to keep the economy on an even keel.

Students of econometrics are often fascinated by the ability of linear multiple regression to estimate economic relationships. Three fundamentals of econometrics are worth remembering.

- First, the quality of the parameter estimates depends on the validity of the underlying economic model.
- Second, if a relevant explanatory variable is excluded, the most likely outcome is poor parameter estimates.
- Third, even if the econometrician identifies the process that actually generated the data, the parameter estimates have only a slim chance of being equal to the actual parameter values that generated the data. Nevertheless, the estimates will still be used because, statistically speaking, they will become precise as more data become available.

Econometrics, by design, can yield correct predictions on average, but only with the help of sound economics to guide the specification of the empirical model. Even though it is a science, with well-established rules and procedures for fitting models to economic data, in practice econometrics is an art that requires considerable judgment to obtain estimates useful for policymaking. **FD**

SAM OULIARIS is a former IMF staff member.

Reference:

Stock, James H., and Mark W. Watson. 2007. *Introduction to Econometrics*, Addison-Wesley Series in Economics. Boston: Pearson Addison Wesley, 2nd ed.

Supply and Demand: Why Markets Tick

When buyers and sellers get together, the key outcome is a price

Irena Asmundson

FOR ECONOMISTS, a market is determined by how supply and demand come together to determine a price. Supply and demand are in turn determined by technology and the conditions under which people operate. At one extreme, the market could be populated by a large number of virtually identical sellers and buyers (for example, the market for ballpoint pens). At the other extreme, there might be only one seller and one buyer (as would be the case if I want to barter my table for your quilt).

Perfect competition

Economists have formulated models to explain various types of markets. The most fundamental is perfect competition, in which there are large numbers of identical suppliers and demanders of the same product, buyers and sellers can find one another at no cost, and no barriers prevent new suppliers from entering the market. In perfect competition, no one has the ability to affect prices. Both sides take the market price as a given, and the market-clearing price is the one at which there is neither excess supply nor excess demand. Suppliers will keep producing as long as they can sell the good for a price that exceeds their cost of making one more (the marginal cost of production). Buyers will go on purchasing as long as the satisfaction they derive from consuming is greater than the price they pay (the marginal utility of consumption). If prices rise, additional suppliers will be enticed to enter the market. Supply will increase until a market-clearing price is reached again. If prices fall, suppliers who are unable to cover their costs will drop out.

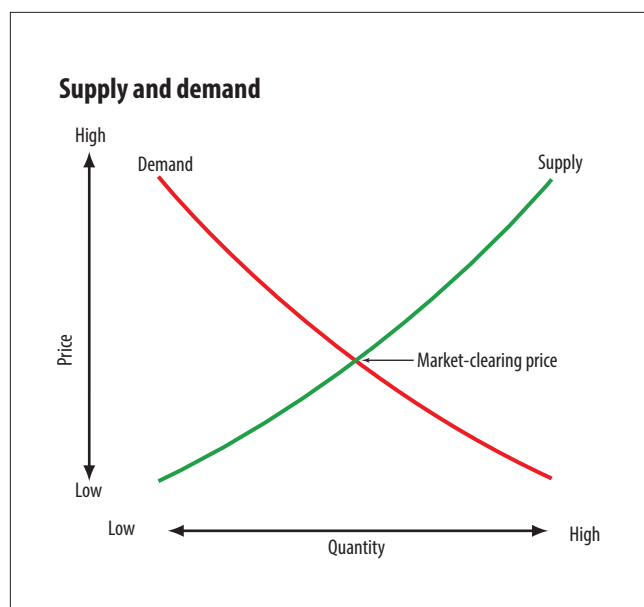
Economists generally lump together the quantities suppliers are willing to produce at each price into an equation called the supply curve. The higher the price, the more suppliers are likely to produce. Conversely, buyers tend to purchase more of a product the lower its price. The equation that spells out the quantities consumers are willing to buy at each price is called the demand curve.

Demand and supply curves can be charted on a graph, with prices on the vertical axis and quantities on the horizontal axis. Supply is generally considered to slope upward: as the price rises, suppliers are willing to produce more. Demand is generally

considered to slope downward: at higher prices, consumers buy less. The point at which the two curves intersect represents the market-clearing price—the price at which demand and supply are the same (see chart).

Prices can change for many reasons (technology, consumer preference, weather conditions). The relationship between the supply and demand for a good (or service) and changes in price is called elasticity. Goods that are inelastic are relatively insensitive to changes in price, whereas elastic goods are very responsive to price. A classic example of an inelastic good (at least in the short term) is energy. Consumers require energy to get to and from work and to heat their houses. It may be difficult or impossible in the short term for them to buy cars or houses that are more energy efficient. On the other hand, demand for many goods is very sensitive to price. Think steak. If the price of steak rises, consumers may quickly buy a cheaper cut of beef or switch to another meat. Steak is an elastic good.

Of course, most markets are imperfect; they are not composed



of unlimited buyers and sellers of virtually identical items who have perfect knowledge. At the other end of the spectrum from perfect competition is monopoly. In a monopoly, there is one supplier of a good for which there is no simple substitute. The supplier does not take the market price as a given. Instead, the monopolist can set it. (Monopoly's twin is monopsony, in which there is only one buyer, usually a government, although there may be many suppliers.)

Barriers to competition

In monopoly situations, there usually is a barrier—natural or legal—to potential competitors. For example, utilities are often monopolies. It would be inefficient for two water companies to manage watersheds, negotiate usage rights, and lay pipes to households. But consumers have no choice except to buy from the monopolist, and access may be unaffordable for some. As a result, governments usually regulate such monopolies to ensure that they do not abuse their market power by setting prices too high. In return for allowing a company to operate as a sole provider, there may be requirements for minimum services to be provided to everyone or a cap on the prices that can be charged. These caps generally allow companies to recover fixed costs.

Monopolists cannot be oblivious to demand—which, as under perfect competition, varies, depending on price. The difference is that a producer in perfect competition fulfills only a portion of total demand, whereas the monopolist benefits from the demand curve of the entire market. So the unregulated monopolist can decide to produce a quantity that maximizes its profit—almost always at a higher price and in a smaller amount than in a perfectly competitive market.

In perfect competition a firm with lower costs can reduce its price and add enough customers to make up for lost revenue on existing sales. Suppose a firm earns 5 cents a unit selling 1,000 units—or \$50—in a total market of 100,000 units. If it lowers its price by 1 cent and gains an additional 1,000 units in sales, its profits will be \$80 on its new level of sales of 2,000 units.

Just one seller

But a monopolist controls all the sales—in this case 100,000 units at a nickel a share, earning a profit of \$5,000. Lowering the price might increase total sales, but likely not enough to offset revenue lost on existing sales. Say it lowered the price (and profit per sale) by a penny, resulting in increased demand of 1,000 units. That would add \$40 to revenues. But the monopolist would also lose a penny in profit on each of the 100,000 units it had been selling—or \$1,000.

The key outcome of a monopoly is prices and profits that are higher than under perfect competition and supply that is often lower. There are other types of markets in which buyers and sellers have more market power than in perfect competition but

less than under a monopoly. In those cases, prices are higher and production is lower than in cases of perfect competition. For example, if an airline had information on how much you valued a flight because it knew you were going to a friend's wedding, it would know it could extract a higher price.

Supply and demand can also be affected by the product itself. In perfect competition, all producers make and buyers seek the same product—or close substitutes. In a monopoly, buyers lack

Supply and demand are determined by technology and the conditions under which people operate.

easy substitutes. Variety, though, allows for substitution across types. For example, the market for tomatoes involves more than simply matching buyers and sellers of an idealized tomato. Consumers may want different types, and producers can respond. Market entrants could compete head to head with an existing producer by applying the same production technology, but they might instead introduce new varieties (cherry, beefsteak, heirloom) to cater to different tastes. As a result, producers have limited market power to set prices when markets are competitive but products are differentiated. Still, varieties of products can be substituted for one another, even if imperfectly, so prices cannot be as high as in monopolies.

A temporary monopoly

Complications arise when the main features distinguishing one product from another are expensive to create but cheap to imitate—for example, books, drugs, computer software. Writing a book can be difficult, but printing a copy has a low marginal cost. Consumers may buy many books, but if one becomes popular, competitors will have an incentive to undercut the publisher and sell their own copies. To allow the author and publisher to recoup fixed costs, governments often grant a temporary monopoly on that book (called a copyright) for the author and publisher. The price exceeds the marginal cost of production, but the copyright motivates authors to keep writing and publishers to produce and market books—ensuring future supply.

The market structures discussed here are a few of the ways supply and demand can differ according to context. Production technologies, consumer preferences, and difficulties in matching sellers with buyers are some of the factors that influence markets, and all play a role in determining the market-clearing price. **FD**

IRENA ASMUNDSON is the chief economist at the California Department of Finance and a former IMF staff member.

Gross Domestic Product: An Economy's All

When it is growing, especially if inflation is not a problem, workers and businesses are generally better off than when it is not

Tim Callen

MANY PROFESSIONS commonly use acronyms. To doctors, accountants, and baseball players, the letters MRI (magnetic resonance imaging), GAAP (generally accepted accounting principles), and ERA (earned run average), respectively, need no explanation. To someone unfamiliar with these fields, however, without an explanation these acronyms are a stumbling block to a better understanding of the subject at hand.

Economics is no different. Economists use many acronyms. One of the most common is GDP, which stands for gross domestic product. It is often cited in newspapers, on the television news, and in reports by governments, central banks, and the business community. It has become widely used as a reference point for the health of national and global economies. When GDP is growing, especially if inflation is not a problem, workers and businesses are generally better off than when it is not.

Measuring GDP

GDP measures the monetary value of final goods and services—that is, those that are bought by the final user—produced in a country in a given period of time (say a quarter or a year). It counts all the output generated within the borders of a country. GDP is composed of goods and services produced for sale in the market and also includes some nonmarket production, such as defense or education services provided by the government. An alternative concept, gross national product, or GNP, counts all the output of the residents of a country. So if a German-owned company has a factory in the United States, the output of this factory would be included in US GDP, but in German GNP.

Not all productive activity is included in GDP. For example, unpaid work (such as that performed in the home or by volunteers) and black-market activities are not included because they are difficult to measure and value accurately. That means, for example, that a baker who produces a loaf of bread for a customer would contribute to GDP, but would not contribute to GDP if he baked the same loaf for his family.

Moreover, “gross” domestic product takes no account of

the wear and tear on the machinery, buildings, and so on (the so-called capital stock) that are used in producing the output. If this depletion of the capital stock, called depreciation, is subtracted from GDP, we get net domestic product.

Theoretically, GDP can be viewed in three different ways.

- The production approach sums the “value added” at each stage of production, where value added is defined as total sales minus the value of intermediate inputs into the production process. For example, flour would be an intermediate input and bread the final product, or an architect’s services would be an intermediate input and the building the final product.
- The expenditure approach adds up the value of purchases made by final users—for example, the consumption of food, televisions, and medical services by households; the investments in machinery by companies; and the purchases of goods and services by the government and foreigners.
- The income approach sums the incomes generated by production—for example, the compensation paid to employees, rent paid to land, interest paid on capital and profit paid to the company owners.

GDP in a country is usually calculated by the national statistical agency, which compiles the information from a large number of sources. In making the calculations, however, most countries follow established international standards. The international standard for measuring GDP is contained in the System of National Accounts, 1993, compiled by the International Monetary Fund, the European Commission, the Organisation for Economic Co-operation and Development, the United Nations, and the World Bank.

Real GDP

One thing people want to know about an economy is whether its total output of goods and services is growing or shrinking. But because GDP is collected at current, or nominal, prices, one cannot compare two periods without making adjustments for inflation. To determine “real” GDP, its nominal value must be

adjusted to take into account price changes to allow us to see whether the value of output has gone up because more is being produced or simply because prices have increased. A statistical tool called the price deflator is used to adjust GDP from nominal to constant prices.

GDP is important because it gives information about the size of the economy and how an economy is performing. The growth rate of real GDP is often used as an indicator of the general health of the economy. In broad terms, an increase in real GDP is interpreted as a sign that the economy is doing well. When real GDP is growing strongly, employment is likely to be increasing as companies hire more workers for their factories and people have more money in their pockets. But real GDP growth does move in cycles over time. Economies are sometimes in periods of boom, and sometimes periods of slow growth or even recession (with the latter sometimes defined as two consecutive quarters in which output declines). In the United States, for example, there were ten recessions of varying length and severity between 1950 and 2017 (see chart). The National Bureau of Economic Research makes the call on the dates of US business cycles.

Comparing GDPs of two countries

GDP is measured in the currency of the country in question. That requires adjustment when trying to compare the value of output in two countries using different currencies. The usual method is to convert the value of GDP of each country into US dollars and then compare them. Conversion to dollars can be done either using market exchange rates—those that prevail in the foreign exchange market—or purchasing-power-parity (PPP) exchange rates. The PPP exchange rate is the rate at which the currency of one country would have to be converted into

that of another to purchase the same amount of goods and services in each country. There is a large gap between market and PPP-based exchange rates in emerging market and developing countries. This is because nontraded goods and services tend to be cheaper in low-income than in high-income countries—for example, a haircut in New York is more expensive than in Bishkek—even when the cost of making tradable goods, such as machinery, across two countries is the same. For advanced countries, market and PPP exchange rates tend to be much closer. These differences mean that emerging market and developing countries have a higher estimated dollar GDP when the PPP exchange rate is used.

The IMF publishes an array of GDP data on its website (www.imf.org). International institutions such as the IMF also calculate global and regional measures of real GDP growth. These give an idea of how quickly or slowly the world economy or the economies in a particular region of the world are growing. The aggregates are constructed as weighted averages of the GDP in individual countries, with weights reflecting each country's share of GDP in the group (with PPP exchange rates used to determine the appropriate weights). So, for example, the IMF's *World Economic Outlook* reported that global real GDP grew by -0.15 percent in 2009, at the height of the global financial crisis and has been growing at an average rate of 3.4 percent since 2012 to date.

What GDP does not reveal

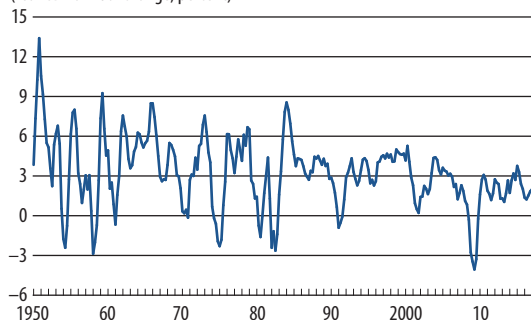
It is also important to understand what GDP cannot tell us. GDP is not a measure of the overall standard of living or well-being of a country. Although changes in the output of goods and services per person (GDP per capita) are often used as a measure of whether the average citizen in a country is better or worse off, it does not capture things that may be deemed important to general well-being. So, for example, increased output may come at the cost of environmental damage or other external costs, such as noise. Or it might involve the reduction of leisure time or the depletion of nonrenewable natural resources. The quality of life may also depend on the distribution of GDP among the residents of a country, not just the overall level. To try to account for such factors, the United Nations computes a Human Development Index, which ranks countries not only based on GDP per capita, but on other factors, such as life expectancy, literacy, and school enrollment. Other attempts have been made to account for some of the shortcomings of GDP, such as the Genuine Progress Indicator and the Gross National Happiness Index, but these too have their critics. [FD](#)

TIM CALLEN is a division chief in the IMF's Middle East and Central Asia Department. **SARWAT JAHAN**, a senior economist in the IMF's Asia and Pacific Department, updated this article.

Growth and gaps

Since 1950, US economic output measured by gross domestic product adjusted for inflation, has mainly been growing, except for ten recessions of varying length and severity.

(real GDP annual change, percent)



Source: National Bureau of Economic Research (NBER).

Note: Light-shaded areas indicate recessions - periods when output declines. Recessions are dated by the NBER, a private organization

Monetarism: Money Is Where It's At

Its emphasis on money's importance gained sway in the 1970s

Sarwat Jahan and Chris Papageorgiou

JUST HOW IMPORTANT is money? Few would deny that it plays a key role in the economy.

But one school of economic thought, called monetarism, maintains that the money supply (the total amount of money in an economy) is the chief determinant of current-dollar GDP in the short run and of the price level over longer periods. Monetary policy, one of the tools governments have to affect the overall performance of the economy, uses instruments such as interest rates to adjust the amount of money in the economy. Monetarists believe that the objectives of monetary policy are best met by targeting the growth rate of the money supply. Monetarism gained prominence in the 1970s—bringing down inflation in the United States and United Kingdom—and greatly influenced the US central bank's decision to stimulate the economy during the global recession of 2007–09.

Today, monetarism is mainly associated with Nobel Prize-winning economist Milton Friedman. In his seminal work *A Monetary History of the United States, 1867–1960*, which he wrote with fellow economist Anna Schwartz in 1963, Friedman argued that poor monetary policy by the US central bank, the Federal Reserve, was the primary cause of the Great Depression in the United States in the 1930s. In their view, the failure of the Fed (as it is usually called) to offset forces that were putting downward pressure on the money supply and its actions to reduce the stock of money were the opposite of what should have been done. They also argued that because markets naturally move toward a stable center, an incorrectly set money supply caused markets to behave erratically.

Monetarism gained prominence in the 1970s. In 1979, with US inflation peaking at 20 percent, the Fed switched its operating strategy to reflect monetarist theory. But monetarism faded in the following decades as its ability to explain the US economy seemed to wane. Nevertheless, some of the insights monetarists brought to economic analysis have been adopted by nonmonetarist economists.

At its most basic

The foundation of monetarism is the Quantity Theory of Money. The theory is an accounting identity—that is, it must be true. It says that the money supply multiplied by velocity (the rate at which money changes hands) equals nominal expenditures in the economy (the number of goods and services sold multiplied

by the average price paid for them). As an accounting identity, this equation is uncontroversial. What is controversial is velocity. Monetarist theory views velocity as generally stable, which implies that nominal income is largely a function of the money supply. Variations in nominal income reflect changes in real economic activity (the number of goods and services sold) and inflation (the average price paid for them).

The quantity theory is the basis for several key tenets and prescriptions of monetarism:

- Long-run monetary neutrality: An increase in the money stock would be followed by an increase in the general price level in the long run, with no effects on real factors such as consumption or output.
- Short-run monetary nonneutrality: An increase in the stock of money has temporary effects on real output (GDP) and employment in the short run because wages and prices take time to adjust (they are sticky, in economic parlance).
- Constant money growth rule: Friedman, who died in 2006, proposed a fixed *monetary rule*, which states that the Fed should be required to target the growth rate of money to equal the growth rate of real GDP, leaving the price level unchanged. If the economy is expected to grow at 2 percent in a given year, the Fed should allow the money supply to increase by 2 percent. The Fed should be bound to fixed rules in conducting monetary policy because discretionary power can destabilize the economy.
- Interest rate flexibility: The money growth rule was intended to allow interest rates, which affect the cost of credit, to be flexible to enable borrowers and lenders to take account of expected inflation as well as the variations in real interest rates.

Many monetarists also believe that markets are inherently stable in the absence of major unexpected fluctuations in the money supply. They also assert that government intervention can often destabilize the economy more than help it. Monetarists also believe that there is no long-run trade-off between inflation and unemployment because the economy settles at long-run equilibrium at a full employment level of output (see “The Output Gap,” p. 22).

The great debate

Although monetarism gained in importance in the 1970s, it was critiqued by the school of thought that it sought to supplant—Keynesianism. Keynesians, who took their inspiration from

the great British economist John Maynard Keynes, believe that demand for goods and services is the key to economic output. They contend that monetarism falters as an adequate explanation of the economy because velocity is inherently unstable and attach little or no significance to the quantity theory of money and the monetarist call for rules. Because the economy is subject to deep swings and periodic instability, it is dangerous to make the Fed slave to a preordained money target, they believe—the Fed should have some leeway or “discretion” in conducting policy. Keynesians also do not believe that markets adjust to disruptions and quickly return to a full employment level of output.

Keynesianism held sway for the first quarter century after World War II. But the monetarist challenge to the traditional Keynesian theory strengthened during the 1970s, a decade characterized by high and rising inflation and slow economic growth. Keynesian theory had no appropriate policy responses, while Friedman and other monetarists argued convincingly that the high rates of inflation were due to rapid increases in the money supply, making control of the money supply the key to good policy.

In 1979, Paul A. Volcker became chairman of the Fed and made fighting inflation its primary objective. The Fed restricted the money supply (in accordance with the Friedman rule) to tame inflation and succeeded. Inflation subsided dramatically, although at the cost of a big recession.

Monetarism had another triumph in Britain. When Margaret Thatcher was elected prime minister in 1979, Britain had endured several years of severe inflation. Thatcher implemented monetarism as the weapon against rising prices and succeeded in halving inflation, to less than 5 percent by 1983.

But monetarism’s ascendance was brief. The money supply is useful as a policy target only if the relationship between money

and nominal GDP, and therefore inflation, is stable and predictable. That is, if the supply of money rises, so does nominal GDP, and vice versa. To achieve that direct effect, though, the velocity of money must be predictable.

In the 1970s velocity increased at a fairly constant rate and it appeared that the quantity theory of money was a good one (see chart). The rate of growth of money, adjusted for a predictable level of velocity, determined nominal GDP. But in the 1980s and 1990s velocity became highly unstable with unpredictable periods of increases and declines. The link between the money supply and nominal GDP broke down, and the usefulness of the quantity theory of money came into question. Many economists who had been convinced by monetarism in the 1970s abandoned the approach.

Most economists think the change in velocity’s predictability was primarily the result of changes in banking rules and other financial innovations. In the 1980s banks were allowed to offer interest-earning checking accounts, eroding some of the distinction between checking and savings accounts. Moreover, many people found that money markets, mutual funds, and other assets were better alternatives to traditional bank deposits. As a result, the relationship between money and economic performance changed.

Relevant still

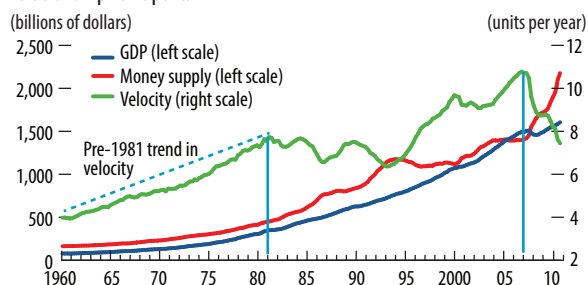
Still, the monetarist interpretation of the Great Depression was not entirely forgotten. In a speech during a celebration of Milton Friedman’s 90th birthday in late 2002, then–Fed governor Ben S. Bernanke, who would become chairman four years later, said, “I would like to say to Milton and Anna [Schwartz]: Regarding the Great Depression, you’re right. We [the Fed] did it. We’re very sorry. But thanks to you, we won’t do it again.” Fed Chairman Bernanke mentioned the work of Friedman and Schwartz in his decision to lower interest rates and increase the money supply to stimulate the economy during the global recession that began in 2007 in the United States. Prominent monetarists (including Schwartz) argued that the Fed stimulus would lead to extremely high inflation. Instead, velocity dropped sharply, and deflation is seen as a much more serious risk.

Although most economists today reject the slavish attention to money growth that is at the heart of monetarist analysis, some important tenets of monetarism have found their way into modern nonmonetarist analysis, muddying the distinction between monetarism and Keynesianism that seemed so clear three decades ago. Probably the most important is that inflation cannot continue indefinitely without increases in the money supply, and controlling it should be a primary, if not the only, responsibility of the central bank. [FD](#)

SARWAT JAHAN is a senior economist in the IMF’s Asia and Pacific Department and **CHRIS PAPAGEORGIU** is a deputy division chief in the IMF’s Research Department.

Varying velocity

When dollars changed hands at a predictable pace before 1981, money and output grew together. But when velocity became volatile, the relationship fell apart.



Source: Board of Governors, U.S. Federal Reserve System.

Note: Quarterly data are seasonally adjusted. Money supply = cash in circulation and checking accounts (M1). Velocity = the number of times a dollar changes hands in a year. GDP is reduced by a factor of 10 to fit on the chart.

II

HOW ECONOMIES

FUNCTION

What Is Direct Investment?

Investors often seek profits from a long-term stake in a foreign operation

Tadeusz Galeza and James Chan

FOREIGN INVESTORS can have myriad motivations for seeking to earn profits in another country. But they have fundamentally two core choices when deciding how to deploy their capital.

They can make a portfolio investment, buying stocks or bonds, say, often with the idea of making a short-term speculative financial gain without becoming actively engaged in the day-to-day running of the enterprise in which they invest.

Or they can choose the long-haul, hands-on approach—investing in an enterprise in another economy with the objective of gaining control or exerting significant influence over management of the firm (which usually involves a stake of at least 10 percent of a company's stock). In the most extreme case, investors may build new facilities from scratch, maintaining full control over operations.

It is the intent of lasting interest that is the crucial component of direct investment. A portfolio investor can sell a stock or bond quickly—whether to cement a gain or avoid a loss. Most corporations entering a foreign market through direct investment expect to substantially influence or control the management of the enterprise over the long haul.

Faces of investment

A number of factors influence a company's decision to engage in direct investment, including analysis of the trade costs with a foreign country. If these costs—including tariffs (taxes on imports), trade barriers such as quotas, and transportation—are higher than the cost, including the costs of production abroad, of establishing presence in the foreign country, the business will maximize its profits through direct investment.

Companies may invest with the idea of producing components that become part of a bigger product. An automaker may invest in a plant to build transmissions that are shipped to a final assembly plant in another country. This so-called vertical direct investment accounts for most of the investment by advanced economies in developing ones. The cost advantages associated with investing in a foreign country—and in many cases performing only a portion of the production process in that country—drive such investment. Abundant or unique natural resources or low labor costs influence the decision to move production overseas and import intermediate or final products from subsidiaries in host economies to the parent company's country (intrafirm trade).

A company may also invest in a foreign country by duplicating there its home country manufacturing processes. This may be done to supply goods or services to a foreign market. That's called horizontal direct investment. In countries with tariffs or other barriers to imports, a foreign firm may find that setting up local operations allows it to circumvent the barriers. Even though trade taxes have been falling over the years, such tariff jumping is still a common way to enter markets where the greatest benefit of direct investment is access to the local market. Another factor driving horizontal direct investment, specifically between advanced economies, is access to a pool of skilled employees and technology. In contrast to vertical direct investment, horizontal direct investment is likely to compete directly with local firms for local market share.

Of course investment need not be purely horizontal or vertical. A foreign subsidiary may provide goods to the parent company and receive services from the headquarters—a clear example of vertical direct investment. But the same subsidiary may also supply the local market, as part of the parent company's horizontal direct investment strategy.

Direct investment takes different shapes and forms. A company may enter a foreign market through so-called greenfield direct investment, in which the direct investor provides funds to build a new factory, distribution facility, or store, for example, to establish its presence in the host country.

But a company might also choose brownfield direct investment. Instead of establishing a new presence the company invests in or takes over an existing local company. Brownfield investment means acquiring existing facilities, suppliers, and operations—and often the brand itself.

Local effects

Countries may encourage inward direct investment to improve their finances. Firms that set up operations in host countries are subject to local tax laws and often significantly boost the host country's tax revenues. Direct investment can also help a country's balance of payments. Because portfolio investments can be volatile, a country's financial circumstances could worsen if investors suddenly withdrew their funds. Direct investment, on the other hand, is a more stable contributor to a country's financial structure. Direct investors do not wish to take actions to undermine the value or sustainability of their investments.

II. HOW ECONOMIES FUNCTION

Other positive effects associated with inward direct investment include increased employment, improved productivity, technology and knowledge transfer, and overall economic growth. Increased competition from foreign firms, whether new or acquired, often forces competitors to increase their productivity so that they don't go out of business. Suppliers and service providers to the direct investment enterprise may also increase their productivity, often because the investor requires higher-volume or higher-quality orders. The increase in volume and variability of products and services in the economy leads to overall improvement in the market's quality and size.

Host countries also benefit from a transfer of knowledge and technology, which often stems from workforce turnover. Incoming firms frequently offer more training opportunities than local employers. This knowledge is later transferred to local companies when trained employees leave the foreign enterprise for local businesses. In addition, there may be some incidental spillover of knowledge through informal networks, when employees exchange ideas and opinions about their workplace practices.

But direct investment may not always be viewed positively from a host country perspective. Because productive companies engage in direct investment, the increased competition they provide may force the least productive local companies out of business. Opponents of direct investment argue that foreign, especially brownfield, investment is a simple ownership transfer that does not generate new jobs. Some critics, moreover, point to the risk of a sudden reversal of the direct investment and a fire sale of assets, drastically reducing their value and, in extreme cases, forcing facilities to close and companies to lay off workers. Direct investment is often restricted in certain companies and industries, such as those involving sensitive high-technology products and in defense-related companies.

Because direct investment depends on the host country's decision to attract and accommodate investments, foreign companies often maintain close relations with the local authorities. This entanglement of business and politics may have an adverse effect on the host country. Perhaps the most common argument against direct investment is the potential power and political influence of foreign investors. The leverage investors have over policymakers becomes troublesome when a foreign company gains significant control over a sector of the economy or becomes a critical, or even the largest, employer in the market.

Attracting direct investment

Despite the potential problems of unregulated direct investment, governments of both advanced and developing economies tend to actively seek foreign investors and the capital they bring.

Advanced economies attract direct investment because of their stable policies, pool of skilled workers, and sizable markets. Developing economies are more interested in greenfield investment, which creates new facilities and jobs. Governments often set up special economic zones, provide the property for

construction of facilities, and offer generous tax incentives or subsidies to attract capital. These special economic zones, if properly designed, allow industries to concentrate in one geographic area, often placing suppliers close to buyers and providing the necessary infrastructure to meet investors' requirements.

Countries with a comparative advantage, such as favorable policies or a significant pool of skilled workers, frequently develop investment-promotion programs, which can include marketing

The overwhelming share of direct investment occurs among advanced economies.

campaigns, information offices, and even bilateral negotiations between governments and foreign firms. Unlike the tax and other fiscal incentives offered to foreign investors, information campaigns do not erode tax revenues from direct investment.

According to the IMF (2014), 63 percent of global direct investment occurs between advanced economies and 20 percent is between advanced and emerging market economies (including low-income countries). Six percent is between emerging market economies, and 11 percent of total direct investment flows from emerging market to advanced economies.

That the overwhelming share of direct investment occurs among advanced economies may seem counterintuitive. But given the large size of these economies, it stands to reason that horizontal direct investment in which advanced economies access pools of skilled workers, advanced technology, and large markets in other advanced economies dominates global direct investment.

Data on direct investment can be hard to interpret because of investments in tax havens. The level of investment in these countries is large, but investors tend to have no physical presence there. Given the pass-through nature of these investments, the usual costs and benefits associated with direct investment, other than collection of fees and taxes, do not apply.

Foreign direct investors may, as their critics claim, buy out domestic assets, pushing local firms out of business or imposing their policies on governments. But the overall benefits to both host and investing economies from foreign direct investment significantly outweigh the costs. Capital inflows from foreign direct investors help finance a country's spending—on investment, for example—and increase tax revenue, create jobs, and produce other positive spillovers for the host economy. **FD**

TADEUSZ GALEZA is a research officer in the IMF's Monetary and Capital Markets Department and **JAMES CHAN** is a senior information management officer in the IMF's Statistics Department.

Reference:

International Monetary Fund (IMF). 2014. *Coordinated Direct Investment Survey*. Washington.

The Output Gap: Veering from Potential

Economists look for the difference between what an economy is producing and what it can produce

Sarwat Jahan and Ahmed Saber Mahmud

DURING ECONOMIC DOWNTURNS an economy's output of goods and services declines. When times are good, by contrast, that output—usually measured as GDP—increases (see “Gross Domestic Product: An Economy's All,” p. 14).

One thing that concerns economists and policymakers about these ups and downs (commonly called the business cycle) is how close current output is to an economy's long-term potential output. That is, they are interested not only in whether GDP is going up or down, but also in whether it is above or below its potential.

The output gap is an economic measure of the difference between the actual output of an economy and its potential output. Potential output is the maximum amount of goods and services an economy can turn out when it is most efficient—that is, at full capacity. Often, potential output is referred to as the *production capacity* of the economy.

Just as GDP can rise or fall, the output gap can go in two directions: positive and negative. Neither is ideal. A *positive output gap* occurs when actual output is more than full-capacity output. This happens when demand is very high and, to meet that demand, factories and workers operate far above their most efficient capacity. A *negative output gap* occurs when actual output is less than what an economy could produce at full capacity. A negative gap means that there is spare capacity, or slack, in the economy due to weak demand (see chart).

An output gap suggests that an economy is running at an inefficient rate—either overworking or underworking its resources.

Inflation and unemployment

Policymakers often use potential output to gauge inflation and typically define it as the level of output consistent with no pressure for prices to rise or fall. In this context, the output gap is a summary indicator of the relative demand and supply components of economic activity. As such, the output gap measures the degree of inflation pressure in the economy and is an important link between the real side of the economy—which produces goods and services—and inflation. All else equal, if the output gap is positive over time, so that actual output is greater than potential output, prices will begin to rise in response to

demand pressure in key markets. Similarly, if actual output falls below potential output over time, prices will begin to fall to reflect weak demand.

The *unemployment gap* is a concept closely related to the output gap. Both are central to the conduct of monetary and fiscal policies. The nonaccelerating inflation rate of unemployment (NAIRU) is the unemployment rate consistent with a constant rate of inflation (see “Unemployment: The Curse of Joblessness,” p. 32). Deviations of the unemployment rate from the NAIRU are associated with deviations of output from its potential level. Theoretically, if policymakers get the actual unemployment rate to equal the NAIRU, the economy will produce at its maximum level of output without straining resources—in other words, there will be no output gap and no inflation pressure.

The output gap can play a central role in policymaking. For many central banks, including the US Federal Reserve, maintaining full employment is a policy goal. Full employment corresponds to an output gap of zero. Nearly all central banks seek to keep inflation under control, and the output gap is a key determinant of inflation pressure.

Because the output gap gauges when the economy may be overheating or underperforming, it has immediate implications for *monetary policy* (see “Money: At the Center of Transactions,” p. 26).

Typically during a recession, actual economic output drops below its potential, which creates a negative output gap. That below-potential performance may spur a central bank to adopt a monetary policy designed to stimulate economic growth—by lowering interest rates, for example, to boost demand and prevent inflation from falling below the central bank's inflation rate target.

In a boom, output rises above its potential level, resulting in a positive gap. In this case, the economy is often described as “overheating,” which generates upward pressure on inflation and may prompt the central bank to “cool” the economy by raising interest rates.

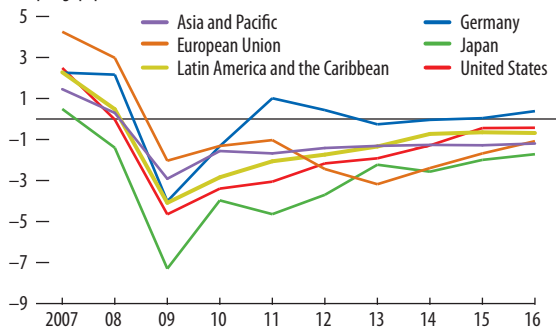
Governments can also use *fiscal policy* to close the output gap (see “Fiscal Policy: Taking and Giving Away,” p. 36). For example, fiscal policy that is expansionary—that raises aggregate

II. HOW ECONOMIES FUNCTION

Wasted potential

The severe recession in 2007–08 caused most economies to go from a positive output gap, exceeding long-run potential, to a negative output gap, in which GDP was below potential.

(output gap, percent of GDP)



Source: IMF, World Economic Outlook, April 2017.

demand by increasing government spending or lowering taxes—can be used to close a negative output gap. By contrast, when there is a positive output gap, contractionary or “tight” fiscal policy is adopted to reduce demand and combat inflation through lower spending and/or higher taxes.

Some policymakers have recently suggested that, in an increasingly integrated world economy, the global output gap can affect domestic inflation. In other words, all else equal, a booming world economy may increase the potential for inflation pressure within a country. For example, stronger global demand for computers raises the price US producers can charge their foreign customers. But because all computer producers are facing a stronger global market, US producers can charge more for their output at home as well. This is known as the “global output gap hypothesis” and calls for central bankers to pay close attention to developments in the growth potential of the rest of the world, not just domestic labor and capital capacity.

But there is so far no conclusive evidence to support the notion that a global output gap influences domestic prices. Still, the global output gap may become increasingly important if the world’s economies continue to integrate.

Hard to measure

Measuring the output gap is no easy task. Unlike actual output, the level of potential output and, hence, the output gap cannot be observed directly. Potential output and the output gap can only be estimated.

Various methodologies are used to estimate potential output, but they all assume that output can be divided into a trend and a cyclical component. The trend is interpreted as a measure of the economy’s potential output and the cycle as a measure of the output gap. The trick to estimating potential

output, therefore, is to estimate trends—that is, to remove the cyclical changes.

A common method of measuring potential output is the application of statistical techniques that differentiate between the short-term ups and downs and the long-term trend. The Hodrick-Prescott filter is one popular technique for separating the short from the long term. Other methods estimate the *production function*, a mathematical equation that calculates output based on an economy’s inputs, such as labor and capital. Trends are estimated by removing the cyclical changes in the inputs.

Any estimate of potential output will have its shortcomings. Estimates are based on one or more statistical relationships and therefore contain an element of randomness. Moreover, estimating the trend in a series of data is especially difficult near the end of a sample. That means, of course, that the estimate is most uncertain for the period of greatest interest: the recent past.

To circumvent these issues, some economists use surveys of producers to infer the extent of excess demand or supply in the economy. But surveys are also imperfect because firms may interpret questions differently, and there is no guarantee that responses will be indicative of demand pressure. Moreover, most surveys have a limited response base.

Regardless of the method used, estimating the output gap is subject to considerable uncertainty because the underlying relationships in the economy—that is, its structure—often change. For example, when the economy is emerging from a deep recession there may be much less spare capacity than anticipated because of such developments as

- unemployed workers who leave the labor market and become economically inactive;
- firms that close, leaving depressed areas and regions; and
- banks that lose money in a recession and become very strict with their lending.

Minding the gap

Because of the difficulties of estimating potential output and the output gap, policymakers need several other economic indicators to get an accurate reading of overall capacity pressure in the economy. Among those indicators are employment, capacity utilization, labor shortages, average hours worked and average hourly earnings, money and credit growth, and inflation relative to expectations.

These alternative measures of capacity can help policymakers enhance their measurement of the output gap. Even though it is difficult to estimate, the output gap has guided and will continue to guide policymakers. [FD](#)

SARWAT JAHAN is a senior economist in the IMF’s Asia and Pacific Department and **AHMED SABER MAHMUD** is the associate director of Applied Economics at Johns Hopkins University.

Structural Policies: Fixing the Fabric of the Economy

Monetary and fiscal policies deal with short-term economic fluctuations, but an economy's problems often go deeper

Khaled Abdel-Kader

ECONOMIES can get out of whack for a variety of reasons. Policymakers, in turn, have a number of ways to try to fix them, depending on what is wrong.

For example, when prices are rising too fast and consumers and businesses are buying at a rate that exceeds an economy's underlying ability to produce goods and services—that is, overall demand is growing too fast—policymakers can take steps to reduce demand. Similarly, during economic downturns, when businesses and consumers close their wallets—aggregate demand is shrinking—governments can take steps to encourage them to open their pocketbooks or substitute government spending for diminished private spending. Such government actions are called demand management or stabilization policies.

Sometimes an economy's problems are deeper and longer lasting than excessive or inadequate demand, usually as a result of government policies or private practices that impede efficient and fair production of goods and services—that is, supply. Fixing such problems can require changes to the fabric of the economy, called structural policies.

Stabilization policies are important in the short run, because it is easier to alter the various components of overall demand for a short time than it is to make a country's resources more productive. Stabilization policies include taxing and spending actions (see “Fiscal Policy: Taking and Giving Away,” p. 36) and changes to interest rates and the money supply (see “Monetary Policy: Stabilizing Prices and Output,” p. 36). When longer-term, structural changes are required to improve aggregate supply, governments must address specific impediments. This may involve the core structure of the economy, such as how prices are set, how public finance is conducted, government-owned enterprises, financial sector regulation, labor market rules and regulations, the social safety net, and institutions.

The recent financial and sovereign debt crises triggered calls for bold structural policies in several euro area countries, while declining growth in many developed and developing countries pointed to a need for fiscal, financial, institutional, and regulatory reforms to enhance productivity and raise growth and employment. Structural policies not only help raise economic

growth; they also set the stage for successful implementation of stabilization policies.

Dealing long term

Structural policies can zero in on a number of areas.

Price controls: Prices in free markets reflect the underlying cost of production. However, governments in some countries set the prices for certain goods and services—such as electricity, gas, and communication services—below production costs, particularly when the goods or services are produced by government-owned companies. These price controls lead to losses that the government must make up—which can cause budget and stabilization problems. Moreover, controls encourage higher consumption than would be the case if the prices of goods and services reflected the true cost of production. Underpricing leads to poor allocation of a society's resources. Were controls eliminated, prices would rise to cover costs, which would promote competition and efficiency.

Management of public finances: Although governments may briefly have to spend more than they take in during a recession—or collect more taxes than needed for a while to dampen spending in a boom—over the long term spending and taxing should be in synch. But complex tax laws and inefficient systems of tax administration, for example, can make it difficult to raise sufficient public revenue, which often leads to large budget deficits and accumulation of debt (a stabilization problem). That in turn can restrict a government's ability to finance development needs such as health care services, education, and infrastructure projects. Tax reforms can facilitate taxpayer compliance and raise revenue by removing exemptions, requiring advance payment of estimated tax liabilities, and simplifying the tax rate structure. Improved tax administration can also increase revenue. For example, better training and higher salaries for tax collectors could reduce corruption and help retain competent staff. Better management of public expenditures could result in more productive use of public funds.

Public sector enterprises: Government-owned enterprises make up a considerable share of the economy in some countries. Some operate efficiently and in the best interest of consumers.

II. HOW ECONOMIES FUNCTION

But often, because there is little if any competition, government-owned businesses deliver low-quality goods and services. Public businesses that compete with private firms often operate at a loss because of political influence or higher operating costs (as a result of unneeded workers, for example), and the government must make up the losses. Stabilization problems can arise if these government enterprises have to borrow from commercial banks to cover the losses. The loans are usually guaranteed by the government, which imposes contingent liabilities on the government budget because the government will have to pay if the enterprises don't. Countries with large state-owned enterprises could sell them to private individuals or firms. Or they could maintain public ownership in general, but take such steps as closing enterprises that are inefficient or have losses, changing their management, or reducing the labor force to align it with business needs—with an appropriate safety net to protect the laid-off staff.

Financial sector: The financial sector's role is to channel funds from savers to borrowers. A sound financial sector helps ensure that such funds are used in the most productive manner, which leads to higher economic growth and development. However, underdeveloped or poorly regulated financial systems in some developing countries could hamper economic growth and make it more difficult to conduct stabilization policies. For example, central banks generally carry out monetary policy by buying and selling on the open market securities that governments have sold to the public. But if there are no so-called secondary markets for government securities, or if they are poorly developed, central banks could be constrained in their attempts to carry out effective monetary policy and may have to resort to inefficient (or unfair) policy tools, such as credit rationing or interest rate controls. Inadequately regulated banks may engage in risky behavior that leads to banking crises—such as a “run,” when worried depositors rush en masse to take out their funds, or a failure, which is generally the result of bad lending practices. But even sound banks can fail if they get caught up in a systemwide run that exhausts the cash they have to pay depositors. Banking crises can interrupt the flow of funds to borrowers, discourage saving, and lead to higher government deficits if the state guarantees deposits or recapitalizes banks. Policymakers can fix underdeveloped financial systems through the introduction of secondary markets, the development of stock markets, and the privatization of government-owned banks. To mitigate crises, policymakers must shore up the financial system through effective regulation and supervision.

Social safety nets: Governments often have programs designed to safeguard a minimum standard of living for the poor and other vulnerable groups. But in many developing countries some costly programs—like fuel and food subsidies—are poorly targeted and benefit the rich more than the poor. In developed countries, pay-as-you-go pension programs have huge unfunded

liabilities because more people are retiring than entering the workforce. In addition, generous unemployment benefits often contribute to high unemployment because employers, who pay unemployment insurance premiums, are reluctant to hire new workers. Governments can change social safety nets to target the needy and achieve considerable savings. To focus on the needy, governments could give low-income households vouchers for basic food items or distribute food only in areas where the poor live. The government could also replace food and fuel subsidies with cash transfers. Pension programs can be changed so that benefits are aligned with projected revenues by raising the retirement age or fully funding pension systems.

Labor market: Unemployment is prevalent in many countries for a variety of reasons and usually rises when the economy is not doing well. But sometimes the cause of unemployment is deeper than the effects of the business cycle. For example excessive social security contributions or a relatively high minimum wage may so boost the cost of hiring that demand for labor shrinks and unemployment rises. Demand for labor may also fall if workers lack the necessary skills because of inadequate training or education. Reforming education and improving on-the-job training programs can help restore demand for labor.

Public institutions: The performance of public institutions can significantly affect a country's economic environment. For example, low government salaries, say in tax administration, can encourage corruption. Also, inefficient legal systems and shortages of courts and judges make it hard for businesses to resolve disputes, which increases costs for businesses and deters investment, especially foreign direct investment—hurting economic growth. Governance and institutions can be improved by simplifying business regulations and licensing, enhancing the country's legal system, streamlining the system of tax administration, and raising salaries for government staff in charge of providing vital services while limiting employment in the public sector to business needs.

Hand in hand

Raising an economy's growth potential requires stabilization and structural policies that complement one another. Stabilization policies lay the foundation for economic growth by helping lower inflation, smooth out consumption and investment, and reduce government deficits. Successful implementation of structural policies is possible only after such macroeconomic imbalances have been resolved. Similarly, though, structural policies enhance the effectiveness of many stabilization measures: promoting competition (a structural policy), for example, can lead to lower prices and, hence, lower inflation (the goal of stabilization policies). [FD](#)

KHALED ABDEL-KADER is a senior economist in the IMF's Institute for Capacity Development.

Money: At the Center of Transactions

Without it, modern economies could not function

Irena Asmundson and Ceyda Oner

MONEY MAY MAKE the world go around, as the song says. And most people in the world probably have handled money, many of them on a daily basis. But despite its familiarity, probably few people could tell you exactly what money is, or how it works.

In short, money can be anything that can serve as a

- **store of value**, which means people can save it and use it later—smoothing their purchases over time;

- **unit of account**, that is, provide a common base for prices; or

- **medium of exchange**, something that people can use to buy and sell from one another.

Perhaps the easiest way to think about the role of money is to consider what would change if we did not have it.

If there were no money, we would be reduced to a barter economy. Every item someone wanted to purchase would have to be exchanged for something that person could provide. For example, a person who specialized in fixing cars and needed to trade for food would have to find a farmer with a broken car. But what if the farmer did not have anything that needed to be fixed? Or what if a farmer could only give the mechanic more eggs than the mechanic could reasonably use? Having to find specific people to trade with makes it very difficult to specialize. People might starve before they were able to find the right person with whom to barter.

But with money, you don't need to find a particular person. You just need a market in which to sell your goods or services. In that market, you don't barter for individual goods. Instead you exchange your goods or services for a common medium of exchange—that is, money. You can then use that money to buy what you need from others who also accept the same medium of exchange. As people become more specialized, it is easier to produce more, which leads to more demand for transactions and, hence, more demand for money.

Many monies

To put it a different way, money is something that holds its value over time, can be easily translated into prices, and is widely accepted. Many different things have been used as money over the years—among them, cowry shells, barley, peppercorns, mobile phone minutes in developing economies, gold, and silver.

At first, the value of money was anchored by its alternative uses, and the fact that there were replacement costs. For example, you

could eat barley or use peppercorns to flavor food. The value you place on such consumption provides a floor for the value. Anyone could grow more, but it does take time, so if the barley is eaten the supply of money declines. On the other hand, many people may want strawberries and be happy to trade for them, but they make poor money because they are perishable and too fragile to transport easily. There is also the problem of divisibility—not everything of value is easily divided, and standardizing each unit is also tricky; for example, the value of a basket of strawberries measured against different items is not easy to establish and keep constant. Not only strawberries make for bad money; most things do.

But precious metals seemed to serve all three needs: a stable unit of account, a durable store of value, and a convenient medium of exchange. They are hard to obtain. There is a finite supply of them in the world. They stand up to time well. They are easily divisible into standardized coins and do not lose value when made into smaller units. In short, their durability, limited supply, high replacement cost, and portability made precious metals more attractive as money than other goods.

Until relatively recently, gold and silver were the main currency people used. Gold and silver are heavy, though, and over time, instead of carrying the actual metal around and exchanging it for goods, people found it more convenient to deposit precious metals at banks and buy and sell using a note that claimed ownership of the gold or silver deposits. Eventually, the paper claim on the precious metal was delinked from the metal. When that link was broken, fiat money was born. Fiat money is materially worthless, but has value simply because people collectively agree to ascribe a value to it. For money that is issued by a government, there is a guaranteed source of demand from requiring taxes be paid in that currency. For other money, such as cryptocurrencies, it only works because people believe that it will. As the means of exchange evolved, so did its source—from individuals in barter, to some sort of collective acceptance when money was barley or mobile phone minutes, to governments in more recent times.

Even though using standardized coins or paper bills made it easier to determine prices of goods and services, the amount of money in the system also played an important role in setting prices. For example, a wheat farmer would have at least two reasons for holding money: to use in transactions (cash in advance) and as a buffer against future needs (precautionary saving). Suppose

II. HOW ECONOMIES FUNCTION

winter is coming and the farmer wants to add to his store of money in anticipation of future expenses. If the farmer has a hard time finding people with money who want to buy wheat, he may have to accept fewer coins or bills in exchange for the grain. The result is that the price of wheat goes down because the supply of money is too tight. One reason might be that there just isn't enough gold to mint new money. When prices as a whole go down, it is called deflation. On the other hand, if there is more money in circulation but the same level of demand for goods, the value of the money will drop. This is inflation—when it takes more money to get the same amount of goods and services (see “Inflation: Prices on the Rise,” p. 30). Keeping the demand for and supply of money balanced can be tricky.

Manufacturing money

Fiat money is more efficient to use than precious metals. Adjustments to its supply do not depend on the amount of precious metal around. But that adds its own complication: Precisely because there is a finite amount of precious metals, there is a limit on the amount of notes that can be issued. If there is no gold or silver to back money, how do governments know how much to print? That gets into the dilemmas governments face. On one hand, the authorities will always be tempted to issue money, because governments can buy more with it, hire

more people, pay more wages, and increase their popularity. On the other hand, printing too much money starts to push up prices. If people start expecting that prices will continue to rise, they may increase their own prices even faster. Unless the government acts to rein in expectations, trust in money will be eroded, and it may eventually become worthless. That is what

Money is anything that can serve as a store of value, unit of account, or a medium of exchange.

happens during hyperinflation. To remove this temptation to print money willy-nilly, most countries today have delegated the task of deciding how much money to print to independent central banks, which are charged with making the call based on their assessment of the economy's needs and do not transfer funds to the government to finance its spending. The term “printing money” is something of a misnomer in itself. Most money today is in the form of bank deposits rather than paper currency (see box).

Belief can fade

Countries that have been down the path of high inflation experienced firsthand how the value of money essentially depends on people believing in it. In the 1980s, people in some Latin American countries, such as Argentina and Brazil, gradually lost confidence in the currency, because inflation was eroding its value so rapidly. They started using a more stable one, the US dollar, as the de facto currency. This phenomenon is called unofficial, or de facto, dollarization. The government loses its monopoly on issuing money—and dollarization can be very difficult to reverse.

Some government policies to restore confidence in a currency nicely highlight the “faith” part of money functioning. In Turkey, for example, the government rebased the currency, the lira, eliminating six zeros in 2005. Overnight, 1,000,000 liras became 1 lira. Brazil, on the other hand, introduced a new currency in 1994, the real. In both countries, citizens went along, demonstrating that as long as everyone accepts that a different denomination or a new currency is the norm, it simply will be. Just like fiat money. If it is accepted as money, it is money. **FD**

IRENA ASMUNDSON is the chief economist at the California Department of Finance and a former IMF staff member and **CEYDA ONER** is a deputy division chief in the IMF's Finance Department.

Reference:

International Monetary Fund (IMF). 2000. *Monetary and Financial Statistics Manual*. Washington.

HOW MONEY IS MEASURED

In official statistics, the amount of money in an economy is generally measured through what is called broad money, which encompasses everything that provides a store of value and liquidity. Liquidity refers to the extent to which financial assets can be sold at close to full market value at short notice. That is, they can easily be converted into another form of money, such as cash. Although currency and transferable deposits (narrow money) are included by all countries in broad money, there are other components that may also provide sufficient store of value and liquidity to count as broad money. Among the things the IMF (2000) says can be counted as broad money are the following:

National currencies (generally issued by the central government).

Transferable deposits, which include demand deposits (transferable by check or money order); bank checks (if used as a medium of exchange); traveler's checks (if used for transactions with residents); and deposits otherwise commonly used to make payments (such as some foreign-currency deposits).

Other deposits, such as nontransferable savings deposits, term deposits (funds left on deposit for a fixed period of time), or repurchase agreements (in which one party sells a security and agrees to buy it back at a fixed price).

Securities other than shares of stock, such as tradable certificates of deposit and commercial paper (which is essentially a corporate IOU).

Price: The Language of Exchange

A price is the amount of money a buyer gives a seller in exchange for a good or a service. But it can be more than that

Irena Asmundson

AT ITS MOST BASIC, a price is the amount of money that a buyer gives to a seller in exchange for a good or a service. When someone hands over \$2.00 and receives a pound of tomatoes, the price is straightforward observation: \$2.00 a pound. When an actual, observable transaction takes place, the price is sometimes called the traded price or the spot price.

But there are many other types of price. Some of them, such as the marginal price, are conceptual. Others are related to the timing of a potential transaction or to the relative power of the buyer and the seller. All of them, however, ultimately have some relation to the spot price.

Suppose that the tomato transaction takes a slightly different form. The seller might indicate a willingness to sell the tomatoes at a certain price, called the selling price or the ask price. The buyer may make it known that he is willing to pay a different price, which is called the bid price. Such a transaction can occur only if the seller values the tomatoes at \$2.00 a pound or less and the buyer values the tomatoes at \$2.00 a pound or more. That is, the bid price must be at least as high as the ask price. If it is not, one or both of the parties would be better off keeping what they already had, whether it is tomatoes or money.

Clearing the market

Most of the time, when economists speak of price, they are referring to a market-clearing price—that is, the price at which the amount of a good or service supplied by all sellers in a market is equal to the amount demanded by all buyers. Generally, economists assume that demand decreases as prices rise, and supply increases with price. The point at which these two prices are the same, or intersect, is the market-clearing price (see chart). If a farmer raised prices to a level greater than the market-clearing price on tomatoes, she would not sell them all, and if she lowered prices, she would have to turn away customers because she would run out of tomatoes before the buyers ran out of demand.

But market-clearing prices are not set in stone. Supply and demand can change. For example, if all customers suddenly decided they liked tomatoes more than they used to and were willing to pay a higher price for the same amount, the market-clearing price

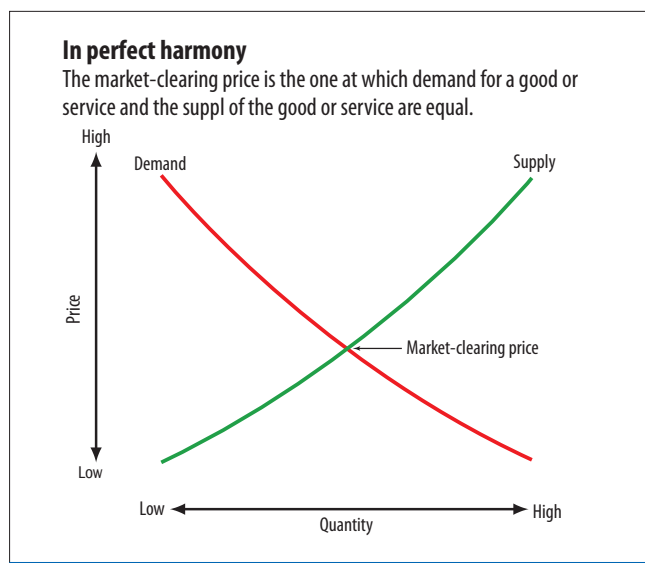
would rise. It could also rise if the supply of tomatoes declined—because of, say, planting decisions or the weather. The clearing price could also decline with changes in demand or supply.

Many prices

The above examples assume a single price that everyone is charged for the same good or service. But in reality many different prices can exist in a market at the same time, depending on the conditions under which a sale takes place.

Suppose the local supermarket has a lot of tomatoes that are likely to go bad in a few days. The market's managers decide to reduce the price to attract buyers and move more tomatoes. One pound of tomatoes is still \$2.00, but if a buyer takes two pounds the cost is \$3.00. There is a difference between a marginal price—the cost of an additional unit of a good, in this case a pound of tomatoes—and the average price. If a buyer takes two pounds, the average price is \$1.50 pound. But the marginal price is \$2 for the first pound and \$1 for the second.

Prices can also differ depending on when the actual transaction takes place and under what conditions. For example, suppose a customer wants to buy 10 pounds of tomatoes and pick them up



the next day. The price for a transaction scheduled for the future is called a forward price. The farmer may be happy to set aside those 10 pounds for the customer. Or she may worry that the customer will forget, leaving her with unsold tomatoes. She may ask for payment in advance or perhaps for a partial payment as a deposit. If the customer pays in advance, he runs the risk that the farmer will forget to set aside the tomatoes, leaving him scrambling to find tomatoes. Once again, the price will depend on the relative values the farmer and the customer place on those tomatoes.

Paying for certainty

If the customer values the certainty of getting those tomatoes, he will be willing to pay a higher price. For example, the tomatoes could be for a birthday dinner for someone who loves tomatoes. But if the tomatoes are for a tomato-tossing game at a picnic, the customer could use water balloons or eggs instead.

Suppose the farmer has enormous fields, and 10 pounds of tomatoes represents a small amount of her daily sales. The value she places on knowing the customer will show up is relatively low. Moreover, there is little risk that she will run out of tomatoes before the customer shows up. Whether the customer buys from her does not matter to her pricing decision, and the customer's forward price should be the same as the expected spot price the next day. If that is the case, the customer might not even bother to settle on a forward price with the farmer. When the seller has the power to set prices and the buyer cannot bargain, the seller is said to be a price setter, and the buyer is said to be a price taker.

If, however, 10 pounds represents half the farmer's daily sales of tomatoes, the value the customer places on those tomatoes matters a great deal. If the customer wants them for his sister's birthday, he may be willing to pay a higher forward price, which the farmer will be willing to accept. If the customer does not care that much about the tomatoes, he may not be willing to pay enough to secure the supply. When there are many farmers from whom to buy, no individual seller is able to set the price; the sellers are said to be price takers.

When there are many buyers and many sellers, a single market-clearing price is most likely to prevail for everyone.

A final type of price relates to future options. One can buy the right to transact at a prespecified price in the future, paying what is called an option price. The prespecified price at which one exercises an option is called the strike price. This is the price that comes into play when there is a great deal of uncertainty about how the spot price might change.

Consider again the case of tomatoes, for example. They need hot, dry weather to ripen properly. If the weather is like that in the near future, tomatoes will be abundant, and their price should be relatively low. If, however, rain is forecast for the near future, buyers and sellers would anticipate fewer tomatoes will be ready for harvest, and the price should rise (assuming the same demand). In the latter case, the customer may be willing to pay

a small amount now (the option price) to secure the right to buy 10 pounds in the future at a prespecified price (the strike price). If the spot price is higher than the strike price, the customer can exercise the option and buy the tomatoes at the strike price.

Many different prices can exist in a market at the same time.

If the spot price is lower, the customer pays the spot price and saves some money. The option price thus has to be both low enough to induce the customer to pay for the certainty and high enough to compensate the seller for honoring the strike price if it is lower than the spot price.

Values and prices

Because so many factors can influence a price, many people try to work out what a good price might be before testing a market. For example, suppose an engineer has invented a machine for picking tomatoes and she sets up a company to make them. She needs funds to build the machines, so she will try to sell stock in her company to raise the money. If she sets a price too high, some of the shares will be left unsold and she may not raise enough money. If she sets a price too low, a lucky buyer will turn around and sell to someone for a higher price. She has to carefully match the value with the price. What factors should she consider when setting the price

Again supply and demand come into play. How many machines can the new factory produce in a year? At what price will they be sold? How many machines will tomato farmers buy? This will depend on the price of tomatoes, the price of wages paid to tomato pickers, and the price of borrowing the money to buy the machine. All these prices must be ascertained before the engineer can figure out the market-clearing price for her machines—which will determine the price of the stock in her company. The stock price will also depend on how investors expect the company to do in the future. Investors might believe she will be able to make a bean-picking machine next year. If one of the prices underlying the inventions moves—because of a bad harvest, or because bean farms start paying higher wages, or because a competitor invents a bean picking machine next year—the company's stock price will move too.

There are many types of prices, representing many types of transactions. Each price represents a bargain struck between a buyer and a seller. The motivations of each party, and the prices that arise, can be as unpredictable as the weather and change as quickly. **FD**

IRENA ASMUNDSON is the chief economist at the California Department of Finance and a former IMF staff member.

Inflation: Prices on the Rise

Inflation measures how much more expensive a set of goods and services has become over a certain period, usually a year

Ceyda Oner

IT MAY BE one of the most familiar words in economics. Inflation has plunged countries into long periods of instability. Central bankers often aspire to be known as “inflation hawks.” Politicians have won elections with promises to combat inflation, only to lose power after failing to do so. Inflation was even declared Public Enemy No. 1 in the United States—by President Gerald Ford in 1974. What, then, is inflation, and why is it so important?

Inflation is the rate of increase in prices over a given period of time. Inflation is typically a broad measure, such as the overall increase in prices or the increase in the cost of living in a country. But it can also be more narrowly calculated—for certain goods, such as food, or for services, such as a haircut, for example. Whatever the context, inflation represents how much more expensive the relevant set of goods and/or services has become over a certain period, most commonly a year.

Measuring inflation

Consumers’ cost of living depends on the prices of many goods and services and the share of each in the household budget. To measure the average consumer’s cost of living, government agencies conduct household surveys to identify a basket of commonly purchased items and track over time the cost of purchasing this basket. (Housing expenses, including rent and mortgages, constitute the largest component of the consumer basket in the United States.) The cost of this basket at a given time expressed relative to a base year is the *consumer price index* (CPI), and the percentage change in the CPI over a certain period is *consumer price inflation*, the most widely used measure of inflation. (For example, if the base year CPI is 100 and the current CPI is 110, inflation is 10 percent over the period.)

Core consumer inflation focuses on the underlying and persistent trends in inflation by excluding prices set by the government and the more volatile prices of products, such as food and energy, most affected by seasonal factors or temporary supply conditions. Core inflation is also watched closely by policymakers. Calculation of an overall inflation rate—for a country, say, and not just for consumers—requires an index with broader coverage, such as the *GDP deflator*.

The CPI basket is mostly kept constant over time for consistency, but is tweaked occasionally to reflect changing consumption patterns—for example, to include new hi-tech goods and to replace items no longer widely purchased. Because it shows how, on average, prices change over time for everything produced in an economy, the contents of the GDP deflator vary each year and are more current than the mostly fixed CPI basket. On the other hand, the deflator includes nonconsumer items (such as military spending) and is therefore not a good measure of the cost of living.

The good and the bad

To the extent that households’ *nominal* income, which they receive in current money, does not increase as much as prices, they are worse off, because they can afford to purchase less. In other words, their *purchasing power* or *real*—inflation-adjusted—income falls. Real income is a proxy for the standard of living. When real incomes are rising, so is the standard of living, and vice versa.

In reality, prices change at different paces. Some, such as the prices of traded commodities, change every day; others, such as wages established by contracts, take longer to adjust (or are “sticky,” in economic parlance). In an inflationary environment, unevenly rising prices inevitably reduce the purchasing power of some consumers, and this erosion of real income is the single biggest cost of inflation.

Inflation can also distort purchasing power over time for recipients and payers of fixed interest rates. Take pensioners who receive a fixed 5 percent yearly increase to their pension. If inflation is higher than 5 percent, a pensioner’s purchasing power falls. On the other hand, a borrower who pays a fixed-rate mortgage of 5 percent would benefit from 5 percent inflation, because the *real interest rate* (the nominal rate minus the inflation rate) would be zero; servicing this debt would be even easier if inflation were higher, as long as the borrower’s income keeps up with inflation. The lender’s real income, of course, suffers. To the extent that inflation is not factored into *nominal interest rates*, some gain and some lose purchasing power.

Indeed, many countries have grappled with high inflation—and in some cases *hyperinflation*, 1,000 percent or more a year.

In 2008, Zimbabwe experienced one of the worst cases of hyperinflation ever, with estimated annual inflation at one point of 500 billion percent. Such high levels of inflation have been disastrous, and countries have had to take difficult and painful policy measures to bring inflation back to reasonable levels, sometimes by giving up their national currency, as Zimbabwe has.

Although high inflation hurts an economy, *deflation*, or falling prices, is not desirable either. When prices are falling, consumers delay making purchases if they can, anticipating lower prices in the future. For the economy this means less economic activity,

Inflation is the rate of increase in prices over a given period of time. Erosion of real income is the single biggest cost of inflation.

less income generated by producers, and lower economic growth. Japan is one country with a long period of nearly no economic growth, largely because of deflation. Preventing deflation during the global financial crisis that began in 2007 was one of the reasons the US Federal Reserve and other central banks around the world kept interest rates low for a prolonged period and have instituted other monetary policies to ensure financial systems have plenty of liquidity.

Most economists now believe that low, stable, and—most important—predictable inflation is good for an economy. If inflation is low and predictable, it is easier to capture it in price-adjustment contracts and interest rates, reducing its distortionary impact. Moreover, knowing that prices will be slightly higher in the future gives consumers an incentive to make purchases sooner, which boosts economic activity. Many central bankers have made their primary policy objective maintaining low and stable inflation, a policy called *inflation targeting*.

What creates inflation?

Long-lasting episodes of high inflation are often the result of lax monetary policy. If the money supply grows too big relative to the size of an economy, the unit value of the currency diminishes; in other words, its purchasing power falls and prices rise. This relationship between the money supply and the size of the economy is called the *quantity theory of money* and is one of the oldest hypotheses in economics.

Pressures on the supply or demand side of the economy can also be inflationary. *Supply shocks* that disrupt production, such as natural disasters, or raise production costs, such as high oil prices, can reduce overall supply and lead to “cost-push” inflation, in which the impetus for price increases comes from

a disruption to supply. The food and fuel inflation of 2008 was such a case for the global economy—sharply rising food and fuel prices were transmitted from country to country by trade. Conversely, *demand shocks*, such as a stock market rally, or *expansionary policies*, such as when a central bank lowers interest rates or a government raises spending, can temporarily boost overall demand and economic growth. If, however, this increase in demand exceeds an economy’s production capacity, the resulting strain on resources is reflected in “demand-pull” inflation. Policymakers must find the right balance between boosting demand and growth when needed without overstimulating the economy and causing inflation.

Expectations also play a key role in determining inflation. If people or firms anticipate higher prices, they build these expectations into wage negotiations and contractual price adjustments (such as automatic rent increases). This behavior partly determines the next period’s inflation; once the contracts are exercised and wages or prices rise as agreed, expectations become self-fulfilling. And to the extent that people base their expectations on the recent past, inflation would follow similar patterns over time, resulting in inflation *inertia*.

How policymakers deal with inflation

The right set of *disinflationary policies*, those aimed at reducing inflation, depends on the causes of inflation. If the economy has overheated, central banks—if they are committed to ensuring price stability—can implement *contractionary* policies that rein in aggregate demand, usually by raising interest rates. Some central bankers have chosen, with varying degrees of success, to impose monetary discipline by *fixing the exchange rate*—tying the value of its currency to that of another currency, and thereby its monetary policy to that of another country. However, when inflation is driven by global rather than domestic developments, such policies may not help. In 2008, when inflation rose across the globe on the back of high food and fuel prices, many countries allowed the high global prices to pass through to the domestic economy. In some cases the government may directly set prices (as some did in 2008 to prevent high food and fuel prices from passing through). Such *administrative price-setting* measures usually result in the government’s accrual of large subsidy bills to compensate producers for lost income.

Central bankers are increasingly relying on their ability to influence *inflation expectations* as an inflation-reduction tool. Policymakers announce their intention to keep economic activity low temporarily to bring down inflation, hoping to influence expectations and contracts’ built-in inflation component. The more credibility central banks have, the greater the influence of their pronouncements on inflation expectations. **FD**

CEYDA ONER is a deputy division chief in the IMF’s Finance Department.

Unemployment: The Curse of Joblessness

The number of people at work is generally closely related to whether an economy is growing at a reasonable rate

Ceyda Oner

BACK IN THE DEPTHS of the global financial crisis in 2009, the International Labor Office announced that global unemployment had reached the highest level on record. More than 200 million people, 7 percent of the global workforce, were looking for jobs in 2009.

It was not a coincidence that the global economy experienced the most severe case of unemployment during the worst economic crisis since the Great Depression. Unemployment is highly dependent on economic activity; in fact, growth and unemployment can be thought of as two sides of the same coin: when economic activity is high, more production happens overall, and more people are needed to produce the higher amount of goods and services. And when economic activity is low, firms cut jobs and unemployment rises. In that sense, unemployment is countercyclical, meaning that it rises when economic growth is low and vice versa.

But unemployment does not fall in lockstep with an increase in growth. It is more common for businesses to first try to recover from a downturn by having the same number of employees do more work or turn out more products—that is, to increase their productivity. Only as the recovery takes hold would businesses add workers. As a consequence, unemployment may start to come down only well after an economic recovery begins. In fact, in the last three recessions, the unemployment rate continued to rise after the end of the recessions; a phenomenon called “jobless recoveries.”

The phenomenon works in reverse at the start of a downturn, when firms would rather reduce work hours, or impose some pay cuts before they let workers go. Unemployment starts rising only if the downturn is prolonged. Because unemployment follows growth with a delay, it is called a lagging indicator of economic activity.

How sensitive is the unemployment rate to economic growth? That depends on several factors, most notably labor market conditions and regulations. One estimate for the strength of this relationship for the US economy is Okun’s Law (named after the late economist Arthur Okun), which postulates that a decline in unemployment by 1 percentage point corresponds to a 3 percent rise in output. More recent estimates find that

the consequent rise in output may be lower, possibly between 2 and 3 percent.

How far does this inverse relationship between growth and unemployment go? If economies kept expanding, would one expect to see unemployment disappear? Actually this is not the case (see Chart 1); even in the 2000s when the global economy was prospering (at least until the 2008–09 crisis), global unemployment declined but never reached zero. This observation raises one basic question: Can unemployment ever fall to zero?

Clearing the market

According to classical economic theory, every market, including the labor market, should have a point at which it clears—where supply and demand are equal. (See “Supply and Demand,” p. 12) Yet the very existence of unemployment seems to imply that in labor markets around the world, the demand for and supply of labor fail to reach an equilibrium. Do labor markets continually fail?

Sometimes it is a matter of wages, or the unit price of labor, not adjusting to clear the market. Some workers, particularly skilled ones, may have reservation wages below which they are not willing to work, but which are higher than what employers are willing to pay. Alternatively, the wage an employer is willing to pay may be lower than the legal minimum wage set by governments to try to ensure that wages can sustain a living. When such rigidities in the labor market lead to a shortage of jobs, it is called structural unemployment and those who are structurally unemployed tend to have longer spells of joblessness, on average.

But the inflexibility of wages does not fully explain the perennial nature of unemployment. Some level of unemployment will always exist, for no other reason than there always will be some people who are between jobs, or just starting out their careers. These people are unemployed not because there is a shortage of jobs in the market, but because finding a job takes time. Such temporary spells of unemployment are referred to as frictional unemployment.

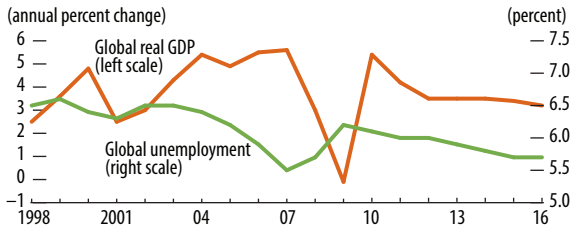
The combination of these factors brings about a long-term average around which the unemployment rate tends to fluctuate, called the

II. HOW ECONOMIES FUNCTION

Chart 1

Jobs and growth

In general, when real global gross domestic product is growing, the unemployment rate declines. The jobless rate generally increases when the world economy is shrinking.



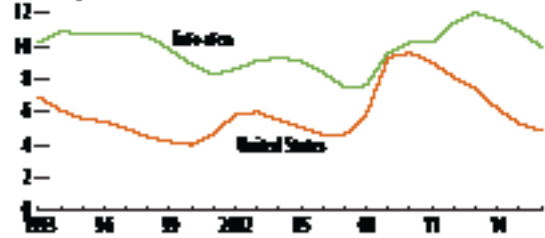
Sources: International Labor Organization; and IMF staff estimates.

Chart 2

Europe's higher unemployment

Europe's labor market rigidities contribute to an unemployment rate in the continent that is generally higher than in the United States.

(unemployment rate, percent)



Sources: Eurostat; and ILO.

natural rate of unemployment (NRU). The term “natural” does not mean it is a constant that cannot be changed; to the contrary, it implies that labor market characteristics, which are mostly driven by policies, determine it. For example, the relatively high rate of unemployment in Europe compared with the United States is in part attributed to Europe’s stronger unions and stricter labor regulations (see Chart 2). These labor market institutions may give European workers a better bargaining position, but they can also render workers too expensive for the employers. In the United States, unionization is lower and labor markets are more flexible, meaning laying off workers are easier, but workers have traditionally enjoyed higher employment rates than their European counterparts.

The natural rate of unemployment is sometimes called the non-accelerating inflation rate of unemployment (NAIRU), because it is consistent with an economy that is growing at its long-term potential so there is no upward or downward pressure on inflation. The flip side of this argument suggests that whenever unemployment temporarily deviates from NAIRU, inflation would be affected. Consider a recession, a period of low economic activity. With lower demand for goods and services, firms would start laying off workers and at the same time not raise their prices as much as they would have. So, one would observe a rise in unemployment and a drop in inflation during recessions (see “Recession,” p. 34). This trade-off between unemployment and inflation—described by the Phillips curve (named after the late economist William Phillips)—is only temporary, though; once prices adjust to a new equilibrium that clears the goods and services market, firms go back to producing at full capacity and unemployment once again falls—to the NAIRU.

Understanding what is behind the long-term equilibrium rate of unemployment helps policymakers understand how they can, and cannot, change it. For example, policies that try to lower unemployment by boosting consumer demand (thereby raising production) can do so only temporarily, and at the cost of higher inflation later. However, policies that are geared toward easing frictional or structural unemployment can boost employment without necessarily affecting inflation.

But NAIRU can also change over time without any explicit policy action: structural changes such as technological advancements or demographic shifts can have long-lasting effects on

unemployment trends. For example, many economists agree that the technology boom of the 1990s increased labor productivity, making each worker more “desirable” to employers, and has therefore reduced the NAIRU—although there was an initial blip of unemployment as workers untrained in using the technologies were displaced. A rapidly aging population—as is occurring in many advanced countries today—is another important factor in reducing the number of people in the job market and bringing down unemployment.

Measuring unemployment

Not all people who don’t work are unemployed. To be considered unemployed for government statistics, a person would not only have to be out of work, but also be actively looking for a job—for example, by sending out résumés. In the United States, unemployment is measured by a monthly survey of households conducted for the Bureau of Labor Statistics and covers a representative sample of more than 100,000 individuals. The labor force includes both those with jobs and those looking for them. The unemployment rate is the percentage of the labor force that is looking for a job. The labor force is only a portion of the total population. The ratio of the labor force to the working age population is called the labor force participation rate.

The labor force excludes people who are of working age but are neither employed nor looking for a job—such as students and homemakers. But the labor force also leaves out jobless people who were in the job market unsuccessfully for so long that they stopped looking for a job. Such discouraged workers are one reason unemployment statistics can underestimate the true demand for jobs in an economy. Another form of hidden unemployment in statistics comes from counting as employed anyone who did any work for pay (or profit, if self-employed) in the week prior to the government survey. This hides the demand for work by people who would have preferred full-time employment, but are working fewer hours only because they could not find full-time jobs. In short, when it comes to measuring unemployment, the devil is well hidden in the details. [FD](#)

CEYDA ONER is a deputy division chief in the IMF’s Finance Department.

Recession: When Bad Times Prevail

It is a sustained period when economic output falls and unemployment rises

Stijn Claessens and M. Ayhan Kose

FOLLOWING THE ONSET of the recent global economic crisis, much of the news, especially in advanced economies, was dire. Unemployment was rising, company profits were falling, financial markets were tumbling, and the housing sector collapsed. Is there a single word to describe these developments? Yes: “recession.”

The crisis was accompanied by recessions in many countries. This pattern is consistent with the historical record. Simultaneous, or synchronized, recessions have occurred in advanced economies several times in the past four decades—the mid-1970s, early 1980s, early 1990s, and early 2000s. Because the United States is the world’s largest economy and has strong trade and financial linkages with many other economies, most of these globally synchronized recession episodes also coincide with US recessions.

Although US recessions had become milder over time, the recent global crisis reversed that trend. The latest episode was one of the longest and deepest recessions since the Great Depression of the 1930s. It led to a sharp increase in unemployment—along with substantial declines in output, consumption and investment.

Calling a recession

There is no official definition of recession, but there is general recognition that the term refers to a period of decline in economic activity. Very short periods of decline are not considered recessions. Most commentators and analysts use, as a practical definition of recession, two consecutive quarters of decline in a country’s real (inflation-adjusted) gross domestic product (GDP)—the value of all goods and services a country produces. Although this definition is a useful rule of thumb, it has drawbacks. A focus on GDP alone is narrow, and it is often better to consider a wider set of measures of economic activity to determine whether a country is indeed suffering a recession. Using other indicators can also provide a timelier gauge of the state of the economy.

In the United States, the private National Bureau of Economic Research (NBER), which maintains a chronology of the beginning and ending dates of US recessions, uses a broader

definition and considers a number of measures of activity to determine the dates of recessions. The NBER’s Business Cycle Dating Committee defines a recession as “a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in production, employment, real income, and other indicators. A recession begins when the economy reaches a peak of activity and ends when the economy reaches its trough.” Consistent with this definition, the Committee focuses on a comprehensive set of measures—including not only GDP, but also employment, income, sales, and industrial production—to analyze the trends in economic activity.

Although an economy can show signs of weakening months before a recession begins, the process of determining whether a country is in a true recession often takes time. For example, it took the NBER committee a year to announce the beginning and end dates of the most recent US recession. This is understandable. The decision process involves establishing a broad decline in economic activity over an extended period of time, after compiling and sifting through many variables, which are often subject to revisions after their initial announcement. In addition, different measures of activity may exhibit conflicting behavior, making it difficult to identify whether the country is indeed suffering from a broad-based decline in economic activity.

Why do recessions happen?

Understanding the sources of recessions has been one of the enduring areas of research in economics. There are a variety of reasons recessions take place. Some are associated with sharp changes in the prices of the inputs used in producing goods and services. For example, a steep increase in oil prices can be a harbinger of a recession. As energy becomes expensive, it pushes up the overall price level, leading to a decline in aggregate demand. A recession can also be triggered by a country’s decision to reduce inflation by employing contractionary monetary or fiscal policies. When used excessively, such policies can lead to a decline in demand for goods and services, eventually resulting in a recession.

II. HOW ECONOMIES FUNCTION

Other recessions, such as the one that began in 2007, are rooted in financial market problems. Sharp increases in asset prices and a speedy expansion of credit often coincide with rapid accumulation of debt. As corporations and households get overextended and face difficulties in meeting their debt obligations, they reduce investment and consumption, which in turn leads to a decrease in economic activity. Not all such credit booms end up in recessions, but when they do, these recessions are often more costly than others. Recessions can be the result of a decline in external demand, especially in countries with strong export sectors. Adverse effects of recessions in large countries—such as Germany, Japan, and the United States—are rapidly felt by their regional trading partners, especially during globally synchronized recessions.

Because recessions have many potential causes, it is a challenge to predict them. The behavioral patterns of numerous economic variables—including credit volume, asset prices, and the unemployment rate—around recessions have been documented, but although they might be the cause of recessions, they could also be the result of recessions—or in economic parlance, endogenous to recessions. Even though economists use a large set of variables to forecast the future behavior of economic activity, none has proven a reliable predictor of whether a recession is going to take place. Changes in some variables—such as asset prices, the unemployment rate, certain interest rates, and consumer confidence—appear to be useful in predicting recessions, but economists still fall short of accurately forecasting a significant fraction of recessions, let alone predicting their severity in terms of duration and amplitude.

Recessions are infrequent but costly

There were 122 completed recessions in 21 advanced economies over the 1960–2007 period. Although this sounds like a lot, recessions do not happen frequently. Indeed, the proportion of time spent in recession—measured by the percentage of quarters a country was in recession over the full sample period—was typically about 10 percent. Although each recession has unique features, recessions often exhibit a number of common characteristics:

- They typically last about a year and often result in a significant output cost. In particular, a recession is usually associated with a decline of 2 percent in GDP. In the case of severe recessions, the typical output cost is close to 5 percent.
- The fall in consumption is often small, but both industrial production and investment register much larger declines than that in GDP.
- They typically overlap with drops in international trade as exports and, especially, imports fall sharply during periods of slowdown.
- The unemployment rate almost always jumps and inflation falls slightly because overall demand for goods and services is

curtailed. Along with the erosion of house and equity values, recessions tend to be associated with turmoil in financial markets.

What about a depression?

The latest US recession—which began in December 2007 and ended in June 2009—was the longest (18 months) and deepest (about a 3.7 percent decline in output) the country has experienced since 1960. The typical US recession prior to 2007 lasted about 11 months and resulted in a peak-to-trough output decline of 1.7 percent. Although investment and industrial

Recession refers to a period of decline in economic activity.

production fell in every recession, consumption registered a decline in only four out of eight episodes since 1960.

One question sometimes asked is how a recession compares with a depression, especially the Great Depression of the 1930s. There is no formal definition of depression, but most analysts consider a depression to be an extremely severe recession, in which the decline in GDP exceeds 10 percent. There have been only a handful of depression episodes in advanced economies since 1960. The most recent was in the early 1990s in Finland, which registered a decline in GDP of about 14 percent. That depression coincided with the breakup of the Soviet Union, a large trading partner of Finland. During the Great Depression, the US economy contracted by about 30 percent over a four-year period. Although the latest recession is obviously severe, its output cost was much smaller than that of the Great Depression. **FD**

STIJN CLAESSENS is head of financial stability policy in the Bank for International Settlements' Monetary and Economic Department and **M. AYHAN KOSE** is director of the Prospects Group of the World Bank's Development Economics Vice Presidency.

References

- Claessens, Stijn, M. Ayhan Kose, and Marco Terrones. 2009. "What Happens During Recessions, Crunches, and Busts?" *Economic Policy*. 60: 653–700. Also IMF Working Paper No. 08/274. Washington.
- , 2011a. "Financial Cycles: What? How? When?" *NBER 2010 International Seminar on Macroeconomics*, Richard Clarida and Francesco Giavazzi, eds. Also IMF Working Paper No. 11/76. Washington.
- , 2011b. "How Do Business and Financial Cycles Interact?" IMF Working Paper No. 11/88. Washington.

Fiscal Policy: Taking and Giving Away

Governments promote stable and sustainable growth through their power to spend and tax

Mark Horton and Asmaa El-Ganainy

FISCAL POLICY influences the economy through government spending and taxation, typically to promote strong and sustainable growth and reduce poverty. The role and objectives of fiscal policy gained prominence during the global economic crisis, when many governments stepped in to support financial systems, jump-start growth, and protect vulnerable groups from the impact of the crisis. In a communiqué following their London summit in April 2009, leaders of the Group of 20 industrial and emerging market economies announced that they were undertaking “unprecedented and concerted fiscal expansion.” What did they mean by fiscal expansion? And, more generally, how can fiscal tools boost the world economy?

How does fiscal policy work?

When policymakers seek to influence the economy, they have two main tools at their disposal—monetary policy and fiscal policy. Central banks target activity indirectly by influencing the money supply through adjustments to interest rates, bank reserve requirements, and purchases and sales of government securities and foreign exchange. Governments influence the economy by changing the level and types of taxes, the extent and composition of spending, and the degree and form of borrowing.

Governments directly and indirectly influence how economic resources are used. An equation of national income accounting that measures an economy’s output—or gross domestic product (GDP)—according to expenditures shows how this happens:

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

On the left side is GDP—the value of final goods and services produced in the economy. On the right side are the sources of aggregate spending and demand—private consumption (*C*), private investment (*I*), purchases of goods and services by the government (*G*), and foreign demand for domestically produced goods—exports minus imports (net exports, *NX*). This equation shows that governments affect economic activity (*GDP*) by controlling *G* directly and influencing *C*, *I*, and *NX* indirectly, through changes in taxes, transfers, spending, and borrowing. Fiscal policy that raises aggregate demand directly through greater government spending is typically called expansionary or “loose.” It is often considered contractionary or “tight” if it reduces demand through lower spending.

Fiscal policy objectives vary. In the short term, governments may focus on macroeconomic *stabilization*—for example, spending more or cutting taxes to stimulate an ailing economy or slashing

spending or raising taxes to rein in inflation or reduce external vulnerabilities. The longer-term aim may be sustainable growth or less poverty through *supply-side* action to improve infrastructure or education. These objectives may be shared broadly across countries, but their relative importance differs with country circumstances. Short-term priorities may reflect the business cycle or response to a natural disaster or global food or fuel price spikes. The longer-term drivers may be development, demographics, or natural resource endowments. Low-income countries might tilt spending toward primary health care in an effort to reduce poverty, whereas advanced economies might favor pension reform to target looming long-term costs related to an aging population. In an oil-producing country, policymakers might gear fiscal policy toward broader macroeconomic developments by moderating procyclical spending—both by limiting bursts of spending when oil prices rise and by refraining from painful cuts when they drop.

Response to the global financial crisis

The global financial crisis that had its roots in the 2007 meltdown in the US mortgage market is a good case study in fiscal policy. The crisis hurt economies around the globe. Financial sector difficulties and flagging confidence hit private consumption, investment, and international trade (all of which affect output, GDP). Governments tried to boost activity through two channels: automatic stabilizers and fiscal stimulus—that is, new discretionary spending or tax cuts. *Stabilizers* kick in as tax revenue and expenditures change and do not depend on government action. They have to do with the business cycle. For instance, as output slows or falls, the amount of taxes collected declines because corporate profits and taxpayers’ incomes are lower, particularly when progressive tax structures place higher-income earners in higher tax brackets. Unemployment benefits and other social spending are also designed to rise during a downturn. These cyclical changes make fiscal policy automatically expansionary during downturns and contractionary during upturns.

Automatic stabilizers are linked to the size of the government and tend to be larger in advanced economies. Where stabilizers are larger, there may be less need for stimulus—tax cuts, subsidies, or public works programs—since both approaches help soften the effects of a downturn. In addition, although discretionary measures can be tailored to stabilization needs, automatic stabilizers are not subject to implementation lags, as discretionary

II. HOW ECONOMIES FUNCTION

measures often are. (It can take time, for example, to design, get approval for, and implement new road projects.) Moreover, automatic stabilizers—and their effects—wind down on their own as conditions improve.

Stimulus may be difficult to design and implement effectively and to reverse when conditions pick up. In many low-income and emerging market economies, however, institutional limits and narrow tax bases mean that stabilizers are relatively weak. Even in countries with larger stabilizers, there may be a pressing need to compensate for the loss of economic activity and to target the government's crisis response to those most directly in need.

Fiscal ability to respond

The exact response ultimately depends on the fiscal space for new spending or tax cuts—that is, how much access the government has to additional financing at a reasonable cost or its ability to reorder its existing expenditures. Some governments were not able to respond with stimulus during the crisis because potential creditors believed additional spending and borrowing would put too much pressure on inflation, foreign exchange reserves, or the exchange rate—or delay recovery by taking away resources from the local private sector (also known as crowding out). Creditors may have doubted some governments' ability to spend wisely, reverse stimulus once in place, or address long-standing structural weaknesses in public finances (chronically low tax revenue because of poor tax structure or evasion, weak control over local government or state-owned-enterprise finances, rising health costs, aging populations). Sometimes, severe financing constraints have forced governments to cut spending in the face of declining revenue (functioning stabilizers). If inflation is high or there is an external current account deficit, fiscal stimulus could be ineffective and even undesirable.

The size, timing, composition, and duration of stimulus matter. Policymakers generally aim to tailor measures to their estimates of the output gap—the difference between expected output and output of the economy at full capacity. A measure of the effectiveness of the stimulus—or, more precisely, how it affects output growth (also known as the multiplier)—is also needed. Multipliers tend to be larger if there is less leakage (for example, only a small part of the stimulus is saved or spent on imports), monetary conditions are accommodative, and the country's fiscal position after the stimulus is considered sustainable. Multipliers can be small or even negative if the expansion raises concern about immediate or longer-term sustainability. In that case the private sector would likely counteract government intervention by saving more or even moving money abroad instead of investing or consuming. Multipliers tend to be higher for spending measures than for tax cuts or transfers and lower for small open economies (in both cases, because of the extent of leakage).

Governments face a trade-off between targeting stimulus to the poor, which tends to yield full spending and a strong economic effect; funding capital investment, which may create

jobs and help bolster longer-term growth; and tax cuts, which may encourage hiring or new capital equipment purchases. In practice, governments have taken a balanced approach that includes measures in all these areas.

As for timing, it can take a while to implement spending measures (program design, procurement, execution), and once in place the measures may outlive the need. However, if the downturn is expected to be prolonged (like the recent crisis), concern over lags may be less pressing: some governments stressed implementation of shovel-ready projects that were vetted and ready to go. This is why stimulus measures should be timely, targeted, and temporary—quickly reversed once conditions improve.

Similarly, the responsiveness and scope of stabilizers can be enhanced—for instance, by a more progressive system that taxes wealthy households at a higher rate than lower-income households. Transfer payments can also be explicitly linked to economic conditions (for instance, unemployment rates or other labor market triggers). In some countries, fiscal rules aim to limit the growth of spending during boom times, when revenue growth—particularly from natural resources—is high and constraints seem less binding. Elsewhere, formal review or expiration (sunset) mechanisms help ensure that new initiatives do not outlive their purpose. Finally, medium-term fiscal frameworks with comprehensive coverage and assessment of revenue, expenditures, assets and liabilities, and risks help improve policymaking over the business cycle.

Big deficits and rising public debt

Fiscal deficits and public-debt-to-GDP ratios have expanded sharply in many countries because of the effects of the crisis on GDP and tax revenue and the cost of the fiscal response. Support and guarantees for the financial and industrial sectors have added to concerns about the financial health of governments. Many countries can afford a moderate fiscal deficit for an extended period when domestic and international financial markets and international and bilateral partners are confident that these economies can meet present and future obligations. But deficits that grow too large and linger too long may undermine that confidence. Aware of these risks, the IMF in late 2008 and early 2009 called on governments to establish a four-pronged fiscal policy strategy to help ensure solvency: stimulus should not permanently affect deficits; medium-term frameworks should include commitment to fiscal correction once conditions improve; structural reforms should be identified and implemented to enhance growth; and countries facing medium- and long-term demographic pressures should commit to health care and pension reform. Even though the worst effects of the crisis are behind us, fiscal challenges remain, particularly in advanced economies in Europe and North America, and this strategy is as valid as ever. [FD](#)

MARK HORTON is an assistant director in the IMF's Middle East and Central Asia Department, and **ASMAA EL-GANAINY** is a senior economist in the IMF's Institute for Capacity Development.

Externalities: Prices Do Not Capture All Costs

There are differences between private returns or costs and the costs or returns to society as a whole

Thomas Helbling

CONSUMPTION, PRODUCTION, AND INVESTMENT decisions of individuals, households, and firms often affect people not directly involved in the transactions. Sometimes these indirect effects are tiny. But when they are large they can become problematic—what economists call externalities. Externalities are among the main reasons governments intervene in the economic sphere.

Most externalities fall into the category of so-called technical externalities; that is, the indirect effects have an impact on the consumption and production opportunities of others, but the price of the product does not take those externalities into account. As a result, there are differences between private returns or costs and the returns or costs to society as a whole.

Negative and positive externalities

In the case of pollution—the traditional example of a negative externality—a polluter makes decisions based only on the direct cost of and profit opportunity from production and does not consider the indirect costs to those harmed by the pollution. The social—that is, total—costs of production are larger than the private costs. Those indirect costs—which are not borne by the producer or user—include decreased quality of life, say in the case of a home owner near a smokestack; higher health care costs; and forgone production opportunities, for example when pollution harms activities such as tourism. In short, when externalities are negative, private costs are lower than social costs.

There are also positive externalities, and here the issue is the difference between private and social gains. For example, research and development (R&D) activities are widely considered to have positive effects beyond those enjoyed by the producer—typically, the company that funds the research. This is because R&D adds to the general body of knowledge, which contributes to other discoveries and developments. However, the private returns of a firm selling products based on its own R&D typically do not include the returns of others who benefited indirectly. With positive externalities, private returns are smaller than social returns.

When there are differences between private and social costs or private and social returns, the main problem is that market outcomes may not be efficient. To promote the well-being of all members of society, social returns should be maximized and social costs minimized. Unless all costs and benefits are

internalized by households and firms making buying and production decisions, market outcomes can lead to underproduction or overproduction in terms of a society's overall condition (what economists call the “welfare perspective”).

Consider again the example of pollution. Social costs grow with the level of pollution, which increases as production increases, so goods with negative externalities are overproduced when only private costs are involved and not costs incurred by others. To minimize social costs would lead to lower production levels. Similarly, from a societal perspective, maximization of private instead of social returns leads to underproduction of the good or service with positive externalities.

Taxation and externalities

Neoclassical economists recognized that the inefficiencies associated with technical externalities constitute a form of “market failure.” Private market-based decision making fails to yield efficient outcomes from a general welfare perspective. These economists recommended government intervention to correct for the effects of externalities. In *The Economics of Welfare*, British economist Arthur Pigou suggested in 1920 that governments tax polluters an amount equivalent to the cost of the harm to others. Such a tax would yield the market outcome that would have prevailed with adequate internalization of all costs by polluters. By the same logic, governments should subsidize those who generate positive externalities, in the amount that others benefit.

The proposition that technical externalities require government regulation and taxation to prevent less than optimal market outcomes was intensely debated after Pigou's seminal work. Some economists argued that market mechanisms can correct for the externalities and provide for efficient outcomes. People can resolve the problems through mutually beneficial transactions. For example, a landlord and a polluter can enter into a contract under which the landlord agrees to pay the polluter a certain amount of money in exchange for a specific reduction in the amount of pollution. Such contractual bargaining can be mutually beneficial. Once the building is less exposed to pollution, the landlord can raise rents. As long as the increase in rents is greater than the payment to the polluter, the outcome is beneficial for the landlord. Similarly, as long as the payment

II. HOW ECONOMIES FUNCTION

exceeds the loss in profit from lower pollution (lower production), the polluting firm is better off as well.

The possibility of overcoming the inefficiencies from externalities through bargaining among affected parties was first discussed in 1960 by Ronald Coase in “The Problem of Social Cost” (among the works that earned him a Nobel Prize in economics in 1991). For bargaining solutions to be feasible, property rights must be well defined, bargaining transaction costs must be low, and there must be no uncertainty or asymmetric information, when one actor knows more than the other about the transaction.

Against this backdrop, optimal government intervention might be the establishment of institutional frameworks that allow for proper bargaining among parties involved in externalities. Property rights—specifically intellectual property rights, such as patents—allow a firm to earn most if not all the returns from its R&D. But it is easier to assign property rights for innovations and inventions. When it comes to basic or general research, property rights are more difficult to define, and government subsidies typically are needed to ensure a sufficient amount of basic research.

Public goods

Problems in defining property rights are often a fundamental obstacle to market-based, self-correcting solutions, because the indirect effects of production or consumption activity can affect so-called public goods, which are a special kind of externality. These goods are both nonexcludable—whoever produces or maintains the public good, even at a cost, cannot prevent other people from enjoying its benefits—and nonrival—consumption by one individual does not reduce the opportunity for others to consume it (Cornes and Sandler 1986). If the private benefits are small relative to the social benefit but private costs to provide them are large, public goods may not be supplied at all. The importance of the public good problem has long been recognized in the field of public finance. Taxes often finance governments’ delivery of public goods, such as law and order (Samuelson 1955).

The public good problem is especially notable in environmental economics, which largely deals with analyzing and finding solutions to externality-related issues. Clean air, clean water, biodiversity, and a sustainable stock of fish in the open sea are largely nonrival and nonexcludable goods. They are free goods, produced by nature and available to everybody. They are subject to no well-defined property rights. As a result, households and firms do not place enough value on these public goods, and efficient market outcomes through bargaining typically are not feasible. In other words, environmental issues often face a collective action problem.

High transaction costs and problems related to uncertainty are other obstacles that prevent parties involved in technical externalities from internalizing costs and benefits through bargaining solutions. Uncertainty problems are far reaching. In fact, the well-known moral hazard is a form of externality in which decision makers maximize their benefits while inflicting damage on others but do not bear the consequences because, for example,

there is uncertainty or incomplete information about who is responsible for damages or contract restrictions. An often-used example is a situation in which an insured entity can affect its insurance company’s liabilities but the insurance company is not in a position to determine whether the insured is responsible for an event that triggers a payout. Similarly, if a polluter’s promised preventive actions cannot be verified because of a lack of information, bargaining is unlikely to be a feasible solution.

Today, the most pressing and complex externality problem is greenhouse gas (GHG) emissions. The atmospheric accumulation of greenhouse gases from human activity has been identified as a major cause of global warming. Barring policies to curb GHG emissions, scientists expect this problem to grow and eventually lead to climate change and its accompanying costs, including damage to economic activity from the destruction of capital (for example, along coastal areas) and lower agricultural productivity. Externalities come into play because the costs and risks from climate change are borne by the world at large, whereas there are few mechanisms to compel those who benefit from GHG-emitting activity to internalize these costs and risks.

The atmosphere, in fact, is a global public good, with benefits that accrue to all, making private bargaining solutions unfeasible. Identifying and agreeing on policies for internalization of the social costs of GHG emissions at the global level are extremely difficult, given the cost to some individuals and firms and the difficulties of global enforcement of such policies (Tirole 2008). Indeed, in the Paris Climate Accord adopted in 2015, member countries of the United Nations Framework on Climate Change agreed on national targets for greenhouse gas emission reductions without any enforcement of those targets or commitment to measures to support them.

Externalities pose fundamental economic policy problems when individuals, households, and firms do not internalize the indirect costs of or the benefits from their economic transactions. The resulting wedges between social and private costs or returns lead to inefficient market outcomes. In some circumstances, they may prevent markets from emerging. Although there is room for market-based corrective solutions, government intervention is often required to ensure that benefits and costs are fully internalized. **FD**

THOMAS HELBLING is an advisor in the IMF’s Research Department.

References:

- Coase, Ronald. 1960. “The Problem of Social Cost.” *Journal of Law and Economics*. 3 (1): 1–44.
- Cornes, Richard, and Todd Sandler. 1986. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge, United Kingdom: Cambridge University Press.
- Pigou, Arthur C. 1920. *The Economics of Welfare*. London: Macmillan.
- Samuelson, Paul A. 1955. “Diagrammatic Exposition of a Theory of Public Expenditure.” *The Review of Economics and Statistics*. 37 (4): 350–56.
- Tirole, Jean. 2008. “Some Economics of Global Warming.” *Rivista di Politica Economica*. 98 (6): 9–42.

International Trade: Commerce among Nations

Nations are almost always better off when they buy and sell from one another

Brad McDonald

IF THERE IS a point on which most economists agree, it is that trade among nations makes the world better off. Yet international trade can be a contentious political issue, both domestically and between governments.

When a firm or an individual buys a good or a service produced more cheaply abroad, living standards in both countries rise. There are other reasons consumers and firms buy abroad that also make them better off—the product may better fit their needs than similar domestic offerings or it may not be available domestically. The foreign producer also benefits from more sales than it could make solely in its own market and by earning foreign exchange (currency) for purchases of foreign-made products.

Still, not every individual or company is better off. When a firm buys a foreign product because it is cheaper, it benefits—but the (more costly) domestic producer loses a sale. Usually, however, the buyer gains more than the domestic seller loses. Except when the costs of production do not include such social costs as pollution, the world is better off when countries import things that are produced more efficiently abroad.

Those who perceive themselves to be harmed by foreign competition have long opposed international trade. Soon after economists such as Adam Smith and David Ricardo established the economic basis for free trade, British historian Thomas B. Macaulay observed the practical problems governments face: “Free trade, one of the greatest blessings which a government can confer on a people, is in almost every country unpopular.”

Two centuries later trade debates still resonate.

Why countries trade

Ricardo observed that trade was driven by *comparative* rather than *absolute* costs (of producing a good). One country may be more productive than others in all goods, in the sense that it can produce any good using fewer inputs (such as capital and labor) than other countries require to produce the same good. Ricardo’s insight was that such a country would still benefit from trading according to its *comparative advantage*—exporting products in which its absolute advantage was greatest, and importing products in which its absolute advantage was comparatively less (even if still positive).

A country may be twice as productive as its trading partners in making clothing, but if it is three times as productive in making steel or building airplanes, it will benefit from making and exporting these products and importing clothes. Its partner will gain by exporting clothes—in which it has a comparative but not absolute advantage—in exchange for these other products (see box). The notion extends beyond physical goods to trade in services—such as writing computer code or providing financial products.

Because of comparative advantage, trade raises the living standards of both countries.

Differences in comparative advantage may arise for several reasons. In the early 20th century, Swedish economists Eli Heckscher and Bertil Ohlin identified the role of labor and capital, so-called factor endowments, as a determinant. The Heckscher-Ohlin proposition maintains that countries tend to export goods whose production uses intensively the factor of production that is relatively abundant in the country. Countries well

COMPARATIVE ADVANTAGE

Even a country that is more efficient (has absolute advantage) in everything it makes would benefit from trade. Consider an example:

Country A: One hour of labor can produce either three kilograms of steel or two shirts. Country B: One hour of labor can produce either one kilogram of steel or one shirt.

Country A is more efficient in both products. Now suppose Country B offers to sell Country A two shirts in exchange for 2.5 kilograms of steel.

To produce these additional two shirts, Country B diverts two hours of work from producing (two kilograms) steel. Country A diverts one hour of work from producing (two) shirts. It uses that hour of work to instead produce three additional kilograms of steel.

Overall, the same number of shirts is produced: Country A produces two fewer shirts, but Country B produces two additional shirts. However, more steel is now produced than before: Country A produces three additional kilograms of steel, while Country B reduces its steel output by two kilograms. The extra kilogram of steel is a measure of the gains from trade.

II. HOW ECONOMIES FUNCTION

endowed with capital—such as factories and machinery—should export capital-intensive products; those well endowed with labor should export labor-intensive products. Economists today think that there are also other important influences on trade patterns (Baldwin 2008).

Recent research finds that episodes of trade opening are followed by adjustment not only *across* industries, but *within* them as well. Greater competition from foreign firms puts pressure on profits, forcing less efficient firms to contract and making room for more efficient firms. Expansion and new entry bring better technologies and new product varieties. Likely most important, trade enables greater selection across different types of goods (say refrigerators). This explains why there is a lot of intra-industry trade (for example, countries that export household refrigerators may import industrial coolers), which the factor endowment approach ignores.

There are clear efficiency benefits from trade that results in *more* products—not only more of the same products, but greater product variety. For example, the United States imports four times as many varieties (such as different types of cars) as it did in the 1970s, while the number of countries supplying each good has doubled. Even more beneficial may be the more efficient investment spending when firms have access to a wider variety and quality of intermediate and capital inputs (think industrial optical lenses rather than cars). By enhancing overall investment and facilitating innovation, trade can bring sustained higher growth.

Economic models that assess the impact of trade typically neglect technology transfer and pro-competitive forces such as greater product variety because these are difficult to model, and results that do incorporate them are subject to greater uncertainty. Where this has been done, however, researchers have concluded that the benefits of trade reforms—such as reducing tariffs and other nontariff barriers to trade—are much larger than suggested by conventional models.

Why trade reform is difficult

Trade contributes to global efficiency. When a country opens up to trade, capital and labor shift toward industries in which they are used more efficiently, and society benefits. But there is more to the story.

Trade also brings dislocation to those firms and industries that cannot cut it. Such firms often seek barriers such as import taxes (called tariffs) and quotas to raise the price or limit the availability of imports. Processors may try to restrict the exportation of raw materials to depress artificially the price of their own inputs. The benefits of trade, though, are not always recognized by those who are helped, and opponents often argue more effectively.

Reforms since World War II have reduced government-imposed trade barriers. But policies to protect domestic industries vary. Tariffs are much higher in certain sectors (such as

agriculture and clothing) and among certain country groups (such as less developed countries) than in others. Many countries discourage trade in services in areas such as transportation, communications, and, often, the financial sector; others welcome foreign competition.

Moreover, trade barriers affect some countries more than others. Often hardest hit are less developed countries, whose exports are concentrated in low-skill, labor-intensive products that industrialized countries often protect. The United States, for example, is reported to collect about 15 cents in tariffs for each \$1 of imports from Bangladesh (Elliott 2009), but only 1 cent for each \$1 of imports from some major western European countries. Yet imports of a particular product from Bangladesh face the same or lower tariffs than similarly classified products imported from western Europe. Although the tariffs on Bangladesh items in the United States may be a dramatic example, World Bank economists calculated that exporters from low-income countries face barriers on average 50 percent higher than those on exports from major industrialized countries (Kee, Nicita, and Olarreaga 2006).

The World Trade Organization (WTO) referees international trade. Agreements since 1948 by its 153 members (of the WTO and its predecessor General Agreement on Trade and Tariffs) promote nondiscrimination and facilitate further liberalization in nearly all areas of commerce, including tariffs, subsidies, customs valuation and procedures, trade and investment in service sectors, and intellectual property (IMF, World Bank, and WTO 2017). These agreements are enforced through a powerful and carefully crafted dispute settlement process.

Under the WTO rules-based international trading system, trade policies are more stable, more transparent, and more open. And the WTO is a key reason why the global financial crisis did not spark widespread protectionism. Restrictive and discriminatory trade policies remain common. Addressing them could yield hundreds of billions of dollars in annual global benefits. But narrow interests seek to delay and dilute further reform. A focus on the greater good and on helping those few who will be harmed can help deliver a fairer and more sensible trading system. **FD**

BRAD MCDONALD is a deputy division chief in the IMF's Strategy, Policy, and Review Department.

References:

- Baldwin Robert E. 2008. *The Development and Testing of Heckscher-Ohlin Trade Models: A Review*. Cambridge, MA: MIT Press.
- Elliott, Kimberley Ann. 2009. "Opening Markets for Poor Countries: Are We There Yet?" Center of Global Development Working Paper 184, Washington.
- IMF, World Bank, and World Trade Organization. 2017. "Making Trade an Engine of Growth for All: The Case for Trade and for Policies to Facilitate Adjustment."
- Kee, Hiau Looi, Alessandro Nicita, and Marcelo Olarreaga. 2006. "Estimating Trade Restrictiveness Indices," World Bank Policy Research Working Paper No. 3840, Washington.

What Are Real Exchange Rates?

What is the value of a country's goods against those of another country, a group of countries, or the rest of the world, at the prevailing exchange rate?

Luis A. V. Catão

HOW DOES ONE DETERMINE whether a currency is fundamentally undervalued or overvalued? This question lies at the core of international economics and many trade disputes.

George Soros had the answer once—in a famous episode back in 1992—when he successfully bet \$1 billion against the British pound, in what turned out to be the beginning of a new era in large-scale currency speculation. Under assault by Soros and other speculators, who believed that the pound was overvalued, the British currency crashed, in turn forcing the United Kingdom's dramatic exit from the European Exchange Rate Mechanism, the precursor to the common European currency, the euro. The United Kingdom never adopted the common currency nor has it since attempted to peg its currency.

But in the ensuing years, neither Soros nor fellow speculators have repeated the feat consistently. Indeed, there is some consensus that the economics profession itself lacks a foolproof method of establishing when a currency is properly valued. This failure is striking given that the exchange rate is a central price in economics and that there is a measure potentially capable of delivering the answer and for which plenty of data exist: the real exchange rate (RER).

What things really cost

Most people are familiar with the nominal exchange rate, the price of one currency in terms of another. It's usually expressed as the domestic price of the foreign currency. So if it costs a US dollar holder \$1.18 to buy one euro, from a euro holder's perspective the nominal rate is €0.85 per dollar (that is, $1/1.18$). But the nominal exchange rate isn't the whole story. The person or firm buying another currency is interested in what can be bought with it. Are they better off with dollars or euros? That's where the RER comes in. It seeks to measure the value of a country's goods against those of another country, a group of countries, or the rest of the world at the prevailing nominal exchange rate.

The real exchange rate between two countries can be measured in terms of a single representative good—say a Big Mac, the McDonald's burger sold in many countries in virtually identical versions. If the real exchange rate for that good is 1, the burger would cost the same in the United States as in, say, Germany, when the price is expressed in a common currency. That would be the case if the Big Mac costs \$5.30 in the United States and €4.50 in Germany. In this one-product world (in which prices equal exchange rates) the purchasing power parity of the dollar and the euro is the same, and the RER is 1 (see box).

But suppose the burger sells for €5.40 in Germany. That would mean it costs 20 percent more in the euro area than in the United

States, suggesting that the euro is thus 20 percent overvalued relative to the dollar. If the real exchange rate becomes that overvalued, there should be pressure on the nominal exchange rate to adjust, because the same good can be purchased more cheaply in one country than in the other. Indeed, it would make economic sense to buy dollars, use them to buy Big Macs in the United States at the equivalent of about 1 euro, and sell them in Germany for 1.2 euros. Taking advantage of such price differentials is called arbitrage. As arbitrageurs buy dollars to purchase Big Macs to sell in Germany, demand for dollars rises, as does the nominal exchange rate, until the price in Germany and the United States is the same—the RER returns to 1. In the real world there are many costs that get in the way of a straight price comparison—such as transportation, trade barriers, and consumption preferences.

But the fundamental notion is that when RERs diverge, the currencies face pressure to change. For overvalued currencies the pressure is to depreciate and for undervalued currencies to appreciate. It can get more complicated if factors such as government policies hinder normal equilibration of exchange rates, often an issue in trade disputes.

Many products

How about comparing purchasing power when countries sell more than one product? To do this, economists usually measure the real exchange rate in terms of a broad basket of goods. Because the price of such a basket normally takes the form of an index number—such as the consumer price index, which includes both goods and services—the RER is also typically expressed as an index that can be benchmarked to any chosen time period. Going back to the dollar-euro example, if the RER index is 1.2, average consumer prices in Europe are 20 percent higher than in the United States, relative to the chosen benchmark. Indices don't measure absolute prices (such as the price of the Big Mac), but changes in overall prices relative to a base year. (If, say, the index is 100 in the year 2000 and 120 in 2017, average prices are 20 percent higher in 2017 than in 2000.)

RER indices between two countries can be important. The massive US trade deficit with China has become a political and economic issue, and whether its roots are in a fundamentally misaligned exchange rate is a point of contention.

But, for the most part, economists and policymakers are more interested in the real effective exchange rate (REER) when measuring a currency's overall alignment. The REER is an average of the bilateral RERs between the country and each of its trading partners, weighted by the respective trade shares

II. HOW ECONOMIES FUNCTION

of each partner. Because it is an average, a country's REER may be in "equilibrium" (display no overall misalignment) when its currency is overvalued relative to that of one or more trading partners so long as it is undervalued relative to others.

To establish when a currency is misvalued and, if so, by how much, a rough assessment can be obtained by the REER series over time. As with the absolute and relative RERs, there should be no changes if the currencies are in equilibrium. But because consumption patterns can change faster than the market baskets statisticians construct—as can trade policies and tariffs and transportation costs—deviations in REERs don't necessarily indicate fundamental misalignment.

One complication is that REERs' fluctuations have intensified, even though transportation costs and tariffs have declined sharply over the past century as a proportion of the final price of goods, and national consumption baskets have grown more uniform. That is, variations in tariffs and transportation costs don't shift goods prices as much as in the past, yet REERs have been moving about by quite a bit over the past three or four decades. Indeed, between the late 19th century and the 1929 financial crash, when changes in transportation costs and tariffs were comparatively large, REER fluctuations were within a 30 percent band among advanced economies (once price spikes from war-related disturbances are excluded). In contrast, in the 1980s, the United States experienced swings as wide as 80 percent, and some advanced economies also experienced variations in their REERs above 30 percent over the past two decades. Some emerging market and developing economies have experienced even larger depreciations, especially during the financial crises in the 1990s and 2000s.

Other things at work

But not all large REER fluctuations are indications of misalignment. Some are remarkably smooth, suggesting that factors besides transportation costs, tastes, and tariffs play a key role in influencing the REER of a currency that is not misaligned.

Technology changes that cause productivity increases in goods commonly traded between countries, called tradables, are thought to be one of those factors. Because productivity increases lead to lower production costs, the REERs would rise to maintain equilibrium. But not all goods in a given market basket are tradables and subject to international competition. Nontradables, such as houses and many personal services, face minimal international price competition. While prices of tradables should tend to equalize across countries in the absence of trade barriers or currency controls, prices of nontradables can differ widely. Economic theory suggests, and data support, that much of the REER variation across countries is accounted for by fluctuations in nontradables prices.

Persistent changes in terms of trade (such as oil producers usually experience) and differences in fiscal policies, tariffs, and even financial development can also help explain why REERs can differ across countries. The IMF and other analysts take such real exchange rate fundamentals into account in estimating the

WHAT IS THE REAL EXCHANGE RATE?

The real exchange rate (RER) between two currencies is the nominal exchange rate (e) multiplied by the ratio of prices between the two countries, P/P^* . The RER therefore is eP^*/P . Consider the case of Germany relative to the United States. Those variables can be defined so that a rise in the RER denotes appreciation (as the IMF typically does) or depreciation (as many economics textbooks do). It is just a matter of convention. Let's define the RER so that a rise denotes appreciation of Germany's real exchange rate. In this case, e is the dollar-euro exchange rate, P the average price of goods in Germany, and P^* the average price of goods in the United States.

In the Big Mac example, we have a price (averaged across McDonald's restaurants) in Germany of about €4.50 and an average price in the United States of about \$5.30 (both as of July 2017). For the Big Mac in Germany, take a dollar-euro exchange rate of 1.18. The RER is then $1.18 \times 4.5/5.3$, which equals 1. So, at the current dollar-euro exchange rate, the euro appears neither undervalued nor overvalued relative to the dollar.

Now consider the RER between China and the United States. One US dollar buys ¥6.8, so the dollar-yuan exchange rate is $1/6.8$, or 0.147. With an average Big Mac price in China of about ¥20, the cost in US dollars of the burger in China (that is, $e \times P$) is $e \times P = 20 \times 0.147 = 2.94$. Since the price of the Big Mac is \$5.30, the RER of the yuan to the US dollar is $2.94/5.3 = 0.55$. This is less than 1, which indicates undervaluation of the yuan. The RER is short of 1 by 45 cents, which means that, by this metric, the yuan is 45 percent undervalued relative to the dollar.

"equilibrium" REER, around which the actual REER should hover if there is no misalignment.

Estimating equilibrium RERs can be difficult because prices are somewhat sticky in the short run, and the nominal exchange rate is not (in countries whose exchange rates are market determined). So REERs typically display considerable short-term volatility in response to news and noisy trading, and it's not surprising that many market participants and policymakers get things wrong—sometimes very wrong. That can lead to massive realignments with devastating consequences. A widely touted reason for the enhanced volatility of exchange rates is much greater liquidity and lower transaction costs in foreign exchange markets since the mid-1980s compared with earlier periods. Because the exchange rate measures the relative price between two currencies but is also an asset price—a foreign exchange rate can be held as a store of value and a vehicle for speculation—swings in financial market sentiment may affect its value. This makes it even harder to use RER computations such as illustrated above—based on the relative prices of two goods or baskets of goods—to gauge misvaluation. Financial market considerations and portfolio decisions by investors based on other considerations can make a difference, so it is necessary to assess the rationale and sustainability of those decisions when assessing misvaluation.

Even though they are imperfect, large variations in REERs have signaled large exchange rate overvaluations that have helped predict many financial crises and that explain many trade imbalances between countries. This is why the IMF and others should closely monitor bilateral real exchange rate and multilateral real effective exchange rate indicators. [FD](#)

LUIS A. V. CATÃO is a senior economist in the IMF's Research Department.

Purchasing Power Parity: Weights Matter

At what rate would the currency of one country have to be converted into that of another to buy the same goods and services in each country?

Tim Callen

HOW FAST IS THE GLOBAL ECONOMY growing? Is China contributing more to global growth than the United States? Is the average person richer in France or in Japan? These types of questions are of great interest to economists and others, and at first blush it appears reasonable to assume that each has a clear-cut answer. But, as with many things in economics, the reality is different.

To answer the questions, one must compare the value of the output from different countries. But each country reports its data in its own currency. That means that to compare the data, each country's statistics must be converted into a common currency. However, there are several ways to do that conversion and each can give a markedly different answer.

Two different yardsticks

International financial institutions produce a wide range of regional and global statistics. The IMF, one of these institutions, publishes many of its statistics—such as real GDP growth, inflation, and current account balances—twice a year in its *World Economic Outlook* (WEO). These statistics combine, or aggregate, the results from many countries into an average. The importance, or weight, of an individual country's data in the overall result depends on the size of its economy relative to the others being compared. To derive these weights, one converts the GDP of a country in national currency terms to a common currency (in practice, the US dollar).

One of the two main methods of conversion uses market exchange rates—the rate prevailing in the foreign exchange market (using either the rate at the end of the period or an average over the period). The other uses the purchasing power parity (PPP) exchange rate—the rate at which the currency of one country would have to be converted into that of another country to buy the same amount of goods and services in each country.

To understand PPP, let's take a commonly used example, the price of a hamburger. If a hamburger is selling in London for £2 and in New York for \$4, this would imply a PPP exchange rate

of 1 pound to 2 US dollars. This PPP exchange rate may well be different from that prevailing in financial markets (so that the actual dollar cost of a hamburger in London may be either more or less than the \$4 it sells for in New York). This type of cross-country comparison is the basis for the well-known “Big Mac” index, which is published by *The Economist* magazine and calculates PPP exchange rates based on the McDonald's sandwich that sells in nearly identical form in many countries around the world.

Of course, any meaningful comparison of prices across countries must consider a wide range of goods and services. This is not an easy task, because of the amount of data that must be collected and the complexities in the comparison process. To facilitate price comparisons across countries, the International Comparisons Program (ICP) was established by the United Nations and the University of Pennsylvania in 1968. PPPs generated by the ICP are based on a global survey of prices. In the last round (2011), each of the participating countries (about 199) provided national average prices for about 1,000 closely specified products.

PPP versus market rates

So which method is better? The appropriate way to aggregate economic data across countries depends on the issue being considered. **Market exchange rates are the logical choice when financial flows are involved.** For example, the current account balance represents a flow of financial resources across countries. It is appropriate to use the market exchange rate to convert these flows into dollars when aggregating across regions or calculating the global current account discrepancy. But **for other variables, the decision is less clear-cut.** Take real GDP growth. International organizations use different approaches. The World Bank uses market-based rates to determine the weights in its regional and global aggregations of real GDP, whereas the IMF and the Organization for Economic Cooperation and Development use weights based on PPP

II. HOW ECONOMIES FUNCTION

rates (although the IMF also publishes a global growth aggregate based on market rates in the WEO). Each methodology has its advantages and disadvantages.

Advantages of PPP. A main one is that PPP exchange rates are relatively stable over time. By contrast, market rates are more volatile, and using them could produce quite large swings in aggregate measures of growth even when growth rates in individual countries are stable. Another drawback of market-based rates is that they are relevant only for internationally traded goods. Nontraded goods and services tend to be cheaper in low-income than in high-income countries. A haircut in New York is more expensive than in Dhaka; the price of a taxi ride of the same distance is higher in Paris than in Jakarta; and a ticket to a cricket game costs more in London than in Lahore. Indeed, because wages tend to be lower in poorer countries, and services are often relatively labor intensive, the price of a haircut in Lima is likely to be cheaper than in New York even when the

PPP is generally regarded as a better measure of overall well-being.

cost of making tradable goods, such as machinery, is the same in both countries. Any analysis that fails to take into account these differences in the prices of nontraded goods across countries will underestimate the purchasing power of consumers in emerging market and developing countries and, consequently, their overall welfare. For this reason, PPP is generally regarded as a better measure of overall well-being.

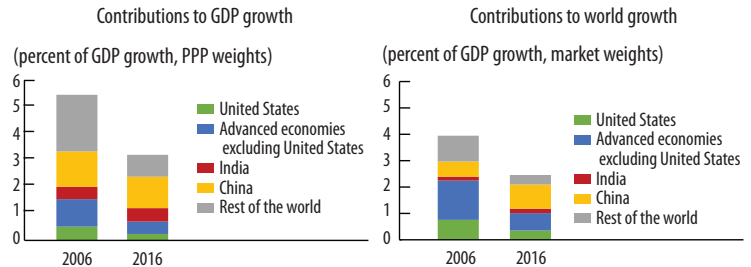
Drawbacks of PPP. The biggest one is that PPP is harder to measure than market-based rates. The ICP is a huge statistical undertaking, and new price comparisons are available only at infrequent intervals. Methodological questions have also been raised about earlier surveys. In between survey dates, the PPP rates have to be estimated, which can introduce inaccuracies into the measurement. Also, the ICP does not cover all countries, which means that data for missing countries have to be estimated.

Does it make a difference?

It depends. There is a large gap between market and PPP-based rates in emerging market and developing economies. But for advanced countries, the market and PPP rates tend to be much closer. As a result, developing economies get a much higher

Who contributes most to global growth?

Contributions to world growth have changed significantly over the past 10 years.



Source: IMF, *World Economic Outlook*, September 2017.

weight in aggregations that use PPP exchange rates than they do using market exchange rates. China's weight in the global economy is about 18 percent using PPP exchange rates, but about 15 percent with market-based weights. For India, the figures are about 7 percent and 3 percent, respectively.

Thus, the choice of exchange rates makes a big difference in calculations of global growth, but little difference to estimates of aggregate growth in advanced economies. The per capita income gap between the richest and poorest countries is modestly reduced under PPP exchange rates (although it remains exceptionally large), and some countries jump up or down the income scale depending on the exchange rate conversion used.

So how fast is the global economy growing? Using PPP, the latest WEO estimates that global growth was 3.2 percent in 2016, but only 2.4 percent at market rates.

Who contributes more to global growth, China or the United States? Using both PPP and market-based weights it's China in 2016 but the situation was much different in 2006. At market rates GDP in China exceeded the combined GDP of the next 12 emerging market and developing economies ranked by size (India, Brazil, Russia, Mexico, Indonesia, Turkey, Saudi Arabia, Argentina, Poland, Iran, Thailand, and Nigeria). In fact, in PPP rates China contributes more to global growth than all advanced economies combined (see chart).

Which country is richer, France or Japan? Even among advanced economies, the choice of exchange rate can matter. Using market rates, per capita income in Japan exceeds that in France, but when PPP rates are used, the situation is reversed. [FD](#)

TIM CALLEN is a division chief in the IMF's Middle East and Central Asia Department. **SARWAT JAHAN**, a senior economist in the IMF's Asia and Pacific Department, updated this article.

Capital Accounts: Liberalize or Not?

There are both benefits and costs to easing restrictions on capital that flows across a country's borders

M. Ayhan Kose and Eswar Prasad

WHAT ACCOUNTS for the surge of cross-border capital flows over the past two and a half decades? Capital account liberalization—that is, easing restrictions on capital flows across a country's borders—provides a big part of the answer. But while the increase in these flows since the mid-1980s—both between industrial countries and from industrial to developing countries—has been associated with a number of benefits, it has also played a role in a number of financial crises.

This raises some fundamental questions. Why have many developing countries followed the advanced economies and signed on to capital account liberalization despite the risks? Is easing the flow of capital among countries really the villain that some opposed to globalization have made it out to be?

Building the capital account

The capital account in a country's balance of payments covers a variety of financial flows—mainly foreign direct investment (FDI), portfolio flows (including investment in equities), and bank borrowing—which have in common the acquisition of assets in one country by residents of another. It is possible, in principle, to control these flows by placing restrictions on flows going through official channels.

Capital account liberalization presumably results in a higher degree of financial integration of that country with the global economy through higher volumes of capital inflows and outflows.

There is, however, a significant difference between financial integration in theory and in practice. Some countries—for example, in Latin America during the 1970s and 1980s—found it difficult to contain capital outflows in times of economic distress despite apparently pervasive controls. In contrast, many developing countries, including a few in Africa, have no significant controls but have experienced only minimal inflows.

This also points to the difficulty of measuring capital controls and, by extension, the degree of capital account liberalization undertaken. The IMF (which has jurisdiction over current account, but not capital account, restrictions) maintains a detailed compilation of member countries' capital account

restrictions. But even these provide, at best, rough indications because they do not measure the intensity or effectiveness of capital controls.

Reasons to control capital flows

Controls on capital account transactions represent a country's attempt to shield itself from risks associated with fluctuations in international capital flows. Capital controls take on special significance in the context of a fixed exchange rate regime. Maintaining such a regime can be made more difficult by unfettered capital flows. This is one reason why even industrial countries had relatively closed capital accounts under the Bretton Woods system of fixed exchange rates, which operated from the end of World War II until 1973.

There could, of course, be various other reasons for maintaining controls, on either inflows or outflows. In a country with a fragile banking system, for instance, allowing households to invest abroad freely could precipitate an exodus of domestic savings and jeopardize the banking system's viability. And short-term capital inflows can be quickly reversed when a country is hit with an adverse macroeconomic shock, thereby amplifying its macroeconomic effect.

Some developing countries also use capital controls to steer the composition of inflows toward more stable forms, such as FDI. Countries favor FDI, among other reasons, because it usually involves flows that are relatively long term and not subject to rapid reversals associated with changes in investor sentiment—as are, say, stocks and bonds. Some countries have also used selective capital controls to try to induce a shift from shorter- to longer-term inflows—in Chile's case, by imposing an implicit tax on capital inflows reversed within less than a year.

Motives for liberalizing

In theory, capital account liberalization should allow for more efficient global allocation of capital, from capital-rich industrial countries to capital-poor developing economies. This should have widespread benefits—by providing a higher rate of return

on people's savings in industrial countries and by increasing growth, employment opportunities, and living standards in developing countries.

Access to capital markets should allow countries to “insure” themselves to some extent against fluctuations in their national incomes such that national consumption levels are relatively less volatile. Since good and bad times often are not synchronized across countries, capital flows can, to some extent, offset volatility in countries' own national incomes.

Capital account liberalization may also be interpreted as signaling a country's commitment to good economic policies. For a country with an open capital account, a perceived deterioration in its policy environment could be punished by domestic and

The benefits of capital account openness seem to be most evident for industrial economies.

foreign investors, who could suddenly take capital out of the country. This provides a strong incentive for policymakers to adopt and maintain sound policies, with obvious benefits in terms of long-term growth. Inflows stemming from liberalization should also facilitate the transfer of foreign technological and managerial know-how and encourage competition and financial development, thereby promoting growth.

What does the evidence say?

The evidence is not quite as compelling as the theory, however. While emerging market countries that have liberalized their capital accounts typically have had higher growth rates, on average, than those that have not, this association does not imply a causal relationship. Statistical analysis suggests that, after controlling for the effects of other factors, the causal effect of capital account liberalization on growth has been weak, at best.

There is also some evidence that emerging market countries have not been able to use international financial markets effectively to reduce consumption volatility. In fact, the financial crises that have occurred in these economies have been associated with sharp falls in both income and consumption. And there appears to be a significant procyclical element to international capital market access for such countries. International investors are willing to lend to them in good times but tend to pull back in bad times, thereby amplifying swings in the domestic macroeconomy.

Is liberalization worth the risk? The answer, as with most such things, is that it depends. Capital account liberalization clearly is not an unqualified blessing and poses major risks if implemented in unfavorable circumstances—particularly without supporting policies.

Opening the capital account while maintaining a fixed exchange rate regime, especially when domestic macroeconomic policies are not consistent with the requirements of the regime, has been followed by crisis in many countries. Countries that have maintained or only gradually eased capital controls while moving toward a more flexible exchange rate regime generally seem to have had better outcomes.

Weak macroeconomic fundamentals can also pose a problem. For instance, capital account liberalization can aggravate risks associated with imprudent fiscal policies by providing access to excessive external borrowing. Premature opening of the capital account also poses serious risks when financial regulation and supervision are inadequate. In the presence of weakly regulated banking systems and other distortions in domestic capital markets, foreign capital inflows could be misallocated and create a host of problems.

What's a country to do?

While the evidence suggests that transitional risks are associated with opening the capital account, resisting liberalization over an extended period may prove futile and counterproductive. As the forces of globalization advance, it becomes harder for countries to maintain closed capital accounts. Increasing openness to international trade expands opportunities for the avoidance of capital account restrictions through under- and overinvoicing of trade transactions. And the increasing sophistication of investors and global financial markets makes it much easier.

One possible strategy is to accept the risks and move forward while controlling them as much as possible. History and international experience provide a guide. Sound domestic policies and institutions, a regulatory framework promoting a strong and efficient financial sector, and effective systems and procedures for monitoring capital flows greatly improve the chances of ensuring that such flows foster sustainable growth.

Notably, the benefits of capital account openness in terms of higher growth and lower volatility seem to be most evident for industrial economies, which also typically have the most open capital accounts. Counterintuitive as it may seem, the relatively more positive experiences of industrial countries therefore suggest that, for developing economies, more—not less—financial integration is the answer. But only if it is done the right way. Perhaps most notably, in all the advanced economies and many other countries, open capital accounts are now taken for granted: no country that has liberalized its capital account in recent decades has reversed the process other than temporarily. [FD](#)

M. AYHAN KOSE is director of the Prospects Group of the World Bank's Development Economics Vice Presidency. **ESWAR S. PRASAD** is professor in the Dyson School at Cornell University and senior fellow at the Brookings Institution.

Current Account Deficits: Is There a Problem?

There can be consequences when the amount a country spends abroad is wildly different from what it receives from the outside world

Atish Ghosh and Uma Ramakrishnan

THE CURRENT ACCOUNT BALANCE seems to be an abstruse economic concept. But in countries that are spending a lot more abroad than they are taking in, the current account is the point at which international economics collides with political reality. When countries run large deficits, businesses, trade unions, and parliamentarians are often quick to point accusing fingers at trading partners and make charges about unfair practices. But do such charges have merit? And are current account deficits necessarily bad anyway? Before trying to answer such questions, we first need to understand what the current account is—and what surpluses and deficits imply.

Measuring the current account

The trade balance is the difference between the value of exports of goods and services and the value of imports of goods and services. A trade deficit means that the country is importing more goods and services than it is exporting; a trade surplus means the opposite. The current account balance is then the trade balance plus net factor income (such as interest and dividends from foreign investments or workers' remittances) and transfers from abroad (such as foreign aid), which are usually a small fraction of the total. Since (for most countries) there is little difference between the trade balance and the current account, a current account deficit often raises the hackles of protectionists, who—apparently forgetting that a main reason to export is to be able to import—think that exports are “good” and imports are “bad.”

The current account can also be expressed as the difference between national (both public and private) savings and investment. A current account deficit may therefore reflect a low level of national savings relative to investment or a high rate of investment—or both. For capital-poor developing economies, which have more investment opportunities than they can afford to take because of low domestic savings, a current account deficit may be natural. A deficit potentially spurs faster output growth and economic development—although recent research does not indicate that developing economies with current account deficits grow faster (perhaps because their less developed domestic

financial systems cannot allocate foreign capital efficiently). Moreover, in practice, private capital often flows from developing to advanced economies. Advanced economies, such as the United States (see chart), run current account deficits, whereas developing and emerging market economies often run surpluses or near surpluses. Very poor countries typically run large current account deficits, in proportion to their GDP, that are financed by official grants and loans.

One point that the savings-investment balance approach underscores is that protectionist policies are unlikely to be of much use in improving the current account balance because there is no obvious connection between protectionism and savings or investment.

Another way to look at the current account is in terms of the timing of trade. We are used to intratemporal trade—exchanging cloth for wine today. But we can also think of intertemporal trade—importing goods today (running a current account deficit) and, in return, exporting goods in the future (running a current account surplus then). Just as a country may import one good and export another under intratemporal trade, there is no reason a country should not import goods of today and export goods of tomorrow.

Intertemporal theories of the current account also stress the consumption-smoothing role that current account deficits and surpluses can play. For instance, if a country is struck by a shock—perhaps a natural disaster—that temporarily depresses its ability to access productive capacity, rather than take the full brunt of the shock immediately, the country can spread out the pain over time by running a current account deficit. Conversely, research also suggests that countries that are subject to large shocks should, on average, run current account surpluses as a form of precautionary saving.

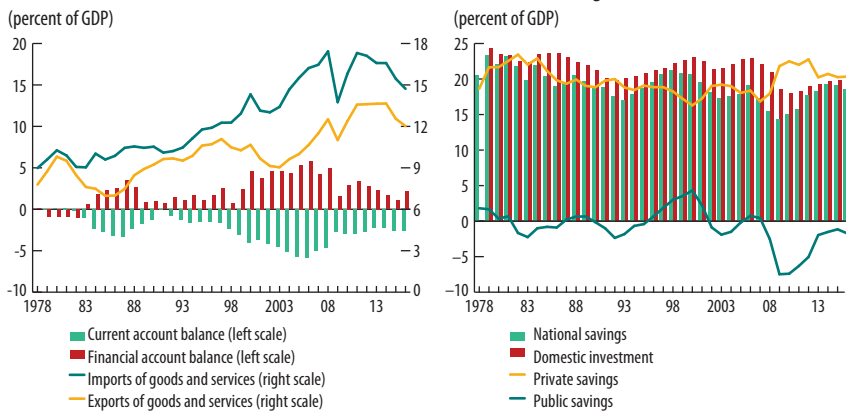
When persistent is too persistent

Does it matter how long a country runs a current account deficit? When a country runs a current account deficit, it is building up liabilities to the rest of the world that are financed by flows in the financial account. Eventually, these need to be

II. HOW ECONOMIES FUNCTION

Net foreign borrower

The United States runs a persistent balance of trade and current account deficit and uses foreign capital to finance the difference between domestic investment and domestic savings.



Source: IMF, *World Economic Outlook*, April 2017.

Note: Financial account = current account plus capital account.

paid back. Common sense suggests that if a country fritters away its borrowed foreign funds on spending that yields no long-term productive gains, then its ability to repay—its basic solvency—might come into question. This is because solvency requires that the country be willing and able to generate (eventually) sufficient current account surpluses to repay what it has borrowed to finance the current account deficits. Therefore, whether a country should run a current account deficit (borrow more) depends on the extent of its foreign liabilities (its external debt) and on whether the borrowing will finance investment with a higher marginal product than the interest rate (or rate of return) the country has to pay on its foreign liabilities.

But even if the country is intertemporally solvent—meaning that current liabilities will be covered by future revenues—its current account deficit may become unsustainable if it is unable to secure the necessary financing. While some countries (such as Australia and New Zealand) have been able to maintain current account deficits averaging about 4 1/2 to 5 percent of GDP for several decades, others (such as Mexico in 1995, Thailand in 1997, and several economies during the recent global crisis) experienced sharp reversals of their current account deficits after private financing withdrew during the financial crisis.

Such reversals can be highly disruptive because private consumption, investment, and government expenditure must be curtailed abruptly when foreign financing is no longer available and, indeed, a country is forced to run large surpluses to repay in short order what it borrowed in the past. This suggests that—regardless of whether a country has a current account deficit (and even if the deficit reflects desirable underlying trends)—large and persistent deficits call for caution, lest a country experience an abrupt and painful reversal of financing.

What determines whether a country experiences such a reversal? Empirical research suggests that an overvalued real exchange rate, inadequate foreign exchange reserves, excessively

fast domestic credit growth, unfavorable terms-of-trade shocks, low growth in partner countries, and higher interest rates in industrial countries influence the occurrence of reversals. More recent literature has also focused on the importance of balance sheet vulnerabilities in the run-up to a crisis—such as the extent to which companies have large liabilities in foreign currencies such as dollars or maturity mismatches that occur when companies have more short-term liabilities than short-term assets and more medium- and long-term assets relative to their liabilities. Recent research has also underscored the importance of the composition of capital inflows—for example,

the relative stability of foreign direct investment compared with more volatile short-term investment flows, such as in equities and bonds. Moreover, weak financial sectors can often increase a country's vulnerability to a reversal of investment flows as banks borrow money from abroad and make risky domestic loans. Conversely, a more flexible policy framework—such as a flexible exchange rate regime, a higher degree of openness, export diversification, and coherent fiscal and monetary policies—combined with financial sector development could help a country with persistent deficits be less vulnerable to a reversal by allowing greater room for better shock absorption.

Judging whether deficits are bad

A common complaint about economics is that the answer to any question is, “It all depends.” It is true that economic theory tells us that whether a deficit is good or bad depends on the factors giving rise to that deficit, but economic theory also tells us what to look for in assessing the desirability of a deficit.

If the deficit reflects an excess of imports over exports, it may be indicative of competitiveness problems, but because the current account deficit also implies an excess of investment over savings, it could equally be pointing to a highly productive, growing economy. If the deficit reflects low savings rather than high investment, it could be caused by reckless fiscal policy or a consumption binge. Or it could reflect perfectly sensible intertemporal trade, perhaps because of a temporary shock or shifting demographics. Without knowing which of these is at play, it makes little sense to talk of a deficit being good or bad. Deficits reflect underlying economic trends, which may be desirable or undesirable for a country at a particular point in time. **FD**

ATISH GHOSH is the IMF historian and **UMA RAMAKRISHNAN** is an assistant director in the IMF's Western Hemisphere Department.

III FINANCE

Shadow Banking: Out of the Eyes of Regulators

Many financial institutions that act like banks are not supervised like banks

Laura E. Kodres

IF IT LOOKS LIKE A DUCK, quacks like a duck, and acts like a duck, then it is a duck—or so the saying goes. But what about an institution that looks like a bank and acts like a bank? Often it is not a bank—it is a shadow bank.

Shadow banking, in fact, symbolizes one of the many failings of the financial system leading up to the global financial crisis. The term “shadow bank” was coined by economist Paul McCulley in a 2007 speech at the annual financial symposium hosted by the Kansas City Federal Reserve Bank in Jackson Hole, Wyoming. In McCulley’s talk, shadow banking had a distinctly US focus and referred mainly to nonbank financial institutions that engaged in what economists call maturity transformation. Commercial banks engage in maturity transformation when they use deposits, which are normally short term, to fund loans that are longer term. Shadow banks do something similar. They raise (that is, mostly borrow) short-term funds in the money markets and use those funds to buy assets with longer-term maturities. But because they are not subject to traditional bank regulation, they cannot—as banks can—borrow in an emergency from the Federal Reserve (the US central bank) and do not have traditional depositors whose funds are covered by insurance; they are in the “shadows.”

Home mortgages

Shadow banks first caught the attention of many experts because of their growing role in turning home mortgages into securities. The “securitization chain” started with the origination of a mortgage that then was bought and sold by one or more financial entities until it ended up part of a package of mortgage loans used to back a security that was sold to investors. The value of the security was related to the value of the mortgage loans in the package, and the interest on a mortgage-backed security was paid from the interest and principal homeowners paid on their mortgage loans. Almost every step from creation of the mortgage to sale of the security took place outside the direct view of regulators.

The Financial Stability Board (FSB), an organization of financial and supervisory authorities from major economies and international financial institutions, developed a broader definition of shadow banks that includes all entities outside the regulated

banking system that perform the core banking function, credit intermediation (that is, taking money from savers and lending it to borrowers). The four key aspects of intermediation are

maturity transformation: obtaining short-term funds to invest in longer-term assets;

liquidity transformation: a concept similar to maturity transformation that entails using cash-like liabilities to buy harder-to-sell assets such as loans;

leverage: employing techniques such as borrowing money to buy fixed assets to magnify the potential gains (or losses) on an investment;

credit risk transfer: taking the risk of a borrower’s default and transferring it from the originator of the loan (or the issuer of a bond) to another party.

Under this definition shadow banks would include broker-dealers that fund their assets using repurchase agreements (repos). In a repurchase agreement an entity in need of funds sells a security to raise those funds and promises to buy the security back (that is, repay the borrowing) at a specified price on a specified date.

Money market mutual funds that pool investors’ funds to purchase commercial paper (corporate IOUs) or mortgage-backed securities are also considered shadow banks. So are financial entities that sell commercial paper (or other short-term obligations) and use the proceeds to extend credit to households (called finance companies in many countries). There are now myriad types of entities performing these intermediation functions, and they are growing all the time.

Why there is a problem

As long as investors understand what is going on and such activities do not pose undue risk to the financial system, there is nothing inherently shadowy about obtaining funds from various investors who might want their money back within a short period and investing those funds in assets with longer-term maturities. Problems arose during the global financial crisis, however, when investors became skittish about what those longer-term assets were really worth and many decided to withdraw their funds at once. To repay these investors, shadow banks had to sell assets. These “fire sales” generally reduced the value of those assets,

forcing other shadow banking entities (and some banks) with similar assets to reduce the value of those assets on their books to reflect the lower market price, creating further uncertainty about their health. At the peak of the crisis, so many investors withdrew or would not roll over (reinvest) their funds that many financial institutions—banks and nonbanks—ran into serious difficulty.

Had this taken place outside the banking system, it could possibly have been isolated and those entities could have been closed in an orderly manner. But real banks were caught in the shadows, too. Some shadow banks were controlled by commercial banks and for reputational reasons were salvaged by their stronger bank parent. In other cases, the connections were at arm's length, but because

Shadow banks raise short-term funds in the money markets and use those funds to buy assets with longer-term maturities.

shadow banks had to withdraw from other markets—including those in which banks sold commercial paper and other short-term debt—these sources of funding to banks were also impaired. And because there was so little transparency, it often was unclear who owed (or would owe later) what to whom.

In short, the shadow banking entities were characterized by a lack of disclosure and information about the value of their assets (or sometimes even what the assets were); opaque governance and ownership structures between banks and shadow banks; little regulatory or supervisory oversight of the type associated with traditional banks; virtually no loss-absorbing capital or cash for redemptions; and a lack of access to formal liquidity (for example, central bank funds) support to help prevent fire sales.

Issues continue

Shadows can be frightening because they obscure the shapes and sizes of objects within them. The same is true for shadow banks. Estimating the size of the shadow banking system is particularly difficult because many of its entities do not report to government regulators. In the run-up to the global financial crisis, the shadow banking system appeared to be largest in the United States, but nonbank credit intermediation was present in other countries—and is growing again—particularly in China. Since 2011, the FSB has conducted a “global” monitoring exercise to examine all nonbank credit intermediation. The exercise, mandated by the 20 major advanced and emerging market economies (the G20), now covers 28 jurisdictions and the euro area. The original results were rough because they used a catch-all category of “other financial institutions,” but now the FSB also examines shadow banks by “function” rather than entity.

Using the entity-based measure, the latest report (end-2015 data) shows that the euro area shadow banking system is now the largest globally, comprising 33 percent of the total (up from 32 percent in 2011), whereas the US shadow banking system has declined from 33 percent to 28 percent. Across the jurisdictions contributing to the FSB exercise, the global shadow system peaked at \$62 trillion in 2007, declined to \$59 trillion during the crisis, and rebounded to \$92 trillion by the end of 2015. The “functional” categorization (a narrower categorization of 27 jurisdictions) shows that of the total \$34.2 trillion, the largest part of shadow banking is made up of asset-management-type activities—some 22 percent of the total.

Even though the FSB's move to examine activities (rather than institutions) comes closer to measuring risks, the measure still falls short of an accurate gauge of risks that shadow banking poses to the financial system. The FSB also does not measure the amount of debt used to purchase assets (often called leverage), the degree to which the system can amplify problems, or the channels through which problems move from one sector to another (although there has been some attempt to gauge these latter linkages using balance sheet data between nonbanks and banks).

Over time, it has been recognized that shadow banking in some countries is encroaching on banks' credit intermediation role. The true risks of these activities and whether they are systemically important are still undetermined, but the ability to monitor their size is improving.

Authorities engage

Since the global financial crisis, the official sector is collecting more and better information and searching for hidden vulnerabilities. Banking supervisors are examining the exposure of traditional banks to shadow banks and trying to contain it through better capital and liquidity regulations—because this exposure allowed shadow banks to affect the traditional financial sector and the economy more generally. Moreover, because many shadow banking entities were either lightly regulated or outside the purview of regulators, many authorities have expanded the scope of information reporting, and some have altered the regulatory perimeter to capture shadow-banking entities and the markets they use. And the authorities are trying to discourage shadow banks from tailoring their behavior to come under the supervision of the weakest (or of no) regulators—domestically or globally.

The authorities are making progress, but they work in the shadows themselves—trying to piece together disparate and incomplete data to see what, if any, systemic risks are associated with the various activities, entities, and instruments that comprise the shadow banking system. **FD**

LAURA E. KODRES is assistant director of the IMF's Institute for Capacity Development.

LIBOR: World Reference Point

The London interbank rate is used widely as a benchmark but has come under fire

John Kiff

EVERY WEEKDAY at about 11 a.m., 17 large banks, under the auspices of the ICE Benchmark Administration, report the rate at which they believe they can borrow a “reasonable” amount of dollars from each other in the so-called London interbank market. They report rates for seven borrowing terms that range from overnight to one year. The four highest and four lowest are thrown out, and the rest are averaged. ICE then announces that average rate at which banks say they can borrow dollars for each of the seven maturities.

A similar process is carried out for five other currencies as well. The average—often referred to in the singular even though there are 35 rates—is called the London interbank offered rate (LIBOR). It is one of the best known and most important interest rates in the world.

But it is not important because banks actually transact business with each other at the announced rate—although that can happen. Rather, LIBOR’s importance derives from its widespread use as a benchmark for many other interest rates at which business is actually carried out. According to ICE, about \$350 trillion in financial contracts are tied to LIBOR.

Because the US dollar is the most important of the world’s currencies, US dollar LIBOR rates are probably the most widely used and cited. Other panels—ranging in size from 7 banks to 17—report daily what it would cost them to borrow British pounds, euros, Japanese yen, and Swiss francs short term in the London interbank market.

Prior to February 2014 LIBOR was administered by the British Bankers’ Association, and LIBOR consisted of 150 separate rates (15 maturities and 10 currencies). Reforms were triggered by controversy over how some banks were reporting the rates at which they “believed” they could borrow and because of some underlying problems with the LIBOR concept. Also, in April 2013 the setting and maintenance of this important benchmark were brought under the purview of the U.K. Financial Conduct Authority.

A recent innovation

Although banks in London have been lending to one another for centuries, LIBOR is a relatively new idea. It dates to 1969, when a syndicate of banks led by Manufacturers Trust (now part of JPMorgan Chase) needed a reference rate for an \$80 million floating rate loan to the Shah of Iran. However, its use took off in the early 1980s on the sudden growth of the use of interest-rate-based financial instruments—such as floating

rate corporate loans, forward rate agreements, and interest rate swaps.

Good standardized and transparent benchmark rates were needed to settle those contracts. Markets turned to the banking industry trade group and the Bank of England to provide such a rate. The British Bankers’ Association launched LIBOR in 1986—initially with only three currencies—the dollar, the yen, and the pound sterling.

LIBOR is supposed to reflect reality—an average of what banks believe they would have to pay to borrow a “reasonable” amount of currency for a specified short period. That is, it represents the cost of funds—although a bank may not actually have a need for the funds on any given day.

But LIBOR has long been dogged by perceptions that the method for setting the rates is flawed and prone to distorted results during periods of market stress when banks stop lending to each other across the full maturity spectrum, from overnight to one year.

A more direct challenge to its authenticity came from attempts to manipulate LIBOR (and other benchmark rates) by a number of big global banks, for which over \$9 billion in fines has been paid to regulators in the European Union, United Kingdom, and United States.

But even before the controversy over manipulation called into question its accuracy, LIBOR was often called a “convenient fiction” because of the disconnect between the LIBORs used as benchmarks and actual borrowing in the London interbank market. Most banks loan each other money for a week or less, so most LIBORs for longer maturities are set on the basis of educated guesses. Yet almost 95 percent of transactions that reference one of the LIBORs—from interest rate derivatives to home mortgages—are indexed to rates for maturities three months or longer. The US three-month maturity period (or “tenor,” as the maturity period is called) is the most popular, according to the U.K. Treasury. A further hint that unsecured term lending has become a fiction was the 2012 decision by ICAP, a large London broker-dealer, to stop publishing its one- and three-month New York Funding Rate (NYFR) indices, an alternative to LIBOR, due to a lack of data from New York-based banks.

Nevertheless, LIBORs have been found to be reasonably accurate, most of the time tracking closely similar benchmarks that are tied to actual unsecured bank funding rates such as those for commercial paper. The glaring exception was the period immediately after the September 2008 failure of the New York

investment banking firm Lehman Brothers, which triggered the global financial crisis. The three-month US LIBOR diverged from two publicly available similar short-term rates—the ICAP NYFR and the three-month rate on Eurodollar deposits, which are US dollar–denominated deposits at banks located outside the United States.

LIBOR was lower than the Eurodollar rate during early 2008 but was markedly lower in the period immediately following the Lehman collapse. LIBOR appears to track the NYFR very closely, except in the immediate aftermath of the Lehman failure, when it too was decidedly lower (see chart).

In part, LIBOR may have been lower after the Lehman failure because of an unintended consequence of a British Bankers' Association rule meant to ensure that banks reported their borrowing costs truthfully: immediate publication of individual banks' reports. While normally this would encourage honesty, in 2007–08 this safeguard may have backfired. Banks were reportedly loath to suggest that they were having trouble obtaining funds by reporting a rate higher than other banks were being charged. So to mask its liquidity problems, a bank with funding problems had an incentive to report lower rates than it really believed it would be offered. Indeed, a number of studies have suggested that banks submitted lowball rates after the collapse of the investment bank Bear Stearns in March 2008 as well as after the Lehman collapse six months later.

Other studies have found situations that suggest many banks were not reporting accurately. But statistical evidence of bank-specific LIBOR collusion and manipulation has been

limited. Nevertheless, criminal investigations eventually uncovered the direct evidence that led to the massive fines.

Following the scandal there were some calls to eliminate LIBOR. But because it is so important and pervasive as a benchmark, the British government decided it could not be junked and should be saved.

First, the British government took over supervision of LIBOR from the bankers' group, which Martin Wheatley, managing director of the U.K. Financial Services Authority, said, "clearly failed to properly oversee the LIBOR setting process." In a 2012 report Wheatley outlined the government's proposed changes that came into force in 2013.

Under the reform, LIBOR continues to be set daily based on reports by panels of banks. But the banks are required to provide data to show that the rates they submit are an accurate reflection of their borrowing costs. And although the submitted rates are still reported publicly, it is done so with a three-month lag so that banks don't have an incentive to lie about their costs during a period of stress. Moreover, criminal sanctions are imposed on banks that misreport.

And to focus the production of LIBORs on interest rates that matter—and for which there are verifiable funding costs—the Australian, Canadian, Danish, New Zealand, and Swedish currencies were phased out and eight maturities eliminated. The number of LIBORs dropped from 150 to the 35 that are most important to market participants. Also, the ICE LIBOR Oversight Committee, composed of rate submitters and users plus other relevant experts, was set up to return credibility to LIBOR and ensure its continued relevance.

Nevertheless, many of the rates are still unsupported by actual interbank transactions, and bank interest in contributing to the LIBOR-setting process is dwindling. The Wheatley report encouraged market participants to consider the need for a backup plan if the rates are no longer produced. So in 2013 G20 leaders called for a fundamental review of LIBOR and other major interest rate benchmarks.

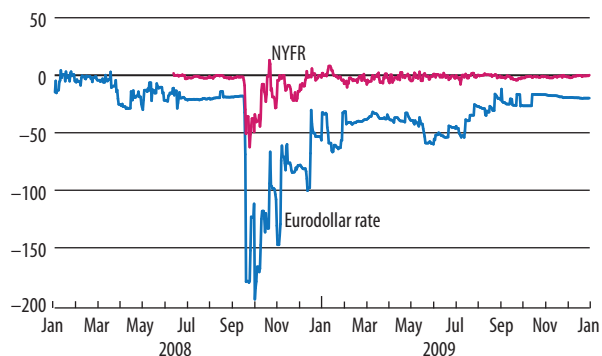
Since 2014 the Financial Stability Board has been leading the charge to strengthen LIBOR and search for transaction-based alternatives. Such alternatives have already been proposed for all five LIBOR currencies. For example, the US Alternative Reference Rates Committee has put forward a benchmark based on the rates at which banks fund purchases of government securities on a secured basis.

But it is not certain that the market will adopt the new rate benchmarks, and if it does, the transition will be a slow and gradual process. New rate setting governance and infrastructures must be put in place, and market participants will have to change legal documentation, systems, and processes. So LIBORs will be with us for many years to come. [FD](#)

Sharp divergence

The three-month U.S. London interbank offered rate (LIBOR) was markedly lower than two similar interest rates—the three-month Eurodollar deposit rate and the three-month New York Funding Rate (NYFR)—after the collapse of Lehman Brothers in September 2008.

(difference between LIBOR and the NYFR and Eurodollar rates, basis points)



Source: Author's calculations.

Note: Eurodollar deposits are U.S. dollars on deposit at banks located outside the United States. The NYFR was compiled by the London broker-dealer ICAP from information reported by prime banks operating in New York and was designed to reflect short-term borrowing costs of those banks. ICAP stopped reporting the NYFR in August 2012. A basis point is 1/100th of 1 percent.

JOHN KIFF is a senior financial sector expert in the IMF's Monetary and Capital Markets Department.

Banks: At the Heart of the Matter

Institutions that match up savers and borrowers help ensure that economies function smoothly

Jeanne Gobat

YOU'VE GOT \$1,000 you don't need for, say, a year and want to earn income from the money until then. Or you want to buy a house and need to borrow \$100,000 and pay it back over 30 years.

It would be difficult, if not impossible, for someone acting alone to find either a potential borrower who needs exactly \$1,000 for a year or a lender who can spare \$100,000 for 30.

That's where banks come in.

Although banks do many things, their primary role is to take in funds—called deposits—from those with money, pool them, and lend them to those who need funds. Banks are intermediaries between depositors (who lend money to the bank) and borrowers (to whom the bank lends money). The amount banks pay for deposits and the income they receive on their loans are both called interest.

Depositors can be individuals and households, financial and nonfinancial firms, or national and local governments. Borrowers are, well, the same. Deposits can be available on demand (a checking account, for example) or with some restrictions (such as savings and time deposits).

Making loans

While at any given moment some depositors need their money, most do not. That enables banks to use shorter-term deposits to make longer-term loans. The process involves maturity transformation—converting short-term liabilities (deposits) to long-term assets (loans). Banks pay depositors less than they receive from borrowers, and that difference accounts for the bulk of banks' income in most countries.

Banks can complement traditional deposits as a source of funding by directly borrowing in the money and capital markets. They can issue securities such as commercial paper or bonds; or they can temporarily lend securities they already own to other institutions for cash—a transaction often called a repurchase agreement (repo). Banks can also package the loans they have on their books into a security and sell this to the market (a process called liquidity transformation and securitization) to obtain funds they can relend.

A bank's most important role may be matching up creditors and borrowers, but banks are also essential to the domestic and international payments system—and they create money.

Not only do individuals, businesses, and governments need somewhere to deposit and borrow money, they need to move

funds around—for example, from buyers to sellers or employers to employees or taxpayers to governments. Here too banks play a central role. They process payments, from the tiniest of personal checks to large-value electronic payments between banks. The payments system is a complex network of local, national, and international banks and often involves government central banks and private clearing facilities that match up what banks owe each other. In many cases payments are processed nearly instantaneously. The payments system also includes credit and debit cards. A well-operating payments system is a prerequisite for an efficiently performing economy, and breakdowns in the payments system are likely to disrupt trade—and, therefore, economic growth—significantly.

Creating money

Banks also create money. They do this because they must hold on reserve, and not lend out, some portion of their deposits—either in cash or in securities that can be quickly converted to cash. The amount of those reserves depends both on the bank's assessment of its depositors' need for cash and on the requirements of bank regulators, typically the central bank—a government institution that is at the center of a country's monetary and banking system. Banks keep those required reserves on deposit with central banks, such as the US Federal Reserve, the Bank of Japan, and the European Central Bank. Banks create money when they lend the rest of the money depositors give them. This money can be used to purchase goods and services and can find its way back into the banking system as a deposit in another bank, which then can lend a fraction of it. The process of relending can repeat itself a number of times in a phenomenon called the multiplier effect. The size of the multiplier—the amount of money created from an initial deposit—depends on the amount of money banks must keep on reserve.

Banks also lend and recycle excess money within the financial system and create, distribute, and trade securities.

Banks have several ways of making money besides pocketing the difference (or spread) between the interest they pay on deposits and borrowed money and the interest they collect from borrowers or securities they hold. They can earn money from

- income from securities they trade; and
- fees for customer services, such as checking accounts, financial and investment banking, loan servicing, and the origination,

distribution, and sale of other financial products, such as insurance and mutual funds.

Banks earn on average between 1 and 2 percent of their assets (loans and securities). This is commonly referred to as a bank's return on assets.

Transmitting monetary policy

Banks also play a central role in the transmission of monetary policy, one of the government's most important tools for achieving economic growth without inflation. The central bank controls the money supply at the national level, while banks facilitate the flow of money in the markets within which they operate. At the national level, central banks can shrink or expand the money supply by raising or lowering banks' reserve requirements and by buying and selling securities on the open market with banks as key counterparties in the transactions. Banks can shrink the money supply by putting away more deposits as reserves at the central bank or by increasing their holdings of other forms of liquid assets—those that can be easily converted to cash with little impact on their price. A sharp increase in bank reserves or liquid assets—for any reason—can lead to a “credit crunch” by reducing the amount of money banks have to lend, which can lead to higher borrowing costs as customers pay more for scarcer bank funds. A credit crunch can hurt economic growth.

Banks can fail, just like other firms. But their failure can have broader ramifications—hurting customers, other banks, the community, and the market as a whole. Customer deposits can be frozen, loan relationships can break down, and lines of credit that businesses draw on to make payrolls or pay suppliers may not be renewed. In addition, one bank failure can lead to other bank failures.

Banks' vulnerabilities arise primarily from three sources:

- a high proportion of short-term funding such as checking accounts and repos to total deposits. Most deposits are used to finance longer-term loans, which are hard to convert into cash quickly;
- a low ratio of cash to assets; and
- a low ratio of capital (assets minus liabilities) to assets.

Depositors and other creditors can demand payment on checking accounts and repos almost immediately. When a bank is perceived—rightly or wrongly—to have problems, customers, fearing that they could lose their deposits, may withdraw their funds so fast that the small portion of liquid assets a bank holds becomes quickly exhausted. During such a “run on deposits” a bank may have to sell other longer-term and less liquid assets, often at a loss, to meet the withdrawal demands. If losses are sufficiently large, they may exceed the capital a bank maintains and drive it into insolvency.

Essentially, banking is about confidence or trust—the belief that the bank has the money to honor its obligations. Any crack in that confidence can trigger a run and potentially a bank failure,

even bringing down solvent institutions. Many countries insure deposits in case of bank failure, and the recent crisis showed that banks' greater use of market sources of funding has made them more vulnerable to runs driven by investor sentiment than to depositor runs.

The need for regulation

Bank safety and soundness are a major public policy concern, and government policies have been designed to limit bank failures and the panic they can ignite. In most countries, banks need a charter to carry out banking activities and to be eligible for government backstop facilities—such as emergency loans from

Banks' primary role is to take in funds—called deposits—from those with money, pool them, and lend them to those who need funds.

the central bank and explicit guarantees to insure bank deposits up to a certain amount. Banks are regulated by the laws of their home country and are typically subject to regular supervision. If banks are active abroad, they may also be regulated by the host country. Regulators have broad powers to intervene in troubled banks to minimize disruptions.

Regulations are generally designed to limit banks' exposures to credit, market, and liquidity risks and to overall solvency risk. Banks are now required to hold more and higher-quality equity—for example, in the form of retained earnings and paid-in capital—to buffer losses than they were before the financial crisis. Large global banks must hold even more capital to account for the potential impact of their failure on the stability of the global financial system (also known as systemic risk). Regulations also stipulate minimum levels of liquid assets for banks and prescribe stable, longer-term funding sources.

Regulators are reviewing the growing importance of institutions that provide bank-like functions but that are not regulated in the same fashion as banks—so-called shadow banks—and looking at options for regulating them. The recent financial crisis exposed the systemic importance of these institutions, which include finance companies, investment banks, and money market mutual funds. [FD](#)

JEANNE GOBAT is a former IMF staff member.

What Are Money Markets?

They provide a means for lenders and borrowers to satisfy their short-term financial needs

Randall Dodd

UNTIL PROBLEMS SURFACED during the global financial crisis, money markets were often taken for granted as plain-vanilla, low-volatility segments of the financial system.

For the most part, money markets provide those with funds—banks, money managers, and retail investors—a means for safe, liquid, short-term investments, and they offer borrowers—banks, broker-dealers, hedge funds, and nonfinancial corporations—access to low-cost funds. The term money market is an umbrella that covers several types of secured transactions, which vary according to the needs of the lenders and borrowers.

One consequence of the financial crisis has been to focus attention on the differences among various segments of money markets, because some proved to be fragile, whereas others exhibited a good deal of resilience.

For the short term

These markets are described as “money markets” because the assets that are bought and sold are short term—with maturities ranging from a day to a year—and normally are easily convertible into cash. Money markets include markets for such instruments as bank accounts, including term certificates of deposit; interbank loans (loans between banks); money market mutual funds; commercial paper; Treasury bills; and securities lending and repurchase agreements (repos). These markets comprise a large share of the financial system—in the United States, accounting for about one-third of all credit, according to the Federal Reserve Board’s Flow of Funds Survey.

These money market instruments, many of them securities, differ in how they are traded and are treated under financial regulatory laws as well as in how much a lender relies on the value of underlying collateral, rather than on an assessment of the borrower.

The most familiar money market instruments are bank deposits, which are not considered securities, even though certificates of deposit are sometimes traded like securities. Depositors, who are lending money to the bank, look to the institution’s creditworthiness, as well as to any government programs that insure bank deposits.

Interbank loans are not secured by collateral, so a lender looks exclusively to a borrower’s creditworthiness to assess repayment probabilities. The most closely watched interbank market is in England, where the London interbank offered rate (LIBOR)

is determined daily and represents the average price at which major banks are willing to lend to each other. That market did not prove to be a reliable source of funding during the crisis. LIBOR rates rose sharply in comparison to other money market rates once the creditworthiness of banks was called into question. Moreover, lending volume decreased significantly as banks struggled to fund their existing assets and were less interested in new lending. Emergency lending by central banks helped make up for the contraction of this funding source. Recent investigations by regulatory authorities have identified serious flaws in integrity of the pricing process by which LIBOR is determined.

Commercial paper is a promissory note (an unsecured debt) issued by highly rated banks and some large nonfinancial corporations. Because the instrument is unsecured (no more than a promise to pay, hence the name), investors look solely to the creditworthiness of the issuer for repayment of their savings. Commercial paper is issued and traded like a security. But because it is short term by nature and not purchased by retail investors, it is exempt from most securities laws. In the United States, for example, commercial paper is issued in maturities of 1 to 270 days, and in denominations that are deemed too large for retail investors (typically \$1 million, but sometimes as small as \$10,000).

The safest investment

Treasury bills, which are issued by the government, are securities with maturities of less than a year. US Treasury bills, sold at a discount from face value and actively bought and sold after they are issued, are the safest instrument in which to place short-term savings. The markets are deep and liquid, and trading is covered by securities laws. US Treasury bills are not only savings instruments; they can be used to settle transactions. Treasury bills, which are issued electronically, can be sent through the payments system as readily as money.

Repos are a large, but more complicated, segment of money markets. Repos offer competitive interest rates for borrowing and lending on a short-term basis—usually no more than two weeks and often overnight. A borrower sells a security it owns for cash and agrees to buy it back from the purchaser (who is in effect a lender) at a specified date and at a price that reflects the interest charge for borrowing over the period. The security at the heart of the transaction serves as collateral for the lender.

Besides making possible secure short-term borrowing and lending in money markets, repo and other securities lending markets are critical to short-selling—when a trader agrees to sell a security he or she does not own. To come up with such a security, the short-seller must borrow it or purchase it temporarily through a repo transaction. When it is time to return the security to the lender, the short-seller again must buy or borrow it. If the price has fallen, the short-seller makes money on the transaction.

Money market mutual funds (MMMFs) are securities offered by companies that invest in other money market instruments—such as commercial paper, certificates of deposit, Treasury bills, and repos. Money market mutual funds are regulated as investment companies in the United States and in the European Union. They offer low-risk return on a short-term investment to retail and institutional investors as well as corporations. A typical MMMF invests in liquid, short-term, highly rated instruments. Although the price is not fixed or guaranteed, the fund is managed so that the price is constant—or in securities parlance, maintains a stable net asset value, usually \$1 a share. (This is in contrast to other mutual funds that invest in stocks or bonds and whose per share value changes daily.) If the value of the underlying MMMF assets rises above \$1 a share, the difference is paid as interest. Until the global crisis, a money market fund with a net value of less than \$1 a share—or breaking the buck, as it is called—was almost unheard of. The few times it happened, the fund’s investment managers used their own resources to keep the price at \$1 a share.

But during the financial crisis, money market funds were threatened by losses on commercial paper and later on notes issued by Lehman Brothers (the broker-dealer that went bankrupt in September 2008). Because MMMFs are important players in other crucial money markets, the US government acted to prevent a panic that might have caused the credit contraction to spread. The US Treasury guaranteed principal and the Federal Reserve created a special lending facility for commercial paper to help MMMFs stave off a run by investors.

Dysfunctional markets

There are some other sectors of the money market that are not so plain and simple. These include asset-backed commercial paper (ABCP) and certain triparty repo transactions.

A firm with hard-to-sell (illiquid) financial assets, such as loans, mortgages, or receivables, might use ABCP to borrow at a lower cost or to move these assets off its balance sheet. It creates a special purpose entity that purchases the illiquid assets from the firm and finances the purchase by issuing ABCP, which—unlike normal commercial paper—is secured or “backed” by the underlying assets. This type of commercial paper can obtain a high credit rating if the assets are rated highly and if the special facility has adequate capital and lines of credit. The capital is

intended to cover unexpected losses on the assets, and the lines of credit take into account the difficulty of selling the underlying assets to meet cash needs.

Some parts of the ABCP market had problems during the crisis. Standard commercial paper issuers—almost exclusively large nonfinancial corporations and banks—file quarterly financial statements that enabled investors to easily assess their credit condition. The credit risk on ABCP depended on, among other things, how the special purpose entity was set up, its credit enhancements, its liquidity backstop, and the value of the underlying assets—all likely to be less transparent and more complex than that of the straightforward commercial paper. In the United States, the ABCP market shrunk by 38 percent from

There are some other sectors of the money market that are not so plain and simple.

August to November 2008.

That hit the MMMF market, which holds more than one-third of outstanding commercial paper. When investors began to withdraw funds from MMMFs, the funds pivoted sharply away from ABCP and into government and agency securities.

The triparty repo market proved to be much less reliable than the ordinary repo market for Treasury and agency securities. The triparty repo market is organized around one or two clearing banks that hold the collateral and transfer ownership from borrower to lender and back again when the loan is repaid.

The triparty repo market was roiled by the collapse of markets for privately issued securities backed by mortgages. These securities made up a large share of the collateral in the triparty repo market. Once the market value and the credit ratings of these securities fell and the trading in these securities dried up, the triparty market suffered from both the higher haircuts (the percentage by which a lender reduces the value of a security for collateral purposes) needed to offset the volatility in the securitized debt market and the difficulty of pricing collateral that no longer had a market price.

Together the crises in the ABCP and triparty repo markets spread funding problems to banks, securities firms, and hedge funds that had used these money markets to fund investments. Today those markets have shrunk dramatically. **FD**

RANDALL DODD works at the US Treasury Department and is a former advisor in the IMF’s Monetary and Capital Markets Department.

Markets: Exchange or Over the Counter

How securities are traded plays a critical role in price determination and stability

Randall Dodd

FINANCIAL MARKETS are complex organizations with their own economic and institutional structures that play a critical role in determining how prices are established—or “discovered,” as traders say. These structures also shape the orderliness and indeed the stability of the marketplace. As holders of subprime collateralized debt obligations and other distressed debt securities found out in the months following the August 2007 onset of the financial turmoil that led to the global economic crisis, some types of market arrangements can very quickly become disorderly, dysfunctional, or otherwise unstable.

There are two basic ways to organize financial markets—exchange and over the counter (OTC)—although some recent electronic facilities blur the traditional distinctions.

Trading on an exchange

Exchanges, whether stock markets or derivatives exchanges, started as physical places where trading took place. Some of the best known include the New York Stock Exchange (NYSE), which was formed in 1792, and the Chicago Board of Trade (now part of the CME Group), which has been trading futures contracts since 1851. Today there are more than a hundred stock and derivatives exchanges throughout the developed and developing world.

But exchanges are more than physical locations. They set the institutional rules that govern trading and information flows about that trading. They are closely linked to the clearing facilities through which post-trade activities are completed for securities and derivatives traded on the exchange. An exchange centralizes the communication of bid and offer prices to all direct market participants, who can respond by selling or buying at one of the quotes or by replying with a different quote. Depending on the exchange, the medium of communication can be voice, hand signal, a discrete electronic message, or computer-generated electronic commands. When two parties reach agreement, the price at which the transaction is executed is communicated throughout the market. The result is a level playing field that allows any market participant to buy as low or sell as high as anyone else as long as the trader follows exchange rules.

Electronic trading has eliminated the need for exchanges to be physical places. Many traditional trading floors are closing, and orders and executions are now all communicated electronically. The London Stock Exchange and the NASDAQ Stock Market

are completely electronic, as is Eurex, a major futures exchange. Many others, offer both floor and electronic trading. The NYSE bought the electronic trading platform Archipelago and is moving increasingly toward electronic trading, as is derivatives exchange CME Group, which maintains both open-outcry and electronic trading. Brazil’s BM&F maintained both until 2009.

Trading over the counter

Unlike exchanges, OTC markets have never been a “place.” They are less formal, although often well-organized, networks of trading relationships centered around one or more dealers. Dealers act as market makers by quoting prices at which they will sell (ask or offer) or buy (bid) to other dealers and to their clients or customers. That does not mean they quote the same prices to other dealers as they post to customers, and they do not necessarily quote the same prices to all customers. Moreover, dealers in an OTC security can withdraw from market making at any time, which can cause liquidity to dry up, disrupting the ability of market participants to buy or sell. Exchanges are far more liquid because all buy and sell orders as well as execution prices are exposed to one another. Some exchanges designate certain participants as dedicated market makers and require them to maintain bid and ask quotes throughout the trading day. OTC markets are less transparent and have fewer rules than exchanges. All of the securities and derivatives involved in the financial turmoil that began with a 2007 breakdown in the US mortgage market were traded in OTC markets.

OTC dealers convey their bid and ask quotes and negotiate execution prices by telephone, mass e-mail messages, and, increasingly, text messaging. The process is often enhanced through electronic bulletin boards where dealers post their quotes. Negotiating by phone or electronic message, whether customer to dealer or dealer to dealer, is known as bilateral trading because only the two market participants directly observe the quotes or execution.

Others in the market are not privy to the trade, although some brokered markets post execution prices and the size of the trade after the fact. But not everyone has access to the broker screens and not everyone in the market can trade at that price. Although the bilateral negotiation process is sometimes automated, the trading arrangement is not considered an exchange because it is not open to all participants equally.

There are essentially two dimensions to OTC markets. In

the customer market, bilateral trading occurs between dealers and their customers, such as individuals or hedge funds. Dealers often initiate contact with their customers through high-volume electronic messages called “dealer-runs” that list securities and derivatives and the prices at which they are willing to buy or sell them. In the interdealer market, dealers quote prices to each other and can quickly lay off to other dealers some of the risk they incur in trading with customers, such as acquiring a bigger position than they want. Dealers can contact other dealers directly so that a trader can call a dealer for a quote, hang up and call another dealer and then another, surveying several in a few seconds. An investor can make multiple calls to the dealers

Some types of market arrangements can very quickly become unstable.

to get a view of the market on the customer side. But customers cannot penetrate the market among dealers.

Interdealer segments

Some OTC markets, and especially their interdealer market segments, have interdealer brokers that help market participants get a deeper view of the market. The dealers send quotes to the broker who, in effect, broadcasts the information by telephone. Brokers often provide trading platforms such as dark pools to give their clients (the dealers) the ability to instantaneously post quotes to every other dealer in the broker’s network. The bulletin boards show bid, ask, and, sometimes, execution prices. The broker screens are normally not available to end-customers, who are rarely aware of changes in prices and the bid-ask spread in the interdealer market. Dealers can sometimes trade through the screen or over the electronic system. Some interdealer trading platforms allow automated algorithmic (rule-based) trading like that of the electronic exchanges. Otherwise the screens are merely informative, and the dealer must trade through the broker or call other dealers directly to execute a trade.

Electronic trading has changed the trading process in many OTC markets and sometimes blurred the distinction between traditional OTC markets and exchanges. In some cases, an electronic brokering platform allows dealers and some nondealers to submit quotes directly to and execute trades directly through an electronic system. This replicates the multilateral trading that is the hallmark of an exchange—but only for direct participants. However dealers resist participation of nondealers and accuse them of taking liquidity without exposing themselves to the risks of providing it. Others criticize dealers for trying to prevent competition that would compress bid-ask spreads in the market. Unlike an exchange, in which every participant has access, these electronic arrangements can treat participants differently based on, say, their size or credit rating. Moreover clearing and

settlements are still left to the buyer and seller, unlike in exchange transactions, where trades are matched up and guaranteed by the exchange.

OTC markets and the financial crisis

The architecture of OTC markets helps explain why structured securities (which divide the risk of the underlying assets into several slices, each of which is sold separately) faced problems during the recent financial crisis. Credit derivatives, commercial paper, municipal bonds, and securitized student loans also faced problems. All were traded on OTC markets, which were liquid and functioned pretty well during normal times. But they failed to demonstrate resilience to market disturbances and became illiquid and dysfunctional at critical times.

That led to two serious complications—the inability to value one’s holdings and the inability to sell them:

- Without liquid and orderly markets, there was no price discovery process and in turn no easy and definitive way to value the securities. The failure of the price discovery process aggravated the problems at banks and other financial firms during the recent crisis by making it more difficult to meet disclosure and reporting requirements. Not only were there no efficient direct market prices, there were often no benchmark prices (prices of assets similar to the one being valued). As a result, the assets and positions once valued at market prices were instead valued through models sometimes not adequately informed by benchmark prices. These problems further depressed prices of affected securities.

- Dealers, facing a crunch on the funding side of their balance sheets and holding excessive illiquid assets on the other, withdrew from the markets. The jump in volatility made it dangerous and expensive for dealers to continue to make markets. Without dealers, there was no trading, especially in securities such as collateralized debt obligations, certain municipal securities, and credit derivatives. With no buyers, investors could not reduce losses by trading out of losing positions and could not sell those positions to meet calls for more margin or collateral to pledge against loans they had taken out to buy those instruments. This illiquidity in OTC markets contributed to the depth and breadth of the financial crisis.

The major regulatory reform underway in the United States, European Union, and other developed financial markets are directly addressing these issues. In some cases trading is shifting from OTC to exchange markets. In others, post-trade clearing of OTC trades is moving to clearinghouses (also known as central clearing counterparties). Trade reporting for OTC transactions is also a part of reform. The role of the dealer in OTC markets is not, however, being explicitly addressed except through possibly higher capital requirements. **FD**

RANDALL DODD works at the US Treasury Department and is a former advisor in the IMF’s Monetary and Capital Markets Department.

IV

ECONOMICS IN ACTION

Financial Services: Getting the Goods

How consumers and businesses acquire financial goods such as loans and insurance

Irena Asmundson

THERE IS A LONG-STANDING DEBATE about how much and what kind of regulations are appropriate for financial services.

Among the things money can buy, there is a distinction between a good (something tangible that lasts, whether for a long or short time) and a service (a task that someone performs for you). A financial service is not the financial good itself—say a mortgage loan to buy a house or a car insurance policy—but something that is best described as the process of acquiring the financial good. In other words, it involves the transaction required to obtain the financial good. The financial sector covers many different types of transactions in such areas as real estate, consumer finance, banking, and insurance. It also covers a broad spectrum of investment funding, including securities (see box).

But distinctions within the financial sector are not neat. For example, someone who works in the real estate industry, such as a mortgage broker, might provide a service by helping customers find a house loan with a maturity and interest rate structure that suits their circumstances. But those customers could also borrow on their credit cards or from a commercial bank. A commercial bank takes deposits from customers and lends out the money to generate higher returns than it pays for those deposits. An investment bank helps firms raise money. Insurance companies take in premiums from customers who buy policies against the risk that a covered event—such as an automobile accident or a house fire—will happen.

Intermediation

At its heart, the financial sector intermediates. It channels money from savers to borrowers, and it matches people who want to lower risk with those willing to take on that risk. People saving for retirement, for example, might benefit from intermediation. The higher the return future retirees earn on their money, the less they need to save to achieve their target retirement income and account for inflation. To earn that return requires lending to someone who will pay for the use of the money (interest). Lending and collecting payments are complicated and risky, and savers often don't have the expertise or time to do so. Finding an intermediary can be a better route.

Some savers deposit their savings in a commercial bank, one of the oldest types of financial service providers. A commercial bank takes in deposits from a variety of sources and pays interest to the depositors. The bank earns the money to pay that interest by lending to individuals or businesses. The loans could be to a

person trying to buy a house, to a business making an investment or needing cash to meet a payroll, or to a government.

The bank provides a variety of services as part of its daily business. The service to depositors is the care the bank takes in gauging the appropriate interest rate to charge on loans and the assurance that deposits can be withdrawn at any time. The service to the mortgage borrower is the ability to buy a house and pay for it over time. The same goes for businesses and governments, which can go to the bank to meet any number of financial needs. The bank's payment for providing these services is the difference between the interest rates it charges for the loans and the amount it must pay depositors.

WHAT DO THEY DO?

These are some of the foremost among the myriad financial services.

Insurance and related services

- Direct insurers pool payments (premiums) from those seeking to cover risk and make payments to those who experience a covered personal or business-related event, such as an automobile accident or the sinking of a ship.
- Reinsurers, which can be companies or wealthy individuals, agree, for a price, to cover some of the risks assumed by a direct insurer.
- Insurance intermediaries, such as agencies and brokers, match up those seeking to pay to cover risk with those willing to assume it for a price.

Banks and other financial service providers

- Accept deposits and repayable funds and make loans: Providers pay those who give them money, which they in turn lend or invest with the goal of making a profit on the difference between what they pay depositors and the amount they receive from borrowers.
- Administer payment systems: Providers make it possible to transfer funds from payers to recipients and facilitate transactions and settlement of accounts through credit and debit cards, bank drafts such as checks, and electronic funds transfer.
- Trade: Providers help companies buy and sell securities, foreign exchange, and derivatives.
- Issue securities: Providers help borrowers raise funds by selling shares in businesses or issuing bonds.
- Manage assets: Providers offer advice or invest funds on behalf of clients, who pay for their expertise.

Another type of intermediation is insurance. People could save to cover unexpected expenses just as they save for retirement. But retirement is a more likely possibility than events such as sickness and auto accidents. People who want to cover such risks are generally better off buying an insurance policy that pays out in the event of a covered event. The insurance intermediary pools the payments (called premiums) of policy buyers and assumes the risk of paying those who get sick or have an accident from the premiums plus whatever money the company can earn by investing them.

Providers of financial services, then, help channel cash from

Governments oversee many financial services because of their importance to the economy.

savers to borrowers and redistribute risk. They can add value for the investor by aggregating savers' money, monitoring investments, and pooling risk to keep it manageable for individual members. In many cases the intermediation includes both risk and money. Banks, after all, take on the risk that borrowers won't repay, allowing depositors to shed that risk. By having lots of borrowers, they are not crippled if one or two don't pay. And insurance companies pool cash that is then used to pay policy holders whose risk is realized. People could handle many financial services themselves, but it can be more cost effective to pay someone else to do it.

Cost of services

How people pay for financial services can vary widely, and the costs are not always transparent. For relatively simple transactions, compensation can be on a flat-rate basis (say, \$100 in return for filing an application). Charges can also be fixed (\$20 an hour to process loan payments), based on a commission (say, 1 percent of the value of the mortgage sold), or based on profits (the difference between loan and deposit rates, for example). The incentives are different for each type of compensation, and whether they are appropriate depends on the situation.

Regulation

Financial services are crucial to the functioning of an economy. Without them, individuals with money to save might have trouble finding those who need to borrow, and vice versa. And without financial services, people would be so intent on saving to cover risk that they might not buy very many goods and services.

Moreover, even relatively simple financial goods can be complex, and there are often long lags between the purchase of a service and the date the provider has to deliver the service. The market for services depends a great deal on trust. Customers (both savers and borrowers) must have confidence in the advice and information they are receiving. For example, purchasers of life insurance count on the insurance company being around when they die.

They expect there will be enough money to pay the designated beneficiaries and that the insurance company won't cheat the heirs.

The importance of financial services to the economy and the need to foster trust among providers and consumers are among the reasons governments oversee the provision of many financial services. This oversight involves licensing, regulation, and supervision, which vary by country. In the United States, there are a number of agencies—some state, some federal—that supervise and regulate different parts of the market.

Financial sector supervisors enforce rules and license financial service providers. Supervision can include regular reporting and examination of accounts and providers, inspections, and investigation of complaints. It can also include enforcement of consumer protection laws, such as limits on credit card interest rates and checking account overdraft charges. However, the recent sudden growth in the financial sector, especially as a result of new financial instruments, can tax the ability of regulators and supervisors to rein in risk. Regulations and enforcement efforts cannot always prevent failures—regulations may not cover new activities, and wrongdoing sometimes escapes enforcement. Because of these failures, supervisors often have the authority to take over a financial institution when necessary.

The role of mortgage-backed securities in the global financial crisis is an example of new financial instruments leading to unexpected consequences. In this case, financial firms looking for steady income streams bought mortgages from the originating banks and then allocated payments to various bonds, which paid according to the mortgages' underlying performance. Banks benefited by selling the mortgages in return for more cash to make additional loans, but because the loan makers did not keep the loans, their incentive to check borrowers' creditworthiness eroded. The mortgages were riskier than the financial firms that bought them anticipated, and the bonds did not pay as much as expected. Borrowers were more likely to default because of their lower income, which reduced the amount bondholders took in—both of which hurt gross domestic product growth. Mortgage-backed securities were initially intended to help mitigate risk (and could have done so under the right circumstances), but they ended up increasing it.

Productive uses

Financial services help put money to productive use. Instead of stashing money under their mattresses, consumers can give their savings to intermediaries who might invest them in the next great technology or allow someone to buy a house. The mechanisms that intermediate these flows can be complicated, and most countries rely on regulation to protect borrowers and lenders and help preserve the trust that underpins all financial services. [FD](#)

IRENA ASMUNDSON is the chief economist at the California Department of Finance and a former IMF staff member.

Strategic Thinking

Game theory analyzes behavior when decisions must take into account the potential actions of opponents

Sarwat Jahan and Ahmed Saber Mahmud

ANYONE WHO HAS HAD to make a strategic decision taking into account what others will do has used game theory. Think of a game of chess. The outcome of the game depends not only on one participant's move, but also on the actions of the opponent. When choosing a course of action—in other words, a “strategy”—a player must take into account the opponent's choices. But the opponent's choices in turn are based on thinking about the course of action the player might take. Game theory studies this interdependent decision making and identifies the optimal strategy—that is, the best course of action—for each player in response to the actions of others and how this leads to an equilibrium outcome, in which no players have a reason to change their strategy.

Because situations involving interdependent decisions arise frequently, so does the potential application of game theory in strategic thinking. Businesses competing in a market, diplomats negotiating a treaty, gamblers betting in a card game, and even those contemplating proposing marriage can use game theory.

The science of strategy

The earliest example of a formal game-theoretic analysis was by Antoine Cournot in 1838, when he studied the business behavior of two firms (a duopoly in economic parlance) with identical costs producing the same products but vying for maximum profits in a limited market. The mathematician Émile Borel suggested a formal theory of games in 1921, which was furthered by Princeton mathematician John von Neumann later in the decade. But game theory became a field in its own right after the publication of *Theory of Games and Economic Behavior* by von Neumann and economist Oskar Morgenstern in 1944. They studied “zero-sum” games, in which the interests of two players are so strictly opposed that the games are pure conflict—with one person's gain always resulting in the other's loss. A good example is chess, which has a winner and a loser. But games do not have to be zero sum. Players can engage in positive-sum games—for example, jointly writing this article generated benefits for both authors/players and was a win-win game. Similarly, games can result in mutual harm (negative sum)—for example, the failure to prevent a war. John Nash treated the more general and realistic case in which a game

involves a mixture of common interests and rivalries and any number of players. Other theorists—most notably Reinhard Selten and John Harsanyi, who shared the 1994 Nobel Prize in economics with Nash—studied even more complex games with sequences of moves, and games in which one player has more information than the others.

What's in a game?

A game is the strategic interaction between two or more players. Each player has a set of possible strategies. For each strategy players pick, they receive a payoff, which is usually represented by a number. That payoff depends on the strategies of all players in the game. Payoffs can also have different meanings. For example, they can signify an amount of money or the number of years of happiness. Game theory presumes that players act rationally—that is, that they seek to maximize their own payoffs.

The prisoner's dilemma is perhaps the best-known example in game theory. Two bank robbers are arrested and are interviewed separately. The robbers can confess or remain silent. The prosecutor offers each the following scenario. If one confesses and the other stays silent, the one who admits the crime will go free while his accomplice will face 10 years behind bars. If both confess, each will go to prison for five years, while if both remain silent each will go to jail for a year.

If Robber A confesses, then it is better for Robber B to confess and receive 5 years in jail than to remain silent and serve 10 years. On the other hand, if Robber A does not confess, it still is better for Robber B to confess and go free than remain silent and spend a year in jail. In this game it is always better for Robber B to confess no matter what Robber A does. That is, the dominant strategy is to confess. Because each player has the same payoff structure, the outcome of the game is that rational players will confess and both will end up in jail for five years. The dilemma is that if neither confesses, each gets one year in jail—a preferable outcome for both. Can this dilemma be resolved? If the game is repeated without a foreseeable end, then both players can reward or punish the other for their respective actions. This can lead to the mutually beneficial outcome in which neither confesses and each spends a year in prison. A real-life example would be collusion between two competing firms to maximize

their combined profit.

Sometimes there is more than one equilibrium in a game. Take the following example: A couple is planning a night out. Above all, they value spending time together, but the husband likes boxing while the wife prefers the ballet. They both must decide independently of the other what they will do, that is they must decide simultaneously. If they choose the same activity, they will be together. If they choose different activities, they will be separate. Spouses get a value of 1 if they get their favorite entertainment; the value 2 is assigned to being together. This

Game theory presumes that players act rationally—that is, that they seek to maximize their own payoffs.

leads to a payoff matrix that maximizes satisfaction when both pick the same activity (see chart, left panel).

If players sacrifice for their partners, they obtain the worst outcome: each goes to the undesired event, but alone, and the payoff is zero. If both choose the event they like, the outcome is better, but neither has the pleasure of the other’s company, so the payoff is 1 for each. If the wife chooses ballet, the optimal result occurs when the husband also chooses ballet. Hence going to the ballet is an equilibrium with a payoff of 3 for the wife and 2 for the husband. By similar logic, when both attend the boxing match, there is also an equilibrium—in which the husband’s payoff is 3 and the wife’s 2. Therefore, this game has two equilibriums.

Modifying this game by letting the players move sequentially—that is, each player is aware of the other’s previous

action—will yield a single equilibrium (see chart, right panel). If the wife moves first and decides to go to the ballet, the husband’s best option would be to go to the ballet. If the wife chooses boxing, the husband would definitely choose to go to the match. The wife’s basic strategy will be to “look ahead and reason backward.” The wife can anticipate where her husband’s decision will lead and use this information to calculate her best decision: in this case choosing ballet. In this type of game, there is a clear advantage to moving first.

Nuclear deterrence

The prisoner and spousal games involve only two players, and each has complete information about the game. Games become more complicated when more players are involved or if players do not all have access to the same information. It is not surprising that game theory has been applied to analysis of the nuclear arms race. The 2005 Nobel Prize winner in economics, Thomas Schelling, showed that the power to retaliate is a more effective deterrent than the ability to withstand an attack and proved that uncertainty about retaliation—which keeps the enemy guessing—may preserve peace more effectively than the threat of certain retaliation.

Game theory has been used to analyze market power and how to regulate monopolies to protect consumers—an avenue of research that earned Jean Tirole the 2014 Nobel Prize in economics. Game theory has also revolutionized the field of information economics by studying games in which some players have more information than others. Three economists earned the Nobel Prize jointly in 2001 for their seminal work on games with asymmetric information: George Akerlof on the market for used cars, Michael Spence on signaling in labor markets through education, and Joseph Stiglitz on self-screening in insurance markets. Strategic thinking plays an important role in the theory of contracts as well, for which Oliver Hart and Bengt Holmstrom won the 2016 Nobel Prize jointly.

Game theory has even been applied in evolutionary biology, where the players (in this case animals) are not necessarily rational beings. The hawk-dove game developed by John Maynard Smith in 1982 involves aggressive and nonaggressive behavior and provides insight into the survival of species. Game theory is being used by some to forecast the fate of the European Union. As long as there are interactive decisions to be made, game theory will be applied to inform them. **FD**

SARWAT JAHAN is a senior economist in the IMF’s Asia and Pacific Department, and **AHMED SABER MAHMUD** is the associate director in the Applied Economics Program at Johns Hopkins University.

THE GAME

Whether the players, in this case, spouses, make entertainment decisions simultaneously or sequentially, they maximize payoff when both attend the same event.

		SIMULTANEOUS MOVES		SEQUENTIAL MOVES			
		Wife		Wife		Husband	
Husband	Ballet	(3, 2)	(0, 0)	Ballet	Boxing	Ballet	Boxing
	Boxing	(1, 1)	(2, 3)	Boxing	(1, 1)	Boxing	(0, 0)

Note: The payoff for the wife is in red, for the husband in black. The payoff amounts for each are 2 points if they attend the same event, 1 point if they attend their preferred event (ballet for the wife, boxing for the husband), and zero if they attend the event they do not like. In the simultaneous game, each makes the decision without knowledge of the other’s choice. In the sequential game, the person picking second knows what the other person chose.

Taxes in Practice

It is hard to design a fair and efficient revenue system

Ruud De Mooij and Michael Keen

TAX POLICY is often guided by simple rules of thumb. Sometimes, they are strikingly right. But sometimes they can be dangerously misleading.

There is an adage, for example, that “an old tax is a good tax.” That may be true for, say, the property tax. But taxes on windows and beards are long gone, import tariffs are in decline, and new levies, such as the value-added tax, have gained ground. Designing a good tax system requires more than a good slogan.

In “Taxing Principles” (p. 72), we deal with the basic principles of taxation. Here we apply them to some central and current tax policy debates.

Personal income

The great appeal of the personal income tax is that it taxes people on an indicator of their ability to pay and can collect progressively more from those with higher incomes. But the indicator is imperfect, because the government cannot be sure whether a high income results from intrinsic talent or luck—which will not be affected by taxation—or hard work and creativity—which might be. Taxing income might not only discourage effort (not just hours worked, but also, for example, entrepreneurial activity and striving for promotion), but can also give rise to tax avoidance and evasion.

The design of the personal income tax, therefore, revolves around a fundamental trade-off: progressive taxes support equity objectives, but can reduce efficiency. As long as people have differing views on what is equitable, there will never be universal agreement on the best tax schedule. But careful theory and empirical evidence have illuminated key considerations.

There is, for instance, the need to consider not only the personal income tax but all taxes and all income support measures—such as the Earned Income Tax Credit in the United States, which gives money to low-wage workers in amounts gradually reduced as income increases. Income support is simply negative income taxation and, when withdrawn as income rises, acts just like a tax on that additional income.

There is a strong case for subsidizing earnings of low-wage workers, because their willingness to work is particularly sensitive to tax, and it is cheaper to ensure their well-being when they are working and so generating some income themselves. But while the average tax rate at the bottom will therefore typically be

negative, the effective marginal rate—the additional tax paid (or benefit not received) when income rises by one dollar—should be positive. Otherwise the subsidy will benefit even more those who need it less. Targeting income support at the poorest limits the revenue cost of earnings subsidies and can be consistent with efficient redistribution, even though it may create high effective marginal tax rates for the poorest.

The proper tax rate structure for high-income earners has always been contentious. Many have concluded that the best-off could be taxed at marginal rates of 60 percent or more without leading to reduced effort or avoidance or evasion large enough to cause the amount of tax they pay to fall. If raising revenue were the only concern, that would be fine. But well-off taxpayers would suffer, which presumably matters for overall social welfare. Moreover, some analysts believe that the calculations underlying the optimal marginal rate fail to capture adverse effects on entrepreneurship.

In broad qualitative terms, the optimal marginal rate structure should thus have a U shape—starting high to recoup support to the very poorest, falling to preserve incentives for the people in the middle, and finally rising to secure revenue from the better-off. This runs counter to the idea that marginal rates should always increase with income, but is consistent with the more basic notion that the average rate should increase with income. All this, however, still leaves plenty of room for debate on the precise shape of that U.

Capital income conundrum

Capital income—interest, dividends, and capital gains—is in most countries received largely by the better-off. High taxes on capital income (or on the underlying wealth) are therefore often viewed as a good way to address inequities. But theory offers further perspectives on this issue.

Capital income enables consumption in the future. Taxing it increases the cost in terms of consumption forgone today. Prudent people who prefer to postpone consumption (or transfer it to their heirs) will be taxed more than those who do not. Some see this as violating horizontal equity (the principle that those who are in all relevant respects identical should be treated the same) on the grounds that time preference is not a legitimate basis on which to differentiate tax liabilities. Moreover, by discouraging future consumption, a tax on capital income may

create relatively large deadweight losses (those incurred from transferring resources out of the private sector).

What all this implies is intensely debated among public finance economists. At one extreme is the view that because it distorts behavior so much, the optimal tax on capital income is zero, with redistribution better achieved by a progressive tax on labor income alone. At the opposite extreme is the view that labor and capital income should be taxed identically—for many years the most popular view. Neither view is on entirely firm theoretical grounds. What has become clear, however, is that the desirable tax rate on capital income, even if not zero, may well differ from that on labor income—not least because capital is more mobile internationally, making it harder to tax without driving the base abroad. Many countries now employ some form of dual income tax, taxing capital income separately from labor income, and at a relatively low rate.

Corporate tax controversies

The notion of tax incidence—who ultimately bears the real burden of a tax—is key when it comes to business taxation—and can lead to the surprising conclusion that much of the incidence might fall on workers. Take an economy that is small in world capital markets, and so must take as given the after-tax rate of return on investment: investors will move their capital abroad if they earn less than this rate. If a country now taxes the returns that investors earn there, the before-tax rate of return will have to rise by enough to leave the after-tax return unchanged. Achieving this requires an outflow of capital. But that outflow leads to a lower domestic capital-labor ratio, which reduces labor productivity—and, in turn, wages. So workers, not shareholders, bear the real incidence of the corporate income tax. Because it is more efficient to tax workers directly than indirectly through the corporate tax, the optimal corporate income tax for such an economy is zero.

But there are important qualifications.

First, normal capital returns (the minimum return required by investors) should be distinguished from above-normal returns—called “rents.” Unlike normal returns, rents that are specific to a particular country can be taxed without affecting investment (think, for instance, of high profits often earned from natural resources). The traditional corporate income tax, however, is not a rent tax because it taxes all returns to equity, both normal and above normal. It can be turned into a rent tax, though—for instance by allowing companies to reduce their taxable income through a deduction for normal equity returns. Some countries have moved in this direction.

Second, practical considerations loom large. The corporate income tax, for example, has the merit that it effectively taxes earnings that companies retain, which are hard to tax at the personal level. Similarly, if there were no corporate tax, small businesses could escape tax by incorporating and labeling their earnings as capital income. Moreover, in many developing

economies it is relatively easy to collect taxes from a few large companies.

Taxing consumption?

A uniform tax on consumption—applying the same rate to all goods and services—is broadly equivalent to a uniform tax on wage and profit income. It simply operates on the other side of an individual’s budget, so its distortions on the labor market should be similar too. Because income taxes fit better with the principle that people should be taxed on their ability to pay, why should governments tax consumption at all?

Progressive taxes support equity objectives, but can reduce efficiency.

There are practical reasons to do so: taxing both income and consumption reduces compliance risk by diversifying the government’s revenue base. But there are also more fundamental justifications, such as taxing particular types of consumption to address externalities, which are effects, good or bad, on those not involved in the underlying transaction—pollution, for example. Such taxes could also address other problematic behaviors, such as drinking and smoking. Another reason is that differential rates on different types of goods and services might help reduce tax-induced disincentives to work. Empirically, however, it has proved hard to identify elements of rate differentiation that are warranted on such efficiency grounds—perhaps with a few exceptions, such as child care services.

Many feel that necessities such as food should be taxed at especially low rates because the poor spend a large proportion of their income on them. But this is a costly way to pursue equity, since while the poor spend a larger proportion of their income on necessities, the rich spend a larger absolute amount and so benefit most from a low rate. Almost all advanced economies, and many others too, should have devices better suited to pursuing their fairness goals—such as income-related transfers or other forms of cash support to the neediest, or public support for housing, health care, and basic education.

There is a fair degree of professional consensus that a uniform, broad-based consumption tax is a sensible benchmark for good policy—with little convincing rationale (externalities aside) for rate differentiation. This is one simple rule of thumb that gives good, practicable advice—but it rests on quite detailed empirical and theoretical reasoning. Policymakers must be wary of the many that don’t. **FD**

RUUD DE MOOIJ is a division chief and **MICHAEL KEEN** is a deputy director, both in the IMF’s Fiscal Affairs Department.

Taxing Principles

Making the best of a necessary evil

Ruud De Mooij and Michael Keen

IT IS HARD to think of anything that some government, at some point, has not taxed. Playing cards, urine, fireplaces, slaves, religious minorities, and windows have all at some point attracted the attention of the tax collector. Nowadays we think of income taxes, value-added taxes, taxes on cigarettes, and the like as the key revenue instruments. But the basic principles for understanding and evaluating all taxes are much the same. In this, the second of two articles on taxation, we examine these principles. In “Taxes in Practice,” we applied them to some current controversies.

The Organisation for Economic Co-operation and Development defines a tax as a “compulsory, unrequited payment to government.” That is, you have to pay it, and you don’t get anything back—at least not directly. (You may derive some benefit from the public spending your payment helps finance, but if not—well, from the perspective of tax collection—that’s just too bad.)

Importantly, however, many policy instruments that are not in a legal sense taxes have much the same effect. Social contributions are a prime example. These are payments linked to an individual’s labor or business income that confer some entitlement to pensions or other social benefits. The personalized benefit means that these are not, strictly speaking, taxes. But the link between payments and contributions is often so far from actuarially fair, and the prospective benefits so remote, that their effect is likely to be very similar to that of an outright tax.

Efficient taxation

A tax transfers resources from the private to the public sector, and so inescapably imposes a real loss on the private sector, leaving aside any benefit from whatever the tax revenue finances. But almost all taxes will cause more harm than that because they typically drive a wedge between the price a buyer pays for something and the amount the seller receives—which may prevent some mutually beneficial trade. Taxing labor income, for instance, means that the cost to an employer of hiring someone exceeds what the employee receives. A worker may be willing to accept a job that pays (at least) \$100 and an employer willing

to pay (no more than) that, but imposing tax on the wage will prevent this trade from happening. This welfare loss from taxation over and above the loss from the direct transfer of real resources out of the private sector is known as deadweight loss (or excess burden) and is what economists have in mind when they talk of tax distortions. (In the example above, because the worker is not hired, no tax is paid, but the deadweight loss is still positive.)

Efficient tax design aims to minimize these losses, whose size depends on two main factors. First, losses are bigger the more responsive the tax base is to taxation. Suppose for instance that the demand for a worker’s labor is completely inelastic, meaning that an employer is willing to pay any price for the worker’s services. Then, with a 20 percent tax rate, as in the example above, the employee would receive \$100, but the employer’s cost would be \$120. The employee is hired, and there is no distortion. But when one side of the market has an alternative to the transaction being taxed, distortions arise, and the easier it is to exercise that alternative, the larger the distortion. And this is true (given a few more assumptions) whether it is decisions like hiring that are affected by the tax or decisions to avoid or evade tax. Second, the loss increases more than proportionately with the tax rate. Increasing a distortion, by setting a higher tax rate, is more harmful when there is a large distortion already in place.

Two prescriptions for efficient tax policy follow: tax at a higher rate things in inelastic demand or supply, and tax as many things as possible to keep rates low. Both of these principles require qualification—because in some cases following these general rules can have adverse consequences.

Taxing a good whose demand is inelastic, for instance, will have little effect on the quantity of that good demanded, but it leaves less to spend on other goods, which can lead to large changes in other markets (more on this in “Taxes in Practice,” p. 68).

And the injunction to seek as broad a tax base as possible must be tempered by one of the most powerful precepts in public finance: transactions between businesses should not be taxed. This is because taxes drive a wedge between buying and selling

prices for intermediate inputs, which is likely to lead firms to choose different inputs than they would in the absence of the tax. As a result, firms end up producing less than they could. Broadening the tax base by including intermediate transactions can, therefore, be very bad news for efficiency. A turnover tax, for instance—charged on all transactions, including business-to-business sales—would have a much bigger base than a tax on final consumption (such as a value-added tax) and could raise the same revenue at a much lower rate. But it would also be much more distortionary.

Another set of qualifications arises from externalities—effects (good or bad) on those not involved in the underlying decisions. Environmental damage, such as climate change, is the leading

Efficient tax design aims to minimize deadweight losses.

example. Here a corrective tax may be called for. Such a corrective tax, also called a Pigovian tax (after economist Arthur C. Pigou, who proposed it), is designed to distort behavior in a desired direction, including, if need be, the actions of businesses—while of course also putting the revenue raised to good use.

Bearing, and sharing, the burden

The person who ultimately bears the real burden of a tax may not be the one legally responsible for remitting payment. For instance, in the example above, when the demand for labor was perfectly fixed, the \$20 loss was suffered by the employer, not the worker—and that would be true regardless of which one was legally responsible for making the payment to the government. This illustrates too the general principle that the burden of a tax—its effective incidence—falls more heavily on the side of the transaction with the least elastic response—that is, the one that finds it more difficult to shift out of the activity being taxed.

These implications are often ignored. Take the current outrage over the small amount of corporate tax paid by many multinational corporations. Corporations are not people, and only real people—shareholders, workers, customers—can bear a real burden from taxes. The debate over corporate tax makes little sense without consideration of who really gains when the effective tax rates are low.

Fairness in taxation is always a major issue, with two main dimensions. Vertical equity concerns the treatment of those with different incomes. The impact of a tax system on this dimension depends on its progressivity—that is, how rapidly the share of income taken by tax increases with the level of income. Horizontal equity holds that those who are in all relevant respects identical should be treated the same.

Each of these concepts is less straightforward than it may seem. Clearly people have different views on the appropriate degree of progressivity. But people may also disagree, for instance, on whether progressivity should be assessed in terms of annual income—a pretty arbitrary period of measurement—or lifetime income. A consumption tax may look regressive relative to annual income but much less so relative to expenditure, which may be a better indicator of an individual’s lifetime income.

And the idea of horizontal equity may not seem controversial, but what does “identical” mean for this purpose? Is it acceptable to differentiate taxes by age, by marital status, across regions, by gender, by height? And what about implicit differentiation? Is a heavy tax on aftershave lotion, overwhelmingly consumed by men, horizontally inequitable?

Collecting taxes

The dividing line between tax evasion (illegal) and avoidance (legal) is not as clear-cut as it may sound—highly paid tax lawyers spend much time testing the distinction. Both are major concerns in all countries. There are challenges here for both the design of taxes and their implementation. On the policy side, tax incentives to encourage particular activities, for instance, all too often provide opportunities for avoidance and evasion.

Tax administrations are on the front line in the fight against failure to pay taxes. Writing tax rules that are easy to understand (though simplicity in tax design is difficult to achieve, given the range of objectives and circumstances to be covered) and easy to find helps make things easy for those who simply want to pay whatever is due. (Roman emperor Caligula’s tax rules were made public only in small type and in an awkward place.) Ultimately, the trick for tax administrations is to ensure that the probability of detecting noncompliance—and the penalties that follow—is high enough to encourage compliance while supporting and reflecting widespread willingness to follow the rules. And a good tax administration must do all that while minimizing both its own expenses (administration costs) and those of taxpayers (compliance costs).

Sometimes the various objectives discussed here point in the same direction—for example, when tariffs (taxes on imports) are replaced by a consumption tax at the same rate. The switch leaves the price of imports to consumers unchanged, but increases government revenue (because the tax is now also collected on domestically sourced sales) and improves efficiency by reducing trade protectionism. But such instances are rare. The real difficulty for taxation arises when the objectives conflict—which we examined in “Taxes in Practice” (p. 68). [FD](#)

RUUD DE MOOIJ is a division chief and **MICHAEL KEEN** is a deputy director, both in the IMF’s Fiscal Affairs Department.

Inflation Targeting: Holding the Line

Central banks use interest rates to steer price increases toward a publicly announced goal

Sarwat Jahan

IN RECENT YEARS, many central banks, the makers of monetary policy, have adopted a technique called *inflation targeting* to control the general rise in the price level. In this framework, a central bank estimates and makes public a projected, or “target,” inflation rate and then attempts to steer actual inflation toward that target, using such tools as interest rate changes. Because interest rates and inflation rates tend to move in opposite directions, the likely actions a central bank will take to raise or lower interest rates become more transparent under an inflation targeting policy. Advocates of inflation targeting think this leads to increased economic stability.

Why inflation targeting?

In general, a monetary policy framework provides a nominal anchor to the economy. A nominal anchor is a variable policy-makers can use to tie down the price level. One nominal anchor central banks used in the past was a currency peg—which linked the value of the domestic currency to the value of the currency of a low-inflation country. But this approach meant that the country’s monetary policy was essentially that of the country to which it pegged, and it constrained the central bank’s ability to respond to such shocks as changes in the terms of trade (the value of a country’s exports relative to that of its imports) or changes in the real interest rate. As a result, many countries began to adopt flexible exchange rates, which forced them to find a new anchor.

Many central banks then began targeting the growth of money supply to control inflation. This approach works if the central bank can control the money supply reasonably well and if money growth is stably related to inflation over time. Ultimately, monetary targeting had limited success because the demand for money became unstable—often because of innovations in the financial markets. As a result, many countries with flexible exchange rates began to target inflation more directly, based on their understanding of the links or “transmission mechanism” from the central bank’s policy instruments (such as interest rates) to inflation.

How does inflation targeting work?

Inflation targeting is straightforward, at least in theory. The central bank forecasts the future path of inflation and compares it with the target inflation rate (the rate the government believes is appropriate for the economy). The difference between the forecast and the target determines how much monetary policy has to be adjusted. Some countries have chosen inflation targets

with symmetrical ranges around a midpoint, while others have identified only a target rate or an upper limit to inflation. Most countries have set their inflation targets in the low single digits. A major advantage of inflation targeting is that it combines elements of both “rules” and “discretion” in monetary policy. This “constrained discretion” framework combines two distinct elements: a precise numerical target for inflation in the medium term and a response to economic shocks in the short term.

Rather than focusing on achieving the target at all times, the approach has emphasized achieving the target over the medium term—typically over a two- to three-year horizon. This allows policy to address other objectives—such as smoothing output—over the short term. Thus, inflation targeting provides a rule-like framework within which the central bank has the discretion to react to shocks. Because of inflation targeting’s medium-term focus, policymakers need not feel compelled to do whatever it takes to meet targets on a period-by-period basis.

What is required?

Inflation targeting requires two things. The first is a central bank able to conduct monetary policy with some degree of independence. No central bank can be entirely independent of government influence, but it must be free in choosing the instruments to achieve the rate of inflation that the government deems appropriate. Fiscal policy considerations cannot dictate monetary policy. The second requirement is the willingness and ability of the monetary authorities not to target other indicators, such as wages, the level of employment, or the exchange rate.

Having satisfied these two basic requirements, a country can, in theory, conduct a monetary policy centered on inflation targeting. In practice, the authorities may also take certain preliminary steps:

- Establish explicit quantitative targets for inflation for a specific number of periods ahead.
- Indicate clearly and unambiguously to the public that hitting the inflation target takes precedence over all other objectives of monetary policy.
- Set up a model or methodology for inflation forecasting that uses a number of indicators containing information about future inflation.
- Devise a forward-looking operating procedure through which monetary policy instruments are adjusted (in line with the assessment of future inflation) to hit the chosen target.

TARGETING INFLATION

Countries across the world have adopted inflation targeting irrespective of their income level.

COUNTRY	INFLATION TARGETING ADOPTION DATE	TARGET INFLATION RATE AT TIME OF ADOPTION	COUNTRY	INFLATION TARGETING ADOPTION DATE	TARGET INFLATION RATE AT TIME OF ADOPTION
New Zealand	1990	1 – 3	Philippines	2002	4 +/- 1
Canada	1991	2 +/- 1	Guatemala	2005	5 +/- 1
United Kingdom	1992	2 (point target)	Indonesia	2005	5 +/- 1
Australia	1993	2 – 3	Romania	2005	3 +/- 1
Sweden	1993	2 (point target)	Serbia, Republic of	2006	4 – 8
Czech Republic	1997	3 +/- 1	Turkey	2006	5.5 +/- 2
Israel	1997	2 +/- 1	Armenia	2006	4.5 +/- 1.5
Poland	1998	2.5 +/- 1	Ghana	2007	8.5 +/- 2
Brazil	1999	4.5 +/- 2	Uruguay ¹	2007	3 – 7
Chile	1999	3 +/- 1	Albania	2009	3 +/- 1
Colombia	1999	2 – 4	Georgia	2009	3
South Africa	2000	3 – 6	Paraguay	2011	4.5
Thailand	2000	0.5 – 3	Uganda	2011	5
Hungary	2001	3 +/- 1	Dominican Republic	2012	3 – 5
Mexico	2001	3 +/- 1	Japan	2013	2
Iceland	2001	2.5 +/- 1.5	Moldova	2013	3.5 – 6.5
Korea, Republic of	2001	3 +/- 1	India	2015	2 – 6
Norway	2001	2.5 +/- 1	Kazakhstan	2015	4
Peru	2002	2 +/- 1	Russia	2015	4

Sources: Hammond 2011; Roger 2010; and IMF staff calculations.

Note: Countries are classified as inflation targeters based on the IMF’s Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER) database.

¹Adoption date is based on the starting point when the interest rate became the main monetary policy instrument.

are exceptions, such as the United States, which explicitly adopted an inflation target of 2 percent in 2012) or have other objectives, such as promoting maximum employment and moderate long-term interest rates, in addition to stable prices.

On target?

It is difficult to distinguish between the specific impact of inflation targeting and the general impact of more far-reaching concurrent economic reforms. Nonetheless empirical evidence on the performance of inflation targeting is broadly, though not totally, supportive of the effectiveness of the framework in delivering low inflation, anchoring inflation expectations, and lowering inflation volatility. Moreover, these gains in inflation performance were achieved with no adverse effects on output and interest volatility.

Inflation targeters also seem to have been more resilient in turbulent environments. Recent studies have found that in emerging market economies, inflation targeting seems to have been more effective than alternative monetary policy frameworks in anchoring public inflation expectations. In some countries, notably in Latin America, the adoption of inflation targeting was accompanied by better fiscal policies. Often, it has also been accompanied by the enhancement of technical capacity in the central bank and improvement of macroeconomic data. Because inflation targeting also depends to a large extent on the interest rate channel to transmit monetary policy, some emerging market economies also took steps to strengthen and develop the financial sector. Thus, the monetary policy outcomes after the adoption of inflation targeting may reflect improved broader economic, not just monetary, policymaking.

Target practitioners?

Central banks from advanced, emerging market, and developing economies and from every continent have adopted inflation targeting (see table). Full-fledged inflation targeters are countries that make an explicit commitment to meet a specified inflation rate or range within a specified time frame, regularly announce their targets to the public, and have institutional arrangements to ensure that the central bank is accountable for meeting the target.

The first country to adopt inflation targeting was New Zealand, in December 1989. The only central banks to have stopped inflation targeting once they started it are Finland, Spain, and the Slovak Republic—in each case after they adopted the euro as their domestic currency. Armenia, the Czech Republic, Hungary, and Poland adopted inflation targeting while they were making the transition from centrally planned to market economies. Several emerging market economies adopted inflation targeting after the 1997 crisis, which forced a number of countries to abandon fixed exchange rate pegs.

There are also a number of central banks in more advanced economies—including the European Central Bank and the US Federal Reserve—that have adopted many of the main elements of inflation targeting but do not officially call themselves inflation targeters. These central banks are committed to achieving low inflation, but some do not announce explicit numerical targets (there

Not a panacea

Inflation targeting has been successfully practiced in a growing number of countries over the past 20 years, and many more countries are moving toward this framework. Over time, inflation targeting has proved to be a flexible framework that has been resilient in changing circumstances, including during the recent global financial crisis. Individual countries, however, must assess their economies to determine whether inflation targeting is appropriate for them or if it can be tailored to suit their needs. **FD**

SARWAT JAHAN is a senior economist in the IMF’s Asia and Pacific Department.

References:

Hammond, Gill. 2011. “State of the Art of Inflation Targeting.” *Centre for Central Banking Studies Handbook—No. 29*. London: Bank of England.
 Roger, Scott. 2010. “Inflation Targeting Turns 20.” *Finance & Development*. March: 46–49.

Regressions: An Economist Obsession

A basic statistical tool for distinguishing between correlation and causality

Rodney Ramcharan

READING IS AN IMPORTANT SKILL, and elementary school teachers have observed that the reading ability of their students tends to increase with their shoe size. To help boost reading skills, should policymakers offer prizes to scientists to devise methods to increase the shoe size of elementary school children? Obviously, the tendency for shoe size and reading ability to increase together does not mean that big feet cause improvements in reading skills. Older children have bigger feet, but they also have more developed brains. This natural development of children explains the simple observation that shoe size and reading ability have a tendency to increase together—that is, they are positively correlated. But clearly there is no relationship: bigger shoe size does not cause better reading ability.

In economics, correlations are common. But identifying whether the correlation between two or more variables represents a causal relationship is rarely so easy. Countries that trade more with the rest of the world also have higher income levels—but does this mean that trade raises income levels? People with more education tend to have higher earnings, but does this imply that education results in higher earnings? Knowing precise answers to these questions is important. If additional years of schooling caused higher earnings, then policymakers could reduce poverty by providing more funding for education. If an extra year of education resulted in a \$20,000 a year increase in earnings, then the benefits of spending on education would be a lot larger than if an extra year of education caused only a \$2 a year increase.

To help answer these types of questions, economists use a statistical tool known as regression analysis. Regressions are used to quantify the relationship between one variable and the other variables that are thought to explain it; regressions can also identify how close and well determined the relationship is. These days, running thousands of regressions has become commonplace and easy—although that was not always the

case (see box)—and, in fact, it is difficult to find an empirical economic study without a regression in it. Other fields, including sociology, statistics, and psychology, rely heavily on regressions as well.

How to run a regression

To illustrate how a regression works, let's take a closer look at the problem of trying to determine the returns to education. The government collects data on people's education level and their subsequent earnings. But people go to school for a variety of reasons—some find it easier to learn than others or are just more motivated to stay in school longer. Others may be successful pursuing nonscholastic careers and may still achieve high earnings. These varied reasons for attending school may affect earnings, making it difficult to know whether the correlation between schooling and earnings represents a causal relationship or is driven by some other factor. People who find it easier to learn in school may also find it easier to learn on the job, resulting in higher earnings. Thus, the positive correlation between higher earnings and education levels may reflect innate aptitude,

THE MAGIC OF COMPUTERS

Initial conceptualizations of regression date back to the 19th century, but it was really the technological revolution in the 20th century, making desktop computers a mainstay, that catapulted regression analysis into the stratosphere. In the 1950s and 1960s, economists had to calculate regressions with electromechanical desk calculators. As recently as 1970, it could take up to 24 hours just to receive the results of one regression from a central computer lab—and that was after spending hours or days punching computer cards. One wrong punch (a misspelled control word or incorrect data value) would invalidate the whole effort.

rather than the effects of education. Before a regression is run, a theoretical model can help explain how and why one “dependent” variable is determined by one or more “independent” or “explanatory” variables. Positing that an individual’s earnings depends on his or her level of education is an example of a simple model with one explanatory variable. A corresponding regression equation, assumed to be linear, would look like:

$$Y = a + bX$$

On the left-hand side is Y , our dependent variable, earnings. On the right-hand side are a , our constant (or intercept), and b , our coefficient (or slope) multiplied by X our independent (or explanatory) variable, education. The regression says in algebra that “earnings depend only on education and in a linear way”; the other explanatory factors, if there are any, are omitted.

But what if we think that the world is much more complicated

Regressions quantify the relationship between one variable and others that are thought to explain it.

and that a variety of factors might explain the impact of education on earnings? In that case, we would run a multiple-variable regression, which would look like:

$$Y = a + b_1X_1 + b_2X_2 + \dots$$

Now, we have several X variables to help explain Y earnings—like ability, intelligence, age, education, marital status, and parental education. The b coefficients simply measure the impact of each of these variables on earnings, assuming the other variables are constant.

Smarter is richer?

Let’s try running a regression on the basis of the theory that hourly wages (our dependent variable) depend on the level of education (our explanatory variable). We’ll assume that another possible explanatory variable—aptitude, as measured by intelligence quotient (IQ) tests—has no effect on wages separate from any effect it may have through education. We plug in all of the data on earnings and education levels. We run the regression and find:

$$Y = 5.40 + 1.06 \text{ EDU}$$

The b coefficient tells us that an additional year of education is associated with a \$1.06 increase in the hourly wage. And for those with no education ($\text{EDU} = 0$), the constant indicates that the average wage is \$5.40 per hour.

But what if we put IQ in the equation—that is, assume that earnings depend on both the level of education and IQ? We

plug in the data on IQ test results and find:

$$Y = 5.40 + 0.83\text{EDU} + 0.001\text{IQ}$$

We learn that individuals who performed better on IQ tests also had higher hourly wages. Moreover, while the impact of education on wages remains positive, it is about 27 percent smaller than if we hadn’t included IQ results (the 27 percent comes from the difference in the coefficients: $100(1.06 - 0.83)/0.83$). The implication is that we previously overestimated the effect of education on wages because we did not take into account the influence of IQ, which is correlated with education.

Potential pitfalls

Despite their benefits, regressions are prone to pitfalls and often misused. Take the following four leading difficulties.

Omitted variables. It is necessary to have a good theoretical model to suggest variables that explain the dependent variable. In the case of a simple two-variable regression, one has to think of the other factors that might explain the dependent variable. In our example, even when IQ is included, the correlation between education and earnings may reflect yet some other factor that is not included. That is, the individuals in the sample may still be different in some “unobserved” way that explains their subsequent earnings, possibly through their education choices. Individuals from wealthy families usually have better access to education, but family wealth may also create more connections in the labor market, leading to higher earnings. Thus, parental wealth may be another variable that should be included.

Reverse causality. Many theoretical models predict bidirectional causality—that is, a dependent variable can cause changes in one or more explanatory variables. For instance, higher earnings may enable people to invest more in their own education, which, in turn, raises their earnings. This complicates the way regressions should be estimated, calling for special techniques.

Mismeasurement. Factors might be measured incorrectly. For example, aptitude is difficult to measure, and there are well-known problems with IQ tests. As a result, the regression using IQ might not properly control for aptitude, leading to inaccurate or biased correlations between education and earnings.

Too limited a focus. A regression coefficient provides information only about how small changes—not large changes—in one variable relate to changes in another. It will show how a small change in education is likely to affect earnings but it will not allow the researcher to generalize about the effect of large changes. If everyone became college educated at the same time, a newly minted college graduate would be unlikely to earn a great deal more because the total supply of college graduates would have increased dramatically. **FD**

RODNEY RAMCHARAN is a professor of finance at the Marshall School of Business, University of Southern California.

What Are Remittances?

For many countries, money transfers from citizens working abroad are a lifeline for development

Dilip Ratha

WHEN MIGRANTS send home part of their earnings in the form of either cash or goods to support their families, these transfers are known as workers' or migrant remittances. They have been growing rapidly in the past few years and now represent the largest source of foreign income for many developing economies.

It is hard to estimate the exact size of remittance flows because many take place through unofficial channels. Worldwide, officially recorded international migrant remittances are projected to reach \$596 billion in 2017, with \$450 billion flowing to developing economies. These are recorded in the balance of payments; exactly how to record them is being reviewed by an international technical group. Unrecorded flows through informal channels are believed to be at least 50 percent larger than recorded flows. Not only are remittances large but they are also more evenly distributed among developing economies than capital flows, including foreign direct investment. Remittances are especially important for low-income countries and account for nearly 4 percent of their GDP, compared with about 1.5 percent of GDP for middle-income countries.

Getting the money there

A typical remittance transaction takes place in three steps:

- The migrant sender pays the remittance to the sending agent using cash, check, money order, credit card, debit card, or a debit instruction sent by e-mail, phone, or through the Internet.
- The sending agency instructs its agent in the recipient's country to deliver the remittance.
- The paying agent makes the payment to the beneficiary.

For settlement between agents, in most cases, there is no real-time funds transfer; the balance owed by the sending agent to the paying agent is settled periodically through a commercial bank. Informal remittances are sometimes settled through goods trade.

The costs of a remittance transaction include a fee charged by the sending agent, typically paid by the sender, and a currency-conversion fee for delivery of local currency to the beneficiary in another country. Some smaller operators charge the beneficiary a fee to collect remittances, presumably to account for unexpected exchange-rate movements. And remittance agents (especially banks) may earn an indirect fee in the form of interest (or "float") by investing funds before delivering them to the beneficiary. The float can be significant in countries where overnight interest rates are high.

Remittances are typically transfers from one person to another person or household. They are targeted to specific needs of

the recipients and thus tend to reduce poverty. Cross-country analyses generally find that remittances have reduced the share of poor people in the population (Adams and Page 2003, 2005; Gupta, Pattillo, and Wagh 2009). World Bank studies, based on household surveys, suggest that international remittance receipts helped lower poverty by nearly 11 percentage points in Uganda, 6 percentage points in Bangladesh, and 5 percentage points in Ghana. Between a fifth and half of the 11 percent reduction in poverty in Nepal between 1995 and 2004, a time of political conflict, has been attributed to remittances.

In poorer households, remittances may buy basic consumption goods, housing, and children's education and health care. In richer households, they may provide capital for small businesses and entrepreneurial activities. They help pay for imports and external debt service; in some countries, banks have raised overseas financing using future remittances as collateral.

More stable than capital flows

Remittance flows tend to be more stable than capital flows, and they tend to be countercyclical—increasing during economic downturns or after a natural disaster when private capital flows tend to decrease. In countries affected by political conflict, they are often an economic lifeline to the poor. The World Bank estimates that in Haiti they represented about 31 percent of GDP in 2017, and in some areas of Somalia, they accounted for more than 70 percent of GDP in 2006.

Remittances proved to be resilient during the financial crisis in source countries such as the United States and western European countries. The crisis affected migrants' incomes, but they tried to absorb the income loss by cutting consumption and rental expenditures. Those affected by the crisis moved to jobs in other sectors. While the crisis reduced new immigration flows, it also discouraged return migration because migrants feared they would not be able to reenter the host country. Thus, the number of migrants—and hence remittances—continued to rise even during the global financial crisis and even more so in recent years in the face of conflicts and natural disasters such as hurricanes and earthquakes.

There are potential costs associated with remittances. Countries that receive remittances from migrants incur costs if the emigrating workers are highly skilled or if their departure creates labor shortages. Also, if remittances are large, the recipient country could face real exchange rate appreciation that may make its economy less competitive internationally. Some argue

that remittances can undercut recipients' incentives to work and thus slowing economic growth. But others argue that the negative relationship between remittances and growth observed in some empirical studies may simply reflect the influence of growth on remittances rather than vice versa.

Remittances also have human costs. Migrants sometimes make significant sacrifices—including separation from family—and incur risks to find work in another country. And they may have to work extremely hard to save enough to send remittances.

High transaction costs

Transaction costs rarely affect large remittances (for the purpose of trade, investment, or aid): as a percentage of the principal amount, they tend to be small, and major international banks are eager to compete for large-value remittances. But for smaller remittances—under \$200, say, which is often typical for poor migrants—fees typically average 7 percent, and can be as high as 15–20 percent in smaller migration corridors (see table).

Cutting transaction costs would help recipient families. Here's how:

First, the fee should be a low fixed amount, not a percentage, because the cost of remittance services does not depend on the amount of principal. The real cost of a remittance transaction—including labor, technology, networks, and rent—is estimated to be significantly below the current level of fees.

Second, competition would bring prices down. New market players can be encouraged by harmonizing and lowering bond and capital requirements and avoiding overregulation (such as requiring full banking licenses for money transfer operators). The intense scrutiny of money service businesses for money laundering or terrorism financing since the 9/11 terrorist attacks on the World Trade Center has made it difficult for them to maintain accounts with their correspondent banks, forcing many in the United States to close. Regulations are necessary, but they should not make it difficult for legitimate money service businesses to maintain accounts with correspondent banks. A risk-based approach to regulation—checking only suspicious transactions and exempting small transactions below, say, \$1,000 from identity requirements—can reduce costs and facilitate flows.

Competition has spurred reductions in fees in the US–Mexico corridor, where remittance fees fell more than 50 percent from over \$26 (to send \$300) in 1999 to about \$12 in 2005, and have leveled off since then at around 5 percent for \$200 in the first half of 2017. Some commercial banks provide free remittance services, hoping to attract customers for their deposit and loan products. And in some countries, new remittance tools—based on cell phones, smart cards, or the Internet—have emerged.

Third, nonexclusive partnerships between providers and existing postal and other retail networks would help expand remittance services without large investments to develop payment networks.

Forth, poor migrants could be given greater access to banks, which tend to charge less. Both sending and receiving

TRANSFER COSTS

Remittance fees could be reduced significantly if they were a flat fee instead of a percentage of the principal transferred. Approximate cost of remitting \$200 (as a percent of principal) between:

	MTOs	Banks
Australia–Fiji	10.0	15.3
Germany–Serbia	6.6	20.9
Japan–Brazil	10.1	18.1
Malaysia–Indonesia	6.5	10.0
Russia–Tajikistan	1.6	—
South Africa–Zimbabwe	14.9	19.2
Saudi Arabia–Pakistan	4.3	4.5
United Arab Emirates–India	3.0	—
United Kingdom–India	3.0	7.3
United States–India	3.0	—
United States–Mexico	5.1	—
United States–Philippines	3.6	—

Source: World Bank Remittance Prices Worldwide database.

Note: — denotes that data are not available. Data are for the third quarter of 2017. Figures include currency-conversions charge. MTOs = money transfer operators.

countries can increase banking access for migrants by allowing origin-country banks to operate overseas; by providing identification cards (such as the Mexican matricula consular) that are accepted by banks to open accounts; and by facilitating participation of microfinance institutions and credit unions in the remittance market.

Boosting flows

Governments sometimes offer incentives to increase remittance flows and to channel them to productive uses. But such policies can be more problematic than efforts to expand access to financial services or reduce transaction costs. Tax incentives may attract remittances, but also may encourage tax evasion. Matching-fund programs to attract remittances from migrant associations may divert funds from other local funding priorities, while efforts to channel remittances to investment have met with little success. Fundamentally, remittances are private funds that should be treated like other sources of household income. Efforts to increase savings and improve the allocation of expenditures should be accomplished through improvements in the overall investment climate, rather than by targeting remittances. **FD**

DILIP RATHA is head of the Global Knowledge Partnership on Migration and Development and lead economist, migration and remittances, in the World Bank.

References:

Adams Richard H., Jr., and John Page. 2003. "International Migration, Remittances, and Poverty in Developing Countries." World Bank Policy Research Working Paper 3179. Washington.

———. 2005. "Do International Migration and Remittances Reduce Poverty in Developing Countries?" *World Development*. 33 (10): 1645–66.

Gupta, Sanjeev, Catherine Pattillo, and Smita Wagh. 2009. "Impact of Remittances on Poverty and Financial Development in Sub-Saharan Africa." *World Development*. 31 (1): 104–15.

