by J. W. Mason, John Jay College - CUNY
January 18, 2017

Some notes on these notes:

- They are very much a work in progress. They cover perhaps 50 percent of the material from an undergraduate intermediate macroeconomics course. I still use them in conjunction with a standard textbook.
- The notes are intended for undergraduates with limited economics background. The goal is mainly to give them the tools to intelligently read the *Financial Times*.
- The focus is on causal relationships between observable economic aggregates, as opposed to economic theory.
- Along with avoiding theory, I've tried to avoid explicitly "heterodox" arguments. Almost all the positive claims here should be acceptable to the great majority of people involved in macroeconomics in a policy, business or media setting.
- The notes are designed to be modular – each subsection heading gives the key point of that subsection, and each section and subsection can be understood on its own.
- Key terms are defined in the margin the first time they appear and again in the glossary.
- At this point there are essentially no references or suggestions for further reading. Those will come in future versions.
- The notes are laid out using a modified version of the Tufte latex package, which tries to reproduce the layout used by Edwin Tufte in *Visual Display of Quantitative Information* and other books.

If you are interested in using these notes in your own classes, feel free to do. If you want the underlying latex files, or if you have any comments on these notes, please email me at jomason@jjay.cuny.edu.

Contents

Macroeconomic Concepts

The economy is made up of economic units, which are grouped into sectors.

An **economic unit** can be a government, a business, a household, or any other entity that can make and receive money payments, own assets and owe debts. Note: In economics, we usually refer to "households" rather than people; a household consists of one or more individuals who pool their income and share ownership of assets, and make many of their purchases together. We do this because families normally function as a single economic unit: They make economic decisions together, and don't engage in market transactions with each other.

The main economic sectors are households, nonfinancial business, finance, and government. Nonprofit institutions are normally grouped with households. Nonfinancial businesses are further divided into corporate and noncorporate business. All units outside the naitonal borders are usually treated as a single sector, referred to as the rest of the world.

Outcomes that policymakers seek to influence are called policy targets.

A policy **target** is a macroeconomic outcome that policymakers seek to influence. A **policy instrument** is a variable that is more or less under control of some public authority – either the executive or legislature of an elected government, or a central bank. At the most general level, macroeconomic policy consists of picking the right values of the instruments to reach the desired levels of the targets. A fundamental challenge is that the sme instrument may have effects on more than one target, and the right value for one target is probably not the right value for the other. This problem is summarized by **Tinbergen's rule**: You must have at least as many instruments as you have *independent* targets, if you want to be able to hit all the targets. If two target variables always move together, they are not independent. For example, because there is such a close relationship between changes in output and changes in employment, many economists would not consider them two separate targets.

Target. An outcome that policymakers wish to influence. The most important macroeconomic targets are output, inflation, unemployment, the government debt ratio, the balance of payments, income distribution, and financial stability.

Instrument. A variable that is directly under the control of policymakers and is adjusted in order to affect other macroeconomic outcomes.

Tinbergen's rule. The principle that to hit a certain number of independent macroeconomic targets, the authorities must have at least that many different policy instruments.

It is important to distinguish between *targets*, which are the real-world outcomes we are concerned with, and **aggregates**, which are the variables we can actually measure. Because statistics are never collected perfectly, and are often collected for different purposes and defined differently than in economic theory, no aggregate is an exact measure of the corresponding target. And there are often a number of different aggregates that might be used to measure the

Aggregate. A variable measured at the level of the economy as a whole. Common aggregates include GDP, the consumer price index (CPI), and the unemployment rate.

same target. For example, today output is normally measured by the aggregate GDP, but in the past it was more often measured by GNP.

Debates about macroeconomic policy come down to three questions:

1. *What* are the appropriate levels for each target? (For example, should inflation be kept to 2%, as most countries currently seek to, or are there reasons to prefer an inflation rate of 4-5%, or of 0%?)

2. *Which* targets are most important, given that there are not enough policy instruments to hit all of them? (For example, some people – often owners of financial assets – think that it is most important to avoid high inflation, while other people – often those who work for a living – think that is most important to avoid high unemployment.)

3. *How* effective are the instruments available at moving the different target variables? (For example, does expansionary monetary policy have a strong effect on investment and thus on output, or only a weak one?)

The most important targets for macroeconomic policy, are output, unemployment, inflation, government debt, distribution, the balance of payments, and financial stability.

Almost all discussions of macroeconomic policy focus on one or more of the following seven targets.

Output - total goods and services produced in the economy.

All else equal, higher output is normally considered desirable – greater market production implies higher living standards and more resources available for public purposes. Keeping output near **potential output** is a central goal for macroeconomic policy, both because it is important in itself and because the behavior of other targets is closely linked to the level of output. Output above potential may be considered "overheating" in the sense that it cannot be sustained for more than a few years, or because it is associated with rising inflation, and with changes in income distribution. In practice, economists and policymakers worry more about output falling below potential than about output rising above potential. Macroeconomic policy is primarily concerned with variations in output over a periods of a few years – recessions, booms, and business cycles.

Output is normally measured by **gross domestic product (GDP)** or some related variable, such as **gross national product (GNP)**. There is no direct way to measure potential output. Statistically,

Output. Total production of goods and services in an economy.

Potential output. The level of output targeted by macroeconomic policy. It is assumed to be the maximum the economy can produce without "overheating" – that is, without rising inflation, shortages of raw materials, etc. Since potential output cannot be directly measured, it may be estimated based either on the level of observable aggregates like unemployment or inflation, or on the long-run trend of output growth.

Gross domestic product (GDP). The most common measure of total output of an economy. It is defined as final goods produced for the market within the borders of the country in a given period.

Gross national product (GNP). An alternative measure of total output of an economy. It is defined as final goods produced for the market by the labor and capital of a country, regardless of where production takes place.

it is estimated based on the trend of output growth in the past. In policy debates, the judgement about whether current output is above or below potential is based on the behavior of unemployment and inflation.

The growth rate of output is often considered a separate target from the current level of output. The majority of macroeconomists believe that the current level of output demands on demand-side factors (how much people and businesses wish to spend) while the long-run growth of output depends mostly or entirely on supply-side factors (the productive capabilities of the country's workers and businesses.) This implies that different kinds of policies may be needed to get output to potential in the short run, and to boost the long-run growth rate of output. For instance, a higher savings rate may reduce current demand for goods and services, but free up resources for productive investment that will contribute to future growth.

Not all economists agree that there is a conflict between raising output in the short term and raising long term growth. Some **Keynesian economics** economists believe that higher demand contributes to long-term growth as well as the current level of output. Many other economists believe that macro policy instruments cannot reliably affect the long-run growth rate one way or the other (since it depends more on technological change) and that policy should therefore focus on stabilizing the economy in the short run.

Keynesian economics. A school of macroeconomics that emphasizes: the determination of output by aggregate demand rather than the productive capacity of the economy; the role of money and finance in shaping economic outcomes; the uncertainty of the future; and the inherent instability of the economy, which must be managed by government.

Unemployment - the fraction of the laborforce unable to find work.

High unemployment is the problem that modern macroeconomics was first developed to address, and the unemployment rate is probably the single economic variable that policymakers pay most attention to. Policy focuses on unemployment partly because it is important in itself – unemployment source of great personal hardship, and when unemployment rate is high it often leads to political instability. Unemployment is also a focus for policy because it is easy to measure the unemployment rate, while potential output cannot be measured directly.

Unemployment rate. The fraction of the laborforce unable to find work. In the US it is normally measured by U-3 – the fraction of the civilian, noninstitutionalized population 16 and older who have zero hours of paid employment and are actively looking for work. But other measures exist.

Unemployment in the US is usually measured by U-3 – the fraction of the civilian, noninstitutionalized population 16 and older who have zero hours of paid employment and are actively looking for work. But broader measures exist, such as U-6, which includes people who have given up looking for work and part-time workers who would prefer to work full-time. Employment can also be measured as the fraction of the population over 16 with jobs; this is the **employment-population ratio**.

Employment-population ratio. The number of people with jobs, divided by the total non-institutionalized population 16 and over.

In general, lower unemployment is better than high unemployment, but no government today seeks to reduce the unemployment

rate to zero. What unemployment rate should be considered **full employment** in practice is debated; in the US, most economists today use a number between 4 and 5 percent. Rising inflation is often taken as a sign that unemployment is too low.

Full employment. The level of employment or, more often, unemployment targeted by macroeconomic policy. In the US today, full employment is often considered to be equivalent to an official unemployment rate of around 5 percent.

Inflation - the average increase in the prices of goods and services.

Most modern central banks are directed by law to focus on maintaining stable prices as their sole or primary task. (The US is an exception – our central bank, the Federal Reserve, is supposed to give equal priority to price stability and to full employment.) It is not always clear why inflation should be the main concern for policymakers, but it has been for at least the past 25 years. Rising inflation is generally taken as a sign that the economy is "overheating" – that output is too high.

Inflation. The average change in prices of goods and services in an economy. It is measured as the annual percentage change in a price index. Negative inflation is called deflation.

Inflation is usually measured by the annual rate of change in a price index, most often the **Consumer Price Index** – the average price of goods consumed by a representative household.

Consumer price index (CPI). An index of the price level. It is supposed to reflect the average price of goods and services consumed by a typical household.

Income distribution – the share of total income received by richer and poorer households.

Until relatively recently, income distribution was not considered a target for macroeconomic policy, in part because it was believed to be quite stable in advanced countries, and in part because it was assumed to depend mostly on microeconomic factors. In recent years, it has become clear that income distribution is not stable – in almost all the advanced countries, there has been a large increase in the share of income received by the rich, and an increase in the share received from property income and decrease in the share from wages. While many economists continue to believe that this shift is mainly due to changes in technology and the supply and demand of various skills, an increasing number of economists believe that macroeconomic variables like interest rates, government budgets, output growth, inflation and unemployment have played an important role in the redistribution of income upward.

Income distribution is measured in various ways. The **Personal distribution of income** is most often measured by the *Gini index*, which ranges from zero in a situation of perfect equality (equal income for all) to 1 in a situation of perfect inequality (one individual has all the income). It also may be measured by the ratio of two percentiles, such as the ratio of the median individual to the income of an individual in the poorest ten percent. The **functional distribution of income** is normally measured by the share of labor income — wages and salaries – in total income.

Personal distribution of income. The distribution of income among households. There are a number of different measures of personal distribution, which describe in different ways the share of income going to high, middle and low-income households.

Functional distribution of income. The distribution of income among the different factors of production – usually this means labor and capital, but it may sometimes be extended to include other factors like land. The most common measure of the functional distribution is the share of labor income – wages, salaries and benefits – in total income.

In public discussions of income distribution, it is most often assumed that a more equal distribution is preferable to a less equal

one, all else equal. Sometimes, it is suggested that policy should simply preserve the existing distribution of income, whatever it is. Either way, the recent rise in the share of property-owners and of the rich is seen as a problem. But in practice, economic policy sometimes seems to favor a redistribution of income upward and from labor to capital, even if few elected officials would state this as a goal.

Government debt ratio – total government debt relative to the size of the economy.

There is no agreement among economists about why, what level, or even whether government debt is economically costly. But most policy discussions take it for granted that it is necessary or desirable to keep government debt from rising too high. Government debt is usually measured as a fraction of the economy. For example, countries in the European Union seek to keep their government debt below 60 percent of GDP.

Debt-GDP ratio. The ratio of government debt to GDP, or $\frac{debt}{GDP}$. This is the usual way to measure government debt in macroeconomics.

Balance of Payments – the total flow of money into the country from the rest of the world, compared with the total flow outward.

The **balance of payments** refers to the all the money payments between a given country and the rest of the world. In other words, it is the net flow of **foreign exchange** into or out of the country.

We sometimes say that a country is running into *balance of payments problems* or facing a *balance of payments constraint*. This means that it is in a situation where the total flows of foreign exchange into the country are not enough to maintain the total flows out of the country. That is, the country as a whole needs to pay more money to the rest of the world, than it is receiving from the rest of the world. This is never a problem for the United States, since US dollars are accepted as payment by the rest of the world. But for many other countries, avoiding large balance of payments deficits is a very important goal of macroeconomic policy.

The main positive contributions to a country's balance of payments are its exports to the rest of the world, foreign investment from the rest of the world, the income it receives from its own foreign investments, and transfers from foreign governments or from its own citizens working abroad. The main negative contributions are imports from the rest of the world, outward foreign investment (including *capital flight*), and payments on foreign debt.

Balance of payments. The total money payments coming into and out of country. The same term is used to refer both to the difference between payments into the country and payments out; and to the system of accounts that records these payments.

Foreign exchange. The money of a country other than one's own. The foreign exchange market is the activity – mostly carried out by large financial institutions – of trading one currency for another.

Financial stability – sustainable growth in asset prices and debt levels.

Unlike other macroeconomic targets, there is no single aggregate variable associated with this target. But an increasing number of policymakers and economists believe that macroeconomic policy must be concerned with excessive swings in assets prices

Financial stability. Sustainable growth in asset prices and private debt. Financial stability is increasingly seen as an important target for macroeconomic policy, but there is no agreement on how to measure it.

(especially *asset bubbles*) and excessive growth of private debt. These concerns are usually presented as reasons for more **contractionary** policy than might otherwise be called for. Low interest rates, it is argued, may encourage too much borrowing by households and businesses, and may inflate the value of stocks, real estate and other assets. So far, however, there is no consensus on how to decide when asset prices or debt are growing too quickly.

Contractionary. Has as its intended or primary effect a reduction in output.

Schools of Thought

Macroeconomics as we know it today begins with the work of John Maynard Keynes.

The origin of macroeconomics as a distinct field is the Great Depression of the 1930s. In the US, the economy shrank by 25%, over a third of the workforce was unemployed, and business investment fell by 90%. In other countries, the collapse was even worse. How could this be the efficient result of the free market? A new theory was needed to explain how the economy could break down so badly.

Macroeconomics begins with John Maynard Keynes and his book *The General Theory of Employment, Interest and Money*. Keynes introduced the idea of aggregate demand and showed that we can't understand the behavior of the economy as a whole if we only think in terms of individuals and particular markets. For example, if an individual decides to spend less and save more he or she will probably become richer. But if everyone tries to spend less at the same time, then everyone will become poorer.

Keynes' ideas were broadly accepted from the 1930s until the 1980s because they helped bring an end to the Depression. But in recent decades, macroeconomics has been deeply divided between different schools of thought.

Classical economists think there is no need for a special field of macroeconomics – they analyze the economy as a whole using the same tools as in microeconomics. This means they look at macroeconomic problems in terms of alternative uses for scarce resources, and

believe that business cycles are temporary glitches, and generally favor *laissez-faire*, or nonactivist policies. They believe that a market economy usually produces the best possible outcome and government interventions do more harm than good. Many economists who work in universities believe in classical macroeconomics.

Keynesian economists believe that business cycles reflect underlying problems that can be addressed with activist government policies

New Keynesian macroeconomics is the most widely held view among economists who work in business and government and central banks. New Keynesian economics, despite the name, is basically the same as Classical economics, but it assumes some "frictions" in the economy that can cause inflation or deflation if not properly managed by the central bank. They think the free market mostly works well, but the financial system sometimes lends too much money, and sometimes lends too little. *Central banks* must step in to set the right interest rate, but after that the economy can mostly look after itself. New Keynesian economists prefer *monetary policy* to *fiscal policy*.

Old Keynesian (or radical Keynesian) economists think the economy is inherently unstable and that markets are often irrational. Government can and should do more to ensure that society's resources are fully utilized and invested in productive ways. Old Keynesians generally prefer fiscal policy to monetary policy.

Monetarist economists believe that almost everything that goes wrong in the economy is because there is the wrong amount of money. They blame bad choices by central banks for most business cycles. If central banks keep the money supply growing steadily, no other policy is needed.

Marxist economists believe the economy is unstable because of a basic conflict between business owners and people who work for wages. Government policy can be effective in the short run, but there will continue to be crises and depressions as long as society is divided between workers and owners.

Do economists agree about anything? Yes.

- Economists agree on terminology (language). All economists use terms like GDP or inflation to mean the same things, and measure them the same way.
- Economists all agree that valid ideas can be stated as *models* and tested using statistics.

Macroeconomic thought can be divided between Classical theories, which focus on the allocation of real resources through markets, and Keynesian theories, which focus on the way production is shaped by money payments and incomes.

Classical economists think that standard microeconomic reasoning also applies at the level of the economy as a whole. Microeconomics

studies the way a fixed stock of resources is allocated between alternative uses by the price mechanism; the Classical view is that this is also the right way to think about problems involving the economy as a whole. Keynesian economics is an effort to create a different kind of economic reasoning specifically for the problems of an economy as a whole. Keynes described what he was doing as replacing an economic theory based on "real exchange" with a theory based on"monetary production. The goal of the Keynesian system is to understand how decisions to spend more or less spending money can affect the total amount of production and employment in the economy.

	Classical	**Keynesian**
What drives macroeconomic outcomes, money flows or real flows?	Real flows – money is just a kind of bookkeeping	Money flows – real decisions about production, employment, etc., depend on money payments
Does the level of output depend more on supply conditions, or on demand?	Supply – the willingness and ability of people to work, and the productivity of businesses	Demand – the willingness of people and businesses to buy things
Is the economy always (or at least normally) at full capacity?	Yes	No
How are imbalances in aggregate markets resolved?	Price adjustments	Quantity adjustments
Are booms and downturns exogenous or endogenous?	Exogenous – caused by "shocks" from outside the economic system	Endogenous – caused by forces within the economic system
Why does output fall in recessions?	The economy's real resources or productive capabilities have declined	Households or businesses have chosen to spend less money
What is the fundamental cause of unemployment?	Wages too high (due to government or unions)	Output too low (due to lack of demand)

We often think of the difference between Keynesian and Classical economics in terms of policy preferences: Classical economists believe that government should not interfere in economic outcomes, while Keynesians believe that active government involvement is needed to keep the economy functioning acceptably.

This is generally the case, but there are exceptions. Monetarist economists, such as Milton Friedman, share most of the Keynesian analysis, but do not support active government intervention in the macroeconomy. Instead, they believe that if the government (in the

form of the central bank) can ensure that the total volume of money payments in the economy rises at a steady pace, the rest of economic life can be let to the private decisions of businesses and households. Marxist economists believe that economic outcomes depend on “real" factors – the division of the product between workers and capitalists, the technology of production – rather than money payments. But Marxists also favor political action to transform the organization of economic life.

Very few economists think the Keynesian or Classical perspectives are 100% true. Some academic economists at places like the University of Chicago believe in "real business cycles" – the idea that even short-term changes in the level of economic activity are driven by changes in technology and people’s desire to work. But most academic economists, and all economists in business and government, accept the Keynesian perspective when it comes to explaining business cycles – changes in the level of economic activity over a few years. The national accounts kept by all modern countries are organized on the assumption that changes in GDP from month to month or year to year, depend entirely on changes in spending, not on changes in productive capacities. But most economists believe the Classical perspective is better suited to describe developments over long periods of time. In other words: If we want to understand why the US economy was booming in 2007, and in recession in 2009, we should look at changes in people decisions to spend money. But if we want to understand why the US is much richer in 2009 than in 1900, we should focus on improvements in the productivity of American businesses.

Output, GDP, and Sectoral Balances

Output is normally measured by Gross Domestic Product (GDP), defined as the total market value of all final goods and services produced within an economy in a given period.

Output is normally measured by the *aggregate* Gross Domestic Product, or GDP. When people talk about the size of the economy – for instance, "the economy grew by 3 percent last year," or "the US economy is twice the size of the UK" – they are usually referring to GDP.

GDP is defined as the total market value of all final goods and services produced for sale on organized markets in a given geographic area within a given period, usually one year. This means:

1. GDP includes only sales of good and services to the **final** purchaser. That is, it includes only things that people or businesses buy for their own use, as opposed to for resale or for use in producing something else. That means **consumption** and **investment** – *housing investment* by households, and **fixed investment** and emphinventory investment by businesses. Most investment by businesses is *fixed investment* in buildings, machinery, research and development, etc., but *inventories* are a special kind of business investment – unsold finished goods, goods in process and unused stocks of raw materials. Final purchases also includes direct expenditure by government on the military, infrastructure, law enforcement, education, and so on, but not **transfers** to individuals like Social Security or Medicare.
2. GDP includes only goods that are bought and sold on markets, and counts them at the price they actually sell at. (There are some exceptions to this rule.) So it does not include anything people produce for their own use, or that is exchanged outside of organized markets. In particular, it does not include domestic labor like childcare, home cooking, etc., and it does not include black markets and other illegal activity.
3. GDP only includes newly produced goods and services. Sales of existing goods or of new or existing **financial assets** are not counted in GDP.
4. GDP includes all production within a geographic area, regardless of who carries it out.
5. GDP is defined for a given period, usually a year but sometimes a quarter (three months).

Final goods. Newly produced goods purchased to be used by the purchaser, as opposed to goods purchased to be resold or used as inputs to make something else. Includes all spending by households on new goods and services (including houses), investment spending by businesses, and spending by government on the direct provision of public services.

Consumption. Spending on goods and services that are used directly to meet people's needs. Includes all spending by households on newly produced goods and services (except new houses), as well as spending by nonprofits and government on services used by households.

Investment. The production of new long-lived means of production like buildings, machines, software, and so on. Unlike in everyday use, "investment" in macroeconomics does not include the purchase of existing real or financial assets.

Fixed investment. Production of new buildings, machinery or other lasting means of production. Includes all investment except for inventory investment.

Transfers. Payments that are made without any good or service being received in return. Transfers include payments through government programs like Social Security and unemployment insurance, as well as private gifts.

Financial asset. An asset like a stock, bond, or loan that does not involve ownership of any concrete object, but instead is a promise of future payment by someone else.

Output may also be measured net rather than gross, national rather than domestic, or as income rather than as product.

GDP is the most common measure of the size of the economy, but it is only one possible way to add up all the flows of money associated with economic activity. There are many debates about what exactly should be counted as final expenditure. In addition, each of the three terms – gross, domestic, and product – reflect choices that could be made the other way.

Gross or net? In economics, the term **net** means that something has been subtracted (or sometimes added) to the original number, while **gross** means that it has not been subtracted. In the case of GDP, we are not subtracting **depreciation**. Depreciation (or capital consumption) refers to the productive resources that were used up during the course of the year – machines that wore out, trees that were cut down, and so on. In principle, it might make sense to subtract depreciation, and the BEA does produce numbers for Net Domestic Product (NDP) with depreciation subtracted. But depreciation is hard to measure accurately, so for most purposes we use Gross Domestic Product instead.

Domestic or national? *Domestic* product includes all economic activity that takes place within the borders of the country, regardless of who carries it out. *National* product includes all economic activity carried out by a country's citizens and businesses, regardless of where it takes place. For example, the whole output of a Japanese auto factory located in the United States is counted in US GDP, but only the part of it "credited" to the American workers would be counted in GNP. The part of the factory's output attributable to the Japanese capital would be counted in Japan's GNP instead. For most countries, GDP and GNP are very similar, but for countries where foreign investment is important, they can look quite different.

Product or income? In principle, the total amount of spending on final goods and services should be exactly equal to the total income received from producing those goods and services. That is the logic of the circular flow. But in practice, some payments are always missed or mismeasured, so the two aggregates will turn out to be different, and we have to decide which one to trust. For most purposes, measurement of products is considered more reliable than measurement of income, but income measures are also reported.

Net. A number from which something has been subtracted. For example, net income for a business means revenue after costs are subtracted, net exports of a country means exports after imports are subtracted, and so on. What a particular net figure is net of depends on context.

Gross. A number from which something has not been subtracted. For example, gross income for a person means income before taxes are subtracted, gross domestic product of a country means that depreciation is not subtracted, and so on. What a particular gross figure is gross of depends on context.

Depreciation. The decline in value of real assets like buildings and machinery, whether from wearing out or from obsolescence.

In macroeconomics, investment means the production of new tools or resources that will be used for production in future periods.

Investment has a different meaning in macroeconomics than in everyday life. For an individual, investment might mean setting aside money with the hope of getting an income from it, for instance by buying shares in a mutual fund. But in macroeconomics, investment refers only to the production of new goods that will contribute to output in future years. That includes:

- New buildings ("plant") and equipment purchased by businesses.
- *Research and development* and other **intellectual property (IP)** spending by business, such as development of a new drug by a pharmaceutical company, or production of a new movie by a studio. This is an important recent change in the definition of investment. This kind of spending was not considered investment before 2012; instead, it as considered a cost of production and was not counted in GDP.

  Plant, equipment and IP investment are grouped together as *fixed investment.*
- Purchases of new houses by people, or *residential investment*. In standard macroeconomic accounting, households do not invest in anything except housing.
- Additions to stockpiles of raw materials and finished goods. When a company produces something and doesn't sell it, that is called **inventory investment**. It might seem strange to call an unsold good an investment, but it makes sense when you consider that the good will contribute to output when it is finally sold. It also makes sense, given that the logic of the accounts requires total spending to be equal to total income. Since the people who produced the unsold goods received incomes, someone must have spent an equivalent amount of money. So we say that the business itself spent the money, by purchasing its own products. Note that only the *change* in inventories is counted as investment. This means that inventory investment is the one kind of investment that can be negative.

Intellectual property (IP). Patents, copyrights, and similar legal claims on creative works and scientific discoveries. In the national accounts, creation of new IP is counted as a form of investment.

Inventories. Unsold finished goods, goods in process, and stocks of raw materials. In the national accounts, the change in inventories is counted as a form of investment.

In a closed economy, total production equals total spending equals total income. output can be measured as any of these.

In a **closed economy** total production must equal total spending on final goods and services, since goods are counted in GDP only when they are sold, and for every sale there must be a purchase. Similarly,

Closed economy. An economy with no trade or financial links to other economies. No economy in reality is perfectly closed (except for the world as a whole), but it is often useful to think about how an economy would behave in isolation.

total production must equal total income, since every dollar of spending is received by someone. In an **open economy** this is no longer true: Production will be different from spending if the country unless the country's **trade balance** is exactly zero, and production will be different from income if some production is carried out using **factors** – labor or capital – of other countries. No economy is completely closed, but it is still useful to think about the closed economy case as a first approximation. In that case, the fundamental rule of national income accounting is:

Total spending = total income = total output

For an individual person or business, the income they receive is not necessarily equal to the value of the output they produce, and the amount they spend probably will not be equal to either. But for an economy as a whole, if everything is counted correctly, these three values will always be the same. When we are talking about an economy as a whole, we can use the terms "income" and "output" interchangeably.

Open economy. An economy connected by trade or financial links to other economies. In reality every economy (except for the world as a whole) is at least somewhat open; we use the term "open economy" to mean cases where the links to the external world are important.

Trade balance. The difference between a country's exports and its imports. If exports are greater than imports, it has a trade surplus; if exports are less than imports, it has a trade deficit.

Factors. Labor, capital and others who must be paid for their contributions to production.

There are some important exceptions to the general rules for what gets counted in GDP.

It's important to understand the basic concepts behind aggregate accounting, and the standard definition of GDP. But you should realize that these numbers do not always mean what they seem to. Here are a few odd rules of the national accounts that many people – even many economists – are not aware of.

"Households" include nonprofits. The **household** sector in the national accounts consists mostly of individuals and families earning income and spending money on their own needs. Every dollar you earn shows up as household income in the national accounts, and every dollar you spend on goods and services shows up as household consumption, except for a purchase of a new home, which shows up as household residential investment. But the household sector also includes nonprofit institutions like churches, charities, and nonprofit hospitals and universities. Any income these institutions receive is counted as household income, and any money they spend is counted as household consumption. Because the output of nonprofits is not sold in markets, it can't be measured like the output of a business. So instead, the value of "consumption" by nonprofits is measured as their total costs – including intermediate goods – minus any revenue from sales. In recent years, consumption by nonprofit institutions comes to about $300 billion, or 2.5% of official household consumption.

Household. People when they are acting on their own behalf, rather than as part of businesses or governments. A household may be an individual or a family or other group of people who pool their incomes and make decisions about earning and spending together.

Homeowners are considered to rent to themselves. In general, the national income and product accounts only count goods and services that are sold in markets. The big exception is the "services" people produce for themselves as homeowners. By the standard conventions of the national accounts, anyone who owns their own home is considered to be renting that home to themselves. The BEA **imputes** (estimates) the value of that rent, and counts it as both income and spending for the household sector – even though no money changes hands. These *owner equivalent rents* currently total $1.2 trillion, accounting for a bit over 10% of official household consumption. Again, this is not actual rental payments, but the BEA's estimate of the value of the "housing services" that people receive from their own homes each year.

Imputation. A variable in the national accounts that can't be measured directly, but has to be estimated based other variables.

Health insurance payments are considered household consumption. All spending on health care for individuals is counted as income and consumption for the household sector, no matter who pays for it. Health benefits you receive from your employer are counted as household income just the same as wages and salaries. More surprisingly, spending through government health insurance programs is also considered income and consumption for the household sector. As far as the BEA is concerned, if your grandmother gets medical treatment and Medicare pays for it, that is exactly the same as if the federal government sent her a check and she decided to buy medical care with it. Employer-provided health insurance plans currently pay for about $600 billion of medical care each year and public health insurance programs (Medicare and Medicaid) pay for about $950 billion. Together, these account for a bit under 15% of total measured consumption.

There are large imputed financial services. Another exception to the rule that only services sold in the market count in GDP, is the "services" people are assumed to receive when they hold assets that pay less than the market interest rate, or borrow money at more than the market interest rate. For example, many people have checking accounts, despite the fact that checking accounts pay little or no interest. The BEA assumes that people are receiving some financial service from the bank that is equal in value to the interest they could otherwise get. These "imputed financial services" are currently estimated at $450 billion per year, or about 4% of total household consumption.

Adding up these four items, you can see that about a third of what the BEA calls household consumption is not what we normally think of as consumption – money people spend on their own needs. (Nonprofit spending is 2.5% of reported consumption, owners' equivalent

rent is 10%, third-party health spending 15%, and imputed financial services are 4%, for 31% in total.) Either no actual money is spent, as with owners equivalent rent and imputed financial services. (These are exceptions to the normal rule that only market transactions are recorded in GDP.) Or money is being spent, but for some social purpose, not people's private needs, in the case of nonprofits. Or money is spent on people's private needs, but not by people themselves, as with third-party health spending. (Third-party education spending – when government pays students' tuition – also falls in this category.) How we count this spending has important implications. By the conventional measure, consumption by households has shown a big increase over the past 50 years, from around 60 percent of GDP in the mid-1960s to over 70 percent today. But it turns out that this increase is entirely due to the four factors described above. If we count as consumption only money actually spent by households on their own needs, there is no long-term increase in consumption spending at all. The two public health programs Medicaid and Medicare alone account for about three quarters of the apparent increase in household consumption as a share of total spending. The figure shows household consumption by the official measure and limited to actual spending by households both as a share of GDP. (Third-party spending (as on healthcare) is still included in GDP; noncash items like owners equivalent rent and imputed financial services are excluded from GDP as well.)

A recent discussion of these issues can be found in "Household Income, Demand, and Saving: Deriving Macro Data with Micro Data Concepts," a working paper by Barry Cynamon and Steve Fazzari.

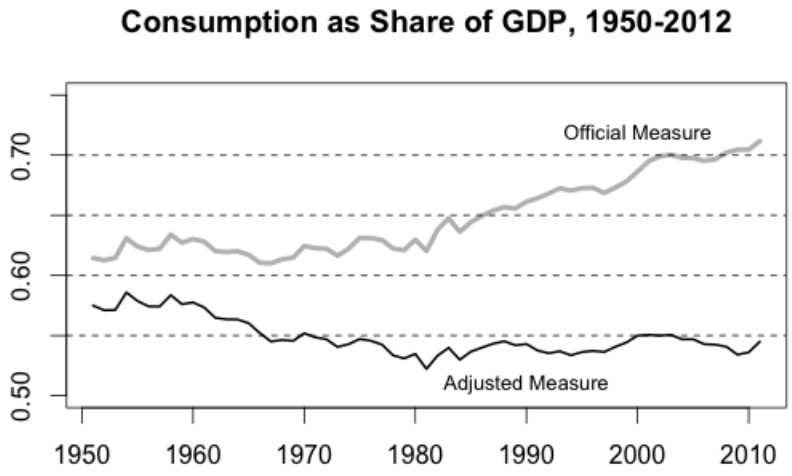


Figure 1: The official measure is the standard definition of consumption. The adjusted measure is limited to money actually spent by households.

Government output is valued at cost. Like nonprofits, governments carry out productive activity but they do not generally sell the goods and services they produce in the market. So their output

cannot be measured by its sale price, the way the output of private businesses is. Instead, the value of government services are computed as the total cost of producing them, including wages and intermediate goods. This means that if labor and other inputs are used as efficiently by government as by business, the value of government services will be underestimated, since the cost of goods produced by private businesses reflect not only labor and other costs, but also profit for the business owners. So if we assumed that government is on average as productive as the private sector, we would have to raise the estimated value of government services by 10 percent or more. Note that this does not say anything about the social value of government spending; it just says that if we estimated a market value for government services the same way we calculate the output of private businesses, the number would be somewhat higher than the official measure.The most important thing for the purpose of this class is simply to know that the value of government output in the national accounts is computed by adding up all the costs of government production.

The national income identity says that total income must be equal to the sum of the various components of GDP.

The **national income identity** states that all spending in the economy can be split into four categories: consumption, investment, final government spending, and net exports – that is, exports minus imports. Or:

National income identity. A fundamental accounting identity that says that total output equals the sum of consumption, investment, final government spending and net exports: $Y = C + I + G + (X - M)$.

$$Y = C + I + G + (X - M)$$

Because this is an **accounting identity**, it always holds exactly. So if we know all but one of the terms in the equation, we can calculate the remaining one. For example, in 2014, US GDP was \$17 trillion. Consumption spending totaled \$12 trillion, final government spending totaled \$3 trillion, exports were \$2 trillion, and imports were \$3 trillion. Knowing this, we can calculate investment spending:

Accounting identity. An equation that must always be true, because of how the terms are defined.

$$17 = 12 + I + 3 + (2 - 3)$$

$$I = 17 - 12 - 3 - 2 + 3 = 3$$

Investment must have been \$3 trillion. And in fact it was.

Note that imports are a subtraction from GDP. This makes sense, since they represent domestic spending that does not fall on domestically produced goods.

The identity also applies to changes in GDP. So it can be used to tell us what kinds of macroeconomic developments are possible.

By itself, the identity is not very useful, since you are unlikely to be in a situation where you know some of the components of GDP but not others. It becomes more interesting when we think about changes in GDP rather than its current level.

Any accounting identity also holds for changes in the variables. If investment spending rises by one dollar, and no other expenditure changes, then GDP must also rise by one dollar. So any change in GDP must involves changes in the various components that add up to the overall change.

Usually, we measure changes in the components in *percent of GDP*, rather than dollars. Note that to say that investment rose by one percent of GDP, is different from saying that it rose by one percent. Using the numbers above, a one percent of GDP increase in investment would be an increase of $170 billion; a one percent increase in investment would be only $30 billion.

If we think that spending determines output – as almost all economists do for short-run changes – then we can say that the changes in the various components explain the change in total GDP. For example, we might say that an economy grew by 1 percent because households increased investment spending by two percent of GDP, while businesses reduced investment spending by one percent of GDP.

The BEA produces tables exactly like this, reporting "Contributions to Percent Change in GDP" by various expenditure categories. For example, in the third quarter of 2015, real GDP grew at a 2 percent annual rate. Of this, consumption contributed 2, investment -0.1, exports 0.1, imports -0.4, and final government spending 0.3. (The numbers don't quite add up because of rounding.) Note that the negative contribution of investment means that investment spending was falling, while the negative contribution of imports means that imports were rising.

This is useful: It tells us that growth in output is currently based on households' willingness to increase consumption relative to their incomes.

The same kind of analysis is also useful when we want to ask what is possible in terms of economic growth. For example, during 2007-2009, residential investment fell by a total of 3 points as the housing bubble collapsed. This decline was partly, but not entirely, offset by an increase in final government spending. Some people argue that this increase in government spending was not needed to maintain demand. But in that case, we can ask, what other component of demand could have increased to make up for the fall in residential

investment? It does not seem plausible that households would have increased consumption expenditure sharply as the value of homes was falling, and household debt was already at high levels. (In fact, real consumption spending was flat during those three years.) it's hard to see how there could have been an increase in investment spending if businesses were seeing falling sales and many were having trouble getting loans. (In fact, business investment fell by 1.2 points.) So someone who claims that additional government spending was not required to maintain demand during 2007-2009 should have some explanation of how US exports might have grown, and/or imports fallen, by an additional 3 percent of GDP during this period.

Similarly, there is concern now in China that high growth has been driven by very high business investment, much of which may turn out to be wasteful or unprofitable. The national income identity reminds us that for investment growth to slow without pulling down GDP growth, some other component of demand must grow faster. Again, to use the national income identity to analyze these issues, we simply recall that $Y = C + I + G + (X - M)$ whether measured in dollars or percentage points of GDP. Suppose that Chinese investment must fall by, say, 5 percent of GDP to get back to a sustainable level. As Martin Wolf discusses in a recent column in the *Financial Times*, there are serious challenges to faster growth of consumption by Chinese households, larger Chinese trade surpluses, or big increases in government spending. But we know that for GDP growth to be sustained, some combination of consumption, government spending and net exports must increase by 5 points.

The national income identity can be rearranged to show that the difference between private saving and investment, plus the government budget balance, must be equal to the trade balance.

Another way of using the national income identity is to introduce taxes, government transfers, disposable income and private savings. We will need to introduce some new variables for this. **Disposable income** is the flow of money available to households. So it includes both current income (that is, wages and profits from the business sector) and *transfers*, less tax payments. So if we write Y_D for disposable income, T for tax payments, and TR for transfers, then:

Disposable income. Income available to households after transfers and tax payments.

$$Y_D = Y - T + TR$$

In macroeconomics, savings simply means that part of total income that is not used for consumption. In other words,

$$S = Y - C$$

In the same way, *private saving* is that part of disposable income that is not used for consumption. Note that private savings includes **retained earnings** of corporations as well as saving by households. So, writing S_P for private saving,

Retained earnings. Profits that are kept by the business that earned them, rather than paid out to shareholders. Retained earnings are an important form of saving in the economy. Historically, corporations have paid out about half their profits and retained about half.

$$S_P = Y_D - C = Y - T + TR - C$$

Or equivalently, consumption is equal to disposable income minus private saving:

$$C = Y_D - S_P = Y - T + TR - S_P$$

We will ignore private transfers and assume all transfers are from government to households. Then total government spending is equal to G plus TR. That is, total government spending is equal to spending on goods and services for public purposes (G) plus spending on transfer payments (TR). For our purposes, we'll consider all government revenue to be taxes. (This is basically true for the US, but not everywhere – in some countries a significant fraction of government revenue comes from the operation of government-owned enterprises.) So the government government budget balance is equal to $T - (G + TR)$. If this number is positive, we say there is a budget surplus; if it is negative, we say there is a budget deficit.

Now we combine the previous equation for consumption with the national income identity and rearrange the terms:

$$Y = C + I + G + (X - M)$$

$$Y = (Y - T + TR - S_P) + I + G + (X - M)$$

$$0 = (I - S_P) + (G + TR - T) + (X - M)$$

$$(M - X) = (I - S_P) + (G + TR - T) \tag{1}$$

This says that the trade deficit must equal the excess of private investment over private saving, plus the government budget deficit. The excess of private saving over private investment is also called the **private balance**. This is the amount of income the private sector (households and business together) has left over after paying for all desired investment. The third term, $X - M$ is the trade surplus. This is the excess of goods and services sold to the rest of the world over goods and services bought from the rest of the world. We can also think of it as net lending to the rest of the world. If a country has a trade surplus, it is lending to the rest of the world; if it has a trade deficit, it is borrowing from the rest of the world. If the private

Private balance. The difference between private saving and private investment.

balance is negative, that means that there is more investment taking place than there is private savings to pay for it. Since savings always equals investment, the remaining saving must come from somewhere else – either the government (via a budget surplus) or the rest of the world (via a trade deficit.)

For example:

In 2015, US GDP was $18.0 trillion. Personal consumption was $12.3 trillion, private investment was $3.1 trillion, exports were $2.3 trillion, imports were $2.8 trillion, and government final expenditure was $3.2 trillion. Transfers totaled $2.7 trillion, and taxes totaled $5.3 trillion. based on this, we can caluclate the sectoral balances for 2015. Net imports were equal to $2.8 trillion - $2.3 trillion or -$500 billion – the US had a trade deficit of $500 billion or, equivalently, it was borrowing $500 billion from the rest of the world. It's more common to give aggregates as a percent of GDP; in this case, the trade deficit was equal to 0.5 / 18 = 0.03 or 3 percent of GDP. Disposable income is equal to total income minus taxes, plus transfers. Recall that total income is just GDP – every dollar spent on domestically-produced goods and services is income for somebody. So disposable income in 2005 was equal to $18.0 trillion - $5.3 trillion + $2.7 trillion = $15.5 trillion. In percent of GDP, this is 15.5/18.0 = 0.858 = 86 percent of GDP. Private saving is disposable income minus consumption. So private saving in 2015 was $15.5 trillion - $12.3 trillion = $3.2 trillion. 3.2 / 18.0 = 0.177 =18 percent of GDP. And the private balance is given by $S_P - I$. So the private balance in 2015 equaled $3.2 trillion - $3.1 trillion = $100 billion or around 1 percent of GDP. The private sector a whole was almost in balance, with private investment just barely covered by private saving. The government balance is given by taxes minus total government spending, or $T - G - Tr$. Finally, the government balance in 2015 was equal to $5.3 trillion - $3.2 trillion - $2.7 trillion = -$700 billion, or -4 percent of GDP. So looking at the balances as a whole, we could say that in 2015 the federal government borrowed an amount equal to 4 percent of GDP, with 1 percent lent by the private sector and 3 percent lent by the rest of the world.

We could just as easily reverse all the terms in Equation 1 (that is, multiply both sides of the equation by negative one). Then it says that the trade surplus must be equal to the excess of private savings over private investment, plus the government budget surplus:

$$(X - M) = (S_P - I) + (T - G - TR)$$

This equation is an *accounting identity*; it is always exactly true. But that raises the question – what if something happens that changes just one of the terms in the equation – how does it balance? Which of the terms in the equation "call the shots'," and which are passive?

For example, in the 1980s, many people believed in the idea of "dual deficits" – that the large budget deficits under the Reagan administration were responsible for the large trade deficits that began around the same time. Later, in the 1990s, the federal government moved back toward budget surpluses, but the trade deficits continued. This made the dual deficits idea less attractive. It remains true, however, that if the government budget deficit increases, either the private sector must be saving more or investing less, or else the trade deficit must increase as well.

Another example: Suppose you think that China ought to have a smaller trade surplus. The national income identity shows us that this is possible only if Chinese households reduce their savings, or Chinese businesses increase their investment, or the Chinese government moves toward a budget deficit. Any successful effort to reduce the Chinese trade surplus must somehow bring about at least one of these outcomes.

In general, classical economists believe that savings calls the tune – that an increase in P_S will lead to an increase in I. They therefore favor measures to discourage consumption and to redistribute income to the rich, who tend to save more. Keynesian economists, on the other hand, believe that investment calls the tune and private savings adjusts in reaction. In this view, an increase in I is more likely to lead to an increase in P_S than the reverse. (This is because not all of the the new income created by increased private investment is consumed.) So Keynesians are more likely to see higher consumption as good for the economy, and to favor redistribution to the poor.

"Keynes's intellectual revolution was to shift economists from thinking in terms of of a model in which a dog called savings wagged his tail labelled investment, to thinking in terms of a model in which a dog called investment wagged his tail labelled savings." – James Meade

Unemployment

The official definition of unemployment is civilian, non-institutionalized people sixteen and over who had no paid work in the past week, want a job and are immediately available for one, and are actively seeking employment.

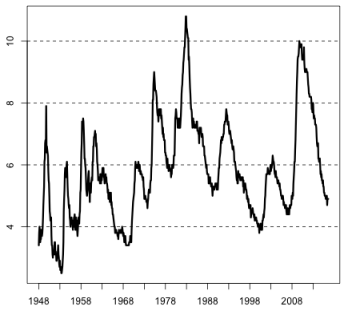


Figure 2: Official US unemployment rate, 1948-2016

Unemployment is one of the main targets of macroeconomic policy. In general, unemployment refers to people who are unable to work for economic reasons. All else equal, low unemployment is preferred to high unemployment. The lowest possible or feasible level of unemployment is called *full employment*.

There are a variety of ways of defining and measuring unemployment. The official or "headline" measure of unemployment is called *U3*. U3 defines someone as unemployed if they (a) are at least 16 years old, and are not in the military, in prison, or otherwise institutionalized; (b) had zero hours of paid work in the past week, and no more than 15 hours of unpaid work in a family business; (c) do not have a regular job from which they are temporarily absent due to vacation, sickness, a strike or lockout, etc.; (d) are currently available for work – that is, would take a job if one were offered to them; and (e) have actively looked for work in the past 4 weeks, by sending out resumes, contacting an employer, visiting a job center, etc. If someone had even one hour of paid work (or 15 hours or more of unpaid work in a family business) they are not counted as unemployed. Other forms of unpaid work are not considered employment, however.

The unemployment rate is defined as the number of unemployed people, divided by the people *in the laborforce* – that is, the sum of those employed, and unemployed. It measures the fraction of people who currently wish to work, but are unable to find jobs.

The unemployment rate is defined as $\frac{\textit{unemployed people}}{\textit{unemployed plus employed people}}$

U3, like other unemployment measures, is released by the Bureau of Labor Statistics each month. When you see a reference to the "unemployment rate" without further detail it normally means U3.[1]

[1] The BLS has a useful guide to exactly how it collects these statistics in "How the Government Measures Unemployment".

Unemployment matters because of the hardship it creates for individuals; because it is a waste of society's productive capacities; because it is an indicator of how close output is to potential; because it affects bargaining power between workers and employers; and because high unemployment is a source of social and political instability.

Unemployment is one of the most-watched macroeconomic aggregates. Along with output and inflation, it is one of the targets that policymakers in the US and most other rich countries focus on most. Unemployment is important for several reasons:

Individual wellbeing. Losing a job, or being unable to find one, is a painful and disruptive event in most people's lives.[2] For most people under capitalism, our main claim on society's resources comes from our paycheck. So loss of a job means a loss of income, and temporary or permanent reduction in living standards. In addition, for many of us, our job is a central part of our identity and our most important connection to the world beyond our immediate family and friends. Unemployment can lead to social isolation and loss of self-respect, and make it hard to sustain other relationships. These "nonpecuniary" costs of unemployment may be even more destructive than the loss of income.

[2] "The nonpecuniary cost of unemployment can be devastating: the unemployed are stripped of their economic identity, left with a searching sense of failure and filled with doubts about their future place in the world ... People take pride in, and are accorded respect for, having a job. ... The unemployed are often seen as suffering the consequence of their own personal shortcomings: a questionable work ethic, shortage of talent, disagreeableness... The jobless themselves often internalize the process, accept the negative judgments, and come to feel 'unworthy, incomplete, and inferior'." - Cristobal Young, "Losing a Job: The Nonpecuniary Cost of Unemployment in the United States"

Potential output. One way of looking at unemployment is that it is a waste of people's capabilities, of their potential to engage in useful work. Someone might be able to contribute to society as a nurse, a cook, truck driver, a musician, etc. – by producing something of value to others. As long as they are unemployed, these productive capacities go unused. So unemployment is costly not only for the individual but for society, which is being deprived of the fruits of the person's labor. Unemployment is by definition a sign that output is falling short of potential; in addition, high unemployment rate is usually seen as a sign that other productive resources – buildings and machines, ideas and technology – are also not being fully utilized.

Labor market slack. For economists, one of the most important effects of unemployment is that it changes the bargaining power between workers and employers. When unemployment is high, there are likely to be many applicants for each new job; this makes it hard for workers to demand higher wages, since they know they could easily be replaced and they will have trouble finding another offer. In other words, there is a great deal of **slack** in the labor market. When unemployment is low, workers are in a stronger bargaining position: Employers will have a harder time replacing them, and they will have an easier time finding a new job if they're not satisfied with this one. This means that businesses are more likely to raise wages when unemployment is low. Low unemployment can also reduce employers' power more generally – when there is not much unemployment, getting fired is less frightening, so workers can exercise more control over the terms of their work. From the point of view of macroeconomic policymakers, excessive wage growth and a breakdown of discipline in the workplace are dangers when unemployment gets too low.

Labor market slack. How easy it is for employers to find new workers.

Links to other aggregates. Unemployment is linked to a number of other important outcomes. It is one of the main variables policy-

makers look at to decide how close output is to *potential output*. Lower unemployment is associated with faster wage growth, because of the stronger bargaining power it creates for workers. This may also mean that low unemployment leads to a more equal income distribution. Lower unemployment is usually associated with faster output growth, through a relationship known as *Okun's law*. Lower unemployment is also linked to higher inflation through a relationship known as the **Phillips curve**. Low unemployment can boost aggregate demand, and in particular household consumption. And it is often associated with a more favorable government **fiscal balance**, since government spending is likely to be lower and tax collections higher when unemployment is low. In part because unemployment is reported monthly – unlike most economic aggregates, which are reported only every quarter or year – and measured precisely, it is often considered a useful guide or forecast to the behavior of these other aggregates.

Phillips curve. One of various relationships between the level or growth rate of output or unemployment on the one hand, and wages or prices on the other. In general, Phillips curves describe how inflation will be higher when the economy is above potential, and lower when it is below.

Fiscal balance. The difference between government revenue and government spending. If revenue is greater than spending, the fiscal balance is positive and we say the government has a budget surplus. If revenue is less than spending, the fiscal balance is negative and we say the government has a budget deficit. If revenue is exactly equal to spending, the fiscal balance is zero and we say the government has a balanced budget.

Social stability. Connected to the nonpecuniary costs described above, high unemployment often undermines social and political instability. At the least incumbent governments are less likely to win reelection when unemployment is high. And in many parts of the world, persistently high unemployment has been followed by increases in crime, civil unrest, political violence and revolutionary upheavals.

When output grows quickly, unemployment falls; when output grows more slowly or falls, unemployment rises. This statistical relationship is known as Okun's law.

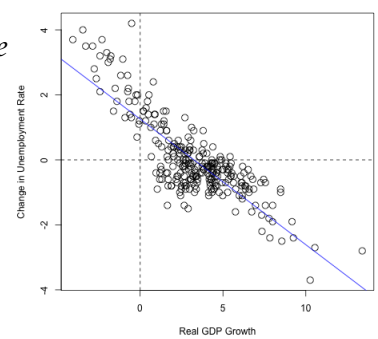


Figure 3: Okun's law, 1948-2016. For every point that growth is above 2 %, unemployment falls by about half a point; for every point that growth is below 2%, unemployment rises by about half a point. As the figure shows, actual growth and unemployment rates in the US fit this rule closely.

Okun's law. An empirical law in economics that says the change in unemployment ΔU is connected to the real growth of output g by a relationship of the form $\Delta U = -a(g - b)$. For the US, a is around 0.6 and b is around 2.

Unemployment and GDP growth are connected by a statistical relationship known as **Okun's law**. This law, which seems to have been very stable in the US over many decades, says that the change in unemployment over one year is normally equal to 0.6 times the difference between 2 and the real GDP growth rate. Or

$$\Delta U = 0.6(g - 2)$$

where ΔU means the change in the unemployment rate, and g is the real (inflation-adjusted) growth rate of GDP. This implies that if a year passes with no growth in real GDP, unemployment will rise by 1.2 points. It takes a real growth rate of 2 percent to hold unemployment constant. And to reduce unemployment by 1 point, requires a year of 4 percent growth, or two years of 3 percent growth, etc. While this relationship is not perfect, it is quite reliable as far as macroeconomic laws go. Similar relationships hold in other countries, but the coefficients are different. In Japan and most European countries,

unemployment is less responsive to economic growth in the US (i.e. instead of an 0.6, the equation would have 0.4 or 0.3 or 0.2), while in a few countries, like Spain and Australia, unemployment seems to be more responsive to economic growth than in the US. The exact values depend on the time period. Population growth is slower in the US today than it was in the 1950s and 1960s, so it takes less economic growth to keep unemployment constant. The *Okun coefficient* of 0.6 does not seem to have changed.

The reason it takes two points of economic growth just to hold unemployment constant is that (1) the population is growing and (2) labor productivity rises over time. So there are more people looking for work each year and, at the same time, it takes fewer people to produce a given amount of goods and services. This means that the amount of goods and services produced must rise just to keep the unemployment rate steady.

Why does unemployment change less than proportionately with output? One reason is that when unemployment is low, more people enter the laborforce, and when it is high, they exit the labor force. Another reason is that companies often cannot, or don't wish to, adjust their staffing every time their sales change. Some jobs are *overhead labor* that the company needs just to operate, whether it is selling a lot or a little; other jobs are costly to recruit and train workers for, so the company tries to avoid hiring and laying off workers every time sales rise or fall. It is also possible that the coefficient is less than one because labor productivity rises more rapidly when unemployment is low. This last relationship is called **Verdoorn's law**; it also implies that low unemployment can improve long run economic growth.

Verdoorn's law. A relationship between output or unemployment and productivity growth: When unemployment is low and output is high, productivity tends to rise faster.

Many people believe the official unemployment rate does not adequately capture the share of potential workers who lack jobs; alternative measures include the broader U-6 unemployment rate and the employment-population ratio.

Because we care about unemployment for different reasons, no one definition will be the right one for all purposes. For instance, should people in prison be counted as unemployed? Currently, they are not. But if we are interested in how much of our collective capacity for work is going to waste, we should certainly count prisoners – everyone in prison is capable of contributing to society in some way, but they are not doing so as long as they are incarcerated. On the other hand, if we are mainly interested in the unemployment rate as a measure of labor-market slack, it makes sense to exclude prisoners – with rare exceptions, employers do not have the option of hiring

them.

In any case, the definition of U3 may not match up with the category of people we want to think of as unemployed. So we may want to use one of the other measures produced by the BLS. The most important of these alternative measures is **U6**. U6 includes everyone in U3, but also includes *involuntary part-time workers* – people who worked fewer than 35 hours in the past week, and say that they would have liked to work more but were unable to get additional hours either at their current job(s) or at an additional one. U6 also includes *discouraged* and *marginally attached* workers – people who are not working, and say that they would take a job if offered one, but have made no effort in the past 4 weeks to get a job, either because they don't think any are available or for some other reason. It is often argued that U6 gives a better measure of "true" unemployment, in the sense of people whose capacity to work is going unused, and who are suffering because of a lack of a job. In practice, it does not always matter which measure we use, since the two tend to move together – U6 is typically about double U3. Besides employed and unemployed people, there are also those *not in the labor force.* This includes all noninstitutionalized civilians 16 or over who neither have any paid employment, nor fit the definition of the unemployed. It includes people who are neither working nor wish to work – retirees, full-time homemakers, and so on. Those under 16, in the military, or in institutions are not counted in the employment statistics at all.

U6. An alternative measure of unemployment that includes everyone who is unemployed by the official definition, plus discouraged and marginally attached and involuntary part-time workers.

As of August 2015, the US population was a bit over 320 million. Of this, 65 million were under 16, 2.5 million were in prison, 1.5 million were in other institutions (mostly nursing homes), and 1.5 million were on active duty in the military. That leaves a civilian noninstitutionalized population of 251 million. Of these 251 million, 149 million were employed, 8 million were unemployed (by U3), and 94 million were not in the labor force. This implies an unemployment rate of $8 \div (8 + 149) = 5.1\%$. But 1.5 million people were discouraged or marginally attached (counted as not in the laborforce by U3, but as unemployed by U6) and another 6 million workers were working part-time because they were unable to find full-time work (counted as employed by U3, but unemployed by U6).[3] So using the U6 definition of unemployment implies an unemployment rate of $(8 + 1.5 + 6) \div (149 + 1.5) = 15.5 \div 150.5 = 10.3\%$.

[3] These figures can be found in "The Employment Situation," a monthly report produced by the BLS.

When we think of unemployment as wasted potential, it is clear that in addition to the forms of unemployment covered by U3 and U6, there is another form of unemployment when someone is engaged in work that is less productive than what they normally do. When someone is stuck in a less productive, lower-paid job because of weak demand, that is called **underemployment** or disguised un-

Underemployment. A situation in which people have jobs that do not make full use of their skills or productive potential. Also called disguised unemployment.

employment. It is unemployment in the sense that some of the person's productive potential is going to waste because of the low level of spending in the economy. There is no generally accepted measure of underemployment.

Economists also sometimes look at the *employment-population ratio* – the ratio of people with paid work of any kind to the total population. This is especially useful for making comparisons across countries, since different countries' statistical agencies may define unemployment differently. Since there were 149 million employed people out of a total civilian noninstitutional population of 251 million, the US employment population ratio as of August 2015 was $149 \div 251 = 0.59$ or 59 percent. Employment population ratios are also calculate for particular demographic groups, such as men between 21 and 64.

High unemployment may be explained either by structural factors – a mismatch between job openings and unemployed workers – or by a lack of demand. Policy responses to structural unemployment focus on the labor market; responses to cyclical or demand-deficiency unemployment focus on the level of public and private spending.

Besides the question of measuring unemployment, economists also classify it based on its causes. Cyclical or **demand-deficiency unemployment** describes people who cannot find jobs because not enough is being produced in the economy to require their labor. **Structural** unemployment is due to a mismatch between workers and employers – either workers don't have the right skills, or they are located in the wrong part of the country, or for some other reason the available workers don't fit the available jobs. Unemployment that results from government regulations or union rules that keep wages "too high" is also considered structural. **Frictional** unemployment is unemployment that results from the normal transitions in a person's worklife – it takes time to find a job after entering the laborforce for the first time, moving to a new area, leaving the military or prison, etc., and during this period of looking the person will be unemployed.

Demand-deficiency unemployment. Unemployment due to a lack of demand for goods and services in the economy. Sometimes also called cyclical unemployment.

Structural unemployment. Unemployment due to a mismatch between workers and the available jobs, such as a lack of appropriate skills or being located in different parts of the country.

Frictional unemployment. Unemployment that results from normal transitions between jobs or in and out of the laborforce.

When there is excessive unemployment, our beliefs about which of these causes is at work will determine what kind of solution we look for. Cyclical unemployment will be reduced by anything that creates more spending in the economy – high government spending, higher business investment, higher exports, lower interest rates, etc. Structural unemployment require interventions to improve the fit between workers and jobs, most often training or education programs, but also potentially help relocating to new areas or subsidies for businesses that match the unemployed workers. Frictional unemployment

is not necessarily a problem – it's often a good thing if people take some time to search for the right job. But it can be reduced through *active labor market policies* that help place unemployed workers in jobs. Direct job creation by government will reduce unemployment whether it is cyclical, structural or frictional.

In general, Keynesian economists argue that most variation in unemployment is explained by cyclical factors, while classical economists believe structural factors are more important. The strength of Okun's Law tends to support the Keynesian position. On the other hand, there are periods when Okun's law seems to hold less well, and then there is a stronger case for looking at structural factors.

Another measure of whether unemployment is structural or cyclical is the **Beveridge curve**. This compares the unemployment rate to the number of vacant jobs listed by employers. If unemployment is mainly caused by weak demand, then we will expect a negative or downward-sloping relationship – unemployment will be high when there are few job vacancies, and unemployment to be low when there are lots of job vacancies. This fits the idea that people are unemployed because there simply not jobs available. On the other hand if we observe lots of vacant jobs at the same time as we see lots of people unemployed, that suggests that there is some mismatch between the unemployed workers and the unfilled jobs. In other words, unemployment in this case is more likely to be structural.

Beveridge curve. A relationship between the number of unemployed workers and the number of job vacancies. When unemployment is high and vacancies are low, unemployment is likely due to deficient demand; when unemployment and vacancies are both high, unemployment is more likely to be structural.

Inflation

Inflation is a general increase in prices.

Equivalently, it is a fall in the purchasing power of money. The opposite of inflation is **deflation** – a general fall in prices, or a rise in the purchasing power of money.

Deflation. Negative inflation, or a decline in the price level.

People sometimes talk about inflation as an increase in the quantity of money circulating the economy, but this is not part of the definition of inflation and is not an especially helpful way of thinking about it. In modern economies, most money takes the form of bank deposits and similar assets, and the amount of bank deposits changes whenever someone takes out or pays back a loan. So there is not really any such thing as a fixed quantity of money. Rather than saying inflation occurs when there is an increase in the money circulating in the economy, it would be more accurate to say that the amount of money in circulation changes based on the level of inflation.

Inflation is defined as the percentage change in the **price index** over a year. The price index itself is an *index number* that is arbitrarily defined as equal to 100 for a certain year. The price index then increases for each subsequent year by the rate of inflation. Economists call numbers expressed in dollars **nominal**; when they are adjusted for inflation, economists call them **real**.

Price index. A measure of the average price of goods and services at a given time and place. If the price index is 1 percent higher in one year than another, that means the price of the "typical" good is 1 percent higher in that year. Since prices don't all change together, a given price index is defined only for a particular basket of goods.

Nominal. Measured in units of money, not adjusted for inflation. Prices and many other numbers in economics are normally measured in money. If we try to adjust a number for changes in the value of money, that gives us a "real" figure. If we don't make any such adjustment but simply use the money value as is, that is a nominal figure.

Real. Economists describe a number or variable as "real" if it has been adjusted for inflation.

Any time we are comparing prices of money quantities from two different years, or the discussing the rate of change of a price or quantity of money, we should consider correcting it for inflation. An annual income of $25,000 meant something very different in 1975 than it does today, because most goods and services were much cheaper then.

There are a number of different measures of inflation.

Inflation is defined as a general rise in prices, but that immediately raises the question: Which prices? Different prices change at different rates, so to construct an index of inflation we must decide which prices we will include and how we will weight them.

In the US, the most important measures of inflation are:

The consumer price index, or CPI. This measures the change in prices of the average basket of goods and services consumed by urban households in the US. The mix of goods in the basket is updated every few years.

The personal consumption expenditure deflator, or PCE. An alternative measure of the price of consumption goods purchased by US households. The main difference between the PCE and the CPI is

that the weight of goods in the CPI basket is based on surveys of households, while the PCE basket is based ons ales reported by businesses. The PCE also includes purchases on households behalf by third parties, such as medical spending paid for by public and private insurance; the CPI includes only direct spending by households.

The GDP deflator. This measures the price of final goods produced in the US – the same set of goods that are included in GDP. So it differs from the CPI and PCE in that it includes investment goods purchased by businesses, and does not include any imported goods.

The producer price index, or PPI. This measures the price received for all goods produced by US companies, including intermediate as well as final goods.

There are also a number of less widely used indexes.

The CPI is often also computed as a "core" measure, which excludes the most volatile prices – typically food and energy – to give a better sense of underlying trends.

All of these measures can be useful, but none of them is the "true" value of inflation. In fact, there is reason to doubt whether the long run change in the price level is something we can measure at all.

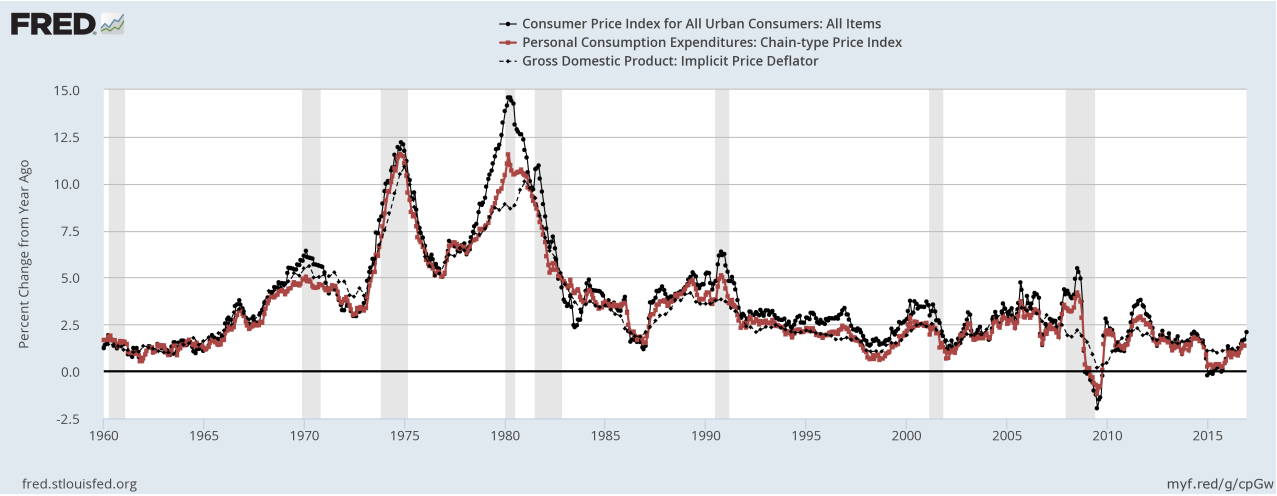


Figure 4: US inflation rates 1960-2016, various measures

As Figure 4 shows, the three main measures of US inflation generally move together, peaking at around 10 percent in the late 1970s and generally staying near the Fed's 2 percent target in recent decades. But as the figure shows, in a given year the different inflation measures may be several points apart. This creates challenges for policy-

makers, since the inflation rate may seem too high by one measure but too low by another.

Construction of price indexes is complicated by changes in the types of goods produced in the economy.

When we are comparing the "average" price level in two widely separated periods, or two countries with very different economies, it is hard to know what to do about the many goods that are produced in one but not in the other.

Another problem is the change in the characteristics of the "same" goods over time. The government agencies that construct price indexes try to adjust for quality changes in goods – for better or worse – since paying the same price for a better good is considered equivalent to paying a lower price to the same good. There is no consensus on how to make these adjustments, and they can be quite large, especially for computers and related goods. If these quality improvements are overestimated, that will tend to reduce the reported rate of inflation.

One reason these issues matter is that many government benefits and taxes are adjusted for inflation, including Social Security. If the government adopts a lower estimate of inflation, that means that Social Security benefits will increase at a slower rate over time.

Today, inflation is almost always positive, but before World War II, deflation was common.

Despite the differences between the various measures of inflation, some general patterns are clear.

Along with interest rates, inflation is probably the macroeconomic variable with the longest recorded history. From the middle ages up until World War II, periods of rising prices alternated with periods of falling prices; over the long run, the average rate of inflation was close to zero. Over the past 70 years, however, the United States and most other countries have experienced only positive inflation; other than Japan, no major country has experienced a significant period of falling prices.

Since the 1990s, 2 percent inflation has been generally defined as "price stability" and macroeconomic policy in most rich countries has tried, generally successfully, to keep it around that level. (Inflation is usually higher in poor countries.) Between the late 1960s and early 1980s, most of the world saw inflation well above this level, with most countries experiencing at least brief periods of inflation rates above 10 percent. A significant number of countries have expe-

rienced episodes of **hyperinflation**, with inflation rates reaching the hundreds, thousands of or even millions of percent.

Hyperinflation. An extremely high rate of inflation. There is no exact cutoff, but most people would consider inflation to become hyperinflation when it is measured in hundreds or thousands of percent per year.

Rapid output growth and low unemployment are generally associated with high or rising inflation. Recessions, stagnant or falling output, and high unemployment are generally associated with low or falling inflation. For smaller countries, there is also a strong link between inflation and the exchange rate. An appreciation (strengthening) of the nominal exchange rate is usually associated with low inflation, while a depreciation (weakening) of the exchange rate is generally associated with higher inflation.

Controlling inflation is a central goal of macroeconomic policy....

Historically, price stability has been the most important goal of macroeconomic policy. Up to about 25 years ago, "price stability" was understood to mean no change in the overall price level – that is, zero inflation on average. Prior to World War II, most rich countries did have average inflation around zero over long periods, though they often experienced rising or falling prices over several years at a time. For most of this period, under the **gold standard**, stability of domestic prices was a secondary goal; the overriding priority was maintaining the purchasing power of domestic currency in terms of gold. From World War II to the 1980s, there was a general agreement that while zero inflation was the ideal, higher inflation might be a necessary cost of achieving other goals of policy, especially low unemployment. In practice, inflation in the single digits or even low double digits was often considered acceptable.

Gold standard. A monetary regime in which the value of the currency is irrevocably set a fixed quantity of gold; the central bank or some other government authority commits to freely buying or selling gold at the official price in any amount required; paper money is backed by gold; and bank lending is strictly limited by the gold reserves available. From the mid-19th century until World War I, most of the world's countries tried to adhere to the gold standard .

Since the 1990s, there has been a general consensus that "price stability" means inflation around 2 percent. In other words, 2 percent inflation is one of the main *targets* of macroeconomic policy. By law the Federal Reserve (the US central bank) must balance the goal of price stability against the goal of low unemployment. For central banks in most other countries, price stability is the only goal that central banks are supposed to pursue. For an **inflation-targeting** central bank, unemployment, output growth, and other macroeconomic variables matter only insofar as they affect inflation.

Inflation-targeting. Describes a macroeconomic policymaker, usually a central bank, whose only goal is to keep inflation at a certain level. An inflation-targeting central bank looks at other macroeconomic targets only insofar as they are thought to affect the inflation rate.

Why 2%? In fact, there is no good reason. Orthodox economic theory suggests that any rate of inflation is as good as any other, as long as it is constant over time (and in particular, as long as it does not change unpredictably.) Statistical studies suggest that high levels of inflation are harmful for economic growth, but only at rates well above 2 percent – most studies do not find a detectable negative effect of inflation on growth unless inflation is over 10 percent. In practice, the 2 percent target was agreed on as a compromise between

economists and policymakers who believed that inflation ought to be zero on average, and others who thought that substantially higher rates could be acceptable or even beneficial. The 2 percent target was also adopted because it happened to be near the actual rate of inflation in a number of countries at the time that they adopted inflation targets in the 1990s. In any case, however it was arrived at, the 2 percent inflation target is now taken very seriously by central banks and other policymakers in most rich countries. (Developing countries usually tolerate higher levels of inflation, and may still may more attention to the exchange rate than to the domestic price level.)

More recently, there has been a renewed discussion of whether higher inflation – either temporary or permanent – might be desirable. A number of economists have suggested that central banks should change their inflation targets to 3, 4 or 5 percent, and/or should be more prepared to tolerate temporary periods of inflation above their target.[4] There is also a difference between those who think that the goal should be inflation as close to 2 percent as possible, with rates both above and below that to be avoided, and those who think of 2 percent as a ceiling, with lower rates also acceptable. The Federal Reserve takes the first view, the European Central Bank takes the second.

[4] For a good discussion of the case for a higher inflation target, see Paul Krugman 2014, "Inflation Targets Reconsidered".

The most important thing to know is that price stability is a central goal of macroeconomic policy and the only goal that many central banks are supposed to pursue; and that in practice price stability means inflation of 2%.

... but there is no agreement about what the costs of inflation actually are.

Despite the broad agreement that preventing high inflation is one of the most important goals of macroeconomic policy, there is surprisingly little agreement on why inflation matters. Orthodox economic theory holds that money is *neutral* in the long run, meaning that the long-run path of "real" variables like employment and output should be the same no matter what happens to the price level. This means that if we want to predict, say, the level of real GDP 20 years from now, we should make the same prediction whether we expect inflation to average 0%, 2%, 10% or -5% over the decade. Even in the short run, the only way inflation can have any real effects is if it unexpected. If people expect inflation, they will simply adjust money contracts for inflation with no other effects. For example, if inflation was formerly 2% but everyone knows it will be 3% in the future, then lenders will demand interest rates one point higher, and borrowers will be willing to pay interest rates one point higher, so exactly the same loans will be made.

The orthodox view is that inflation itself does not matter. However, unexpected changes in inflation do matter, because they mean that people enter into contracts that, after the fact, turn out to be mistakes. Uncertainty about future inflation also matters, since it may prevent people from making contracts that would have been mutually beneficial, or may trick people into making contracts that turn out to leave one party worse off. So while the level of inflation doesn't matter, it does matter that people feel confident about it, whatever it is. And since many important economic decisions are forward-looking, it is important that people also feel confident about the inflation rate in the future. For example, if you sign a 30-year mortgage loan, the "real" interest rate on the mortgage depends not just on inflation today, but on inflation over the next 30 years. So if we think that long-term loans serve an important economic function, we need people to be confident that inflation rates will not change unexpectedly. This means that no matter how the current inflation target – today, 2% – was arrived at, central banks need to stick to it. What's more, they need to demonstrate their commitment to the target, for instance by accepting a higher rate of unemployment than is otherwise socially desirable.

Economic variables measured in dollars (or other currency) are called nominal variables. "Real" variables are nominal variables adjusted for changes in the price level.

Most economic outcomes that we observe are measured in currency. The income of an individual, the price of a house, the GDP of a country or the balance of trade between two countries are all quantities of dollars or of some other currency. But since prices change over time, the amounts of real goods and services that a given number of dollars can purchases also changes over time. So if we want to compare the values of some economic variable in two different time periods, it is often desirable to adjust it for changes in the price level, or to "correct for inflation." The **price level** refers to the average price of goods and services in a country; *inflation* is the rate of change in the price level. Negative inflation (a fall in the prices of goods and services) is called *deflation*. A variable corrected for inflation is often called an *real* variable, while a variable that has not been corrected for inflation is a *nominal* variable.

Price level. The average price of goods and services at a given time and place, as measured by a price index.

A change in the price level is equivalent to a change in the value of money. If goods and services today cost more than they did last year, that is the same as saying that a dollar this year is worth less than a dollar last year. So another way of thinking of inflation, is as a decline in the value of money. Because the value of money changes

over time, we can't simply talk about prices in dollars (or other currency) if we want to compare prices at different times. We have to say "2016 dollars" or "2010 dollars", etc., since dollars were worth different amounts in those different years. When we measure a price in 2010 dollars, we are measuring it in terms of the goods and services a dollar could buy in 2010, which will be somewhat more than a dollar can buy today.

The inflation rate is defined as the annual percentage change in a *price index*. So correcting for inflation first requires picking a particular price index to use. Every country's national statistical agency produces a number of different price indexes. The most widely used price index in the US is the *Consumer price index (CPI)*, produced by the Bureau of Labor Statistics in the Labor Department, but there are many others, including the personal consumption expenditure deflator, the producer price index, and the GDP deflator. (Most US price indexes other than the CPI are produced by the Bureau of Economic Analysis in the Commerce Department.) Different indexes will be suitable for different purposes – there is no one true price level or inflation rate. But if someone talks about "inflation" in the US without specifying a price index, they are probably talking about the change in the CPI.

The procedure for adjusting for inflation depends on whether we are adjusting a price or quantity; a rate or change; or an exchange rate.

There are three different ways to correct for inflation, depending on the kind of variable we are adjusting. If the variable is a quantity of currency, we divide by the price index in the current year and multiply by the price index in the base year. If the variable is a rate or a percentage change, we subtract the inflation rate. And if the variable is an exchange rate, then subtract the inflation rate of one country and add the inflation rate of the other.

1. A variable with units of dollar (or of some other currency), such as the income of a person, the GDP of a country, or the value of an asset.

   To correct for inflation for a value in dollars, you first must choose which year we will convert the value to. For instance, if we want to compare a price in 2010 to a price in 2015, we can either convert the 2010 price to 2015 dollars, or convert the 2015 price to 2010 dollars. Then we look up a price index that includes both years, and apply the following formula:

$$\textit{price in year 2 dollars} = \textit{price in year 1 dollars} * \frac{\textit{index in year 2}}{\textit{index in year 1}}$$

For example, suppose a house was purchased for $400,000 in December 2010 and sold for $420,000 in December 2010. So over the past five years, the nominal value of the house has increased by 5 percent. We want to know how much the real value increased (or decreased, as the case may be). We can look up the CPI in many places online, such as the FRED website, where it is at `https://research.stlouisfed.org/fred2/series/CPIAUCSL`. We see there that the value of the index for December 2015 is 238. In December 2010 it was 220. So if we want to convert the $400,000 price of five years earlier to 2015 dollars, we calculate:

$$\textit{price in year 2010 dollars} * \frac{\textit{index in 2015}}{\textit{index in 2010}} = \textit{price in year 2015 dollars}$$

$$\$400,000 * \frac{238}{220} = \$400,000 * 1.08 = \$433,000$$

So in 2015 dollars, the house has gone from a price of $433,000 to $420,000 - while its nominal value has increased, its real value has actually declined. To be exact, the change in its real value is equal to $(\$420,000 - \$433,000)/\$433,000 = -3\%$. Note that this is very close to the percentage increase in the percent change in the nominal value of the house (5%) minus in the price index (8%). For small changes this will always be true, but it becomes less so as the price changes get bigger.

Note that we can just as easily convert the 2015 price to 2010 dollars, rather than converting the 2010 price to 2015 dollars. In that case, 2010 would be year 2 and 2015 would be year 1. So we would calculate:

$$\textit{price in year 2015 dollars} * \frac{\textit{index in 2010}}{\textit{index in 2015}} = \textit{price in year 2010 dollars}$$

$$\$420,000 * \frac{220}{238} = \$400,000 * 0.92 = \$388,000$$

In 2010 dollars, the house has increased from $400,000 to $388,000. So the change in its real value is equal to is equal to $(\$388,000 - \$400,000)/\$400,000 = -3\%$ As you can see, while the two dollar values are different depending which base year we choose, the percentage change in the real value is the same. No matter what base year we choose, we will come to the same conclusion, that the value of the house declined by 3 percent in real terms. Published price indexes always have a value of 100 for the **base year**, but

Base year. The year for which a price index is defined to be equal to 100. Every index must have a base year, but it makes no difference which year is chosen.

it makes no difference which year is used for this purpose. We follow the same procedure for converting two prices to a common year regardless of the base year for the index we are using.

The -3% figure we calculated above is the *total* real change over the period. To find the annual change, an easy approximation is to simply divide by the number of years, in this case five. So we would say that between 2010 and 2015, the real value of the house declined at an average rate of 3/5 percent, or 0.6 percent, per year. Note: This is an approximation. To get the exact annual rate of change, you need to calculate $(1+p)^n - 1$, where p is the percentage change expressed as a decimal (0.03 in this case) and n is the number of years. For the problems we will solve in this class, you do *not* need to use this formula; just dividing by the number of years is close enough.

2. A variable with units of percent, such as an interest rate or a growth rate.

   To correct for inflation for a rate or a percent change, you can simply subtract inflation. This is not exactly correct; it is an **approximation** that is very close to the correct value as long as we are talking about inflation rates of just a few percent a year, and periods of time of no more than a few years. For example, nominal US GDP today is about 3.7 percent higher than it was a year ago. (GDP has been consistently growing at between 3.5 and 4 percent since the recession ended in 2010.) In other words, the growth rate of GDP is currently 3.7 percent. But this is a nominal growth rate; it does not take account of the fact that dollars are worth somewhat less today than they were a year ago. If we are interested in the change in the amount of goods and services produced in the US over the past year, we may want to correct the nominal growth rate for inflation. Inflation over the past year has averaged 1.6 percent. So to find the real growth rate of GDP, we calculate:

$$\textit{real growth rate} \approx 3.7\% - 1.6\% = 2.1\%$$

   The real growth rate of GDP is just a bit over 2 percent. Again, this is an approximation, but for most inflation rates we see in the real world (and for all problem in this class), it will be good enough.

   Another example: Suppose you were thinking of buying a house, and find that you can get a mortgage loan at a 5 percent rate of interest. Your parents tell you that when they bought their first house in the early 1990s, they had to borrow at a 7 percent interest rate, so you are getting a good deal. But are you? In the early 1990s, inflation was as high as 5 percent. That means that the

Approximation. A method of calculating a number that gives something close to the right result, and is easier or more convenient than calculating the number exactly. When an equation is approximation, we use $\approx$ rther than a standard equals sign.

real interest rate paid by your parents was only $7 - 5 = 2$ percent, while you are facing a real interest rate of $5 - 1.6 = 3.4$ percent. In other words, while nominal interest rates were higher then, the burden of the loan was less, because each year its value was eroded more by inflation than it will be today.

To get the exact correction for a rate or change over a large number of years, you will need to convert the inflation rate to a price index and use the rate or change to compute the levels at the beginning and end of the period, as described below. But for most purposes, the approximation of subtracting the inflation rate is good enough.

3. The change in the exchange rate between two countries.

   Correcting an exchange rate – the price of one currency in terms of another – for inflation is somewhat more complicated because you must take into account inflation in both of the countries concerned. We will discuss this when we study exchange rates.

To convert between a price index and an inflation rate, just remember that inflation is the change in the index between two dates.

You can think of the price index as being the price of a typical or representative good in the economy. So if the price index is, say, 100 in year 1, 104 in year 2, 108 in year 3, and so on, that means that a good that cost \$100 in year 1 would cost \$104 in year 2, \$108 in year 3, and so on. Since inflation just means the average change in price of goods between two years, the inflation rate is the percentage change in the index. So if you have an index with two years, you can calculate the inflation rate as:

$$\textit{inflation between year 1 and year 2} = \frac{\textit{index in year 2}}{\textit{index in year 1}} - 1$$

This is the same as

$$\textit{inflation between year 1 and year 2} = \frac{\textit{index in year 2} - \textit{index in year 1}}{\textit{index in year 1}}$$

If we have an inflation rate and want to go to an index, again we just remember that the inflation rate is the change in the index. So if we have two dates a year apart, then:

$$\textit{index at date 2} = \textit{index at date 1} * (1 + \textit{inflation rate between dates 1 and 2})$$

What if the inflation continues at the same rate for a number of years? Well, each year the price index will be multiplied by $(1 +$

inflation). So if the same inflation rate continues for n years, then the price index at the end of that time will be equal to multiplying by $(1 + \textit{inflation})$ a total of n times. So:

$$\textit{price index index after n years of inflation rate } i = \textit{initial price index} * (1+i)^n$$

If we want to go from two values of the price index to the average inflation rate in the intervening period, we just reverse this and write:

$$i = (\frac{\textit{price index index at end}}{\textit{price index at start}})^{\frac{1}{n}} - 1$$

In math this is called the **geometric mean**, as opposed to the more familiar arithmetic mean. As you can see, when $n = 1$ this is the same as the formula for two successive years given above.

Geometric mean. An average of numbers that are multiplied together. (The more familiar arithmetic mean is the average of numbers that are added.) A common use of the geometric mean is to convert between a total change over some period and the average annual rate of change during the period. If x_1 is the value at the beginning of the period and x_2 is the value at the end, the average annual rate of change is $(x_2/x_1)^n - 1$, where n is the number of years in the period.

When inflation rates are low and we are looking at just a few years, we can get an approximately correct answer by using the arithmetic mean instead:

$$i \approx (\frac{\textit{price index index at end}}{\textit{price index at start}} - 1)/n$$

Using the example above, the average inflation rate between 2010 and 2015 was $(\frac{238}{220})^{\frac{1}{5}} - 1 = 1.08^{0.2} - 1 = 1.0159 - 1 = 0.0159 = 1.59\%$. The approximation gives us $(1.08 - 1)/5 = 1.63\%$. So in this case they are very close.

Again, for the problems in this class, the arithmetic mean – dividing the total change by the number of years – will be good enough. You do *not* need to learn the geometric mean formula. But you should be aware that if we were dealing with very high rates of inflation, or very long periods, it would become less accurate, and it would be better to use the geometric mean instead.

Despite the names, nominal variables are the ones we directly observe in the world, while "real" variables are constructed by economists and depend on various assumptions.

Remember, a price level is the average price of a *basket* of goods and services. But many different goods and services are produced, and their prices do not all change at the same rate, so correcting for inflation requires choosing the most relevant basket of goods. For household income and goods and assets purchased by households, the Consumer Price Index (CPI) is normally used – it counts the prices of a basket of goods consumed by a representative urban household. If someone talks about "inflation" without saying which index, they are probably referring to the CPI. But for GDP and similar aggregate

variables, the **GDP deflator** is more relevant – it counts the prices of the same goods counted in GDP. Other indexes, such as the *Personal Consumption Expenditure Deflator*, or the *Producer Price Index*, may be used for other purposes. These different indexes do not always behave the same way, and it is not always obvious which is the right one for a given question. For example, Social Security benefits are **indexed** (increased each year) to their real value constant using the CPI. But some economists argue that they should be indexed using the basket of goods typically consumed by retirees, rather than the basket of goods consumed by all households. since retirees consume more of goods whose prices rise rapidly, such as health care, and less of goods whose prices rise more slowly or even fall over time, like computers, it is arguable that they face a higher level of inflation than the general population, and the price index used for their income should reflect that.

GDP deflator. A price index used to convert nominal Gross Domestic Product (GDP) to real GDP. It includes all goods and services that are counted in GDP, as opposed to the goods and services consumed by a typical household which are used for CPI.

Price indexing. Automatically adjusting some ongoing payment for inflation, so that its real value is constant over time.

In addition, when we are correcting an interest rate for inflation, we have to pick inflation over the right time period. For example, if you are taking out a 30-year mortgage today, the real burden of that loan depends not just on inflation today, but on inflation over the whole 30-year life of the loan. But of course, we do not know what inflation will be in future years. So while the nominal interest rate on the loan is a hard fact, written into the contract, the “real" rate is a more or less uncertain guess. (And the borrower and lender may have different guesses.) So despite the name, “real" variables are not really real – while nominal quantities really exist out there in the world, converting them to real quantities always involves a judgment call.

Links Between Macroeconomic Aggregates

Demand and Output: The Multiplier

Fiscal policy refers to the use of the government budget as a tool, or instrument, to change the level of GDP or other macroeconomic aggregates.

Fiscal policy means adjusting government spending and/or taxes and in order to change the level of output. In other word, it means using the government budget as an *instrument* to affect the *target* of output, instead of (or in addition to) the target of government debt. If a government, for example, raises taxes to try to restrain private spending because of fears of inflation, that would be an example of fiscal policy. If the government raised taxes for some other reason – like to close a deficit, or to discourage some undesirable behavior, or to redistribute income – that would not be fiscal policy – though it would affect output just the same.

Fiscal policy. The use of government spending and/or taxes as a tool to change the level of output.

Government policies that tend to raise output are called **expansionary**. Policies that tend to reduce output are called *contractionary*. **Stimulus** is another term for expansionary fiscal policy; **austerity** is another term for contractionary fiscal policy. Note that stimulus is always enacted in order to raise output, while austerity may be enacted with the goal to reducing output, but is more likely to be aimed at some other target, usually reducing government debt.

Expansionary. Has as its intended or primary effect an increase in output.

Stimulus. Expansionary fiscal policy.

Austerity. Contractionary fiscal policy, usually with the goal of reducing the ratio of government debt to GDP.

Because government spending and taxes are normally set by legislatures, fiscal policy must pass through the same legislative process as the passage of any other law. This is different from **monetary policy**, which is set by an independent authority, the central bank.

Monetary policy. Actions taken by the central bank to change the level of output or other macroeconomic outcomes. Often consists of changing a single short-term interest rate (the "policy rate", or in the US, the federal funds rate) but can also include all kinds of decisions by the central bank that affect the price or availability of credit.

The underlying idea behind fiscal policy is that there is not a fixed amount of production in the economy. An increase in money expenditure can, at least potentially, increase total incomes and output. For fiscal policy to be effective, an increase in government spending, or a reduction in taxes, should cause households and businesses to also spend more – or at least, it must not cause them to reduce spending by as much as the government increased it.

The multiplier describes the response of output (GDP) to an autonomous change in spending.

The most important question about fiscal policy is the size of the **multiplier**. The multiplier is the ratio between the change in the government budget position and the resulting change in output or GDP. For example, if we think that an increase in government spending of $10 billion will result in GDP rising by $15 billion, then the multiplier would be 1.5. (In this case, there would be $5 billion

Multiplier. The relationship between a change in investment, government spending, or other autonomous expenditure, and the change in output that results from it. Mathematically, the multiplier can be expressed as $\frac{\Delta Y}{\Delta A}$ where Y is output, A is autonomous expenditure, and Δ means change.

of additional private spending, on top of the government spending.) The multiplier will depend on the country, on economic conditions, and on the specific budget changes involved. But there are some things we can say about it in general.

Here is one way to think about the multiplier. Suppose the city of New York begins some major new expenditure – on expanding the subway, let's say. The subway expansion itself counts in G, government final expenditure, and adds to GDP. But each person who receives an income from the project – individual employees on the project, obviously, but also the owners of businesses that contract with the city – will spend some of that income. (The employee might buy a new winter coat; the contractor might buy a new summer house.) Some of the income will be saved, or spent on imported goods, or paid in taxes but some will be spent on local goods and services. The sellers of those goods and services will then receive income, some of which will be spent locally in turn. And the people who receive *that* income will spend some of it, and so on. Because there are **leakages** at each stage, the increase in spending will come to an end eventually. But it will be greater than original spending. The multiplier tells us, if we add up all the additional spending how does it compare to the original government spending that started the process? It is clear that people spend a lot of their income on locally produced goods and services, the multiplier will be high; if large parts of income are saved, taxed, and spent on imported goods, the multiplier will be low.

Leakage. Uses of income that do not contribute to aggregate demand, and do not directly create income for other units in the economy. The most important leakages are saving, imports and tax payments.

The most common use of the multiplier is to estimate the effects of fiscal policy on output. The multiplier used in this way is referred to as the **fiscal multiplier**. But the same analysis applies to any change in **autonomous spending** – spending that changes for reasons unrelated to current income. The multiplier describes how much GDP changes in response to a change in investment, or a change in exports, just as much as it describes how much GDP changes in response to a change in government spending.

Fiscal multiplier. The multiplier applied to changes in government spending.

Autonomous spending. Spending that does not depend on current income.

The multiplier is the ratio of the final change in GDP to the initial change in spending.

Mathematically, the multiplier is simply the ratio of the resulting change in output to the initial change in spending. Using Y for output, A for autonomous spending, and Δ to mean change, we can write:

$$multiplier = \frac{\Delta Y}{\Delta A}$$

Or equivalently,

$$\Delta Y = multiplier * \Delta A$$

For example, suppose you think the multiplier is 2. Then if there is an *autonomous* increase in spending of \$ 100 billion (for example, an increase in government spending, an increase in exports, or an increase in investment), then you would expect GDP to increase by \$100 billion $*2 =$ \$200 billion.

Similarly, if we can observe or estimate the change in output that results from a change in autonomous spending, we can calculate the multiplier. For example, one study of the effects of 2009 stimulus bill looked at how much federal health care spending had increased in various states. They found that for each \$2 million of additional government spending, output in that state was about \$3 million higher. This implies a multiplier of $\frac{3}{2}$, or 1.5.

We can estimate the multiplier if we know what fraction of income is saved, what fraction is spent on imports, and what fraction is taken in taxes.

We can make a **quantitative** estimate of how large the multiplier might be, and what factors will make it larger or smaller.

Quantitative. Measured or estimated numerically. The opposite is qualitative, which refers to something that can be described but not given as a number.

We start with the national income identity:

$$Y = C + I + G + (X - M)$$

Next we make some **behavioral** assumptions. First, we will take government spending, exports and (for the moment) investment to be **exogenous** – that is, fixed or determined outside the model. (To say something is exogenous means that we are taking it as given – our model does not try to explain it. The variables a model does explain are called **endogenous**. In this case, Y, C and M are endogenous.)

Behavioral. Describing the choices of the actors within a model. A behavioral equation is one with parameters that have to be estimated on the basis of data. Behavioral equations may be more or less accurate approximations of the phenomena they describe, but they will never be exactly correct.

Exogenous. Determined outside the model. Variables that a model does not try to explain, but simply takes as given.

Endogenous. A variable that is determined by other variables, as opposed to an exogenous variable that is fixed by policy or by nature.

Another way of saying that a variable is exogenous is to say that our model is treating it as given, or fixed. Economists often convey the idea that a variable is fixed by writing a bar over it. So if x is fixed, we write it as $\bar{x}$. Using this notation, we write:

$$I = \bar{I}$$

$$G = \bar{G}$$

$$X = \bar{X}$$

Consumption, on the other hand, we think depends strongly on current income. There is also an exogenous component, $\bar{C}$, but most consumption spending, we think, depends on current income.

We write this relationship as

$$C = \bar{C} + c * Y_D$$

Lower case c is the **marginal propensity to consume**, sometimes written as *mpc*.

Marginal propensity to consume. The fraction of each additional dollar of income that is spent on consumption.

c and *mpc* can be used interchangeably as abbreviations for the marginal propensity to consume.

For example, if $c = 0.75$ then each dollar of additional disposable income would cause 75 cents of additional consumption spending. Note that the equation includes Y_D, or *disposable income*, not Y. t is the **marginal tax rate** on income. This is included because household consumption does not depend on total income but on *disposable income*, that is, income after taxes.

Marginal tax rate. The fraction of each additional dollar of income that is taken in taxes.

Finally, we think a fixed share of spending goes to imports:

$$M = mY$$

For example, if $m = 0.2$, then 20 cents out of each new dollar spent in the economy goes to imports, leaving the circular flow. So in this case, if GDP increases by \$50 billion, we would expect imports to increase by \$50 billion $*0.2 =$ \$10 billion.

Let's combine the exogenous terms into a single variable, *autonomous spending*, or A for short:

$$A = G_0 + C_0 + I_0 + X_0$$

And instead of using c, the marginal propensity to consume, let's use s, the marginal propensity to save. Since all disposable income is either consumed or saved,we know that:

$$s + c = 1$$

Or equivalently,

$$s = 1 - c$$

Then we have

$$Y = A + (1 - t)(1 - s)Y - mY$$

$$sY + tY - stY + mY = A$$

$$Y = A(\frac{1}{s + t - st + m})$$

The term in parentheses is the *multiplier*.

The multiplier is given by

$$\frac{1}{s + t - st + m}$$

where s is the marginal propensity to save, t is the marginal tax rate, and m is the marginal propensity to import.

What we can see here is that the larger are the various leakages, the smaller will be the multiplier. If savings, taxes and imports are low, the multiplier will be large; if savings, taxes and imports are

high, the multiplier will be small. It's quie possible for the multiplier to be less than one, but, in this simple model, it will always be positive. In other words, no matter how high are tax rates, imports and saving, an increase in government spending will always lead to some increase in total output.

Let's fill in some plausible numbers for the US. Imports in the US are about 15% of national income, but we know that the import share reliably rises in booms and falls in recessions. In other words, the *marginal* import propensity is higher than the *average* import propensity. Statistical evidence suggests that in the US, a 1% rise in income typically leads to a 2% rise in imports. (In other words, the *income elasticity* of imports seems to be around 2 in the US.) So a reasonable value for m is around double the import share, or 0.3. Savings are quite low in the US, and consumption responds strongly to current income. In fact, some people will respond to an increase in income by increasing their consumption by even more than the change in income, implying a *negative* savings propensity. (This may happen because a higher income makes it easier to borrow money, or because some purchases, especially durable goods, are *lumpy* – you have to buy them all in one piece.) But overall, s is certainly positive, especially since an important component of savings is *retained earnings* – corporate profits that are not paid out to shareholders. Overall, $s = 0.2$ is a reasonable first guess. Finally, federal taxes are around 15% of GDP, and state and local taxes are another 10%. Income taxes of course vary with income. Some other taxes, like corporate profit taxes, vary more than proportionately with income, while others, like inheritance taxes, don't vary much with current income at all. Transfers also include payments that vary with income, like Medicaid and unemployment benefits, and payments that don't vary with income, including the two largest transfers, Medicare and Social Security. Overall, a value of 0.2 seems reasonable for t as well.

Put these estimates together and we have:

$$Y = A(\frac{1}{0.3 + 0.2 + 0.2 - 0.04}) = A(\frac{1}{0.66}) \approx A \times 1.5$$

And in fact, 1.5 is quite close to many recent **econometric** estimates of the multiplier based on historical data.

Econometric. Statistical analysis of economic data.

In general, the higher the fraction of income that is spent on domestic goods and services, the greater will be the multiplier.

There are two reasons for this link. First, many households (and businesses) are **liquidity-constrained** – they would like to be spending more than they currently are, but they cannot because they don't have liquid savings available and they cannot (or do not want to)

Liquidity-constrained. Describes a household, business or other economic unit that is spending less than it would otherwise choose to because of a lack of current access to cash or credit. This is distinct from spending that is low because of low income or wealth.

borrow more.For anyone in this position, higher income will increase spending simply because it allows them spend more. Second, even people who have savings or could borrow more, still have to decide what is a reasonable level of spending. Since we can't predict the future, most people or businesses use current income as a rough guide to what income is likely to be in the future. So when income increases, people are likely to want to spend more, unless they have strong reason to think that the current increase is only temporary.

Productivity and Factor Shares

Wages, productivity, employment, inflation and distribution are linked by accounting identities.

Labor productivity (or just productivity, for short) is defined as the amount produced, divided by the amount of labor used to produce it. The labor share is defined as the fraction of total income paid out to labor. We can use these two *identities* to analyze changes in output, employment, wages and prices. This doesn't tell us what will happen in the economy, but it does tell us something about what *can* happen. Using these identities also helps us describe developments in the economy more precisely, and clarifies what assumptions are needed for various stories or predictions about the economy to be true.

There is a mathematical rule that lets us convert equations to linear form, which is easier to work with.

A **linear equation** is one in which the variables are only added or subtracted. None of the variables are multiplied by each other, and none are raised to a power (that is, there are no expressions like x^2). Linear equations are generally easier to work with, so it's convenient to be able to change other kinds of equations to linear ones if possible.

Linear equation. An equation in which the terms are only added or subtracted. None of the variables are multiplied or divided, and none have exponents.

One useful tool for making linear equations is: If $a = b * c$ then

$$\textit{percentage change in } a \approx \textit{percentage change in } b + \textit{percentage change in } c \tag{2}$$

Similarly, if $a = b/c$ then

$$\textit{percentage change in } a \approx \textit{percentage change in } b - \textit{percentage change in } c \tag{3}$$

The Greek letter Δ (delta) is often used to mean the change in a variable. So to save space, I will write $\%\Delta$ when I mean "percent change in..." For example " $\%\Delta$ employment" means "percent change in employment."

So we can rewrite Equation 2 as:

$$\%\Delta a \approx \%\Delta b + \%\Delta c$$

This is linear – the variables are simply added. Whereas $a = b * c$ is not linear, since the variables are multiplied. Note that the original equation described the *levels* of the variables, while the new, linear one describes the *changes* in them.

This works the same if we have more than two variables on the right hand side.

Labor productivity is defined as output divided by employment.

When economists talk about "productivity", they mean either **labor productivity** or **total factor productivity**. Labor productivity is the output produced by a given amount of labor; total factor productivity is the amount of output produced by a given amount of labor and capital. Total factor productivity is important for economic theory, but it is hard to apply in practice, since measuring capital is difficult and you need to make additional assumptions about how the labor and capital are combined. For most practical purposes, labor productivity is more relevant. Whenever someone refers to "productivity" by itself, they almost always mean labor productivity.

Labor productivity. Total output divided by total employment. The most common measure of productivity.

Total factor productivity. Total output divided by the labor and capital used. Important in economic theory but hard to apply in the real world.

Labor productivity is defined as output divided by the amount of labor used. Labor can be measured either in hours of work, or number of people employed. Here we will measure it by number of people employed. So labor productivity is defined by:

$$productivity = \frac{output}{employment} \tag{4}$$

We can measure productivity for the economy as a whole, for an industry or sector, or for a single business. If we are measuring it for the economy as a whole, then "output" is GDP; for an industry or business, it is value added. When we are talking about changes in productivity, we normally measure output in real (or inflation-adjusted) terms.

We can rearrange Equation 4 to get

$$output = employment * productivity$$

In other words, total production in an economy (or an industry or business) is equal to the number of people employed, times the average amount produced by each one.

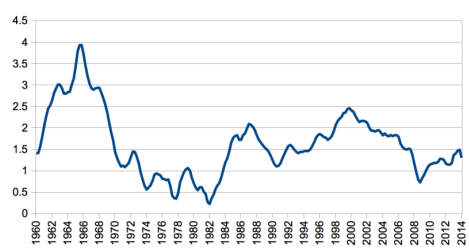


Figure 5: US labor productivity growth rates, 5-year moving averages. In recent years, productivity has grown at less than 1.5 percent per year, which is lower than in much of the postwar period.

The change in employment over some period of time is equal to the change in output minus *the change in labor productivity.*

We can analyze changes in employment in terms of changes in output and productivity. Using Equation 3, we can write:

$$\%\Delta employment \approx \%\Delta output - \%\Delta productivity \tag{5}$$

The percent change in employment is equal to the percent change in output minus the percent change in employment. For example, in 2014, total employment in the US rose by 2.2 percent, output rose by 2.9 percent, and productivity rose by 0.7 percent. (This is an exceptionally low rate by historical standards.) We can apply this equation to a change over one year or over several years. But if we apply it to a very long period (say, 50 years), the approximation may be less accurate.

Equation 5 is an accounting identity: it is true by definition. But it still shows us a couple of things that might not be obvious.

First of all, changes in employment can be due to either changes in total output, or to changes in productivity – that is, either changes in how much is produced, or in how much labor is used for a given amount of production. Over short periods, changes in output growth are much more important. For example, Table 1 shows the average annual change in employment during the expansion of 2002-2007 and the recession of 2008-2009.

Period	Employment	Output	Productivity
2002-2007	0.8%	2.7%	1.9 %
2008-2009	-3.1%	-1.5%	1.6%

Table 1: Average Annual Change in Employment, Output and Productivity

Employment grew at an average rate of 0.8 percent per year over 2002-2007, and fell at a rate of 3.1 percent per year during 2008-2009. This difference is entirely explained by the fact that output was rising during the first period, and falling in the second period. As you can see, labor productivity actually grew somewhat slower during the recession than during the expansion, but the change is quite small compared with the changes in output and employment growth.

Over long periods, faster labor productivity growth could contribute to slower employment growth. This is called "technological unemployment," but it is not clear that it is a real problem.

Over longer periods, however, changes in the speed of labor productivity growth may be more important. The second thing that

Equation 5 tells us is that if output is growing at a constant rate, than faster productivity growth must men slower growth in employment. If productivity grew fast enough, you might even see a situation where output continued to grow while employment fell.

The idea that rapid improvements in labor productivity might lead to a fall in employment is a familiar one in the media and in policy discussions. It is often referred to as **technological unemployment** – the idea that "robots will take our jobs." Obviously, there are many specific cases of work jobs that become obsolete through technological change. But Equation 5 helps us think about this possibility more systematically. To begin with, it highlights the point that a prediction that technological progress will reduce employment is simply a claim that we will see an acceleration of productivity growth but not of GDP growth. But this raises two questions. First, productivity growth has been slowing down in recent years, not accelerating. Since 2000, productivity growth has averaged barely one percent a year, compared with around 1.5 percent per year during the period between 1950 and 2000 (and as high as 3 percent a year during the 1960s). So the "robots will take our jobs" story is not just extrapolating from what is already happening; someone telling this story has to explain why the recent trend of declining productivity will reverse itself. Second, Equation 5 makes it clear that faster productivity growth can lead to lower employment *or* to faster growth in output. Someone telling the "robots will take our jobs" story also has to explain why faster productivity growth will not simply lead to faster growth of GDP. After all, 120 years ago most Americans worked in agriculture. Technological change has resulted in the disappearance of almost all of those jobs. But the result has not been mass unemployment, but rather increased production in the other (secondary and tertiary) sectors of the economy.

Technological unemployment. Unemployment that results from labor productivity rising faster than total output, so that fewer workers are needed.

While the statistics are not decisive either way, there is some evidence that labor productivity rises faster when output is growing more rapidly. In this case, we might see the opposite of technological unemployment – employment and productivity moving together. For example, the early 1930s, when employment fell very steeply, labor productivity actually declined – one of the only periods on record when this occurred. And when employment rose in the recovery from the Depression, productivity rose as well. Note that in this case Equation 5 was still true – as an accounting identity, it is always true – but the big changes in output overwhelmed the effect of productivity on employment.

The technological unemployment issue is an example of why accounting identities are useful. They can't prove that a certain story about the economy is true. But they can clarify what that story

means, and show what assumptions it involves.

The labor share is the fraction of total income going to wages.

Another useful accounting identity is that the **labor share** is the fraction of total income that comes as wages and salaries. In the simplest story, all income is either labor income or capital income. But we can talk about the labor share even if there are other kinds of income. It simply means the fraction of total income that is received by labor. We can write:

$$\textit{labor share} = \frac{\textit{total wages}}{\textit{output}} \tag{6}$$

As with productivity, we can apply this to the economy as a whole or to a particular sector or industry. Note that in the national accounts, the labor share in government and nonprofits is 100% by definition. So if we look just at the business sector, the labor share will always be somewhat lower than for the economy as a whole. Wages here include fringe benefits, like health insurance or pension contributions. In this equations, wages and output are measured in *nominal* terms.

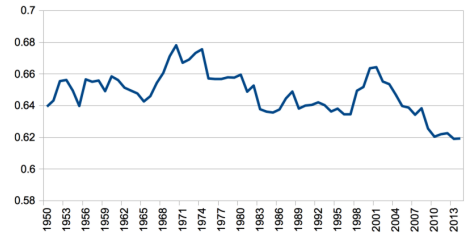


Figure 6: US labor share. During the 1980s, and again over the past 15 years, there was a fall in the share of national income going to labor.

Labor share. The fraction of output going to workers, calculated as total wages and salaries plus benefits divided by total income. Also called the wage share.

The change in the average nominal wage is equal to the change in average productivity, plus the inflation rate, plus the change in the wage share.

Equation 6 is an accounting identity. We also know that total wages equals the average wage times total employment, the real wage equals the nominal wage divided by the price level, and the change in the price level is inflation.

Substituting these identities into Equation 6 and applying the linear approximation gives us[5]

$$\%\Delta \textit{nominal wage} \approx \%\Delta \textit{productivity} + \%\Delta \textit{prices} + \%\Delta \textit{wage share}$$

In other words, the percentage increase in the average nominal wage must be equal to the sum of the percentage increases of labor productivity, the price level, and the wage share. And since inflation is just the percentage change in the price level, we can rewrite this as:

$$\%\Delta \textit{nominal wage} \approx \%\Delta \textit{productivity} + \textit{inflation} + \%\Delta \textit{wage share} \tag{7}$$

To think about what this means, imagine a business that for whatever reason decides to increase wages. What can happen as a result? It might be that profits will fall – that is a rise in the labor share.

[5] Exactly how this is derived is shown in the box nearby.

It might be that it will increase its prices – over the economy as a whole, that is the same as inflation. Or it might be that the higher wages will cause the business to become more productive, perhaps because workers will be more concerned about losing such a good job or because they will feel a greater sense of loyalty. Any of these outcomes are possible. But what we know for sure is that a one percent increase in wages *must* result in some combination of a higher wage share, higher prices, and/or higher productivity, that add up to one percent. This is true both at the level of an individual business and for the economy as a whole.

Equation 7 lets us think systematically about many things that happen in the economy. For instance, what happens if labor productivity grows more rapidly, while nominal wage growth is unchanged? The equation says that in this case either inflation or the wage share must fall. The fact that more rapid productivity growth is *deflationary* – tends to lead to lower prices – is not obvious, but the equation makes it clear.

Macroeconomic theory often assumes that the wage share is fixed. This implies that the increase in real wages (the increase in nominal wages less inflation) must be just equal to the growth of productivity. But in the real world, this is often not the case. Over the past 15 years, nominal wages have increased by an average of 2.9 percent a year, inflation has averaged 2.4 percent a year, and productivity has increased by 1.1 percent per year. This means that the labor share has fallen by a bit over half a percent per year. ($2.9 - 2.4 - 1.1 = -0.6$.) While this might not seem like much, over 15 years that adds up to a 9 percent total decrease – a substantial fall in the share of income going to workers.[6]

[6] This is a *9 percent decrease*, not a *9 percentage point* decrease.

We can turn Equation 7 around and ask what happens when productivity increases.

$$\%\Delta productivity \approx \%\Delta nominal\ wage - \%\Delta prices - \%\Delta wage\ share$$

Again, think of an individual business: Rising labor productivity means they are now able to produce the same quantity of goods with fewer workers. Let's say the number of workers required per unit of output has fallen by 10 percent. What happens? Either the remaining, more productive workers can each be paid 10 percent more; or the company can cut its prices by 10 percent; or the gains from increased productivity can go to higher profits (i.e. a lower labor share). Again, the accounting identity doesn't tell us which of these outcomes will happen. But it does tell us that one of them – or some combination – *must* happen whenever productivity rises.

Note: Deriving Equation 7

We know that the wage share is equal to total nominal wages divided by nominal output. This implies that

$$\textit{total nominal wages} = \textit{nominal output} * \textit{wage share}$$

We know that nominal output is equal to real output times the price level. So substitute that in:

$$\textit{total nominal wages} = \textit{real output} * \textit{price level} * \textit{wage share}$$

We know that real output is equal to employment times productivity, so substitute that in:

$$\textit{total nominal wages} = \textit{productivity} * \textit{employment} * \textit{price level} * \textit{wage share}$$

Finally, we know that the average wage is equal to total wages divided by total employment, so we divide both sides by employment to get:

$$\textit{nominal wage} = \textit{productivity} * \textit{price level} * \textit{wage share}$$

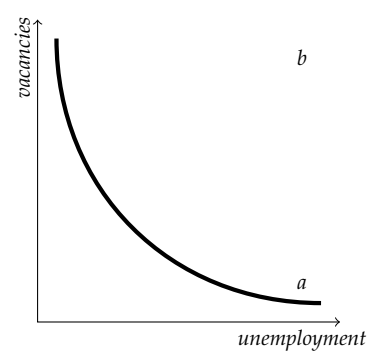
And now we apply the linear approximation of Equation 2, and that gets us Equation 7.

Output and Unemployment: The Beveridge Curve and Okun's Law

The Beveridge curve relates the unemployment rate to the number of job vacancies posted by employers. It is used to help decide whether high unemployment is the result of structural factors or deficient demand.

Whenever unemployment is high, there will be debates about whether this is due to *deficient demand* or to *structural* factors. In other words, are people unable to find work because businesses are not hiring, because there is no demand for additional output; or are people unable to find work because something is stopping them from taking the jobs that are available?

One tool used to answer this question is the *Beveridge curve*. This is a graph with the unemployment rate on one axis and the number of job vacancies being advertised by employers on the other. The idea is that, in a situation where there simply are no jobs available, we should see high unemployment and very few jobs openings posted (point *a* in the figure). While in a situation where the problem is that the unemployed workers are unable to do the available jobs, we should see both high unemployment and a large number of vacancies (point *b* in the figure).



Beveridge curve. At a point like *a*, unemployment is likely to be due to deficient demand. At a point like *b*, unemployment is more likely to be structural.

The idea behind the Beveridge curve is simple: In a situation where people are unemployed because businesses have no need for additional workers, given a lack of demand for their products, relatively few businesses will be trying to hire new workers. (Of course some will be, to replace departing workers and because even in a situation of generally weak demand some businesses will still be expanding.) On the other hand, in a situation where people are unemployed because they can't take the jobs that are available, businesses will be having trouble filling open positions. So there should be a large number of job vacancies.

In the US, and in most other rich countries, unemployment tends to rise when vacancies fall, and vice versa. This is why we normally see a downward-sloping Beveridge curve. In other words, most *short-term* changes in unemployment seem to be driven by demand rather than by structural factors. But the case is less clear when it comes to longer-term changes in unemployment, and to very large changes in unemployment as we have seen since 2008.

Figure 7: US Beveridge Curve

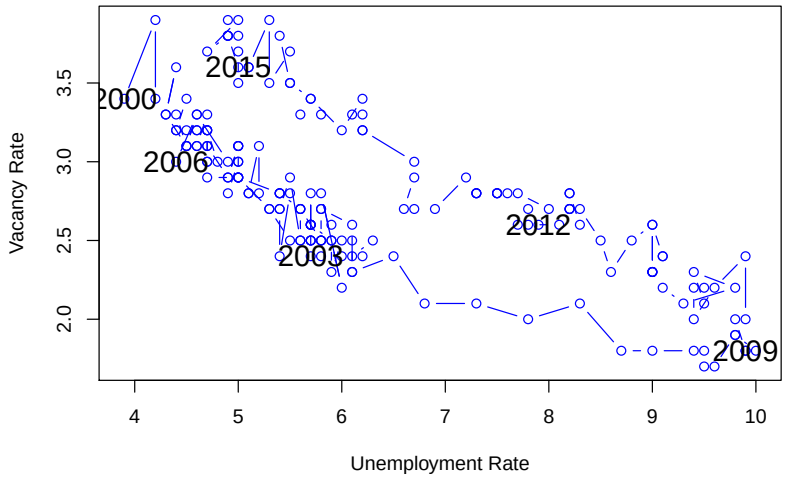
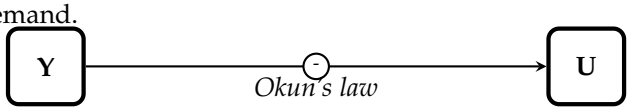


Figure 7 shows the US Beveridge curve from December 2000 to July 2016. So for instance in November 2009, the unemployment rate was 9.9 percent and the vacancy rate was 1.8 percent. As you can see, the points almost all lie near a line running from the upper left to the lower right; in other words, the unemployment rate is high when the vacancy rate is low, and vice versa. This suggests that most of the ups and downs in unemployment since 2000 have been driven by shifts in aggregate demand, rather than by shifts in the fit between workers

and available jobs. However, the curve since 2009 lies above and to the right of the pre-2009 curve. This does suggest that there has been a worse fit since the recession. In other words, while most of the high unemployment since the recession was due to weak demand, some may also have been due to structural factors.

While we most often think of a mismatch between workers and jobs in terms of skills, it can also reflect factors like geographic location or the pay being offered versus what unemployed workers are willing to accept. It can also reflect legal or institutional barriers that prevent people from taking certain jobs.

The Beveridge curve is not the only tool to assess the relative importance of structural versus demand deficiency explanations for high unemployment. Other pieces of evidence we might look at is whether there is low employment and few vacancies across the board, or whether certain sectors are growing while others are shrinking. We also might look at whether wages are rising in some sectors or industries, as they should be if the problem is a lack of suitable workers to fill them. If employment and wages are growing slowly almost everywhere, that suggests the problem is weak demand.



Output and employment are linked via Okun's law.

Okun's law says that when output grows rapidly, unemployment will fall, and when output grows more slowly or falls, unemployment will rise. The exact relationship varies between countries, but within countries it seems to be quite stable over time. It is one of the more reliable empirical rules in economics: In the US, about two-thirds of variation in unemployment rate can be explained by Okun's law.

If we write the change in unemployment as ΔU and the real (inflation-adjusted) growth rate of output as g, then the general form of the law is:

$$\Delta U = -a(g - b) \tag{8}$$

Again, ΔU is the change in the unemployment rate compared with a year ago, and g is the real (inflation-adjusted) growth rate of output. a and b are the **parameters** of the law – the numbers that describe the relationship between the *variables* unemployment and growth. The values of a and b are **estimated** statistically based on historical data. The exact values will depend on the time and place we look at and the details of the estimation process.

When economists estimate Equation 8 using data for the US since

Parameters. Numbers in an equation that describe the relationships between the variables.

Estimation. The process of using statistics to determine the parameters of an equation. The goal is to find the parameter values that give the best fit to the observed data on the variables in the equation.

World War II, they usually find values for a around 0.6 and for b around 2.[7] So for the US, we can write:

$$\Delta U = -0.6(g - 2) \tag{9}$$

[7] In other sources, you will encounter different values for the parameters. The important thing is not the exact values, but the logic of the relationship.

In other words, the change in unemployment is equal to negative 0.6 times the percentage growth rate minus 2 . For example, if GDP grew by 4% in one year, we would expect the change in the unemployment rate to be $-0.6(4-2) = -0.6 * 2 = -1.2$. We would expect unemployment to fall by 1.2 points. On the other hand, if real GDP were to fall by one point, we would expect the change in unemployment to be $-0.6(-1-2) = -0.6 * -3 = 1.8$. We would expect the unemployment rate to increase by 1.8 points. This means that it takes two points of real GDP growth just to hold unemployment constant. In a year when real growth is zero (real GDP is constant), we should expect unemployment to rise by 1.2 points.

Note that this equation doesn't say what the change in the unemployment rate (ΔU) or the growth rate (g) actually are. Rather, it describes a *function* linking the two. It says that *if* growth is high, unemployment is probably falling; and *if* growth is low or negative, unemployment is probably rising. So if you have an idea about what will happen to one of the variables, you can make a good guess about what must happen to the other. The numbers that appear in a function like this are called its *parameters*. In the case of Okun's law, while the basic form of the law is the same across countries, the parameters vary. In general, the first parameter (-0.6) depends on labor market institutions, while the second (2) depends on the growth rates of the laborforce and labor productivity. In the US, both the laborforce and productivity grow by roughly 1 percent per year, so it takes a total of 2 percent additional production just to hold the unemployment rate constant.

In the US, and in many other countries, actual changes in unemployment and GDP growth follow Okun's law fairly closely.

Figure 8 shows real GDP growth rates and change in unemployment in the US for every year from 1950 through 2015. The diagonal line shows the version of Okun's law based on **regressing** GDP growth on the change in unemployment using quarterly data for 1950Q1 through 2016Q2. In this case, the exact parameters are $a = 1.95$ and $b = 0.62$. (While estimated parameters will always depend on the exact data used, as this example shows, in the case of Okun's law, estimated values don't vary too much.) As you can see, most of the points (labeled by year) fall quite close to the diagonal line. For example, in 1960 real GDP growth was negative 0.7 percent, and

Regression. A statistical technique for describing the relationship between some variables. The goal is to minimize 'errors" – differences between the actual data and the estimated relationship. .

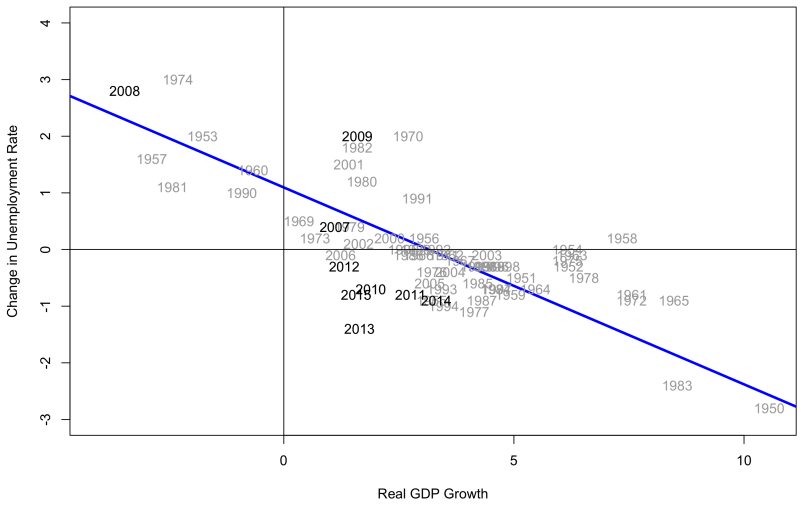


Figure 8: Changs in Unemployment and GDP Growth, 1950-2015

unemployment rose by 1.4 points – almost exactly what we would predict from Okun's law. In the postwar US (and in most other times and places), we never see GDP and unemployment fall together – every year of negative GDP growth is a year of rising unemployment. And almost all the years of strong GDP growth are years of falling unemployment (1958 is the one exception).

Still, as the figure makes clear, there is important variation in unemployment that is not directly linked to output growth, and so is not explained by Okun's law. For example, look at the most recent **business cycle**, including the recession of 2007-2009 and the expansion from 2010 to the present. (These years are printed heavier in the figure.) The 2.8 point rise in unemployment in 2008 was close to what we would predict given the 3.5 percent fall in GDP in that year. But unemployment rose almost as much in 2009, despite the positive GDP growth in that year. And in the years since then, the unemployment rate has consistently declined faster than we would have predicted based on Okun's law. This faster fall in unemployment has been due to a mix of slower, population growth, slower productivity growth (so that more workers are required to produce the same amount of output), and unemployed workers becoming discouraged and exiting the laborforce, but in any case they are not explained by changes in output. It remains to be seen how well Okun's law will describe future changes in unemployment and GDP growth.

Business cycle. Periodic shifts in the level of economic activity. Business cycle expansions see high output growth, low unemployment, and high or rising inflation; business cycle downturns or recessions see output growing slowly or falling, high unemployment, and low or falling inflation. Smoothing out business cycles is a central goal of macroeconomic policy.

The first parameter in Okun's law reflects how easy it is for employers to adjust the size of their workforce. The second parameter reflects the average growth rates of productivity and the workforce.

Why do changes in output produce less than proportional changes in unemployment? At first glance, it seems like the first parameter *a* should be closer to -1. All else equal, producing twice as much goods and services should require twice as many workers. And those new workers will presumably be drawn from the ranks of the unemployed. So we might expect a one point rise in GDP to lead to a one point fall in unemployment. Why does it, instead, usually lead to only half a point fall? There are several reasons:

- Some of the additional labor used for higher production comes from more increased hours by those who already have jobs. Since the unemployment rate only counts people who do not have jobs at all, changes in hours don't show up in it.
- When businesses increase hiring, more people enter the laborforce. And when businesses reduce hiring, more people exit the laborforce. So even if a one percent rise in output always meant a one percent rise in employment, it wouldn't necessarily mean a one percent fall in unemployment. Some of the new hiring would be from people who were not counted as unemployed.
- Hiring and firing workers is costly. So when sales drop, businesses don't immediately lay off workers, even if they currently don't need them. And when sales rise, businesses first try to get extra work out of their current employees, before hiring new ones.

This last factor may be the most important, and is often the most interesting. In many countries, there are legal restrictions on businesses' ability to lay off workers – for example, they may be required to follow seniority in layoffs, or to pay severance benefits. Where workers are represented by unions, union contracts may also limit layoffs. Even in the US, where employers generally have complete freedom to hire and fire as they wish, it may still be costly. New workers must be recruited, vetted, trained, and so on. So if sales fall, it is often cheaper to keep currently unneeded workers on payroll than to get rid of them and then hire them back when sales pick up. This is especially true for jobs that require specific skills, or where workers may be hard to replace for other reasons.

If the parameter *a* reflects how easy or hard it is for employers to hire and lay off workers, that means that estimates of Okun's law for different countries tell su something about labor markets. A country where unions are strong, where there are tight legal restrictions

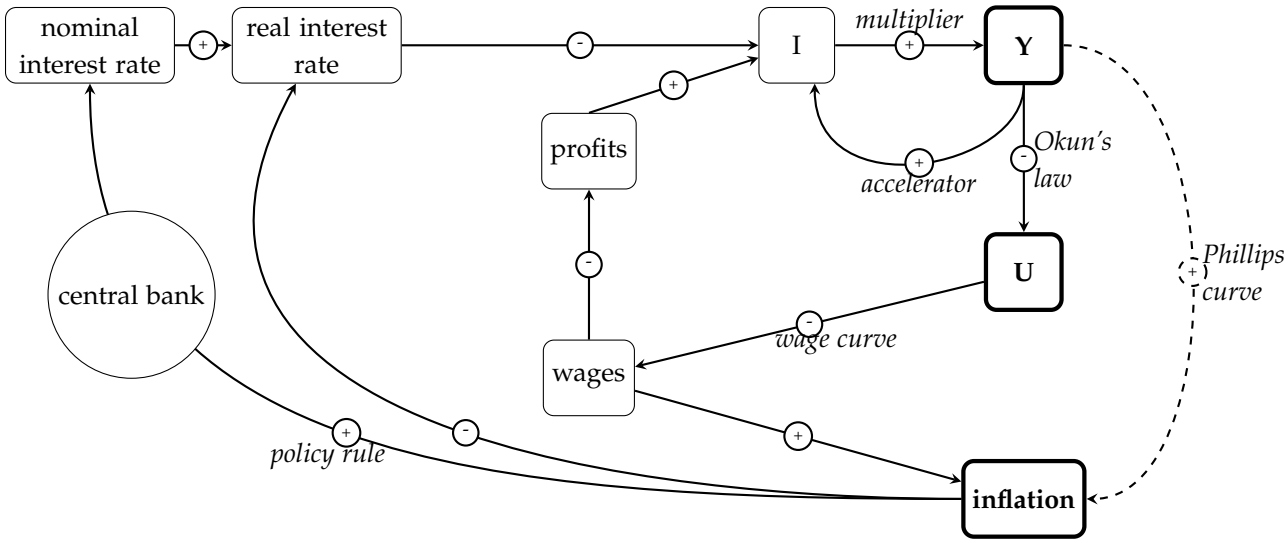
on businesses' ability to fire people, or where important industries require specific skills, the parameter should be close to zero. In a country where unions are weak, where businesses can hire and fire as they choose, and where workers are generally interchangeable, the parameter should be close to one. Of course many factors influence this parameter. But it is worth noting that in the the US it's relatively close to one, while for instance in Japan – a country famous for its "lifetime employment" model, where most workers spend their entire careers at one company – it is close to zero.

The first parameter in Okun's law says how much changes in unemployment respond to GDP growth. The second parameter says how much GDP growth is need just to hold unemployment constant. (In other words, the first parameter tells us the slope of the line in Figure 8, while the second parameter tells us its level.) The second parameter reflects two factors: the long-term average growth of the laborforce, and the long-term average growth of productivity. Because both of these are normally positive, output needs to grow just to hold unemployment constant. The fact that population that expects to work is increasing, means that the number of jobs also needs to increase. And the fact that productivity is rising means that the same goods can be produced each year with less labor than the year before. So to keep the number of jobs constant, the amount of stuff produced needs to increase. The second parameter in Okun's law (b in Equation 8) is equal to the sum of average productivity growth and average laborforce growth.

In the US, population and probably productivity are growing more slowly today than in the decades after World War II. So if we were to estimate the equation using only data from the 1950s and 1960s, we would get a higher value for the second parameter – maybe 3 rather than 2. If we estimated it using only recent data, we would get a lower value. So the relationship is not constant, but it changes only gradually over time.

Business Cycles

The figure below shows some of the links between macroeconomic aggregates that are most important for business cycles in a closed economy.



I is investment.
Y is output (usually measured by GDP).
U is unemployment.
A + in the line between two aggregates means there is a positive relationship between them, that is, a rise in the first will cause a rise in the second, and a fall in the first will cause a fall in the second.
A - in the line between two aggregates means there is a negative relationship, that is, a rise in the first will cause a fall in the second, and a fall in the first will cause a rise in the second.

Much of macroeconomics consists of establishing cause-and-effect relationships between various economic aggregates. It can be helpful to think about these relationships using a flowchart.

When we think there is some regular pattern linking two observable quantities, we say there is a *functional relationship* between them. Much of macroeconomics consists of describing and explaining the most important functional relationships between *economic aggregates* like output, inflation, and unemployment. It's important to distinguish statements about functional relationships among variables, from statements about the variables themselves. When we say that one variable tends to rise when another falls, that does not mean that either one actually is rising or falling. Rather, it describes a pattern we can observe over an extended period in which the variables sometimes rise and sometimes fall.

With most functional relationships, we have an idea about which variable is cause and which is effect. In the case of output and unemployment, we think that unemployed people get jobs because more stuff is being purchased and produced. The *direction of causality* is from output to unemployment. It can be helpful to present these causal links in a flowchart, so we can see at a glance how a change in one aggregate affects others, both directly or indirectly. You often find flowchart diagrams similar to the one here in the documentation of the macroeconomic models used by professional forecasters in government and business.

In this class, we will look at a number of flowcharts presenting the causal links between various macroeconomic aggregates. In these flowcharts, the *targets* of macroeconomic policy are in bold to help focus attention on the outcomes that normally guide policy decisions. where an increase in one aggregate causes an increase in another, there is a small plus sign (+) in the line. Where an increase in one variable causes a decrease in another, there is a small minus sign (-) next to the line. This kind of *negative relationship* between two variables also means that a decrease in the first one, causes an increase in the second.

Many causal relationships have names. Here we see that the link from investment to output is called the *multiplier*, the link from output to investment is called the **accelerator**, and the link from output to unemployment is called *Okun's law*.[8] The *Phillips curve* is often used to refer to the link from output to inflation, which may take place via unemployment and wages, or by some other channel. Some of these links, like Okun's law, can be *quantified* – that is, we can make a definite prediction, based on statistical evidence, for how much one variable will change in response to a given change in the other. For other links, we have an idea of the direction of the resulting change but we can't put an exact number on it.

Accelerator. The link from output growth to private investment.

[8] Strictly speaking, the accelerator describes a link from the growth of output to investment, but we can ignore that detail here.

When a change in one variable produces changes in other variables that induce further change in the first variable, that is called a feedback loop. Feedback loops may be negative or positive.

When we see a loop on a flowchart, that means that a change in one variable will affect other variables in a way that results in a further change in the first variable. This is called a *feedback loop*. Positive feedbacks are cases when an increase in a variable leads, via other variables, to a further increase; negative feedbacks are cases where an initial increase leads, via other variables, to a decrease back toward the original variable. Another way of describing this is that when

there is a positive feedback, a change in the variable is *amplified*, while when there is a negative feedback, a change in the variable is *dampened*.

A fundamental challenge in thinking about the economy is that, in reality, everything is connected to everything else. but to be able to tell a coherent story or build a usable model, we need to focus on a few relationships and ignore the others. And this task is made harder by the fact that relative strength of the different relationships varies depending on the country and the historical period, and on the length of time we are interested in. So it is never a question of finding the "right" model, but only the best one for a particular purpose.

"Economics is a science of thinking in terms of models joined to the art of choosing models which are relevant to the contemporary world. ... Good economists are scarce because the gift for using 'vigilant observation' to choose good models ... appears to be very rare. "

- John Maynard Keynes

Business cycles are alternating periods in which output, inflation, and employment rise and fall together.

Business cycles are more or less regular changes that we can observe cross a number of macroeconomic aggregates. A business cycle **expansionary** generally involves rising output, falling unemployment, and high and/or rising inflation. A business cycle **recession** involves falling (or more slowly growing) output, rising unemployment, and lower and/or falling inflation or even deflation. Business cycles are generally understood to involve changes in desired spending by households and businesses. They are not the result of changes in the productive capacity of the economy, but rather, of changes – for whatever reason – in people's willingness to spend money.

Expansionary. A period of rising output.

Recession. A period in which economic activity is declining. It is widely believed that a recession is defined as two consecutive quarters of declines in real GDP. But in fact, in the US recession dates are determined by the National Bureau of Economic Research based on a variety of economic indicators.

Many other factors affect output growth, unemployment and inflation. Not all of the changes in these aggregates are linked to each other. In the case of unemployment, cyclical or demand-deficiency unemployment is linked to output but structural and frictional unemployment are not.

The goal of macroeconomic policy is generally understood to be the smoothing out of business-cycle fluctuations. The idea is to adjust aggregate spending such that output is at potential, unemployment is at the *full employment* level, and inflation is at the **price stability** level. One important question in macroeconomics is whether these three goals are always compatible.

Price stability. The level of inflation desired by policymakers. Historically, price stability meant literally stable prices, that is, zero inflation on average; but since the 1990s, it has generally been understood as a low but positive level of inflation – 2% in the US and most other rich countries.

Investment affects output via the multiplier.

When businesses decide to expand, they must purchase new capital equipment, buildings and other structures, and software. These purchases are sales for other businesses. As a result, the businesses producing the capital equipment will earn more (increasing capital income), must hire new workers (increasing labor income), and purchase inputs from still other businesses, creating sales for them in turn. The labor and capital income are spent on consumption goods, causing those businesses to earn more, hire more workers, and increase purchases from still other businesses. As this process works its way through the economy, each dollar of new investment spending may eventually result in several dollars of additional final goods purchases. The ratio between the initial increase in investment spending and the eventual increase in GDP is known as the *multiplier*. The multiplier will be larger when more goods are produced domestically and when people consume a large fraction of additional income, rather than saving it. It will be smaller when imports and/or savings are high.

The multiplier applies to any **autonomous** increase in spending, but for purposes of thinking about business cycles, investment is usually most important.

Autonomous. Describes a change in spending that is independent of current income.

Investment is influenced by the growth rate of output (via the accelerator), by the profit rate, and by the availability of credit.

In general, a business will expand when its existing capacity cannot produce as much as it could potentially sell; when its business is profitable; and when it can finance the expansion with its own *retained earnings* or with borrowed funds. Unless all three of these conditions are met, a business is unlikely to undertake new investment spending.[9] For example, a restaurant is unlikely to expand into a new space unless it is regularly filling all the tables in its existing space, *and* its current operations are making money, *and* it can get the funds needed for expansion on reasonable terms.

These influences are represented by the three lines leading to investment ("I") on the flowchart. The link from output ("Y") to investment is called the *accelerator*. Strictly speaking, this is not a link from the level of output, but from the *change* in output. When output is rising rapidly, more businesses are likely to find their existing capacity is insufficient to produce as much as they can sell; expanding will require investment spending. When output is rising more slowly or falling, more businesses are likely to find that they have more capacity than they need, and have no need to invest. Higher profits also make investment more likely, for two reasons. First, when profits are

[9] Of course many other factors - like the need to compete with rivals, the desire to enter new markets, or the obsolescence of existing assets, also influence investment. But these are more microeconomic factors, having to do with the specific characteristics of specific business or industry. For macroeconomic purposes the three listed here are most important.

high in general, businesses will see good opportunities to expand, introduce new products, or enter new markets. Second, profits are an important source of finance for investment. A business that can may for expansion out of its own retained earnings does not need to seek a loan. This link is shown by the line from profits to investment.

The third factor influencing investment is the available of credit. Since most businesses cannot finance all their desired investment projects out of their own retained earnings, they have to borrow money from banks or by issuing bonds. When credit is abundant and cheap, more investment will be carried out than when it is scarce and expensive. The availability of credit includes many factors in addition to the interest rate – the terms on which loans will be made, the collateral and other conditions demanded, and whether are banks are willing to lend at all. But to keep things simple, we often focus on the interest rate alone. In this story, scarce credit means high interest rates, and abundant credit means low interest rates. This is represented as the link from real interest rates to investment. Since high interest rates discourage investment, there is a minus sign in this link. Note that it is *real* interest rates that matter for business. The borrower doesn't care about how many dollars they will have to pay back, but how much that will be relative to their own sales or earnings. If high inflation means that future dollars are worse less than today's, paying back the loan will be easier, so it will be more tempting to carry out investment funded with debt.

Output and employment are linked via Okun's law.

Okun's law says that when output grows rapidly, unemployment will fall, and when output grows more slowly or falls, unemployment will rise. The exact relationship varies between countries, but within countries it seems to be quite stable over time. If we write the change in unemployment as ΔU and the real (inflation-adjusted) growth rate of output as g, then for the US Okun's law is:

$$\Delta U = -0.5(g - 2.5)$$

In other words, the change in unemployment is equal to negative 0.5 times the percentage growth rate minus 2.5. So it takes around 2.5 points of real GDP growth to hold unemployment constant. For example, if GDP grew by 4.5% in one year, we would expect the change in the unemployment rate to be $-0.5(4.5 - 2.5) = -0.5 * 2 = -1$ – we would expect unemployment to fall by one point. On the other hand, if real GDP were to fall by one point, we would expect the change in unemployment to be $-0.5(-1 - 2.5) = -0.5 * -3.5 = 1.75$ – we would expect the unemployment rate to increase by 1.75

point.

Note that this equation doesn't say what the change in the unemployment rate (ΔU) or the growth rate (g) actually are. Rather, it describes a *function* linking the two. It says that *if* growth is high, unemployment is probably falling; and *if* growth is low or negative, unemployment is probably rising. So if you have an idea about what will happen to one of the variables, you can make a good guess about what must happen to the other. The numbers that appear in a function like this are called its *parameters*. In the case of Okun's law, while the basic form of the law is the same across countries, the parameters vary. In general, the first parameter (-0.5) depends on labor market institutions, while the second depends on the growth rates of the laborforce and labor productivity. In the US, the laborforce grows by roughly 1 percent a year while labor productivity grows at around 1.5 percent, so it takes a total of 2.5 percent additional production just to hold the unemployment rate constant.

Output and inflation are linked via the Phillips curve.

The *Phillips curve* usually refers to the link from output to inflation, which may take place via but it is also often used to mean the link from unemployment to inflation. When the Phillips curve is drawn on a graph, it is drawn with the inflation rate (or price level) on the vertical axis, and either output (GDP) or unemployment on the horizontal axis. If output is on the horizontal axis, the curve slopes upward, to show that inflation tends to rise with output; if unemployment is on the horizontal axis, the curve slopes downward, to show that inflation tends to fall when unemployment rises.[10] A steep Phillips curve means that inflation will change a lot in response to a small change in output or unemployment; a flat or shallow Phillips curve means that inflation changes on a little in response to output or unemployment.

Wage curve. A relationship between unemployment wages: When unemployment is low, wages rise more quickly; when unemployment is high, wages rise more slowly or fall.

Unlike Okun's law, the Phillips curve there does not seem to have stable parameters. How much additional inflation you get for a one point acceleration in GDP growth, or a one point fall in unemployment, depends on the current values of the variables, as well as on the country and historical period we are looking at, and how long a time period we are interested in. Many economists believe that the curve is steeper over long periods. That is, one year of high GDP growth may not raise inflation very much at all, but if the high growth rate is sustained year after year, eventually inflation will rise. Some economists believe that the Phillips curve is vertical in the very long run – that is, there is only one level of unemployment that is consistent with constant inflation. This unique unemployment rate is

called the *natural rate of unemployment* or **Non-accelerating inflation rate of unemployment (NAIRU)**. In the 1990s, economists and policymakers put a lot of energy into trying to determine the NAIRU so that central banks could try to hold unemployment at that level. But the concept has become less popular since then – most economists no longer believe that there is one unique level of unemployment at which inflation is stable. But the more general idea of the Phillips curve – that lower unemployment tends to lead to higher inflation, and high unemployment to low inflation or deflation – continues to be widely accepted.

[10] When the curve was first drawn by New Zealand economist A. W. Phillips in the 1950s, he was looking only at the link from unemployment to nominal wage growth. Today, the link from unemployment to wages is sometimes called the **wage curve**.

The multiplier-accelerator cycle is a positive feedback loop between investment and output.

The link from output to investment is called the *accelerator*. Strictly speaking, the accelerator relationship says that the level of investment tends to be determined by the increase in output, as opposed to the level – businesses are likely to expand only when demand for their products increases beyond the level they can meet with their existing capacity. As you can see, there is a positive feedback loop linking investment to output via the multiplier, and output to investment via the accelerator. In this story, a rise in investment spending by businesses increases incomes for their workers, and for other businesses they purchase inputs from. This leads to higher spending, and businesses increase investment to meet the new demand. Higher investment increases incomes in turn. This process continues until something interrupts the increase in investment – perhaps **supply constraints**, perhaps a fall in profitability or a shift toward more pessimistic **expectations** about future demand or profits. But when investment falls, the multiplier means that total spending falls more, leaving other businesses with excess capacity and causing investment to fall still further. This process continues until either new investment drops to zero and cannot fall any further (as happened in the 1930s) or until something intervenes to boost demand – perhaps a shift toward more *expansionary* policy by the central bank. Then investment and output begin to rise again.

This loop was first described by the British economist Roy Harrod in the 1940s, and was the most important theory of business cycles in the 1950s and 1960s. Harrod pointed out that there will be some combination of investment output growth that can remain constant, but if the economy moves away from that stable point, the multiplier-accelerator feedback loop tends to carry it even further away. Harrod described this problem of instability as the *knife edge*. While the multiplier-accelerator feedback loops is not as central in macroeco-

Non-accelerating inflation rate of unemployment (NAIRU). The rate of unemployment at which inflation neither rises nor falls. Sometimes referred to as the "natural unemployment rate." Whether there is a unique NAIRU is debated by macroeconomists; many believe that stable inflation is possible with many different rates of unemployment.

Supply constraints. Limits on the productive capacity of the economy due to natural resources, capital, technology, worker skills, or other "real" factors.

nomic thought, as it once was, it can still be a useful way of thinking about why modern economies tend to go through recurring booms and busts rather than growing steadily.

The Goodwin cycle is a negative feedback loop involving investment, employment and profits.

An important factor affecting investment, in addition to credit conditions and demand, is profitability. This is the focus of a story about business cycles and macroeconomic instability often – but not only – proposed by Marxist economists. In this story, we are interested in the feedback loop from investment, to output, to unemployment, to the distribution between wages and profits, and back to investment. The idea is that a rise in investment leads to higher investment, which in turn brings down unemployment and, by improving workers' bargaining position, raises wages. Then what? It could be that the change in nominal wages is fully passed on to higher prices, leaving real wages and the wage share unchanged. (This is what happens in the standard textbook model.) But even if faster wage growth is associated with higher inflation, it's unlikely that *all* wage increases are passed on to higher prices. It seems likely that when low unemployment leads to bigger wage gains, at least some of the higher wages come at the expense of capital owners – in other words, that they increase the share of the total product going to workers and reduce the share going to capital-owners. A lower share of output going to owners normally means a lower profit rate. And if profits fall enough, that will discourage further investment. A fall in investment, in turn, will bring output back down and unemployment back up. Here, we are looking at a negative feedback loop. But that does not guarantee that the system will reach a stable **equilibrium**; instead, it may show repeated cycles.

This type of feedback loop is called a **Goodwin cycle**, after Richard Goodwin. Goodwin cycles are primarily discussed by Marxist economists, since they focus more on the conflict between workers and owners than most other economists do. But there is good reason to think that these type of cycles play an important role in real economies. It is a well-established statistical fact in the US and most other advanced economies that the profit share rises early in expansions but falls in the year or two before the recession begins. And changes in investment spending normally follow changes in profitability.

The logic of a Goodwin cycle is shown in the diagram. Let's start with an economy at the height of boom, represented in the diagram by a point like *a*. During the boom, high investment leads to high output and low unemployment. Low unemployment causes wages

Expectations. People's beliefs about the future, as reflected in current prices and behavior.

Equilibrium. A situation where, given the actions of everyone else, no one wants to change their own actions. Or, a situation that does not have any tendency to change on its own – that will persist until disturbed from outside.

Goodwin cycle. One of several possible sources of instability in a capitalist economy. In a Goodwin cycle, high growth leads to lower unemployment, which causes wages to rise at the expense of profits, which reduces investment and brings growth back down.

to rise – shown as a move to the right in the diagram. But as wages rise, profits decline, and falling profits eventually cause investment to fall, moving the economy downward in the diagram to a point like *b*. This is the start of a recession. The decline in investment has now led unemployment to rise, weakening workers bargaining position. During the recession, both wages and investment decline, bringing the economy to point *c*. The decline in wages is eventually sufficient to restore profitability, and investment begins to rise again, even while unemployment remains high. Eventually, unemployment falls enough that wages can again begin to rise – this is point *d*. Finally, during the expansion, both investment and wages are rising, until the economy reaches the peak of the cycle at point *a* once again.

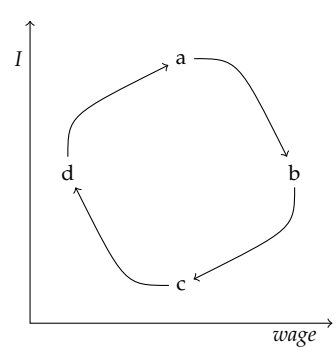
The Wicksell cycle, or "cumulative process," is a positive feedback loop involving investment, inflation and real interest rates.

Along with profits and demand, business investment may also be influenced by the interest rate. For businesses, what matters is the *real* interest rate, since they are not concerned with the absolute number of dollars they sell but with the the repayment relative to their expected earnings. This creates another potential source of instability, first identified by the Swedish economist Knut Wicksell in the late 19th century. The problem Wicksell saw is that because inflation leads to lower real rates, it may feed on itself rather than dying out.

The key point for Wicksell is that the nominal interest rate is set by the banking system. While the banks may have good reasons for setting a particular rate, there is no reason to think that the rate that maximizing their profits will be the rte consistent with price stability for the economy as whole. Suppose, for instance, that banks decide to reduce the interest rates they charge (presumably in the hopes of attracting more borrowers.) The result will be more borrowing, and more investment. The increased investment will increase total spending in the economy via the multiplier, and the increased spending will sooner or later lead to higher inflation via the Phillips curve. As inflation rises, real rates will fall further, encouraging even more borrowing and investment, leading to more spending and still higher investment, leading to higher inflation and still lower real rates. This positive feedback loop could continue indefinitely.

In Wicksell's view, the solution to this problem was to have a **central bank** that would adjust interest rates up or down to guarantee price stability. When inflation rises, the central bank should raise nominal rates by even more, so that the real rate goes up, not down.

Wicksell was mostly worried about runaway inflation, but the same positive feedback loop can also operate in reverse: A rise in



Goodwin cycle. High investment leads to low unemployment and rising wages, but rising wages eventually reduce profits so much that investment declines.

interest rates leads to a fall in investment, which leads to a fall in output and employment, which leads to lower inflation or deflation, which leads to a further rise in the real interest rate, causing a further fall in investment, and so on. It is widely believed that this negative version of Wicksell's "cumulative process" played an important role in the Great Depression of the 1930s. Between 1929 and 1933, prices in the US fell by about 7 percent per year. (Price declines were similar in most European countries.) This *deflation* meant that even very low nominal interest rates implied quite high real interest rates. These high real rates made it difficult for households and businesses to service existing loans, and made new borrowing prohibitively expensive. The result was declining spending, declining incomes, and a wave of bankruptcies and bank failures. According to Irving Fisher – perhaps the leading American economist in the early 20th century – this **debt deflation** process explained the economic collapse of the 1930s.

Central bank. The bank for other banks, which is responsible for stabilizing the financial system. Almost all modern economies have a public central bank, which is also responsible for managing the level of activity in the economy as a whole.

Money and Finance

The monetary side of an economy can be thought of in terms of an interlocking set of balance sheets – records of all payments that each economic unit must make and expects to receive.

Every economic unit has a **balance sheet**. Governments, businesses, households, all have balance sheets. You have a balance sheet, even if you never write it down. A balance sheet is simply a list of the unit's **assets** and **liabilities**.[11] We write balance sheets with assets on the left and liabilities on the right. An asset is anything you own that has a market price, and that you expect to receive income or other benefits from in the future. A liability is an obligation to make a payment at some time in the future. For our purposes, we can think of a liability as being the same as a debt. For example, a typical household might have a balance sheet like this:

Balance sheet. A record of all the assets and liabilities of an economic unit.

Asset. Anything that is owned, has a market value, and will provide some monetary or money-like benefit to the owner in the future. Assets are divided into real assets, like land, buildings, and intangible property like patents and copyrights; and financial assets, which are payments commitments by some other unit.

Liability. A binding commitment to make some payment in the future. Every liability is a financial asset for some other unit. The most common form of liability is a debt, which is an asset for the lender.

[11] For businesses, balance sheets are one of three basic forms of financial record keeping. The other two are the income statement and the cashflow statement.

Assets	Liabilities
Checking account $2,000	Credit card balance $1,000
Retirement savings $30,000	Student debt $15,000
Automobile $8,000	Auto loan $4,000
Home $120,000	Mortgage $70,000
Total assets $160,000	Total liabilities $90,000
	Net worth $70,000

Net worth is defined as total assets minus total financial liabilities, and is recorded as a liability. It may sound strange that net worth is a liability, but that is how it is treated in accounting. With net worth included, total assets are always equal to total liabilities.

Net worth. A unit's total assets minus its financial liabilities. Net worth is treated as a liability on balance sheets, ensuring that total assets and total liabilities are always equal.

Assets can be divided into real and financial assets. A financial asset is any kind of claim that requires someone else to make payments to you in the future. For example, if you borrow $20 from me and give me an IOU, that IOU represents your obligation to pay me $20 sometime in the future. So it is a liability for you, and an asset for me.Various kinds of debt and equity (stocks) are the most familiar financial assets, but there are many others. A *real asset* is something you expect to receive income or services from not because anyone is required to pay you, but because you can use it in some way that will benefit you. The machines and buildings owned by a business are examples of real assets. In the balance sheet above, the automobile and the home are real assets, and the checking account and retirement savings are financial assets.

Every liability is an asset for someone else. (If you owe money, there must be someone you owe it to.) Every financial assets is a

liability for someone else. (Someone must pay you the money.) Real assets are not liabilities for anyone. Everything on a balance sheet is a **stock**, as opposed to a flow.

Stock. In economics, a stock is anything that can be measured as a quantity at a moment in time. Wealth, population, and total employment are examples of stocks. Flows are things that can be measured only over a period of time, such as income or GDP.

Financial transactions involve changes in balance sheets. Every real transaction has a corresponding financial transactions. But many financial transactions have no connection with real transactions.

Financial transactions cannot change the net worth of any of the units involved. This means that every financial transaction must include at least four balance sheet entries – two (or more) for each of the two (or more) units involved.

For example, suppose you buy a car for $8,000, paying $2,000 from your checking account and getting a loan for the remaining $6,000. Then the changes in your balance sheet look like this:

Assets	Liabilities
-$2,000 checking account	
+$8,000 automobile	+$6,000 auto loan

Your assets have increased by $6,000 – you've added the $8,000 car as an asset, but reduced your checking account asset by the $2,000 down payment. Meanwhile, your liabilities have increased by the $6,000 of the loan you took out. So your net worth is unchanged.

How can we be sure that the car is worth the same as the $8,000 that was paid for it? The reason is that the value of an asset is simply its market price – the only way of knowing what something is worth, is what was paid for it. This is the result of the same asset being used as **medium of exchange** and **unit of account**. Goods and services are exchange only for money, and their value is measured by the amount of money they exchange for.

Medium of exchange. A good or asset that is used in purchases of goods and services – instead of being traded directly for each other, goods and services are traded only for the medium of exchange. One of the functions of money.

Unit of account. The good or asset that is the standard by which the value of other goods and assets is measured. One of the functions of money.

The term "money" refers both to a type of asset and a unit of measurement.

Defined in terms of its functions, money is the asset in which payments are made, income is saved, values are measured, and debts must be paid.

There are four functions of money:

1. *Medium of exchange*. Money is something you can give in return for a good or service.
2. *Unit of account*. We measure the value of things in terms of money.
3. *Store of value*. If you have more income right now than you want to spend, you can hold money to give you the option of spending later.

4. *Means of payment*. If you have a debt to someone, they will accept money to settle it. This is the function of money that is connected to an asset's status as **legal tender**. By law, anything designated as legal tender (dollar currency in the US) must be accepted in payment as a debt. Legal tender does *not* have to be accepted as payment for goods or services.

Legal tender. A form of money that by law must be accepted in payment of a debt. Legal tender does not have to be accepted as payment for goods and services.

The first three of these functions are discussed more than the fourth. But historically, the use of money to settle debts seems to be older than the use of money to exchange goods and services.The argument that money originates with debt-settlement, and not with the exchange of goods, is made persuasively in the book *Debt*, by David Graeber.

Many assets perform some of the functions of money, but not all of them. Different assets perform the functions of money at different times. For example, in a country with high inflation, people might use the local currency to buy goods and services and settle debts, but not as a store of value. Even in the US, it is not always clear which asset exactly is "money". For example, the price of a good may vary depending on whether payment is made in cash, by check, by debit card or direct debit from a bank account. So there is no hard and fast rule about whether a given asset is or is not money. In some times and places, no single asset performs all of these functions, any it does not make sense to say that anything is "money".

Economists used to talk about the "quantity of money," and describe central banks as "printing money" or "controlling the money supply," but that kind of language is no longer used very much. There is too much fuzziness about what does and does not function as money, for there to be a meaningful quantity of it.

Assets can be more or less liquid – that is, it can be easier or harder to use them to make unexpected payments. Liquidity is the "moneyness" of an asset.

The **liquidity** of an asset refers to how quickly and reliably it can be used to make unanticipated payments. Another way of thinking of liquidity is how easy it is to trade it or convert an asset into money, by selling it, borrowing against it, etc. Liquid assets should have low **transaction costs** a predictable value, thick markets, and come in standard form. And it should be easy to buy or sell them without having to change whatever else you are doing. A financial asset like a bond is very liquid – there are well-organized markets with lots of buyers and sellers, one bond is just like another, and owning a bond and collecting interest from it doesn't require you to do anything else. A stock is less liquid because its value is unpredictable – you

Liquidity. The degree to which an asset can be used to make payments reliably and at short notice.

Transaction costs. Costs of carrying out a sale or exchange. They include any fees, taxes or payments to third parties associated with the sale; the costs of finding the two parties, bringing them together, and transferring the good between them; and any adverse price changes that result from a purchase or sale.

might not be able to convert into as much cash as you were expecting, when you need it. A piece of machinery, a building, a house are much less liquid – each one is a little different, so it's harder to find buyers, and there are often significant costs involved in transferring ownership. And to get an income from it, you have to actively do something with it.

Most of the functions of money can only be performed by an asset that is very liquid. Another way of thinking of liquidity is the "moneyness" of an asset. Something you can easily convert into money when you need it, is almost as good as holding money itself.

As applied to a balance sheet of a business, household, or other economic unit, liquidity refers to how easily they can acquire money when they need it. Holding money obviously makes your balance sheet liquid, as does holding liquid assets. But an economic unit may also be liquid if it is easy for them to borrow money.

Many different assets can function as money. They can be divided into commodity money, fiat money and credit money.

One way of classifying the various assets that can function as money is into **commodity money**, **fiat money** and **credit money**. Commodity money is something that has intrinsic value – that is desired for its own sake – as well as functioning as money. The classic example of commodity money is gold, silver and other precious metals, but many other goods have functioned as money. If the asset is to be used in routine transactions as medium of exchange and means of payment, it must have certain properties: It must be easy to transport, it must be reliably measured and distinguished from non-money assets to avoid fraud, it must be easily subdivided to allow for transactions of different amounts. If the asset is only functioning as a unit of account, a broader range of assets may be used. Commodity money is not used in modern economies.

Fiat money is a token with no intrinsic value that is used as money because it has been designated as money by a government. Coins and bills are examples of fiat money. Governments normally designate a given asset as *legal tender* within their territory, meaning it must be accepted as settlement of debts and other obligations. In particular, governments decide what form they will accept tax payments in. In some cases, assets come to be used as medium of exchange or in payment for private debts, solely because they are needed to make tax payments. Even someone who doesn't owe taxes to the government, may accept the asset if they know others who do owe taxes will value it for that purpose. The idea that taxes are the original reason why money is used, and the fundamental source of its value, is

Commodity money. An asset that serves as money based on its intrinsic value. Commodity money consists of some physical asset (often precious metals) that would be desired for its own sake even if it did not function as money.

Fiat money. An asset that serves as money only because a government has declared it legal tender. Fiat money consists of tokens (paper bills, coins, etc.) that have no intrinsic value, and that can be legally created only by the government.

Credit money. The liability of a bank or similar financial institution that functions as money. Credit money includes checking accounts and other deposits that can be used to make payments. Credit money is the main form of money in modern capitalist economies; it is created when banks make loans.

known as **chartalism**. There are some clear historical cases where a new currency is introduced through the a tax and then becomes more widely used. For example, European countries imposed taxes in their own currencies on their colonies in Africa and elsewhere as a way to compel people to work in European plantations or other businesses, which were the only local source of the currency. It's not clear how much this story applies to currencies more generally.

Chartalism. The idea that the ultimate source of money's value is that a government accepts it payment of taxes.

Finally, credit money is an account at a bank or similar financial institution that can be used to make payments. This is by far the most widely-used form of money in most modern economies. A bank account is a *liability*, or debt, of the bank to the account-holder. It functions as money because it can be transferred to someone else as payment for a good or service or to settle or debt. Imagine a friend owes you $1,000, while you owe $1,000 to your landlord for rent. You cannot normally go to your landlord and tell them that while you can't give them cash, they can have your friend's IOU, and collect the money from them. But if your "friend" is a bank, then you can. In fact, this is the form most of our large transactions take. As with other forms of money, bank deposits function as money simply because they are widely regarded as money – people accept them as payment because they know that others will accept them as payment in turn. So anyone whose IOUs are widely accepted as payment, can create money. Historically, many businesses have created money in this way. For example, mining companies and others with "company towns" have sometimes paid workers in tokens or scrip that can be used to make payments in their own stores. But in most modern economies, only banks and similar financial institutions normally create money in this way.

One important exception to the rule that only banks create credit money is corporate shares, which function as money in certain particular cases. When corporations purchase other corporations in mergers or acquisitions, some or all of the purchase price is often paid in shares newly issued by the acquiring company. Corporations may also pay some of employees' salaries in the form of shares in the corporation, usually through the use of stock options. Many companies pay top executives this way; some, especially in high-tech sectors, pay many of their employees this way. Finally, some corporate debt contracts allow repayment to be paid partly or entirely in the form of the company's shares, instead of in legal tender. In the first two cases, the shares are functioning as medium of exchange; in the third, as means of payment. Households, governments and small businesses do not normally make payments with corporate shares, but only with bank liabilities.

In modern economies, most money is credit money. It is created by banks in the process of making loans.

Banks create money by lending.

People often think that banks lend out money that others have deposited it with them. We think of banks as vaults where some people first bring in money to leave with the bank for safe-keeping, and which the bank then lends out to someone else. This is not how banks work. No one needs to deposit money with a bank in order for it to make loan. The money lent by the bank is newly created at the time it makes the loan. This doesn't mean that there is no limit on how much banks can lend – there are limits. But it is definitely wrong to imagine, as many people do, that there is a fixed pool of money or savings for banks to lend out.

We can see how this works by looking at the changes in balance sheets in the process of lending. Let's imagine that someone gets a $250,000 mortgage in order to buy a house. To make it more concrete, we will call the house buyer Bert and the house seller Sara.

The first step is for the borrower and bank to agree on a loan. When the loan contract is signed, the bank makes two entries in its accounts. It records the $250,000 loan to Bert – this is an asset for the bank, and a liability for Bert. And it records a $250,000 deposit – this is an asset for Bert and a liability for the bank. The deposit created by a loan is the money that the borrower can now spend. It is a liability of the bank, because the bank is now legally obliged to pay out the $250,000 to the borrower, or to someone else designated by the borrower. Here is how the two balance sheets change in this first step:

Bert		Bank A	
Assets	Liabilities	Assets	Liabilities
+$250,000 deposit	+ $250,000 mortgage	+$250,000 mortgage	+ $250,000 deposit

Notice that both the loan and deposit appear twice. Both are financial assets, so they must be liabilities for another unit. In this case, the deposit is the asset for Bert – it is the money he has borrowed – and it is a liability for the bank. The mortgage is a liability for Bert, and an asset for the bank.

Any loan transaction creates two offsetting debts, or IOUs – one from the borrower to the bank, and one from the bank to the bor-

rower. The borrower's debt to the bank has to be paid off on a fixed schedule – over 30 years, in the case of a typical mortgage. And it has to be paid back with interest, on top of the original amount. (The liability recorded on the balance sheet normally is only the **principal** or face value of the loan, and not the interest payments, but this depends on the type of loan.) The bank's debt to the borrower does not carry any interest, but the borrower can claim it at any time. (In other words, the borrower gets liquidity, while the bank gets income.) We don't normally think of bank deposits as debts. But they are, in the precise sense that they are a legal obligation by the bank to make a payment to the deposit-holder or to someone else that the holder designates.

Principal. The principal of a loan is the amount the borrower actually receives when the loan is incurred.

In the case of a mortgage, the borrower does not keep the deposit, but immediately uses it to make a payment to the seller of the house. This means that the bank's debt to the house-buyer becomes a debt to the house-seller instead.

In other words, the loan creates both an IOU from the borrower to the bank, and from the bank to the borrower. The borrower then signs over the IOU from the bank to the seller of the house – Sara, in our example. In effect, Bert tells the bank, "I'm transferring your debt to me over to this other person. Now you owe the money to them, not to me." In return Sara transfers title of the house to Bert. Here is what happens to the balance sheets in this step.

Sara		Bert		Bank A	
Assets	Liabilities	Assets	Liabilities	Assets	Liabilities
		+$250,000 deposit	+ $250,000 mortgage	+$250,000 mortgage	+ $250,000 deposit
+ $250,000 deposit - $250,000 house		- $250,000 deposit + $250,000 house			- $250,000 deposit (payable to Bert) + $250,000 deposit (payable to Sara)

This is the end of the story for Bert. He now has a new $250,000 house, and a $250,000 mortgage debt to Bank A. He has traded the future flow of housing services from the house for a less liquid overall financial position (since he is now legally obliged to make payments on the mortgage). His net worth has not changed. Presumably he now feels better off, since otherwise he would not have made the transaction. Sara and Bank A must also feel better off: Sara must place a greater value on the liquidity of cash than on the services of the house, while Bank A is happy with interest income it will get

from the mortgage.

In some cases, the seller and buyer will use the same bank. In that case, the process stops here – the payment between them just consists of a change in the ledgers of the bank they both use. In this case, if the seller of the house, Sara, uses the same bank as Bert, then the bank will simply replace an entry marked "liability – $250,000 to Bert," with a new entry marked "liability – $250,000 to Sara."

More often, though, the seller and buyer use different banks. Suppose Sara has a checking account at Bank B and wants to deposit the money from the sale there. Then the deposit must be transferred between the two banks. Instead of Bank A having an entry on its books "liability – $250,000 to Sara," that entry will now appear on the liability of side of the balance sheet of Bank B. We can't stop there, though, because this transaction by itself would reduce the net worth of Bank B by $250,000 and increase the net worth of Bank A by $250,000. There is no reason for Bank B to accept the new liability – in effect, to take over Bank A's debt – unless Bank A gives Bank B something of equal value. This something is a **settlement asset**. In most modern banking systems the settlement asset is **reserves** at the central bank. So in step 3 Bank A transfers both a $250,000 liability and $250,000 of reserves to Bank B.

Settlement asset. Banks periodically must settle any outstanding balances among themselves. Banks that have lost deposits on net must make a payment to banks that have gained deposits. An asset used to make these payments between banks is called a settlement asset.

Reserves. Liabilities of the central bank used by other banks to make payments to each other. Reserves are also used by the central bank to buy assets from private banks. In some banking systems, banks are required by law to hold a certain amount of reserves.

The fact that banks behave this way is one reason why their liabilities are used as money. If it were not quick and easy to transfer deposits between banks, or if one dollar of deposits at one bank did not always equal a dollar of deposits at a different bank, people would be less willing to accept banks' IOUs as payment for goods and services or as settlement for debts.

This next step is shown below:

Sara		Bank A		Bank B	
Assets	Liabilities	Assets	Liabilities	Assets	Liabilities
		+$250,000 mortgage	+ $250,000 deposit		
+ $250,000 deposit - $250,000 house			- $250,000 deposit (payable to Bert) + $250,000 deposit (payable to Sara)		
- $250,000 deposit at Bank A + $250,000 deposit at Bank B		-$250,000 reserves	- $250,000 deposit	+$250,000 reserves	+$250,000 deposit

It may be helpful to think of it this way: After receiving the payment from Bert, Sara could withdraw the $250,000 in cash from Bank A, physically transport it to Bank B, and deposit it there. The transaction would be recorded exactly the same way as the one above, except that the entries marked "reserves" would say "vault cash" instead. In the real world, this is not what happens – in fact a major purpose of banks is to free the payments system from reliance on physical cash. But it can be useful to think of the physical-cash case to understand the logic of why reserves move from the bank where deposits are being withdrawn, to the bank where they are going.

You can see that at each step, each unit's assets and liabilities change by the same amount, so their net worth remains constant.

The process could end here, but it usually doesn't.[12] In normal times, Bank A does not have the $250,000 reserves on hand. Reserves do not pay interest, so there is no reason for banks to keep them around. So the bank borrows the reserves it needs from some other bank, which happens to have extra at the moment. They normally borrow the reserves "overnight" – for one business day – and pay an interest rate on them that is agreed on by the two banks involved. If we call the bank with excess reserves Bank C, then this last part of the process looks like this:

[12] The text from this point to the end of the section describes the US banking system up until 2009. Since the financial crisis of 2008-2009, there have been some changes in the way the payments system operates. In particular, most banks now have substantial excess reserves, and the central bank now pays interest on those reserves.

Bank A		Bank B		Bank C	
Assets	Liabilities	Assets	Liabilities	Assets	Liabilities
+$250,000 mortgage	+ $250,000 deposit				
	- $250,000 deposit (payable to Bert) + $250,000 deposit (payable to Sara)				
+$ 250,000 reserves	+ $250,000 loan from Bank C			-$ 250,000 reserves + $250,000 loan to Bank A	
-$250,000 reserves	- $250,000 deposit	+$250,000 reserves	+$250,000 deposit		

Reserves do not have to move between banks every time a deposit is transferred. Instead, what happens is that every so often – usually at the end of each business day – banks add up all the transfers of deposits between them. Since in general, people are equally likely to be making payments from accounts at one bank to a second bank or

from the second bank to the first, most of these transfers will cancel out. Only the excess, or net, transfers, have to be settled with reserves. Still, at the margin, each additional loan that a bank makes implies an additional dollar of deposits, and most of those deposits will quickly be transferred to other banks. So the cost of borrowed reserves is an important factor for banks in deciding whether to make a loan.

In the US, the market for borrowed reserves is called the federal funds market. The interest rate on reserve loans, like the one Bank C makes to Bank A here, is called the federal funds rate. In other countries there are different names, but the market for overnight reserve loans functions in a similar way. In normal times, at least between 1980 and 2008, the federal funds rate (or equivalent) is the main tool of *monetary policy*. There is a centralized market where banks that need reserves announce how much they need and what interest rate they will pay for them, and banks that have excess reserves announce how much they have to lend and what interest rate they will accept for them. If the interest offered by one bank matches the interest acceptable to a bank with reserves to lend, a loan takes place. If not, banks on one or both sides will have to adjust their offers. Employees of the central bank are monitoring this market, keeping track of the interest rate at which these overnight loans take place. (In the US, this is the job of the New York federal reserve bank.) They have a goal, or target, for this interest rate. If the interest rate starts to rise, that means reserves are getting more expensive, suggesting they are too scarce. So the central bank provides more reserves, normally by buying government bonds.[13] The next set of balance sheets shows what would happen in this case:

[13] In the US, the Fed traditionally makes these trades using Treasury bills, the shortest-maturity government debt.

This is what we would see if Bank A had to pay an interest rate above the Fed's target for reserves in the interbank market, when it transferred the deposit to Bank B. On the other hand, if the overnight interest rate falls below the range targeted by the central bank, they conclude reserves are too plentiful, and remove some by selling government bonds. In this way they keep the actual interest rate close to the level set by policymakers. So there is a relationship between the amount of reserves supplied by the central bank, and the volume of lending; but it is not the direct, automatic link that existed under the fractional reserve system that is still described by more economic textbooks. It's important to understand that central banks do not "set the interest rate" by decree; the central bank cannot force private banks to change the interest rate they charge on loans. Rather, central banks seek to influence market interest rates by changing the supply of reserves.

Debt Dynamics

One challenge in measuring government debt is that some of it is owed to other government bodies, including public pension systems and the central bank.

In most countries, a large fraction of government debt is owed to other parts of the government, including public pension funds and the central bank. For example, as of the end of 2015,the US federal government has a total debt of$18.2 trillion, or just over 100% of GDP. But $2.8 trillion of this is held by the Social Security Administration, $2.2 trillion is held by other government trust funds, and $2.8 trillion is held by the Federal Reserve. Since this kind of debt is owed by one part of government to another, it doesn't involve any obligation for the government as a whole. It is more of a bookkeeping device. For example, in 2015, Social Security taxes exceeded Social Security payments by $26 billion. This is recorded as a $26 billion increase in the **Treasury securities** held by Social Security trust fund. In this sense, Social Security is "lending" money to the rest of the government, and $26 billion was added to the total amount of federal debt. But since Social Security is part for the government, no new debt was actually issued, and there was no additional borrowing by government as a whole.

Treasury securities. The official name for the financial instruments that make up government debt. Treasury bills are the shortest-maturity debt, Treasury bonds are the longest maturity, and Treasury notes are in between.

For this reason, the number most people pay attention to is not total government debt, but *government debt held by the public*. This figure excludes debt held by public pension systems and other trust funds. It is intended to include only debt that represents genuine payment obligations by the government. Somewhat confusingly, however, debt held by the public does include debt held by the central bank. Historically this did not make much difference, since central banks owned only a small fraction of government debt. In recent years, however, central banks have purchased government debt on a much larger scale. In the US, for instance, the Federal Reserve now owns about 20 percent of debt "held by the public." We can debate how government debt held by the central bank should be counted. But at present, it is counted with debt owed to the private sector, rather than debt owed to government trust funds. So as of the end of 2015, federal debt held by the public was just over $13 billion – $10.3 billion held by private businesses, households, and foreign governments, plus the $2.8 billion owned by the Federal Reserve.

For purposes of macroeconomic policy and analysis, the relevant measure of government debt is the debt-GDP ratio – that is, government debt divided by GDP.

Whether we look at the total debt, the debt held by the public, or the debt held by the public excluding the central bank, this is a *nominal* quantity – one measured in units of money. But for most purposes this is not very meaningful – the value of money varies over time, the size of the economy also varies over time, and some countries are much larger and richer than others. So the debt measured in money does not tell us anything about how large it is in any useful sense. What matters is the size of the debt relative to the government budget or to the size of the economy. So in macroeconomic contexts, we focus on the *debt-GDP ratio* rather than the level of the debt itself. This is simply the debt divided by GDP:

$$\textit{debt-GDP ratio} = \textit{debt}/\textit{GDP}$$

Since both debt and GDP are in units of the national currency, the ratio is simply a number, and is often written as a percent. For example, for the US in 2016, the debt-GDP ratio is around 0.75, or 75%. (This is measuring debt in the usual way as debt held by the public, including debt owned by the central bank.)

As Figure 9 shows, the debt-GDP ratio in the US was over 100 percent at the end of World War II, then fell steadily to around 20 percent in the mid-1970s. It then rose in the 1980s, declined in the 1990s, and rose sharply over 2008-2012. Today's debt-GDP ratio is high by historical standards, but well below the levels seen during World War II.

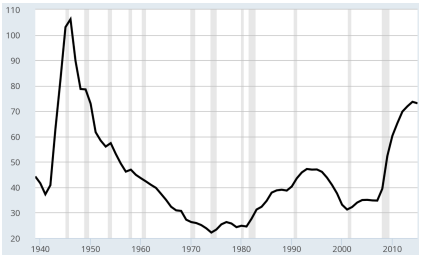


Figure 9: US debt-GDP ratio in percent, 1939-2016. This is based on the standard definition of debt held by the public, which excludes debt owed to Social Security and other trust funds but includes debt owed to the Federal Reserve.

While changes in the debt-GDP ratio are usually discussed in terms of government borrowing, the nominal growth rate of GDP is equally important.

The debt-GDP ratio is a ratio, or fraction. A fraction increases when the numerator rises, or when the denominator falls. Government borrowing is an increase in debt, so it raises the numerator of the fraction. But a decrease in GDP reduces the denominator, so it also increases the fraction. Since debt is measured in dollars, the denominator of the fraction must also be in dollars – that is, it is nominal GDP. So the ratio tends to fall when real GDP growth or inflation is high, and tends to rise when real GDP growth and inflation. If the budget is balance, so the government does not engage in any new borrowing – that is, if taxes are exactly equal to spending – then the debt ratio will fall as long as nominal GDP growth is positive. Even if

the government is running a deficit – spending more than it raises in taxes – the ratio may still fall if nominal GDP growth is high.

For example: Between 1970 and 1980, federal debt held by the public went from $280 billion to $710 billion, an increase of $430 billion. In other words, over the decade of the 1970s, the federal government spent $430 billion more than it collected in taxes. Over the same period, nominal GDP increased from $1.05 trillion to $2.8 trillion. So the debt-GDP ratio was $280/1050 = 0.27$ in 1970 and $710/2800 = 0.25$ in 1980. In other words, despite the fact that there were substantial budget deficits (averaging over 2 percent of GDP) over the whole decade of the 1970s, the debt ratio actually declined, thanks to the steady growth of real GDP and the fairly high inflation of the decade.

On the other hand, Greece is an example of a country that successfully reduced its government debt, but saw the debt ratio rise anyway because of an even larger fall in GDP.

	Debt	GDP	Debt-GDP Ratio
2005	199	195	0.98
2010	330	226	1.46
2015	314	176	1.78

Table 2: Government debt, GDP, and the debt-GDP ratio for Greece, selected years. GDP and debt are in billions of euros.

From 2005 to 2010, there was a large increase in both the absolute amount of debt (131 billion euros) and a large increase in the debt-GDP ratio (48 percentage points). During this period, the government of Greece was running large budget deficit, so the large rise in the debt-GDP ratio was the result of high levels of borrowing by the Greek government. This is the way people usually think about changes in the debt-GDP ratio – that they reflect the borrowing choices of the government. From 2010 to 2015, on the other hand, the size of Greek government debt fell somewhat from 330 billion euros to 314 billion euros. (This was the result both of much smaller budget deficits and some **writeoff** of Greek government debt by creditors.) But the debt-GDP ratio continued rising, by 38 percentage points, almost as much as in the previous period. This was because GDP fell much more than debt did – from 226 billion euros in 2010 to just 176 billion euros in 2015. This over 20 percent fall in Greek GDP involved a deep depression in Greece, with falling incomes and mass unemployment. To the extent that fiscal austerity contributed to this downturn, it may actually have raised the debt ratio rather than reducing it.

Writeoff. A decision by a creditor to cancel or reduce an outstanding debt. Debt may be written off when the debtor defaults on the loan or declares bankruptcy, or (especially in the case of sovereign governments) it may result from a negotiated agreement.

The key point here is that as a fraction, the debt ratio has a denominator as well as numerator. One implication of this is that high

inflation will rapidly reduce the debt ratio. This does not have to be true – in principle, interest payments on the debt might rise along with inflation, and keep the real value of the debt constant. But in practice, interest rates seldom keep up with inflation, and countries that undergo prolonged high inflation almost always end up with very low debt-GDP ratios. Another implication is that changes in a country's *fiscal balance* may have a larger or smaller effect on the debt ratio depending on what effects they have on GDP. If a government borrows to increase spending, the increase in the debt ratio will be less insofar as this *expansionary* policy raises GDP. On the other hand, tax increases cuts in government spending will be less effective in reducing the debt ratio if they cause GDP to decline. In extreme cases, as in Greece today, policies of *austerity* may even increase the debt ratio, if they reduce GDP by more than they reduce government debt. One last implication is that if we want to understand why debt ratios have increased in some countries and some periods, and fallen in other countries and at other times, we have to pay attention to differences in inflation and real GDP growth rates and not just to government spending and taxes.

One percent less growth of nominal GDP always increases the debt ratio by about one percentage point. But the same amount of new borrowing implies a bigger percentage increase in the size of debt when the debt is low, than it does when the debt is already high. (Ten billion dollars in new borrowing means a 50% increase in a debt of $20 billion, but the same $10 billion of new borrowing means only a 5% increase in a debt of $200 billion and 0.5% increase in a debt of $2 trillion.) This means that changes in borrowing are usually the most important factor in changes in the debt ratio when it is small, but inflation and real GDP growth become more important when debt is large.

When discussing changes in the debt-GDP ratio, it is useful to separate out interest payments from other expenditures. The fiscal balance net of interest payments is called the primary balance.

The law of motion of government debt is an accounting identity describing changes in the debt ratio in terms of the primary balance, interest rates, real GDP growth, inflation and a stock-flow adjustment term.

Using the law of motion of government debt, we can decompose historical changes in the debt ratio into the into the contributions of the primary balance, interest rates, real GDP growth and inflation.

Many historical movements in the debt-income ratio, such as the long decline between World War II and the 1970s, and the sharp rise in the 1980s, are mainly explained by changes in interest rates, GDP growth and inflation, rather than by changes in government borrowing.

Using the law of motion of government debt, we can calculate the long-run trajectory of the debt-ratio on given assumptions. A fiscal balance is sometimes described as "sustainable" or "unsustainable" based on the long-run path it implies for the debt ratio.

Changes in household debt-income ratios can be broken down into the contributions of the household primary balance, interest rates, real income growth and inflation.

For household debt, defaults can also play an important part in changes in the debt ratio.

The rise in the household debt-income ratio over the past 30 years is explained mainly by lower inflation and higher interest rates compared with the postwar period, rather than by increased borrowing by households.

One important lesson from the law of motion of debt and from historical debt dynamics is that higher interest rates can lead to higher debt ratios, rather than to lower ratios as is often assumed.

Central Banks and Monetary Policy

Central banks both conduct macroeconomic policy, and serve as bankers to governments and to private banks.

Central banks perform many functions in modern economies. Many of these can be though of as stabilizing the four prices of money.[14]

We don't usually think of money as having a price, let alone several different prices. But money is used in all kinds of transactions, and in each of them, the terms on which money exchanges with

[14] The idea of four prices of money comes from Perry Mehrling.

something else can be thought of as a price of money. The most important four of these prices are:

Par. The price of money in different forms in terms of each other. To say that payment is at par means that a dollar in one form – say, a bill or deposit in one bank – can be exchanged for exactly one dollar in a different form – say, coins or a deposit at a different bank.

Price level. We usually think of changes in the price level (inflation or deflation) as a change in the price of goods and services in terms of money, but we can also think of it as a change in the price of money in terms of goods and services.

Interest. Interest rates can be thought of as the price of current money in terms of future money. This goes for any financial contract that involves a payment today in exchange for repayment later.

Exchange rate. An exchange rate is just the price of one country's money in terms of another country's money.

All of these prices are set in markets; central banks cannot normally set them by law. (Elected governments sometimes do, as when they adopt price controls.) But these prices can be influenced in various ways. Much of what a central bank does can be thought of as managing these four prices.

Maintain price stability. By law or by convention, most central banks' primary macroeconomic responsibility is to keep prices stable. Today, that usually means a low but positive rate of inflation (2% in the US). Historically, many central banks sought to keep inflation at 0 on average, so that there was no long-term trend in the price level. Today, most economists see maintaining price stability as part of the larger task of stabilizing aggregate demand, but some (especially **monetarists**) see price stability as a distinct target. We can think of this function as stabilizing the price of money in terms of goods and services.

Monetarism. A school of economic thought that believes that there is a fixed quantity of money. set by the central bank and that this quantity of money is tightly linked to aggregate demand and inflation. For monetarists, instability in the economy comes mainly from too fast or too slow growth of the supply of money, so the most important goal for macroeconomic policy is ensure steady money growth.

Conduct countercyclical demand management via monetary policy. In capitalist economies, there is no central coordination between desired saving and investment, and total spending will often exceed or fall short of the economy's productive capacity. When people try to buy more than the economy can produce, the result is inflation and perhaps shortages; when people choose to buy less than the economy can produce, the result is unemployment and perhaps deflation. For various reasons, desired spending will periodically

decline; sometimes it may also speed up. These periodic changes in desired spending are what we call *business cycles*.

In recent decades, the central bank is usually seen as the agency responsible for correcting these departures from potential output, encouraging households and businesses to spend more when output falls short of *potential output* and forcing them to spend less when output rises above potential. In order to do this, they may intervene in a variety of different ways in a variety of different financial markets. Today, most central banks conduct **countercyclical** policy by adjusting the level of a very short-term interbank interest rate (the federal funds rate in the United States). But historically, they have used a number of other tools, and focused on changes in the supply of money and/or the volume of lending rather than the level of interest rates. Whichever tool is currently being used, the goal is to find a financial "lever" that can reliably change the overall level of spending in the economy.

Countercyclical. Describes any economic policy or relationship that leads to higher spending when demand is weak and lower spending when demand is strong. Countercyclical policy is a deliberate effort to smooth out business cycles by boosting spending (for instance by lower interest rates or lower taxes) in recessions, and holding spending down in booms.

Control the quantity of money. At one point, it was common to think of central banks as "printing money" or setting the money supply. The idea was that there was a tight link between the amount of **base money** – reserves and currency – created by the central bank and the and the amount of credit money created by banks. So the central bank was supposed to have tight control over the total amount of money in the economy.For this to be true, bank deposits would have to be tightly constrained by reserve requirements, and it would have to be impossible to use financial assets other than bank deposits as money. Both of these conditions were more or less true in the early 19th century when bank money was limited by gold reserves, and again in the mid-20th century, when finance was tightly regulated. But neither has been even close to true for a long time. Today, it does not make sense to think of the central bank as exercising any kind of direct control over the quantity of money in the economy. Money today is *endogenous*: It is private decisions about lending and borrowing that determine the amount of money in circulation, not vice versa. On the other hand, until 2008, it was still true that monetary policy often operated by changing the supply of bank reserves. And "tight money" and "loose money" are still common ways of referring to *contractionary* and *expansionary* monetary policy.

Base money. Money directly created by the government or central bank, as opposed to money created by banks.

Control the level of interest rates. Central banks do not normally tell private banks what interest rates they may or must charge. (At one time there were legal ceilings on the interest rates banks lenders could charge, but they were mostly eliminated in the US and other rich countries in the 1980s.) Instead, central banks try to control

the overall level of interest rates by intervening in financial markets, so as to create incentives for private lenders to rise or lower rates. They may do this by intervening in the markets in which banks lend to each other overnight, or by paying banks interest on reserves, or by lending directly to banks, or by buying or selling government debt or other securities. Whatever specific form intervention takes, the goal is to shift supply and/or demand in financial markets to bring interest rates to a level deemed socially desirable. Normally, central banks focus on short-term interest rates, but in some cases – during wartime, or when engaged in **quantitative easing** – it may attempt to move longer rates as well. Interest rates can be thought of as the price of money today in terms of money tomorrow, and in the U.S., maintaining "moderate long-term interest rates" is part of the legal mandate of the Fed. In practice, though, interest rates themselves are not normally a target of macroeconomic policy. Rather, the central bank seeks to use them to hit some other target.

Quantitative easing. A form of unconventional monetary policy in which a central bank buys large amounts of long-term government debt or other longer maturity assets. The goal is to reduce longer interest rates. It is called "quantitative" because the central bank normally picks a quantity of assets to buy, rather than announcing a target interest rate as in conventional monetary policy.

Control the growth of credit. In order to stabilize demand, the central bank may seek to regulate the growth of credit – the total amount of lending in the economy. Reserve requirements are one way of limiting bank lending. In some cases, like Japan during much of the 20th century, the central bank may directly instruct banks on how much to lend. In other cases, the central bank does not directly set targets for credit growth, but monitors it as one factor in setting monetary policy.

In many ways, controlling interest rates, controlling the money supply, and controlling credit growth are three sides of the same coin. In a credit-money economy, money and loans move together, since they are created by the same transactions; and interest rates are important mainly because of the way they affect choices about lending or borrowing. So to some extent, these are not three different functions of the central bank, but three ways of looking at the same function. A shift toward contractionary monetary policy, say, can be described as raising interest rates, or reducing the money supply, or restricting credit growth. Still, they don't always move together. And even when they do, it makes a difference how the central bank – and banks and the public – thinks of monetary policy as doing.

Manage the exchange rate and hold foreign exchange reserves. In countries that seek to manage their **exchange rate** – either by keeping at a fixed level, or allowing it to float within certain limits – it is normally the central bank that is responsible for carrying out this policy. If the currency is too weak, the central bank seeks to raise

Exchange rate. The value of one currency in terms of some other currency.

its price by buying it in *foreign exchange markets*; if the currency is too strong, the central bank seeks to reduce its price by selling it. Central banks may also hold **foreign exchange reserves**, to use in managing the exchange rate or for some other purpose. Unusually among central banks, the Fed in the U.S. makes no effort to influence the exchange rate and does not hold foreign exchange reserves. It does, however, sometimes intervene in foreign exchange markets to assist the central banks of other countries.

Foreign exchange reserves. Money of a foreign country, held by a central bank in order to manage the exchange rate, to have available in a crisis, or for some other purpose.

Handle payments between banks. One way of thinking of the central bank is that is the bank for banks. When individuals or businesses make payments to each other, they usually do so by changing the entries on the ledger of a bank. Fund's are transferred from one account to another, or in other words, the bank records the payment by debiting the payer's account and crediting the the payee's account. In the same way, when banks make payments to each other, they do so with their accounts at the central bank. (Strictly speaking, only large banks make payments through the central bank; smaller banks make payments through some large bank.) The bank making a payment has its account at the central bank debited by the amount of the payment, and the bank receiving the payment has its account credited. These entries on the ledger of the central bank are called bank *reserves*. At the end of each business day, banks typically settle all outstanding balances between each other by transferring reserves.

The role of the central bank in handling payments between banks is one of its less visible functions, but it is important. By providing a universally accepted *settlement asset*, the central bank ensures that all deposits at different banks exchange at one for one, or at par. The use of reserves for interbank settlement is also what allows the central bank to influence interest rates. Banks that do not have sufficient reserves to make required payments to other banks must borrow them. By increasing or decreasing the supply of reserves, the central bank can make them cheaper or more expensive to borrow, which in turn may affect other interest rates st by the banks. Because reserves are simply an IOU from the central bank, it can never run out of them, and no one else can supply them. The use of central bank reserves for payments between banks also contributes to *financial stability*. If some private asset were used for interbank payments, it is possible the payments system could break down in a financial crisis.

Act as lender of last resort for the banking system. A critical function of central banks is to lend to private banks when they are unable or unwilling to lend to each other. One of the defining features

Lender of last resort. An institution, usually a central bank, that prevents bank runs by lending to banks in a crisis.

of banks is that they borrow short and lend long. Bank loans to small businesses average around seven years; corporate bonds average around ten years; and home mortgages are most often 30 years. A bank's liabilities, on the other hand, are very short-term: deposits may be withdrawn at any time, while other forms of bank borrowing, like commercial paper, are often a few weeks or months at most. Banks make money because the interest they pay on their short-term borrowing is lower than the interest they charge on their long-term lending. But this means they have to be able to keep rolling over their debts; if there is even a short period in which a banks is unable to borrow, it will fail. The problem is worse because depositors and other banks will not want to be the last ones to withdraw their money from a failing back. So if there is any doubt about the **solvency** of a bank, depositors may rush to withdraw their funds and other banks will refuse to roll over their loans. This means that even if a bank is fundamentally sound, fears that it might fail can cause it to actually fail. And since banks depend on other banks for loans, the failure of one bank can cause others to fail in turn. This periodically leads to **bank runs** in which a country's entire financial system can collapse.

One job of the central bank is to prevent runs from getting out of control, by stepping in to lend to banks when they are unable or unwilling to lend to each other. The challenge in doing this is to lend to banks that are solvent, but facing immediate liquidity problems, while allowing insolvent banks to fail. Another challenge is to avoid creating **moral hazard** – if banks know that they can borrow from the central bank in an emergency, they will be tempted to take excessive risks.

Solvency. Having a net worth greater than zero. An economic unit is solvent if, over time, its income will be sufficient to meet its liabilities. When a unit is having trouble making required payments, it's important to know if that is because it is insolvent, or because it is illiquid An illiquid unit has assets worth more than its liabilities, but cannot convert them to means of payment immediately.

Bank run. A financial crisis in which fears that banks may fail leads people to withdraw their money, in order not to be left with worthless deposits in a bankrupt bank. Since banks depend on other banks for loans, and since the failure of one bank can raise doubts about the soundness of others, runs often spread from one bank to others.

Moral hazard. The danger that by protecting people or businesses from bad outcomes, the authorities will encourage carelessness or excessive risk-taking in the future.

Regulate financial institutions. All central banks regulate the financial system to some degree, though the exact responsibilities vary a great deal between different central banks. Central banks cannot perform their other functions without setting rules for banks. Their ability to control interest rates and credit conditions depends on maintaining appropriate restrictions on bank activities. And the lender of last resort function will lead to moral hazard unless it is combined with regulatory limits on risk-taking by banks. And because the central bank is responsible for financial stability, it may need to restrict bank lending and asset positions in order to control speculation in asset markets.

Act as banker to the government. Historically, an important function of the central banks is to be the bank of the government. This means acting as, in effect, the government's investment bank or under-

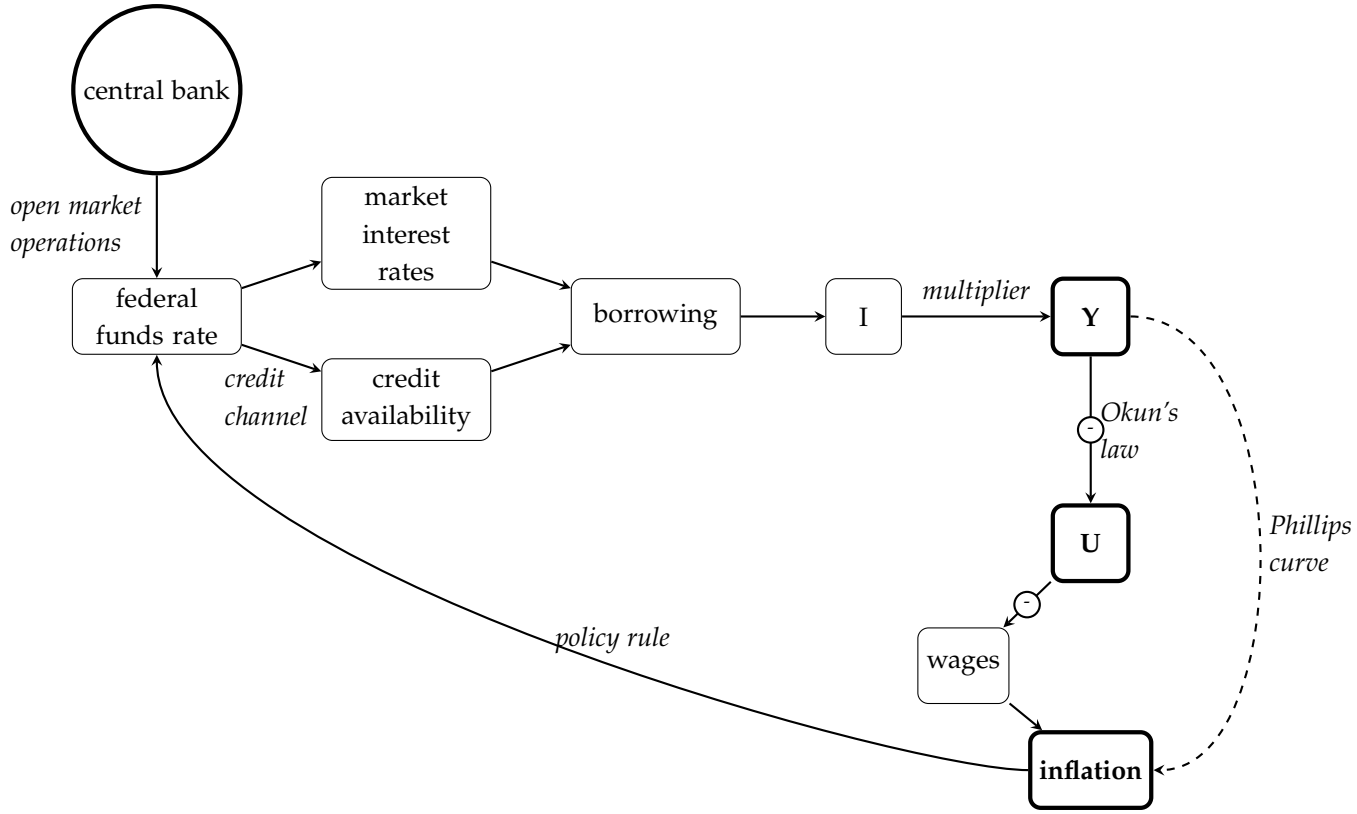
writer: When a government needs to borrow, it can get the funds first from the central bank, with the central bank then taking responsibility for selling the government's bonds to the public. The central bank also normally holds the government's operating balances. These functions are less important in developed countries today. The central bank also guarantees the government's debt, ensuring that there is no possibility of the government defaulting on its debt. A government with its own central bank, which borrows in its own currency, is a risk-free borrower. It normally borrows at the lowest interest rate of any borrower in that currency.

Most central banks are more or less independent from elected government; many began their existence as private banks.

Today, we usually think of the main function of central banks as setting monetary policy, with lender of last resort and, in most countries, managing foreign exchange as important secondary roles. But historically, the banker to the government role came first. The earliest central banks – such as the Swedish Riksbank, founded in 1668, and the bank of England, founded in 1694 – were set up primarily as bankers to their governments, to help them borrow more easily. Central banks gradually came to handle payments between other banks, and the lender of last resort function was formalized in the mid-19th century.[15] Countercyclical monetary policy developed only in the 20th century. And it was not until the 1980s that there was general agreement that the central bank should be the main body responsible for for stabilizing demand.

[15] Walter Bagehot gave one of the first descriptions of the Bank of England's role as lender of last resort in his 1873 book *Lombard Street: A Description of the Money Market* .

The figure below shows the way monetary policy has been understood to operate since the 1980s.



The macroeconomy may be stabilized by a central bank following an appropriate policy rule.

The possibility of macroeconomic instability has been recognized since at least the 19th century. Many solutions have been proposed to stabilize spending at a level consistent with price stability and full employment. But since the 1980s, the dominant view has been that all that is needed to eliminate macroeconomic instability, is to have a central bank follow an appropriate **policy rule**. While the details vary, the basic idea is that when inflation rises (or more generally, when the economy is "overheating"), the central bank should raise interest rates by enough to bring output back down to sustainable level.

Policy rule. A strict rule that is supposed to guide central bank decisions about interest rate policy. In the US, the Federal Reserve is sometimes described as following a policy rule called the "Taylor rule".

The overwhelming consensus among economists and policymakers today is that the macroeconomy is not stable. Economists disagree on many questions. But only a small minority believe that economic outcomes would stay within acceptable bounds without a central bank actively managing the availability of credit.

While very few economists believe that full employment and price stability are possible without active management of the economy by central banks, there are more economists, especially since 2007, who believe that the tools normally used by central banks are inadequate for this purpose. One concern is that central banks cannot reliably control the terms on which banks lend to the private sector. Another is that interest rates don't have a strong enough effect on business investment. A third concern is that central bank intervention may do more harm than good. This concern is based on the idea, expressed by Milton Friedman among others, is that there are "long and variable **lags**" in the effects of monetary policy. As a result, by the time the central bank's actions influence the real economy, conditions may have changed so much that the bank may be pushing in the wrong direction. In this view, the negative feedback loop from investment, to output, to inflation, to the interest rate, to investment, produces cycles rather than convergence to equilibrium.

Lag. The amount of time it takes a change in some policy instrument or other exogenous variable to produce its effects on other economic variables.

In the US, the Federal Reserve tends to follow a policy rule called the **Taylor rule**, which gives equal wight to divergences of inflation and of unemployment from their target values. In most other rich countries, central banks follow policy rules that focus exclusively on inflation. But no central bank follows a rule strictly and mechanically; there is always room for **discretion** by the authorities.

Taylor rule. A formula supposedly followed by the Federal Reserve in setting interest rates. It says the bank should raise interest rates when unemployment falls below its target level or inflation rises above it, and should give the two targets equal weight.

Discretionary. Describes policy decisions that are made by the authorities based on their judgement or preferences, rather than according to a fixed rule.

Orthodox macroeconomics focuses on the causal chain running from interest rates, to investment, to output, to unemployment, to wages, to inflation.

In recent years, macroeconomic policy has been conducted primarily by central banks. By law or in practice, central bankers' top concern in normal times is low and stable inflation. And while central banks have many tools with which to influence the financial system and the larger economy, their primary tool in recent decades has been changes in the short-term interest rate that they more or less directly control. As a result, macroeconomics textbooks have come to focus on one particular causal chain – from interest rates, to investment, to output, to unemployment, to wages, to inflation.

1. The central bank takes actions to change the **policy rate** of interest – in this case, the **federal funds rate**, the interest rate banks charge each other for very short-term loans.
2. Changes to the Federal funds rate are passed on to other nominal interest rates, including the rates offered to nonfinancial businesses and households. With inflation given, changes in nominal rtes are also changes in real rates.

Policy rate. The interest rate that is controlled most directly by the central bank, with the hope that other interest rates will move with it. Between the 1980s and 2008, the policy rate in the US was the Federal funds rate.

Federal funds rate. The interest rate large banks charge each other for very short-term ("overnight") loans of reserves at the Fed. Between the 1980s and 2008, changes in the federal fund rate were the main tool for monetary policy in the United States.

3. Changes in the rate of interest affect businesses' decisions about how much to borrow and invest. When interest rates fall, businesses borrow and invest more; when interest rates rise, they borrow and invest less.
4. Changes to business investment affect the total lev el of spending in the economy. An additional dollar of investment normally produces more than one additional dollar of total spending. The ratio between the increase in investment and the total resulting increase in GDP is called the *multiplier*.
5. Higher GDP reduces unemployment, as described by Okun's law. Lower GDP, similarly, increases unemployment.
6. Lower unemployment tends to raise wages, as workers have more bargaining power relative to employers. Higher unemployment tends to reduce wages.
7. Changes in wages tend to get passed on to other prices. In the simplest version of this story, businesses simply set their prices as a fixed **markup** over wages. So when wages rise faster, inflation will be higher; when wages rise more slowly, inflation will be lower.

Markup. The difference between the marginal cost of producing a good and its final price.

The whole process is typically assumed to involve a *lag* of one to two years – that is, it will take between and one and two years for a change in monetary policy by the central bank to have its full effect on employment, output and inflation.

This standard story captures several important facts about the world. First, it is true that there is often a close link between output, unemployment and inflation. *Expansionary* policy tends to raise output and inflation, and reduce unemployment. *Contractionary* policy does the opposite, reducing output and inflation and raising unemployment. Second, the component of output that varies most of the business cycle is investment. Booms and busts usually result from rises and falls in investment spending. Consumption and government spending tend to follow the overall state of the economy; they don't normally drive it. And in the US, net exports are too small to play a central role in the business cycle. (In other countries they are more important.) Third, credit conditions are one of the three main factors affecting business investment. And finally, macroeconomic policy is normally carried out by a central bank trying to make credit more or less available, as measured by the prevailing interest rate.

But while the standard story describes one important piece of the picture, in the real world things are more complicated. First, the central bank does not have perfect control over the actual terms on which businesses and households can borrow. There are many interest rates in the economy, and they do not all move in lockstep.

And the interest rate is not the whole story – most businesses and households cannot borrow as much as they want at the prevailing interest rate, so the terms on which credit is available matter as well as its price. Second, monetary policy may work through other channels besides business borrowing. Households also borrow; in fact, mortgage borrowing may be more sensitive to monetary policy than is borrowing for investment. And changes in the policy interest rate also can have important effects on asset prices and on the *exchange rate* with other currencies. Third, there are many other factors beside credit conditions that influence flows of spending in the economy. Investment often rises or falls for reasons that have nothing to do with monetary policy.

Nonetheless, it's important to understand the standard story, partly because of the important element of truth in it, and partly because it is how most macroeconomic policymakers, at central banks and elsewhere, talk about their decisions.

The central bank may change the policy rate by trading securities for reserves through open-market operations. Or it may lend directly to banks through the discount window, or pay interest on reserves.

Historically, central banks have sought to control prices and aggregate demand through a number of policy instruments. But today, most central banks normally carry out monetary policy by setting a single short-term interest rate. The interest rate most directly controlled by the central bank is referred to as the *policy rate*. The policy rate is normally an interest on loans between banks.

In the US, the policy rate from the 1980s up until 2008 was the *federal funds rate*. Prior to 1980, policy worked primarily through the quantity of reserves rather than an interest rate. Since 2008, the federal funds rate has not been a tool of policy; it is not clear whether the federal funds rate will return to being the policy rate in the future, or if it will be replaced by interest on reserves or by some other instrument.

Regardless of the specific instrument used, changes in the policy rate are intended to change the marginal cost to banks of expanding their balance sheets. When the central bank buys and sell securities through **open market operations**, it changes the supply of reserves available to banks for lending, which changes the interest rate banks that need to borrow reserves will have to pay. A bank making a loan will generally lose reserves, as the newly created deposits are transferred to other banks, so the cost of lending an additional dollar will be close to the cost of borrowing an additional dollar of reserves. Changes in the rate at the **discount window** work the same way,

Open market operations. One of the main tools through which monetary policy is conducted. When the central bank wishes to raise interest rates, it sells government securities for reserves, reducing the quantity of reserves available for banks to borrow. When it wishes to lower interest rates, it buys securities, increasing the quantity of reserves available.

Discount window. An arrangement by which the central bank provides additional reserves to private banks, in exchange for some other asset.

except here we are talking about loans made directly by the central bank to private banks, rather than loans from one bank to another. More recently, central banks have introduced **interest on reserves** as a new policy rate. Since 2008, abundant reserves in the banking system have made it impossible to raise the federal funds or discount rates above zero – there is no reason for banks to pay interest to borrow reserves when banks have far more reserves available than are needed for settlement purposes. So central banks instead pay interest on banks' excess reserves. Banks that lose reserves as a result of new lending, give up the interest on the reserves. So interest on reserves, like the federal funds and discount rates, allow the central bank to control, or at least influence, the cost to banks of making additional loans.

Interest on reserves. Interest on reserves, or interest on excess reserves (IOER) is interest paid by the central bank to private banks on their holdings of reserves. Historically, central banks did not normally pay interest on reserves but since 2008 it has been introduced as an instrument of monetary policy.

Other interest rates are set at the policy rate plus a spread. The spread is not necessarily fixed, however, so interest rates do not all move together.

Interest rates may be higher or lower depending on many factors – the creditworthiness of the borrower, the length of the loan, whether or not there is collateral, and so on. So we cannot speak of "the" interest rate – there are many different interest rates. In October 2016 in the US, for example, 1-year government bonds were at 0.7 percent, 10-year government bonds were at 1.8 percent, a typical 30-year mortgage was 3.5 percent, a Baa-rated corporate bond was at 4.3 percent, and the typical credit-card loan was at 12.5 percent. While these rates broadly move up and down together, they do not move in lockstep with each other or with the federal funds rate. As of October 2016, the Federal funds rate was at 0.5 percent. Ten years earlier, in October 2006, it was at 5.25 percent, so it is nearly 5 points lower than a decade ago. But mortgage rates are only about 3 points lower than a decade ago, corporate bond rates are about 2 points lower, and credit card rates are only one point lower.

As Figure 10 shows, the spreads between the federal funds rate and the various market rates vary quite a bit over time. Market interest rates do move broadly in line with the federal funds rate, but they don't do so quickly or completely. Short-term government bonds (the thin gray line) track the federal funds rate closely. But changes in the federal funds rate are passed through to the other rates only partially and gradually. A simple *regression* suggests that a one point rise or fall in the federal funds rate over one year is associated with a change of 0.4 percent in the average interest rate on bank lending. The important thing for monetary policy is that when the central bank raises the policy rate, other interest rates do eventually rise, and when the central bank cuts the policy rate, other rates do eventually fall.

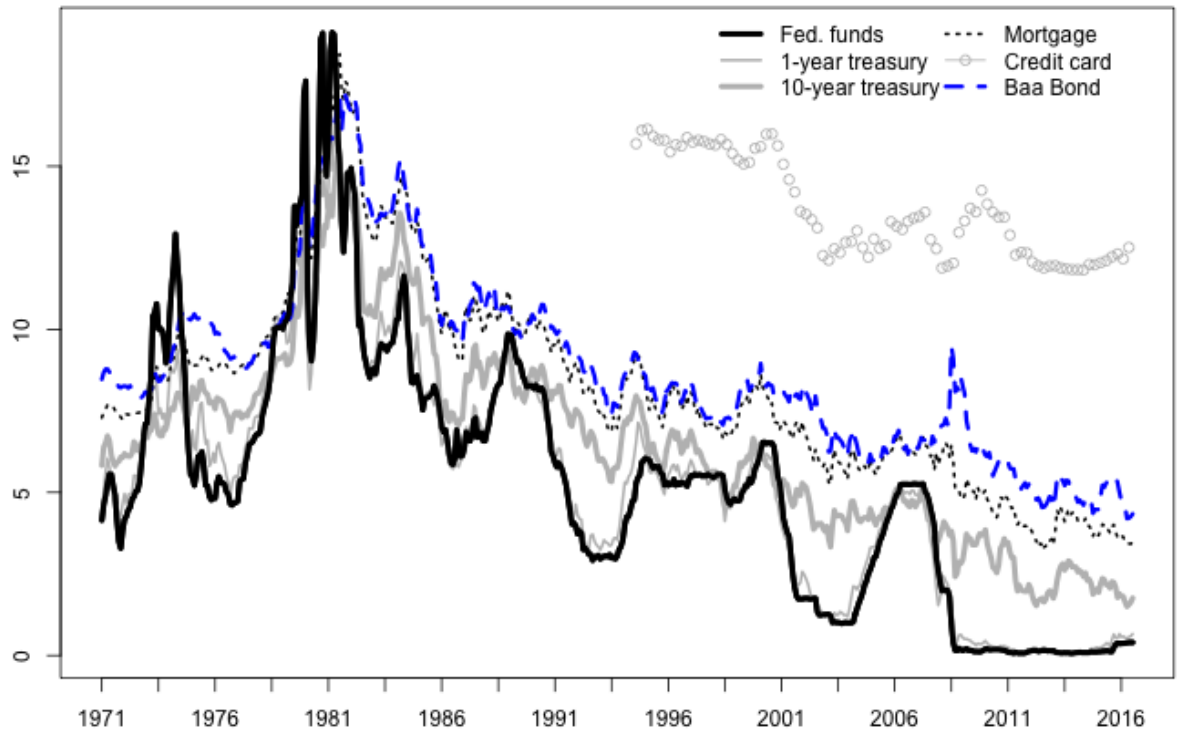


Figure 10: Selected interested rates, 1971-2016. Source: FRED.

The supply of credit depends both on interest rates and on banks' willingness to lend.

The interest rate is, in effect, the price of a loan. But bank loans are different from other goods and services. Normally, the relationship between the buyer and seller ends once the sale takes place, but with a loan, the lender is expecting repayment over many years. So unlike most businesses, which are happy to sell to anyone who is willing to buy, a bank must be selective about who it makes loans to – it must look carefully at their finances and other characteristics to decide if they will be able to pay back the loan. And it will normally impose conditions on the loan, such as requiring the borrower to pay part of the purchase price of whatever is being bought with the loan, or to provide **collateral** as protection against **default**. So the cost of a loan only tells us part of what we need to know about the supply of credit. It also matters how easy it is to get loans at the given interest rate. This can be important for monetary policy: Changes in the

Collateral. When a person or business takes out a loan, they may agree to put up an asset they own as collateral. If the borrower fails to repay the loan on schedule, the lender gets the collateral.

Default. Failure to make the required payments on a loan. When a borrower defaults, the loan is normally written down or written off by the lender.

policy rate may affect the availability of loans as well as the market interest rate.

Monetary policy that operates by changing the availability of credit, rather than via market interest rates, is said to be working through the **credit channel**. The idea is that when it becomes more expensive for banks to make loans, they may be reluctant, for various reasons, to raise interest rates to their borrowers, but will become more selective about who they lend to instead. And when lending becomes cheaper for banks, they may not reduce interest rates, but instead may become more willing to make loans. Many macroeconomists – including former Fed chair Ben Bernanke – believe that the credit channel may be more important than market interest rates for the transmission of monetary policy.[16] The credit channel explains why investment seems to respond to changes in the policy rate even though most investment is financed with long-term loans, and long rates are the least responsive to monetary policy.

Credit channel. One way that monetary policy can affect the real economy. Monetary policy that operates through the credit channel affects the availability of loans, rather than interest rates.

[16] Ben Bernanke 1995, "Inside the Black Box: The Credit Channel of Monetary Policy Transmission," *Journal of Economic Perspectives*.

For monetary policy to be effective, at least some market interest rates must move with the policy rate, or changes in the policy rate must affect the availability of credit.

Investment decisions are based, at least in part, on how cheaply and easily firms and households can borrow. Changes in investment in turn lead to changes in output, employment and inflation.

For households and businesses, current spending (consumption for households, costs of production for businesses) are normally paid for out of current income. Investment spending is more often made using borrowed funds. So the cost and availability of credit is an important factor in determining how much investment takes place. Consumption spending, especially on **durables**, may also be responsive to changes in credit conditions, but normally less than investment spending. In the US, residential investment – purchase of newly constructed housing – has generally been the spending most sensitive to changes in credit conditions.

Durables. Consumption goods that will provide a flow of services for many years after they are purchased, such as cars or appliances. While purchases of durables are normally classed with consumption, in some ways they resemble investment spending.

Changes in investment spending lead to changes in output via the *multiplier*. So low interest rates, by boosting investment spending, will eventually lead to higher GDP. A common estimate is that the **interest elasticity** of output is around one, meaning that a one point decline in the policy interest rate will eventually raise output by one percent. But it can take considerable time for these effects to be felt. According to the widely-used forecasting model of the Organization for Economic Cooperation and Development (OECD), a sustained one point increase in the federal funds rate would re-

Interest elasticity. How responsive output (or some other variable) is to changes in the interest rate. For example, if we say output has an interest elasticity of 0.5, that means we think that a 1 point fall in the interest rate will cause output to rise by 0.5 points.

duce US GDP by just 0.1 percent (a tenth of a point) in the first year. Over two years, it would reduce GDP by a total of 0.4 points, over three years, 0.7 points, over four years 0.9 points, and over five years 1.1 points. Estimates for other countries suggest somewhat smaller effects of interest-rate changes, but similar *lags*. So a change in monetary policy may well take five years to produce its full effects on the economy. This makes monetary policy a somewhat awkward tool for responding to short-run fluctuations in demand.

Eventually, however, a one-point increase in interest rates is generally expected to reduce GDP by about one point, more or less. This also implies higher unemployment, as described by *Okun's law*, and lower inflation, as described by the *Phillips curve*. Similarly, lower interest rates should eventually lead to higher GDP, lower unemployment and higher inflation. Note that these effects are on the level of GDP, not on the growth rate of GDP. Only a few economists believe that the central bank can directly affect the long-run growth rate of the economy. But most believe that if it makes large enough changes in the policy rate, and sustains them for long enough, it can successfully achieve its targets for output and price stability.

The central bank sets monetary policy based on the behavior of inflation, unemployment, output or other macroeconomic variables. It may follow a policy rule, which dictates exactly how it should change the policy rate in response to changes in the target variables.

Macroeconomic stabilization requires shifting policy in an *expansionary* direction when output is below potential, when unemployment is high, and/or when the economy is facing deflation; and shifting policy in a *contractionary* direction when output is above potential, unemployment is very low, and/or when inflation is high or rising. Central banks can adjust policy either using their *discretion*, or by following a *policy rule*. Discretionary policy depends on the judgement of the bank's leadership, weighing a variety of factors to make a decision in each particular case. Following a policy rule means announcing in advance exactly how policy will be set, and sticking to that rule no matter what happens. Policy rules usually involve just one or a few targets. Today, policy rules are usually understood to be rules for setting an interest rate, but there can be rule for whatever instrument the central bank uses for monetary policy. Historically, the *gold standard* functioned as a kind of policy rule for central banks before World War I.

In practice, central banks never follow rules exactly – there will be situations that seem to require a different policy than the one specified by the rule, and views will change about how policy should be

conducted. But many economists believe that central banks ought to try as best they can to follow a fixed rule. Milton Friedman, the famous *monetarist* economist, argued for a rule requiring the central bank to ensure a steady growth of the money supply. Michael Woodford, one of the most prominent American economists currently writing about monetary policy, argues for a strict policy rule:

> Why is it not enough to appoint central bankers with a sound understanding of the way the economy works, and then grant them complete discretion to pursue the public interest in the way that they judge best? ... There are two important advantages of commitment to an appropriately chosen policy rule. One is that the effectiveness of monetary policy depends as much on the public?s expectations about future policy as upon the bank?s actual actions. Hence a bank must not only manage to make the right decision as often as possible; it is also important that its actions be predictable. The second ... reason is that if a bank .... is not bound by any past commitments, it will choose a systematic pattern of behavior that is suboptimal.[17]

[17] Michael Woodford, *Interest and prices: Foundations of a theory of monetary policy*, p. 17-18.

Woodford, like many other economists, argues that discretionary policy will suffer from **inflationary bias** – central banks with full discretion will tend to allow inflation to rise higher than is socially optimal. During the 1970s, when inflation was very high in the US and many other rich countries, this argument for a strict policy rule seemed persuasive. Not all economists agree that central banks or other policymakers suffer from inflationary bias. In recent years an increasing number of economists have come to favor central bank discretion. For example, Narayana Kocherlakota, a prominent macroeconomist who recently served on the Fed's governing body, formerly favored a strict rule but now says that "societies will achieve better outcomes if central banks are given complete discretion to pursue well-specified goals."[18] He attributes his change of views in part to his first-hand experience at the Fed.

Inflationary bias. A tendency of policymakers to allow inflation to rise higher than is socially optimal.

[18] Narayana Kocherlakota 2016, "Rules versus discretion: A reconsideration," Brookings Papers on Economic Activity. A summary of Kocherlakota's argument and a video interview with him can be found here.

The Federal Reserve does not officially follow a policy rule, but in practice it often seems to be following a rule known as the Taylor rule. This rule says that for every point inflation rises above its target level, the policy rate should be raised by 1.5 points, and for every point output falls below potential, the policy rate should be reduced by 0.5 points.

Officially, monetary policy in the United States is entirely *discretionary* – the central bank can adjust the policy rate and other instruments however it thinks best to hit its ultimate targets of full employment and price stability. But in recent decades, the Federal Reserve has often seemed to be as following a *Taylor rule*. This is a policy rules that says that interest rates should be raised by a fixed amount for each point that inflation is above, or output below, their target levels.

There are various forms of the Taylor rule, but one common one is:

$$i = inf + r^* + 1/2(inf - inf^*) + 1/2(Y - Y^*)$$

In this equation, i is the policy rate set by the central bank, *inf* is the inflation rate, *inf** is the target inflation rate, r^* is the central bank's estimate of the neutral real interest rate, Y is output (GDP), and Y^* is an estimate of potential GDP. For example, suppose the economy is exactly where the central bank thinks it should be – output at potential, inflation at its target level, and the policy rate at the neutral level.

For example, suppose that the economy is initially in what the central bank considers its ideal state – output at potential, inflation at its target level, and the real rate at the neutral level. For concreteness, let's say the inflation target is 2 percent, as it is in the US and many other rich countries, and the central bank considers the neutral real rate of interest to be 3 percent. Then the central bank would set its policy rate at $2 + 3 + 1/2(0) + 1/2(0) = 5$ percent. Notice that the neutral rate is defined as a *real* rate, while the policy rate is a *nominal* rate. So in a situation where the central bank wishes to be neither expansionary nor contractionary, it sets its policy rate at the neutral rate plus inflation.

Now suppose that inflation rises by one point, to 3 percent. Since inflation is now above its target, the central bank will want to shift policy in a *contractionary* direction, so it will raise the policy rate. The Taylor rule says that the new rate should be equal to $2 + 3 + 1/2(3 - 2) + 1/2(0) = 4.5$ percent. On the other hand, suppose that, again starting from a situation where all variables were at their target levels, output growth slowed, leaving GDP one point below the central bank's estimate of potential. Then the central bank would want to adjust policy in an expansionary direction. In this case, the Taylor Rule says that the policy rate should be equal to $2 + 2 + 1/2(0) + 1/2(-1) = 2.5$ percent. Between 1980 and 2008, the Fed seemed to respond roughly this way in practice. But like most central banks, it tries to adjust the policy rate slowly and steadily. So if the rule implies an interest rate very different from the current policy rate, the Fed moves to the new rate in small steps, rather than in one big jump.

The Taylor rule implies that the Fed puts equal weight on deviations of inflation from target and on deviations of output from potential. Other central banks might use different rules. For example, some central banks claim to set policy based solely on inflation, without looking at the output gap at all. On the other hand, some economists have suggested that a rule that place a greater weight on output might be preferable.

Any viable policy rule must ensure that the policy rate changes at least in proportion to inflation – that is, when inflation rises or falls, the policy rate must rise or fall by at least as much. The reason for this is that the policy rate is a nominal rate, but what matters for economic behavior is the real rate. For a given nominal rate, each point that inflation rises increases the real interest rate by one point, and each point that inflation falls reduces the the real rate of interest by one point. So if the nominal rate does not change in the same direction as inflation, by at least as much, the real rate will fall when inflation rises and rise when inflation falls. Since lower real rates are *expansionary* and higher real rates are *contractionary*, this effect will tend to amplify the original change in inflation – higher inflation will lead to lower real rates leading to even higher inflation, and lower inflation will lead to higher real rates leading to even lower inflation and then to deflation. This form of instability was first described by the Swedish economist **Knut Wicksell** around the turn of the 20th century. It is believed to have played a role in the Great Depression of the 1930s: Deflation in the US and a number of other countries led to very high real interest rates, discouraging investment and forcing many businesses to default on their debts, which depressed demand and led to further deflation. To avoid this kind of instability, when inflation changes, central banks try to adjust nominal rates in the same direction, by at least as much.

Wicksell cycle. A form of instability in capitalist economies first described by the Swedish economist Knut Wicksell, who called it the "cumulative process." It describes a situation where high inflation leads to lower real interest rates which lead to even higher inflation, or where low inflation or deflation leads to higher real interest rates which lead to even more deflation.

There are a number of points where the transmission of monetary policy can break down.

Conventional monetary policy cannot force the policy interest rate below zero. This limit is known as the zero lower bound.

The central bank normally controls the level of the *policy rate* of interest – in the US, the *federal funds rate* – by changing the supply of reserves available for lending between banks, by buying or selling government securities in *open market operations*. Buying securities increases the quantity of reserves available for borrowing by banks, while selling securities in return for reserves reduces the quantity of reserves available. Central banks may also supply reserves to banks by lending them directly through the *discount window* or through **repurchase agreements**. All of these forms of monetary policy depend on the idea that reducing the supply of reserves will increase the interest rate that banks with excess reserves can charge for lending them, and increasing the supply of reserves will reduce the interest rate that banks with excess reserves can charge for them. Interest rates are, in effect, the price of reserves; as with other goods, increas-

Zero lower bound. A practical limit on monetary policy: conventional tools like open market operations cannot move the policy rate below zero.

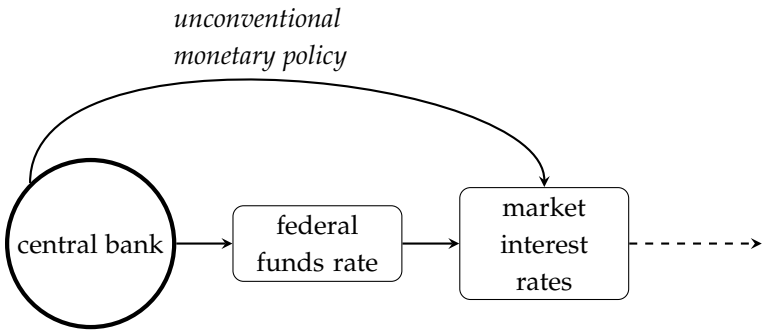
Repurchase agreements. Repurchase agreements, also called "RPs" or "repos", are a tool the central bank uses to make loans to banks and other financial institutions. In effect they are loans, but they are structured as the sale of a security, combined with an agreement to buy it back in a certain amount of time. In a reverse repo or "RRP?, the Fed borrows money from primary dealers.

ing the supply reduces the price, and reducing the supply increases the price. Bu this kind of policy cannot reduce interest rates below zero. Even if there is an infinite supply of something available, it will be normally be free – that is, it will have a price of zero. Negative interest rates mean that the lender is paying the borrower. But as long as holding excess reserves is costless for banks, they will never pay to lend them out.

One solution to this problem is to make holding reserves costly, by imposing a tax on reserves. This is equivalent to a negative policy rate. A number of central banks have experimented with taxes on reserves as a way of creating a negative policy rate. So far, the Federal Reserve has never tried to create negative rates this way in the US. One concern is that if the tax gets too high, banks will trade in their reserves for physical currency instead. At the level of taxes that have been imposed in other countries, this has not been a problem.

Monetary policy normally has its direct effects on short-term, interbank interest rates. Changes in these rates may not be passed on to longer-term rates or to rates faced by nonfinancial borrowers.

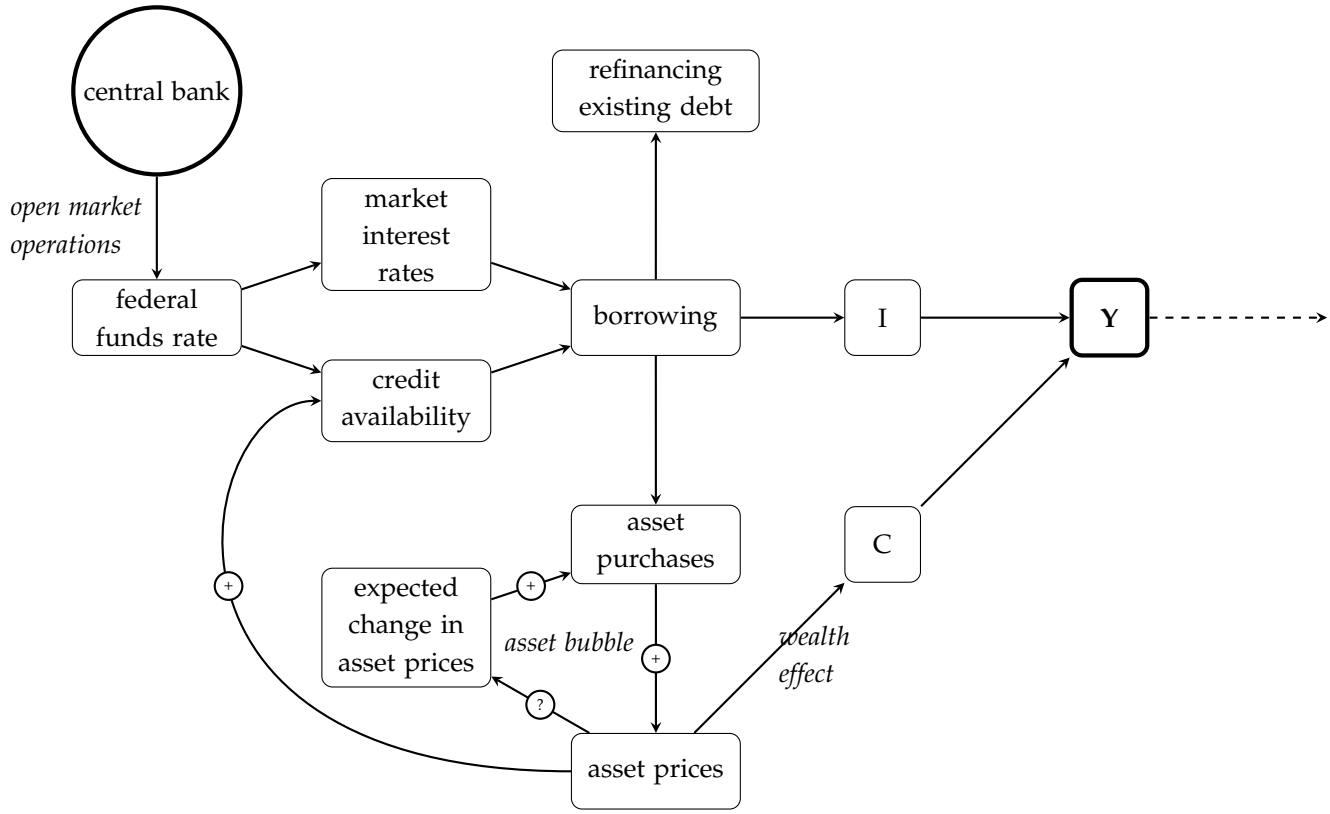
If open market operations cannot be used or are ineffective, central banks may seek to influence longer interest rates via quantitative easing or other forms of unconventional monetary policy.



Even if the central bank can move market interest rates and the availability of credit, that may not affect the actual volume of lending, which depends on demand for loans as well. For this reason, monetary policy is more reliable in a contractionary than in an expansionary direction.

One problem with using monetary policy to control the level of economic activity is that borrowed funds are not always used for the purchase of newly produced goods and services.

Easy credit may encourage purchases of existing assets. If investors see the resulting price increases as signs that prices will continue to increase in the future, an asset bubble can result.



I is investment.
Y is output (usually measured by GDP).
U is unemployment.

A + in the line between two aggregates means there is a positive relationship between them, that is, a rise in the first will cause a rise in the second, and a fall in the first will cause a fall in the second. A - in the line between two aggregates means there is a negative relationship, that is, a rise in the first will cause a fall in the second, and a fall in the first will cause a rise in the second.

Monetarists believe that the money supply is controlled by the central bank, and that there is a direct link from the money supply to output or prices, without any need for transmission through the financial system.

Exchange Rates

A nominal exchange rate is the price of one country's money in terms of another's. A real exchange rate is the nominal exchange rate adjusted for the price levels in the two countries. An exchange rate index is the value of one country's currency in terms of a basket of others.

Foreign exchange refers to the currency of a country other than our own.

The *nominal* exchange rate between two countries indicates how many units of the first country's currency you can buy for one unit of the second country's currency. For example, the nominal exchange rate between the US and the UK might be 2 dollars per pound. Since this is a price of two moneys, we can give any exchange rate in two ways. For example, we can say that one British pound is worth 2 dollars, or we can say that one dollar is worth 0.5 pounds.

Because an exchange rate can be expressed in two different ways, it is sometimes convenient to use the notation $E_{a/b}$ when writing about exchange rates. This means "the exchange rate in currency a per units of currency b." For example, if we write $E_{\$/e} = 1.14$, that means the dollar-exchange rate is equal to 1.1 dollars per euro. If one euro is worth 1.1 dollars, then one dollar must be worth 0.91 euros. So we can write the same exchange rate as $E_{e/\$} = 0.91$. In general, for any two currencies a and b, $E_{a/b} = \frac{1}{E_{b/a}}$. We can think of $E_{a/b}$ as the value of currency b measured in units of currency a, and $E_{b/a}$ as the value of currency a measured in units of currency b. Exchange rates are usually, but not always, expressed in units of the weaker currency per unit of the stronger currency. We are more likely to see "one dollar equals 90 yen" then "one yen equals 0.011 dollars."

We say that a country's currency is **appreciating** or getting *stronger* when one unit of that currency can buy more of a foreign currency than it was formerly able to. We speak of a currency **depreciating** or getting *weaker* when one unit can buy fewer units of a foreign currency than formerly. Since the value of one currency is measured in terms of the other currency, an appreciation of one is the same as a depreciation of the other.

Appreciate. An increase in value of one currency relative to another one. We can also say it has gotten stronger.

Depreciate. An decline in value of one currency relative to another one. We can also say it has gotten weaker.

When $E_{a/b}$ goes up, that means it takes more of currency a to buy one unit of currency b. So a is depreciating and b is appreciating. We can describe the same change as $E_{b/a}$ going down – it takes less of currency b to buy one unit of currency a. Either way, a has gotten weaker and b has gotten stronger.

A higher value of $E_{a/b}$ means that b is stronger relative to a, and a is weaker relative to b.

We normally measure appreciations and depreciations in percentages. For example, suppose the exchange rate between the dollar and the euro changes from 1.1 dollars per euro to 1.2 dollars per euro.

This means that one euro now buys more dollars, so the euro has become more valuable – it has gotten stronger, or appreciated, against the dollar. Or equivalently, the dollar has depreciated against the euro.

How much has the exchange rate changed? We would *not* say that "the euro has appreciated by 0.1" or "the dollar has depreciated by 0.1." This would not be meaningful. Rather, we say that the euro has appreciated by 9 percent. We calculate this in the same way we calculate any percentage change: new value minus old value, divided by old value. In this case, $\frac{1.2-1.1}{1.1} = 0.09$ = 9 percent. The euro has appreciated by 9 percent against the dollar. Similarly, the dollar depreciated by about 9 percent against the euro. (Actually the percentage change in the dollar is slightly smaller, but we don't need to worry about that here.)

Similarly, if we know the exchange rate at the start (or end) of a period and we know the change during the period, we can calculate the exchange rate at the end (or start) of the period. For example, as of December 2016 the US-Mexico exchange rate is around 20 pesos per dollar. In other words, $E_{peso/\$} = 20$, and $E_{\$/peso} = 0.05$. (Since $1/20 = 0.05$.) Suppose the peso were to appreciate by 10 percent against the dollar during 2017. This means that at the end of 2017, the peso can buy 10 percent more dollars than at the end of 2016. So $E_{\$/peso}$, the exchange rate expressed in dollars per peso, must be 10 percent higher. So we multiply the old exchange rate by one plus 10 percent, or 1.1. At the end of 2017, then, we will have $E_{\$/peso} = 0.05 * 1.1 = 0.055$, or equivalently, $E_{peso/\$} = 1/0.055 = 18.2$ – one dollar will equal 18.2 pesos.

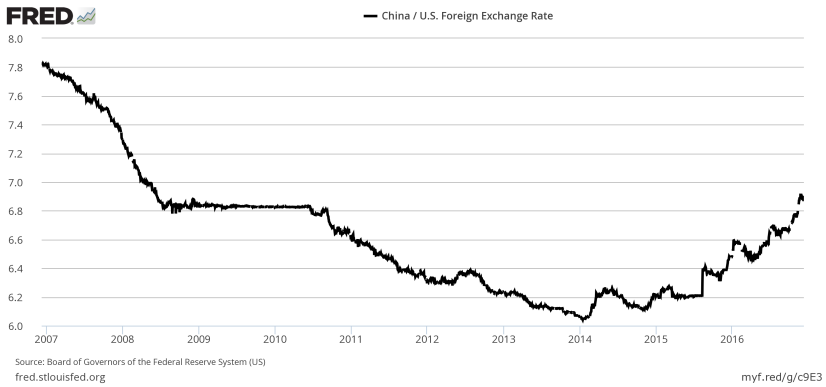
All this applies to a change in exchange rate over any period of time. But if the period is longer or shorter than a year, we will sometimes want to convert the total percentage change to an annual percentage rate.

When the value of a **fixed exchange rate** is changed by deliberate government policy, that is called a *devaluation* or *revaluation*, rather than a depreciation or appreciation.

Fixed exchange rate. An exchange rate whose value is set by a government rather than in foreign exchange markets.

Because exchange rates can be quoted either as units of the home currency per unit of the foreign currency, or as units of the foreign currency per unit of the home currency, the terms "higher" and "lower" are ambiguous when applied to exchange rates. For example, suppose the value of the dollar increased to 0.8 pounds; we could just as well describe that as the value of the pound falling to 1.25 dollars. The same change could be described as the exchange rate going up or down, depending which way the it is quoted. So we try to avoid speaking of currencies going up or down; rather than saying a currency "went up," we say it got stronger or appreciated.

When you see a reference to an exchange rate going up, for instance in a newspaper article, you should take a moment to figure out if higher means stronger, or if it could mean weaker. For example, in the nearby figure shows the value of the Chinese currency, the renminbi. This exchange rate is usually expressed in renminbi per dollar. So higher values mean a weaker renminbi, and lower values mean a stronger renminbi. As you can see, the Chinese currency appreciated against the dollar in 2007-2008, and again from 2011 through 2013 but over the past two years it has mostly depreciated against the dollar.



The real exchange rate is the exchange rate adjusted for inflation in both of the currencies involved.

The *real exchange rate* is the nominal exchange rate adjusted for inflation. Or equivalently, the real exchange rate is the price of a basket of goods in one country relative to the price of the same basket in another country. Real exchange rates are affected by changes in the nominal exchange rate, and by the difference in the inflation rates in both countries.

For example: Suppose you are a real estate speculator and you decide to sell 100 houses in the United States and buy a bunch of similar houses in the UK. How many houses you can buy there will depend on:

1. How many dollars you get for each house you sell in the US.
2. How many pounds you get for each dollar.
3. How many pounds a similar house costs in the UK.

So, the number of British houses you can get for your American houses will increase if:

1. The price of an American house in dollars rises by a lot. (Inflation is high in the US.)

2. The number of pounds you get for each dollar increases. (There is a nominal appreciation of the dollar.)

3. The cost of a British house in pounds does not rise by a lot. (Inflation is low in the UK.)

Any of these changes is equivalent to a *real appreciation* of the dollar.

Let's write the *nominal* dollars-per-euro exchange rate as $E_{\$/e}$. In other words, if one euro is worth 1.2 dollars, we can write $E_{\$/e} = 1.2$. We could equally well describe the same exchange rate as $E_{e/\$} = 0.83$, since if one euro is worth 1.2 dollars then one dollar is worth 0.83 euros. If we write the exchange rate as $E_{\$/e}$ (dollars per euro) then a higher number means an appreciation of the euro, and a depreciation of the dollar. If we write the exchange rate as $E_{e/\$}$ (euros per dollar), then a higher number means an appreciation of the dollar, and a depreciation of the euro.

We'll write the price level in the US as $P_{\$}$, the price level in the euro area as P_e, and the *real* exchange rate as $RE_{\$/e}$. Then

$$RE_{\$/e} = E_{\$/e}(\frac{P_e}{P_{\$}})$$

In other words, if everything in the euro area costs exactly 20 percent more than the same good in the US, and the euro is worth 1.2 dollars, that is the same real exchange rate as if everything in the euro area had the same price as in the US, and the nominal exchange rate was one dollar for one euro.

An increase in $E_{\frac{a}{b}}$ is a nominal depreciation of currency a, and a nominal appreciation of currency b. An increase in $RE_{\frac{a}{b}}$ is a real depreciation of currency a, and a real appreciation of currency b.

In practice, we are usually interested in the change in the real exchange rate, not its absolute level. For this, we can write:

$$\Delta RE_{\frac{a}{b}} \approx \Delta E_{\frac{a}{b}} - \mathit{inflation}_a + \mathit{inflation}_b$$

Fo example, suppose we are talking about dollar-euro exchange rate. This exchange rate is usually expressed as dollars per euro, so we write:

$$\Delta RE_{\$/e} \approx \Delta E_{\$/e} - \mathit{inflation}_{\$} + \mathit{inflation}_e$$

where Δ means change in percentage points. An increase in this number is an appreciation of the euro and a depreciation of the dollar. So each additional point of inflation in Europe over one year implies a one percentage point real appreciation of the euro. And each additional point of inflation in the US over one year implies a

one point real appreciation of the dollar. If inflation in Europe and the US is the same over a year, then the change in the real exchange rate will just be the same as the change in the nominal exchange rate.

For example, suppose that in a given year, the inflation rate in the US is 2 percent, the inflation rate in Mexico is 10 percent, and the peso depreciates by 4 percent against the dollar. So we can calculate the change in the real exchange rate as follows:

$$\Delta RE_{\frac{peso}{\$}} \approx \Delta E_{\frac{peso}{\$}} - inflation_{Mex} + inflation_{USA} = 4 - 10 + 2 = -4$$

Since the exchange rate is being expressed here in pesos per dollar, the negative value means a 4 percent real *appreciation* of the peso. In other words, even though the American currency became more expensive relative to the Mexican currency, American *goods* became cheaper relative to Mexican goods.

For a given nominal exchange rate, higher inflation in our country than in our trade partners means a real appreciation for currency; lower inflation than in our trade partners means a real depreciation for us.

In some cases, we may expect the real exchange rate to be more or less fixed. In that case, as you can see from the previous equation, higher inflation in our country than in our trade partners must lead to a nominal depreciation of our currency, and lower inflation must lead to a nominal appreciation. In the extreme case, where **relative purchasing power parity** holds, the real exchange rate is fixed and does not change – in other words, $\Delta RE_{\frac{a}{b}} = 0$. This implies that

$$\Delta E_{\frac{a}{b}} = inflation_a - inflation_b$$

Relative purchasing power parity. The theory that differences between the price of a representative basket of goods in different countries depend on economic fundamentals and therefore should be stable over time. Equivalent to claim that real exchange rates will be constant.

In this case, the change in the nominal exchange rate should be just equal to the difference between the two countries' inflation rates, with the currency with higher inflation depreciating and the country with lower inflation appreciating.

In other cases, the nominal exchange rate may be fixed, but policymakers might consider a change in the real exchange rate to be desirable. If the nominal exchange rate can't change, moving the real exchange rate will require achieving a different level of inflation. When a country tries to weaken its real exchange rate via lower inflation (or deflation) relative to its trade partners, that is sometimes called an *internal devaluation*, as opposed to the normal devaluation that involves a change in the nominal exchange rate.

Higher inflation in a country relative to its trade partners often leads to a nominal depreciation of its currency. So inflation has two contradictory effects on exchange rates. First, higher inflation implies

a *real appreciation* of the currency, for a given nominal exchange rate. But second, higher inflation tends to lead to a *nominal depreciation* of the currency.

Exchange rate indexes measure the value of a currency against a basket of others.

So far, we have been talking about the value of one currency in terms of one other currency, that is, *bilateral* exchange rates. But there are many different currencies in the world, and they do not all move together. Normally, a given currency will be appreciating against some other currencies, and depreciating against others. An *exchange rate index* is the average value of a currency against a "basket" of other currencies.

Exchange rate indexes may be either *trade-weighted* or *effective*. Effective weights take into account not just trade between a given pair of countries, but also the extent to which they export to the same markets.

The absolute value of an index does not mean anything, only the *change* in the index matters. For instance, you could say that the trade-weighted exchange rate index of the US dollar rose (or the dollar appreciated) by 10% between 1985 and 1986, but it makes no difference whether you call that a change from 1 to 1.1, or 10 to 11, or 100 to 110. When people say that a country's currency appreciated or depreciated without saying against what, they often mean that its exchange rate index appreciated or depreciated. But for non-US currencies, they also may mean that it appreciated or depreciated against the dollar.

Relative prices of goods and services are most important in determining exchange rate changes over the long run; over medium periods, interest rates play a larger role, and over the short run exchange rate changes are dominated by speculation in the foreign exchange market.

Floating exchange rates are determined by market demand – by the desire to use one currency to buy a different one. There are three main reasons to buy a foreign currency. First, one might want to purchase goods or services from a country where that currency is used. Second, one might wish to buy an asset in a country where the currency is used, in order to receive an income from it – a bond which will yield interest payments, a stock which will yield dividends, a business which will yield profits, or real estate which will yield rents.

Floating. Describes an exchange rate that is determined by private trading in the foreign exchange markets. The alternative is a fixed exchange rate, which is set by the government.

	Reason to Hold Currency	**Important in...**	**Currency Will Appreciate If...**	**Equilibrium Condition**
Goods market	To acquire goods and services from that country	Long run (years to decades)	Lower prices Lower costs, esp. wages Higher productivity Higher quality/more desirable exports Greater natural resources Country has trade surplus	*Absolute PPP.*: Same price in all countries: Or *relative PPP*: prices move together across countries Balanced trade
Asset market	To acquire income-yielding asset in that country (bond, stock, business, land etc.)	Medium run (months to years)	Higher interest rates Higher profits (actual or expected) Asset prices expected to rise Assets seen as liquid, safe, secure, *and* greater fear/insecurity in markets	*Interest parity*: Equal interest rate across countries
Foreign exchange market	To profit from appreciation of currency	Short run (minutes to months)	It is expected to appreciate	Exchange rate follows *random walk*: impossible to predict future changes from current value

Third, one might wish to hold the currency itself, in the hopes that it will increase in value, or appreciate. So we can think of demand for a currency coming from the goods market, from the asset market, and from the *foreign exchange* market. The exchange rate will tend to change whenever there is the opportunity for **arbitrage** in any of these markets. Arbitrage means buying something where it is cheap and selling it where it is more expensive.

Arbitrage. Taking advantage of two different prices for the same good or asset by buying it where the price is low and selling it where the price is high.

A currency will *appreciate* when there is more demand for it relative to other currencies, and *depreciate* when there is less demand for it. So looking at the goods market, the asset market, and the foreign exchange market shows different reasons why a currency might change in value. When demand for different currencies is just balanced, there is no reason for any of them to change in value. We describe this as an *equilibrium*. We can also think of it as a "no-arbitrage" condition. For example, if a car from Japan costs less than an equivalent one made in the US, there is an "arbitrage opportunity" to buy cars in Japan and sell them in the US. Since Japanese cars are sold in yen, importing Japanese cars to the US requires selling dollars and buying yen. This increases demand for yen, and reduces demand for dollars, so over time, the yen will tend to gain value (or appreciate) and the dollar will tend to lose value. Since a rise in the value of the yen makes Japanese cars more expensive, eventually this exchange rate movement will make the cars from both countries equally expensive, eliminating the arbitrage opportunity. So one no-arbitrage or equilibrium condition for exchange rates is that similar goods have the same price everywhere. There are several different possible equilibrium conditions, as described in the table.

It is important to remember that no theory can predict foreign exchange movements with any precision. At best we can describe general tendencies.

In the long run, a country's currency should appreciate when its goods are cheaper than similar goods elsewhere, or when it runs a trade surplus.

Demand from the goods market depends on demand for a country's products. A country that produces higher-quality goods, or goods that are more desired in higher-income countries; that sees its prices fall, or at least rise more slowly than in its trade partners; a country that is more **competitive** – that is, that produces goods at lower cost thanks to lower wages or greater productivity; or a country that gains access to new natural resources, will see more demand for its goods, and will tend to see its currency appreciate. Another way of looking at it is that if for whatever reason a country sells more to the rest of the world than it buys from it – that is, runs a trade surplus – there will be more demand for its currency, and it will appreciate. While

Competitiveness. The cost of producing a good in one country compared with the cost of producing similar goods elsewhere. A country will be more competitive if its costs – especially wages – are lower than elsewhere, or if its industries are more productive.

greater competitiveness and a trade surplus often go together, they are not the same thing, since other factors also influence the trade balance.

If we think of the goods market in terms of competitiveness, we will predict that exchange rates will adjust so that the same goods have the same real price everywhere. This prediction is **absolute purchasing power parity** (PPP). It says that over the long run, prices should converge to the same level elsewhere. (In other words, all **real exchange rates** should be equal to one.) This does not seem to be borne out in reality. A weaker prediction is *relative purchasing power parity*. This says that while price difference may exist between countries (for example, more labor-intensive goods will be cheaper in poor countries) these differences depend on economic "fundamentals" and will be stable over time. Relative PPP predicts that real exchange rates will be constant in the long run – that changes in nominal exchange rates will just offset differences in inflation rates between countries. There is better support for this version, especially when differences in inflation rates are large. A country with very high inflation will almost always have a depreciating currency in nominal terms, and vice versa.

Absolute purchasing power parity. The theory that over the long run, a representative basket of goods should have the same price in every country, regardless of what currency is used.

Real exchange rate. The price of a basket of goods in one country relative to the same basket in a different country. Or equivalently, the nominal exchange rate adjusted for inflation.

If we think of the goods market in terms of the trade balance, we will predict that exchange rates will adjust to eliminate trade surpluses and deficits. Most economists believe that a weaker currency will boost exports and reduce imports, because it makes the country's goods cheaper relative to the rest of the world's. So if surplus countries see their currencies appreciate, and deficit countries see theirs depreciate, then eventually trade will be balanced everywhere. We can observe these tendencies in the real world, but they operate very slowly. A country that has a trade deficit for ten years will probably see its currency depreciate, but a country that has a trade deficit for just one year is as likely to see its currency gain or lose value.

In the medium run, a country's currency will appreciate when its assets becomes more attractive to foreign investors, and in particular when its interest rate rises relative to other countries'.

Demand from the asset market depends on the attractiveness of a country's assets to foreign investors. Businesses, banks and other financial institutions, and wealthy individuals have a choice about what country's assets to hold. In general, they will seek out assets which promise the highest return, but this may be balanced against other factors – investors may accept a lower **yield** on assets that are perceived as safer, more secure, or more *liquid*. For example, a pension fund or insurance company may have government bonds from

Yield. The income received by the owner of an asset – interest on a bond, dividends from a stock, rent from real estate, and so on. The total return on an asset is the yield plus capital gains.

a number of countries in its portfolio. When the managers of the fund decide which bonds to hold, they will look at the interest rates available in various countries. If bonds in one country now offer a higher interest rate, they may decide to increase their holdings of bonds from that country. This will require them to first purchase the country's currency; increasing demand for it. Note that the current exchange rate does not matter in this case, since both the bond and its interest payments are denominated in the same currency.

While there are various different kinds of assets, we think that interest rates are most important for exchange rate movements. The decision to buy or establish a business in a foreign country depends on many factors, and takes a lot of preparation and planning; it can't be easily changed with every change in expected yield. Bonds, on the other hand, are held simply for income, not as part of a larger business plan, and they are easily bought and sold, so investors may try to change their bond holdings every time interest rates change. This means that higher interest rates will usually cause a currency to appreciate, as investors try to buy more of that country's bonds. Note that what matters is *relative* interest rates – interest rates in this country compared with the alternative. In Europe, this may mean bonds of other European countries, but elsewhere it usually means US treasury bonds. So an increase in interest rates in the US is likely to make bonds in countries like Turkey, Brazil, Indonesia less attractive, and cause their currencies to depreciate. And a decrease in interest rates here is likely to make those currencies appreciate.

Borrowing in a currency is the opposite of lending it – people try to borrow in countries where interest rates are low, and that reduces demand for their currency. Borrowing in a currency with low interest rates and lending in a different currency where interest rates are higher is known as the **carry trade**.

Carry trade. Borrowing in a currency with low interest rates and lending in a different currency where interest rates are higher.

Interest rates are not the only thing that affects asset demand. Safety (against financial risk), liquidity (the ability to easily sell the bonds, or convert them to some other currency), and security (against unfavorable changes in tax or regulation, or other government actions) make a country's assets more attractive. In recent decades, these factors have particularly favored the US – largely for these reasons, foreign investors are willing to hold US asset at an average return about 3 points below the return American investors receive on their investments abroad. A few other countries – Germany, Switzerland, Japan – benefit from a similar "safety premium." This factor is most important when investors are most frightened. So the dollar and similar currencies tend to appreciate when there is an increased perception of danger in financial markets. For example, there was a steep appreciation of the dollar (20 percent in one year) during the

financial crisis of 2008-2009. Foreign investors may also buy assets with lower yield because they expect **capital gains** – an increase in the assets' price in the future. This is especially important for stock purchases. This factor is a source of instability in international financial markets, because expectations about capital gains can change rapidly.

Capital gain. Profit that comes from an increase in the price of an asset, rather than from the income it generates.

If we focus on interest rates, then the simplest version of the equilibrium condition will be equal interest rates everywhere – as long as bonds yield more in one country than in another, investors will be selling the lower-yielding currency and buying the higher-yielding one. While the **interest parity condition** does not hold in this strong form, it is widely believed that there is some tendency for interest rates to converge between countries.

Interest parity. The theory that arbitrage across countries will result in expected returns being the same on similar bonds in all countries. In some versions, differences in interest rates should reflect expected exchange rate changes.

In the short run, changes in exchange rates are driven by speculation in foreign exchange markets. An implication of this is that short-run changes in exchange rates are random and unpredictable.

Speculation means buying something not in order to use it or get an income from it, but in the hopes of selling it later at a higher price. Many participants in the foreign exchange market are speculators – they hold foreign exchange not in order to use it to buy something else, but in the hopes that the currency itself will appreciate. Over the short run (periods much less than a year) this speculation is the dominant factor in exchange rate movements. In other words, the main reason why the dollar strengthened today, or the yen weakened over the past month, is changes in investors' beliefs about what those currencies are likely to do in the future.

Speculation. Buying an asset in order to resell it later at a higher price, rather than to use it or get an income from it.

Speculation dominates exchange rate movements because the gains from correctly guessing exchange rate changes are very large. For example, on March 30, 2016 the dollar **index** declined in value by one percent. Someone who sold dollars for euros on March 21 and bought euros for dollars on April 1 would have made the equivalent of a 3,700 percent annual return. So if you are buying and selling foreign exchange every day, it makes sense to focus on predicting future exchange rates. For instance, if you are confident that the dollar-euro exchange rate next week will be $1.20 per euro, then you will want to buy euros and sell dollars if the exchange rate today is $1.10 per euro, and sell euros and buy dollars if the exchange rate today is $1.30 per euro. In the first case you expect the euro to appreciate against the dollar, and in the second case you expect it to depreciate.

Exchange rate index. The average of a country's exchange rate against a number of other currencies.

The dominance of speculation in foreign exchange markets implies that short-term exchange rate changes follow a **random walk** – that is, they are fundamentally unpredictable. Consider the euro-dollar

Random walk. A statistic whose future changes cannot be predicted from its current or past values. If a variable follows a random walk, then the best guess for its future value is always its current value, whatever that may be.

case just described. If the dollar-euro exchange rate today is $1.30 per euro, but most market participants think that next week it will be $1.20 per euro, then they will try to sell euros and buy dollars (since the euro is worth more dollars today than they expect it to be next week). But if most people in the market are trying to sell euros and buy dollars, then the value of the euro will fall against the dollar. Similarly, if most market participants expect the euro to appreciate next week, they will buy euros today – causing the euro to appreciate today. If markets are dominated by speculation, then the only exchange rate than can be stable is the same exchange rate that is expected to hold in the near future. Some traders may expect a change in one direction, some may expect a change in the other direction, but the market as a whole must expect the same exchange rate tomorrow as today. And if today's exchange rate is always equal to the best guess of tomorrow's, it follows that the best guess for tomorrow's exchange rate is today's – that is, we cannot predict appreciation or depreciation based on today's exchange rate. This is the definition of a random walk – you cannot predict future changes based on today's value.

The view that exchange rates follow a random walk in the short run is well supported by the data. The same behavior is found in other markets where most trades are for speculation, such as stock markets. But the fact that exchange rates or stock prices follow random walks in the short run does not mean they are unpredictable over longer periods. While it is easy to bet on the value of a stock or exchange rate tomorrow, or next week, or next month, it is harder to speculate on prices many years from now. So there is no contradiction between saying that exchange rates are completely unpredictable in the short run, and that they respond predictably to interest rates, competitiveness and the trade balance over longer periods. Speculators may also take these longer-run factors into account when guessing about future exchange rate movements. For instance, the value of a currency is often affected by changing expectations about interest rates in that country over coming months.

In practice, many factors are influencing exchange rates at the same time. But in the short run, all the other factors affect exchange rates only insofar as they change expectations in foreign exchange markets.

For example, between July and August 2016 the euro *appreciated* from $1.10 to $1.13. In an article on why the euro was getting stronger the *Financial Times* noted that strong job growth in the US should have led to a stronger dollar, by making it more likely that the Federal Reserve would increase interest rates. But according to a study cited

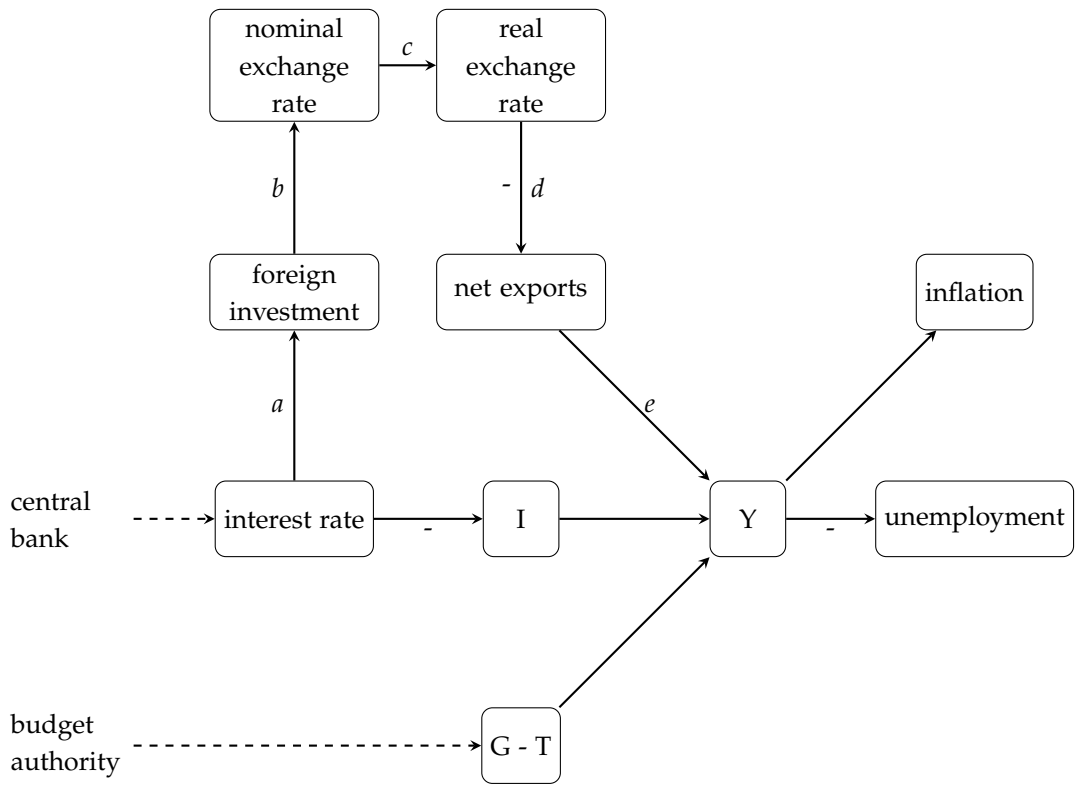
"Euro enjoys strongest run since April," August 19, 2016

in the article, "the July payrolls surprise [unexpectedly strong jobs growth] led to just a 40 basis point [0.4 percent] fall in the euro's value ... ordinarily they would have expected a decline of around 70-100 basis points." One reason that the euro continued to strengthen may be that

> investors seem to be paying increasing attention to the rate of GDP growth in the eurozone and the US. With the European continent picking up momentum over the last 12 months, the rate of US growth has actually fallen behind the continent since the first quarter of the year. That's all supportive for the euro.

The important points here: First, faster growth tends to lead a stronger currency, both because a rapidly growing economy wis more attractive to foreign investors and because faster growth makes it more likely that central banks will raise interest rates. Second, what matters is not absolute growth but growth relative to trade partners – slower growth in the US will cause the euro to appreciate just as much as faster growth in Europe. Finally, these effects don't happen automatically, but only insofar as they are reflected in the trades in foreign exchange markets. Note that the article attributes the stronger euro not simply to faster growth in Europe, but to the fact that investors are paying attention to it. This is a reminder that short-term movements in exchange rates are driven by *speculation* in foreign-exchange markets, not by "fundamentals".

Macroeconomics in an Open Economy



In an open economy, monetary policy can affect the level of economic activity via trade flows. Higher interest rates cause the currency to appreciate, reducing net exports.

In a *closed economy*, we normally think that monetary policy works mainly by changing interest rates and credit conditions, which in turn affects the amount of investment spending by businesses and households. We can see this in the flowchart: The central bank raises interest rates; higher interest rates reduce investment; lower investment leads to lower GDP; and lower GDP leads to higher unemployment and lower inflation. (This is *contractionary* monetary policy; *expansionary* policy would be the opposite changes in each variable.) We could add further details, but in a closed economy the transmission mechanism of monetary policy looks essentially like this.

In an *open economy*, monetary policy has a second channel by which it can affect the level of economic activity.

a An important part of foreign investment is lending and borrowing across borders, including purchases of foreign bonds. Owners of financial assets prefer a higher return to a lower one, all else

equal, so they will prefer to lend where interest rates are high. So a higher interest rate leads to an increase in net foreign investment into our country. This includes both increased purchases of our assets by foreigners, and reduced purchases of foreign assets by domestic wealth owners.

b Purchases of our assets have to be made using our currency. So higher foreign investment increases the demand for our currency, which tends to raise its price. In other words, increased foreign investment will normally cause a nominal appreciation of our currency.

c An increase in the value of our currency relative to other currencies will, all else, cause our goods to become more expensive relative to foreign goods. In other words, when there is a nominal appreciation, there will tend to be a real appreciation as well.

d An increase in the price of our goods and services will lead people to purchase foreign substitutes instead. This is likely to lead to lower net exports, as our newly cheap exports fall and our newly cheap imports rise.

e Net exports are a component of GDP, so lower net exports will reduce GDP just as lower investment or lower government spending would. The fall in GDP may be greater or less than the fall in net exports, depending on the *multiplier*.

An increase in interest rates leads to a stronger currency, which lowers net exports (or equivalently, causes the trade balance to move toward deficit), reducing aggregate demand. So higher interest rates are contractionary by this channel as well; the two transmission mechanisms do not conflict, they reinforce each other.

The importance of the exchange-rate channel of monetary policy transmission depends on how easy it is to lend money and trade assets across borders; how large trade is relative to domestic production of goods and services; and how sensitive trade is to changes in the exchange rate.

How powerful the exchange-rate channel of monetary policy is in practice depends on several factors.

First, foreign investment flows must reliably respond to interest rate differentials. This requires, first of all, that there is a high degree of **capital mobility**, without prohibitive legal or institutional barriers to loans and asset sales across borders. But it also requires that investors be very quick to change their portfolios in response to interest rate differentials. And, it requires that factors outside the central

Capital mobility. How easy it is to buy and sell assets and lend and borrow money across national borders.

bank's control don't move foreign investment and/or demand for the currency too much for it to counteract.

Second, trade flows must reliably respond to exchange rate changes. This may or may not be the case. Trade normally involves contracts that are signed long in advance of the delivery of the goods, and even in the absence of contractual commitments it takes time to fnd new suppliers for imported goods and new markets for exports. And there are not always good domestic substitutes for imported goods. Exporters may also be willing to accept lower profits rather than give up market share. As a result, the short-run response of trade to exchange rates may be quite small. Many economists believe that the trade balance follows a **J-curve** after a devaluation. Initially, net exports actually fall, since trade volumes have not changed much and imported goods are now more expensive. But eventually, perhaps after a year or two, people's buying patterns will change in response to the new prices and net exports will rise.

J-curve. Describes the tendency for a depreciation of the currency to be followed first by a move toward trade deficit, then by a larger move toward trade surplus.

Finally, trade flows must be large relative to the economy. In the United States in the 1950s and 1960s, imports and exports made up less than 5 percent of GDP. In those conditions,the exchange-rate channel could not be an important part of the transmission mechanism for monetary policy even if it had a big effect on trade, because trade flows were just too small.

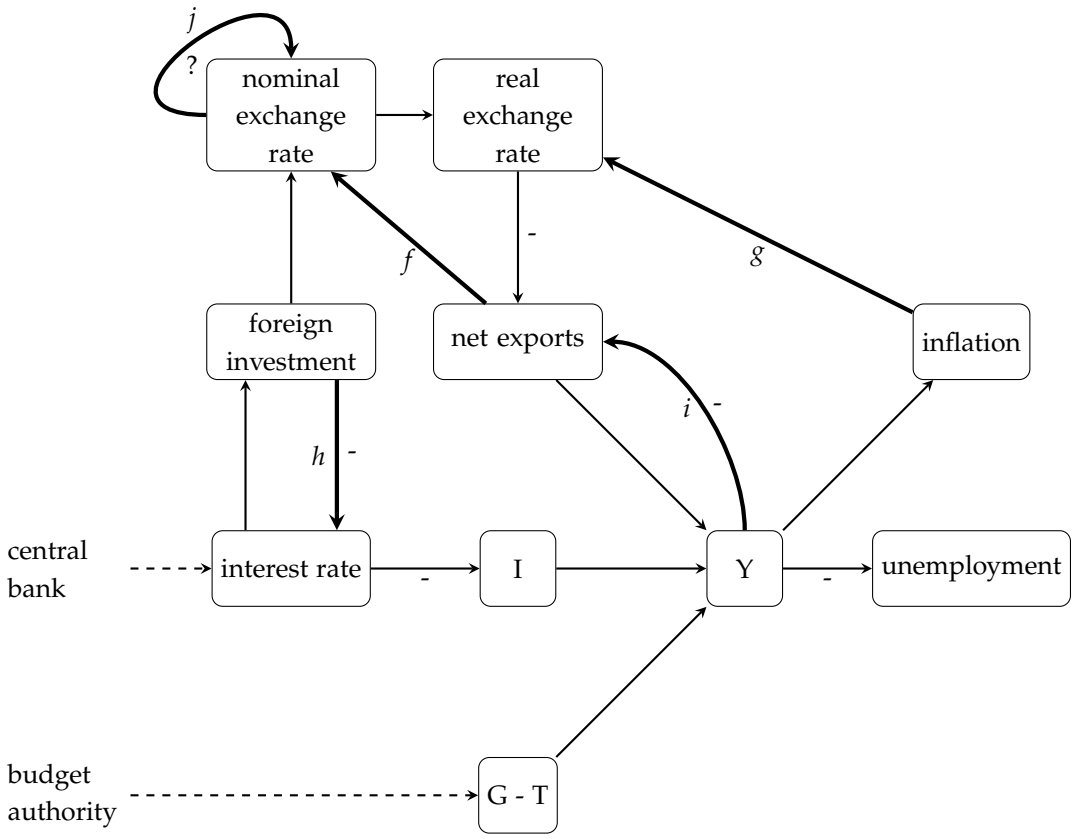
There are a number of feedback mechanisms in the open economy that can limit the effectiveness of macroeconomic policy.

The second flowchart adds five feedbacks, shown in bold.

The currency may appreciate in response to a trade surplus, and depreciate in response to a trade deficit. (f)

Purchases of our goods create demand for our currency, causing it to appreciate. Similarly, our purchases from foreigners reduce demand for our currency relative to ours, causing ours to depreciate. So higher net exports should lead to appreciation, and lower net exports to depreciation. If this effect is strong enough, we should not see persistent trade imbalances; they will automatically correct themselves through the adjustment of nominal exchange rates.

When *floating* exchange rates were first widely adopted in the 1970s, many economists expected them to lead to balanced trade for this reason.



Higher inflation implies a real appreciation of a country's currency. (g)

Faster GDP growth is likely, all else equal, to cause prices to rise more quickly. A depressed economy is likely to experience low inflation or even *deflation*. This will affect the price of the country's goods relative to those produced elsewhere.

Both the response of prices to output, and the response of trade to prices, are quite slow. So this effect can only be important over long horizons – periods of several years or more.

If this effect is strong enough, it can help stabilize output and/or trade flows. In principle, as long as an economy remains depressed long enough, its prices will fall relative to its trade partners. Eventually, this fall in relative prices should boost net exports enough to bring the country back to full employment, without the need for any use of *expansionary* policy. This is sometimes described as an *internal devaluation*, in contrast to a normal devaluation that involves a change in the nominal exchange rate.

In the 18th century, David Hume described a similar mechanism to this in his *specie-flow mechanism*. (Specie is an old word for gold.)

Hume imagined that countries with trade surpluses would gain gold, causing their prices to rise, while countries with trade deficits would lose gold, causing their prices to fall. Eventually these relative price movements would always restore trade to balance, Hume argued, so there was no need for governments to concern themselves with the level of imports and exports.

Inflows of foreign investment reduce domestic interest rates. (h)

If this effect is strong enough, there will be a single "world interest rate," and it will be impossible for national central banks to move their individual countries' rates away from the world rate.

Former Fed Chairman Alan Greenspan has suggested that this is the situation faced by central banks today, even in the United States. But while foreign investment flows may limit the divergence of different countries' interest rates, it is clear there is still some space for it. For example, interest rates on governments bonds in the US are currently about two points higher than rates on similar bonds in Germany.

A fraction of every additional dollar is spent on imports. So faster GDP growth leads to lower net exports. (i)

In addition to relative prices, our imports depend on our income, while our exports depend on incomes in the rest of the world. That means that faster income growth here will tend to reduce net exports (X-M). The degree to which imports respond to changes in income is referred to as the *income elasticity* of imports. The degree to which imports and exports respond to changes in the real exchange rate is referred to as their *price elasticity*.

This relationship is the most important factor in short-run changes in trade flows, and it may be the dominant factor in the long run as well. While income elasticities of imports vary across countries, they seem quite stable for a given country. Differences in income growth seem to explain a large fraction of historical trade shifts.

Some economists combine this observation with the assumption that trade does not respond much to relative prices (either because price elasticities are low, or because real exchange rates do not change much), and the assumption that trade must be balanced in the long run (that is, countries cannot run trade deficits indefinitely). If this is the case, it follows that GDP must adjust to whatever level is consistent with balanced trade. This is the theory of *balance of payments constrained growth*, and it seems to explain a large part of the variation in economic growth across countries.

Fiscal policy is weaker in an open economy than in a closed one, because some stimulus leaks away as imports.

The same link from GDP to imports limits the effectiveness of fiscal policy in an open economy. Since some of each dollar spent goes abroad to pay for imports, less is left to circulate domestically and raise spending here. This means that the *multiplier* will normally be smaller in an economy with a large share of imports, and larger in an economy where the import share is high.This does not mean that stimulus (or austerity) has a lower total effect on GDP in an open economy, it just means that some of that effect shows up in the country's trade partners rather than in the country itself.

In the short run, foreign exchange markets are driven by expectations of future exchange rates. This may be either stabilizing or destabilizing, depending how exchange rates are formed. (j)

In the short run (over periods much less than a year, perhaps several years) exchange rates are determined almost entirely by *speculation* – investors trying to guess how exchange rates will change. That means that whether a currency appreciates or depreciates depends on whether speculators expect it to appreciate or depreciate in the future. This can be stabilizing, if investor expectations are *anchored* – that is, if they have a definite idea of the normal or usual level of exchange rates. In this case, if a currency becomes unusually strong, speculators are likely to expect a depreciation and will sell it, bringing its price back toward the normal level. But expectations can also be *extrapolative*, meaning that when investors see a change in value, they expect that change to go further. In this case, when a currency becomes unusually strong, speculators will expect it to appreciate further, and will buy it – which will ensure that it does in fact appreciate. Since speculators are mainly trying to guess what other speculators will do, many different beliefs about future exchange rates can be "self-confirming" – over short periods, exchange rates will just be at whatever level speculators expect them to be at.

The central role of speculation in foreign exchange markets means that, at least over periods less than a year or so, exchange rates do not respond reliably to any macroeconomic variables. Central banks that wish to control short-run movements in their exchange rates must be able to influence the expectations of market participants. In other words, they must have *credibility*. Over the longer run, interest rates, relative prices and trade flows do influence exchange rates, but it may take several years for these factors to overcome the "noise" of speculation.

Glossary

Glossary

absolute purchasing power parity The theory that over the long run, a representative basket of goods should have the same price in every country, regardless of what currency is used.

accelerator The link from output growth to private investment.

accounting identity An equation that must always be true, because of how the terms are defined.

aggregate A variable measured at the level of the economy as a whole. Common aggregates include GDP, the consumer price index (CPI), and the unemployment rate.

appreciate An increase in value of one currency relative to another one. We can also say it has gotten stronger.

approximation A method of calculating a number that gives something close to the right result, and is easier or more convenient than calculating the number exactly. When an equation is approximation, we use $\approx$ rther than a standard equals sign.

arbitrage Taking advantage of two different prices for the same good or asset by buying it where the price is low and selling it where the price is high.

asset Anything that is owned, has a market value, and will provide some monetary or money-like benefit to the owner in the future. Assets are divided into real assets, like land, buildings, and intangible property like patents and copyrights; and financial assets, which are payments commitments by some other unit.

austerity Contractionary fiscal policy, usually with the goal of reducing the ratio of government debt to GDP.

autonomous Describes a change in spending that is independent of current income.

autonomous spending Spending that does not depend on current income.

balance sheet A record of all the assets and liabilities of an economic unit.

balance of payments The total money payments coming into and out of country. The same term is used to refer both to the difference between payments into the country and payments out; and to the system of accounts that records these payments.

bank run A financial crisis in which fears that banks may fail leads people to withdraw their money, in order not to be left with worthless deposits in a bankrupt bank. Since banks depend on other banks for loans, and since the failure of one bank can raise doubts about the soundness of others, runs often spread from one bank to others.

base money Money directly created by the government or central bank, as opposed to money created by banks.

base year The year for which a price index is defined to be equal to 100. Every index must have a base year, but it makes no difference which year is chosen.

behavioral Describing the choices of the actors within a model. A behavioral equation is one with parameters that have to be estimated on the basis of data. Behavioral equations may be more or less accurate approximations of the phenomena they describe, but they will never be exactly correct.

Beveridge curve A relationship between the number of unemployed workers and the number of job vacancies. When unemployment is high and vacancies are low, unemployment is likely due to deficient demand; when unemployment and vacancies are both high, unemployment is more likely to be structural.

business cycle Periodic shifts in the level of economic activity. Business cycle expansions see high output growth, low unemployment, and high or rising inflation; business cycle downturns or recessions see output growing slowly or falling, high unemployment, and low or falling inflation. Smoothing out business cycles is a central goal of macroeconomic policy.

capital mobility How easy it is to buy and sell assets and lend and borrow money across national borders.

capital gain Profit that comes from an increase in the price of an asset, rather than from the income it generates.

carry trade Borrowing in a currency with low interest rates and lending in a different currency where interest rates are higher.

central bank The bank for other banks, which is responsible for stabilizing the financial system. Almost all modern economies have a public central bank, which is also responsible for managing the level of activity in the economy as a whole.

chartalism The idea that the ultimate source of money's value is that a government accepts it payment of taxes.

closed economy An economy with no trade or financial links to other economies. No economy in reality is perfectly closed (except for the world as a whole), but it is often useful to think about how an economy would behave in isolation.

collateral When a person or business takes out a loan, they may agree to put up an asset they own as collateral. If the borrower fails to repay the loan on schedule, the lender gets the collateral.

commodity money An asset that serves as money based on its intrinsic value. Commodity money consists of some physical asset (often precious metals) that would be desired for its own sake even if it did not function as money.

competitiveness The cost of producing a good in one country compared with the cost of producing similar goods elsewhere. A country will be more competitive if its costs – especially wages – are lower than elsewhere, or if its industries are more productive.

Consumer price index (CPI) An index of the price level. It is supposed to reflect the average price of goods and services consumed by a typical household.

consumption Spending on goods and services that are used directly to meet people's needs. Includes all spending by households on newly produced goods and services (except new houses), as well as spending by nonprofits and government on services used by households.

contractionary Has as its intended or primary effect a reduction in output.

countercyclical Describes any economic policy or relationship that leads to higher spending when demand is weak and lower spending when demand is strong. Countercyclical policy is a deliberate effort to smooth out business cycles by boosting spending (for instance by lower interest rates or lower taxes) in recessions, and holding spending down in booms.

credit channel One way that monetary policy can affect the real economy. Monetary policy that operates through the credit channel affects the availability of loans, rather than interest rates.

credit money The liability of a bank or similar financial institution that functions as money. Credit money includes checking accounts and other deposits that can be used to make payments. Credit money is the main form of money in modern capitalist economies; it is created when banks make loans.

debt deflation A form of economic instability first described by Irving Fisher in the 1930s. A fall in the price level increases the real burden of debt, which forces business and households to cut back spending, which causes the pice level to fall even more.

debt-GDP ratio The ratio of government debt to GDP, or $\frac{debt}{GDP}$. This is the usual way to measure government debt in macroeconomics.

default Failure to make the required payments on a loan. When a borrower defaults, the loan is normally written down or written off by the lender.

deflation Negative inflation, or a decline in the price level.

demand-deficiency unemployment Unemployment due to a lack of demand for goods and services in the economy. Sometimes also called cyclical unemployment.

depreciate An decline in value of one currency relative to another one. We can also say it has gotten weaker.

depreciation The decline in value of real assets like buildings and machinery, whether from wearing out or from obsolescence.

discount window An arrangement by which the central bank provides additional reserves to private banks, in exchange for some other asset.

discretionary Describes policy decisions that are made by the authorities based on their judgement or preferences, rather than according to a fixed rule.

disposable income Income available to households after transfers and tax payments.

durables Consumption goods that will provide a flow of services for many years after they are purchased, such as cars or appliances. While purchases of durables are normally classed with consumption, in some ways they resemble investment spending.

econometric Statistical analysis of economic data.

employment-population ratio The number of people with jobs, divided by the total non-institutionalized population 16 and over.

endogenous A variable that is determined by other variables, as opposed to an exogenous variable that is fixed by policy or by nature.

equilibrium A situation where, given the actions of everyone else, no one wants to change their own actions. Or, a situation that does not have any tendency to change on its own – that will persist until disturbed from outside.

estimation The process of using statistics to determine the parameters of an equation. The goal is to find the parameter values that give the best fit to the observed data on the variables in the equation.

exchange rate index The average of a country's exchange rate against a number of other currencies.

exchange rate The value of one currency in terms of some other currency.

exogenous Determined outside the model. Variables that a model does not try to explain, but simply takes as given.

expansionary Has as its intended or primary effect an increase in output.

expansionary A period of rising output.

expectations People's beliefs about the future, as reflected in current prices and behavior.

factors Labor, capital and others who must be paid for their contributions to production.

federal funds rate The interest rate large banks charge each other for very short-term ("overnight") loans of reserves at the Fed. Between the 1980s and 2008, changes in the federal fund rate were the main tool for monetary policy in the United States.

fiat money An asset that serves as money only because a government has declared it legal tender. Fiat money consists of tokens (paper bills, coins, etc.) that have no intrinsic value, and that can be legally created only by the government.

final goods Newly produced goods purchased to be used by the purchaser, as opposed to goods purchased to be resold or used as inputs to make something else. Includes all spending by households on new goods and services (including houses), investment spending by businesses, and spending by government on the direct provision of public services.

financial asset An asset like a stock, bond, or loan that does not involve ownership of any concrete object, but instead is a promise of future payment by someone else.

financial stability Sustainable growth in asset prices and private debt. Financial stability is increasingly seen as an important target for macroeconomic policy, but there is no agreement on how to measure it.

fiscal multiplier The multiplier applied to changes in government spending.

fiscal policy The use of government spending and/or taxes as a tool to change the level of output.

fiscal balance The difference between government revenue and government spending. If revenue is greater than spending, the fiscal balance is positive and we say the government has a budget surplus. If revenue is less than spending, the fiscal balance is negative and we say the government has a budget deficit. If revenue is exactly equal to spending, the fiscal balance is zero and we say the government has a balanced budget.

fixed exchange rate An exchange rate whose value is set by a government rather than in foreign exchange markets.

fixed investment Production of new buildings, machinery or other lasting means of production. Includes all investment except for inventory investment.

floating Describes an exchange rate that is determined by private trading in the foreign exchange markets. The alternative is a fixed exchange rate, which is set by the government.

foreign exchange reserves Money of a foreign country, held by a central bank in order to manage the exchange rate, to have available in a crisis, or for some other purpose.

foreign exchange The money of a country other than one's own. The foreign exchange market is the activity – mostly carried out by large financial institutions – of trading one currency for another.

frictional unemployment Unemployment that results from normal transitions between jobs or in and out of the laborforce.

full employment The level of employment or, more often, unemployment targeted by macroeconomic policy. In the US today, full employment is often considered to be equivalent to an official unemployment rate of around 5 percent.

functional distribution of income The distribution of income among the different factors of production – usually this means labor and capital, but it may sometimes be extended to include other factors like

land. The most common measure of the functional distribution is the share of labor income – wages, salaries and benefits – in total income.

GDP deflator A price index used to convert nominal Gross Domestic Product (GDP) to real GDP. It includes all goods and services that are counted in GDP, as opposed to the goods and services consumed by a typical household which are used for CPI.

geometric mean An average of numbers that are multiplied together. (The more familiar arithmetic mean is the average of numbers that are added.) A common use of the geometric mean is to convert between a total change over some period and the average annual rate of change during the period. If x_1 is the value at the beginning of the period and x_2 is the value at the end, the average annual rate of change is $(x_2/x_1)^n - 1$, where n is the number of years in the period.

gold standard A monetary regime in which the value of the currency is irrevocably set a fixed quantity of gold; the central bank or some other government authority commits to freely buying or selling gold at the official price in any amount required; paper money is backed by gold; and bank lending is strictly limited by the gold reserves available. From the mid-19th century until World War I, most of the world's countries tried to adhere to the gold standard .

Goodwin cycle One of several possible sources of instability in a capitalist economy. In a Goodwin cycle, high growth leads to lower unemployment, which causes wages to rise at the expense of profits, which reduces investment and brings growth back down.

gross A number from which something has not been subtracted. For example, gross income for a person means income before taxes are subtracted, gross domestic product of a country means that depreciation is not subtracted, and so on. What a particular gross figure is gross of depends on context.

gross national product (GNP) An alternative measure of total output of an economy. It is defined as final goods produced for the market by the labor and capital of a country, regardless of where production takes place.

gross domestic product (GDP) The most common measure of total output of an economy. It is defined as final goods produced for the market within the borders of the country in a given period.

household People when they are acting on their own behalf, rather than as part of businesses or governments. A household may be

an individual or a family or other group of people who pool their incomes and make decisions about earning and spending together.

hyperinflation An extremely high rate of inflation. There is no exact cutoff, but most people would consider inflation to become hyperinflation when it is measured in hundreds or thousands of percent per year.

imputation A variable in the national accounts that can't be measured directly, but has to be estimated based other variables.

inflation The average change in prices of goods and services in an economy. It is measured as the annual percentage change in a price index. Negative inflation is called deflation.

inflationary bias A tendency of policymakers to allow inflation to rise higher than is socially optimal.

inflation-targeting Describes a macroeconomic policymaker, usually a central bank, whose only goal is to keep inflation at a certain level. An inflation-targeting central bank looks at other macroeconomic targets only insofar as they are thought to affect the inflation rate.

instrument A variable that is directly under the control of policymakers and is adjusted in order to affect other macroeconomic outcomes.

intellectual property (IP) Patents, copyrights, and similar legal claims on creative works and scientific discoveries. In the national accounts, creation of new IP is counted as a form of investment.

interest parity The theory that arbitrage across countries will result in expected returns being the same on similar bonds in all countries. In some versions, differences in interest rates should reflect expected exchange rate changes.

interest elasticity How responsive output (or some other variable) is to changes in the interest rate. For example, if we say output has an interest elasticity of 0.5, that means we think that a 1 point fall in the interest rate will cause output to rise by 0.5 points.

interest on reserves Interest on reserves, or interest on excess reserves (IOER) is interest paid by the central bank to private banks on their holdings of reserves. Historically, central banks did not normally pay interest on reserves but since 2008 it has been introduced as an instrument of monetary policy.

inventories Unsold finished goods, goods in process, and stocks of raw materials. In the national accounts, the change in inventories is counted as a form of investment.

investment The production of new long-lived means of production like buildings, machines, software, and so on. Unlike in everyday use, "investment" in macroeconomics does not include the purchase of existing real or financial assets.

J-curve Describes the tendency for a depreciation of the currency to be followed first by a move toward trade deficit, then by a larger move toward trade surplus.

Keynesian economics A school of macroeconomics that emphasizes: the determination of output by aggregate demand rather than the productive capacity of the economy; the role of money and finance in shaping economic outcomes; the uncertainty of the future; and the inherent instability of the economy, which must be managed by government.

labor share The fraction of output going to workers, calculated as total wages and salaries plus benefits divided by total income. Also called the wage share.

labor productivity Total output divided by total employment. The most common measure of productivity.

labor market slack How easy it is for employers to find new workers.

lag The amount of time it takes a change in some policy instrument or other exogenous variable to produce its effects on other economic variables.

leakage Uses of income that do not contribute to aggregate demand, and do not directly create income for other units in the economy. The most important leakages are saving, imports and tax payments.

legal tender A form of money that by law must be accepted in payment of a debt. Legal tender does not have to be accepted as payment for goods and services.

lender of last resort An institution, usually a central bank, that prevents bank runs by lending to banks in a crisis.

liability A binding commitment to make some payment in the future. Every liability is a financial asset for some other unit. The most common form of liability is a debt, which is an asset for the lender.

linear equation An equation in which the terms are only added or subtracted. None of the variables are multiplied or divided, and none have exponents.

liquidity The degree to which an asset can be used to make payments reliably and at short notice.

liquidity-constrained Describes a household, business or other economic unit that is spending less than it would otherwise choose to because of a lack of current access to cash or credit. This is distinct from spending that is low because of low income or wealth.

marginal tax rate The fraction of each additional dollar of income that is taken in taxes.

marginal propensity to consume The fraction of each additional dollar of income that is spent on consumption.

markup The difference between the marginal cost of producing a good and its final price.

medium of exchange A good or asset that is used in purchases of goods and services – instead of being traded directly for each other, goods and services are traded only for the medium of exchange. One of the functions of money.

monetarism A school of economic thought that believes that there is a fixed quantity of money. set by the central bank and that this quantity of money is tightly linked to aggregate demand and inflation. For monetarists, instability in the economy comes mainly from too fast or too slow growth of the supply of money, so the most important goal for macroeconomic policy is ensure steady money growth.

monetary policy Actions taken by the central bank to change the level of output or other macroeconomic outcomes. Often consists of changing a single short-term interest rate (the "policy rate", or in the US, the federal funds rate) but can also include all kinds of decisions by the central bank that affect the price or availability of credit.

moral hazard The danger that by protecting people or businesses from bad outcomes, the authorities will encourage carelessness or excessive risk-taking in the future.

multiplier The relationship between a change in investment, government spending, or other autonomous expenditure, and the change in output that results from it. Mathematically, the multiplier can be expressed as $\frac{\Delta Y}{\Delta A}$ where Y is output, A is autonomous expenditure, and Δ means change.

national income identity A fundamental accounting identity that says that total output equals the sum of consumption, investment, final government spending and net exports: $Y = C + I + G + (X - M)$.

net A number from which something has been subtracted. For example, net income for a business means revenue after costs are subtracted, net exports of a country means exports after imports are subtracted, and so on. What a particular net figure is net of depends on context.

net worth A unit's total assets minus its financial liabilities. Net worth is treated as a liability on balance sheets, ensuring that total assets and total liabilities are always equal.

nominal Measured in units of money, not adjusted for inflation. Prices and many other numbers in economics are normally measured in money. If we try to adjust a number for changes in the value of money, that gives us a "real" figure. If we don't make any such adjustment but simply use the money value as is, that is a nominal figure.

Non-accelerating inflation rate of unemployment (NAIRU) The rate of unemployment at which inflation neither rises nor falls. Sometimes referred to as the "natural unemployment rate." Whether there is a unique NAIRU is debated by macroeconomists; many believe that stable inflation is possible with many different rates of unemployment.

Okun's law An empirical law in economics that says the change in unemployment ΔU is connected to the real growth of output g by a relationship of the form $\Delta U = -a(g - b)$. For the US, a is around 0.6 and b is around 2.

open market operations One of the main tools through which monetary policy is conducted. When the central bank wishes to raise interest rates, it sells government securities for reserves, reducing the quantity of reserves available for banks to borrow. When it wishes to lower interest rates, it buys securities, increasing the quantity of reserves available.

open economy An economy connected by trade or financial links to other economies. In reality every economy (except for the world as a whole) is at least somewhat open; we use the term "open economy" to mean cases where the links to the external world are important.

output Total production of goods and services in an economy.

parameters Numbers in an equation that describe the relationships between the variables.

Personal distribution of income The distribution of income among households. There are a number of different measures of personal distribution, which describe in different ways the share of income going to high, middle and low-income households.

Phillips curve One of various relationships between the level or growth rate of output or unemployment on the one hand, and wages or prices on the other. In general, Phillips curves describe how inflation will be higher when the economy is above potential, and lower when it is below.

policy rate The interest rate that is controlled most directly by the central bank, with the hope that other interest rates will move with it. Between the 1980s and 2008, the policy rate in the US was the Federal funds rate.

policy rule A strict rule that is supposed to guide central bank decisions about interest rate policy. In the US, the Federal Reserve is sometimes described as following a policy rule called the "Taylor rule".

potential output The level of output targeted by macroeconomic policy. It is assumed to be the maximum the economy can produce without "overheating" – that is, without rising inflation, shortages of raw materials, etc. Since potential output cannot be directly measured, it may be estimated based either on the level of observable aggregates like unemployment or inflation, or on the long-run trend of output growth.

price stability The level of inflation desired by policymakers. Historically, price stability meant literally stable prices, that is, zero inflation on average; but since the 1990s, it has generally been understood as a low but positive level of inflation – 2% in the US and most other rich countries.

price indexing Automatically adjusting some ongoing payment for inflation, so that its real value is constant over time.

price level The average price of goods and services at a given time and place, as measured by a price index.

price index A measure of the average price of goods and services at a given time and place. If the price index is 1 percent higher in one year than another, that means the price of the "typical" good is 1 percent higher in that year. Since prices don't all change together, a given price index is defined only for a particular basket of goods.

principal The principal of a loan is the amount the borrower actually receives when the loan is incurred.

private balance The difference between private saving and private investment.

quantitative Measured or estimated numerically. The opposite is qualitative, which refers to something that can be described but not given as a number.

quantitative easing A form of unconventional monetary policy in which a central bank buys large amounts of long-term government debt or other longer maturity assets. The goal is to reduce longer interest rates. It is called "quantitative" because the central bank normally picks a quantity of assets to buy, rather than announcing a target interest rate as in conventional monetary policy.

random walk A statistic whose future changes cannot be predicted from its current or past values. If a variable follows a random walk, then the best guess for its future value is always its current value, whatever that may be.

real Economists describe a number or variable as "real" if it has been adjusted for inflation.

real exchange rate The price of a basket of goods in one country relative to the same basket in a different country. Or equivalently, the nominal exchange rate adjusted for inflation.

recession A period in which economic activity is declining. It is widely believed that a recession is defined as two consecutive quarters of declines in real GDP. But in fact, in the US recession dates are determined by the National Bureau of Economic Research based on a variety of economic indicators.

regression A statistical technique for describing the relationship between some variables. The goal is to minimize 'errors" – differences between the actual data and the estimated relationship. .

relative purchasing power parity The theory that differences between the price of a representative basket of goods in different countries depend on economic fundamentals and therefore should be stable over time. Equivalent to claim that real exchange rates will be constant.

repurchase agreements Repurchase agreements, also called "RPs" or "repos", are a tool the central bank uses to make loans to banks and other financial institutions. In effect they are loans, but they

are structured as the sale of a security, combined with an agreement to buy it back in a certain amount of time. In a reverse repo or "RRP?, the Fed borrows money from primary dealers.

reserves Liabilities of the central bank used by other banks to make payments to each other. Reserves are also used by the central bank to buy assets from private banks. In some banking systems, banks are required by law to hold a certain amount of reserves.

retained earnings Profits that are kept by the business that earned them, rather than paid out to shareholders. Retained earnings are an important form of saving in the economy. Historically, corporations have paid out about half their profits and retained about half.

settlement asset Banks periodically must settle any outstanding balances among themselves. Banks that have lost deposits on net must make a payment to banks that have gained deposits. An asset used to make these payments between banks is called a settlement asset.

solvency Having a net worth greater than zero. An economic unit is solvent if, over time, its income will be sufficient to meet its liabilities. When a unit is having trouble making required payments, it's important to know if that is because it is insolvent, or because it is illiquid An illiquid unit has assets worth more than its liabilities, but cannot convert them to means of payment immediately.

speculation Buying an asset in order to resell it later at a higher price, rather than to use it or get an income from it.

stimulus Expansionary fiscal policy.

stock In economics, a stock is anything that can be measured as a quantity at a moment in time. Wealth, population, and total employment are examples of stocks. Flows are things that can be measured only over a period of time, such as income or GDP.

structural unemployment Unemployment due to a mismatch between workers and the available jobs, such as a lack of appropriate skills or being located in different parts of the country.

supply constraints Limits on the productive capacity of the economy due to natural resources, capital, technology, worker skills, or other "real" factors.

target An outcome that policymakers wish to influence. The most important macroeconomic targets are output, inflation, unemployment, the government debt ratio, the balance of payments, income distribution, and financial stability.

Taylor rule A formula supposedly followed by the Federal Reserve in setting interest rates. It says the bank should raise interest rates when unemployment falls below its target level or inflation rises above it, and should give the two targets equal weight.

technological unemployment Unemployment that results from labor productivity rising faster than total output, so that fewer workers are needed.

Tinbergen's rule The principle that to hit a certain number of independent macroeconomic targets, the authorities must have at least that many different policy instruments.

total factor productivity Total output divided by the labor and capital used. Important in economic theory but hard to apply in the real world.

trade balance The difference between a country's exports and its imports. If exports are greater than imports, it has a trade surplus; if exports are less than imports, it has a trade deficit.

transaction costs Costs of carrying out a sale or exchange. They include any fees, taxes or payments to third parties associated with the sale; the costs of finding the two parties, bringing them together, and transferring the good between them; and any adverse price changes that result from a purchase or sale.

transfers Payments that are made without any good or service being received in return. Transfers include payments through government programs like Social Security and unemployment insurance, as well as private gifts.

Treasury securities The official name for the financial instruments that make up government debt. Treasury bills are the shortest-maturity debt, Treasury bonds are the longest maturity, and Treasury notes are in between.

U6 An alternative measure of unemployment that includes everyone who is unemployed by the official definition, plus discouraged and marginally attached and involuntary part-time workers.

underemployment A situation in which people have jobs that do not make full use of their skills or productive potential. Also called disguised unemployment.

unemployment rate The fraction of the laborforce unable to find work. In the US it is normally measured by U-3 – the fraction of the civilian, noninstitutionalized population 16 and older who have

zero hours of paid employment and are actively looking for work. But other measures exist.

unit of account The good or asset that is the standard by which the value of other goods and assets is measured. One of the functions of money.

Verdoorn's law A relationship between output or unemployment and productivity growth: When unemployment is low and output is high, productivity tends to rise faster.

wage curve A relationship between unemployment wages: When unemployment is low, wages rise more quickly; when unemployment is high, wages rise more slowly or fall.

Wicksell cycle A form of instability in capitalist economies first described by the Swedish economist Knut Wicksell, who called it the "cumulative process." It describes a situation where high inflation leads to lower real interest rates which lead to even higher inflation, or where low inflation or deflation leads to higher real interest rates which lead to even more deflation.

writeoff A decision by a creditor to cancel or reduce an outstanding debt. Debt may be written off when the debtor defaults on the loan or declares bankruptcy, or (especially in the case of sovereign governments) it may result from a negotiated agreement.

yield The income received by the owner of an asset – interest on a bond, dividends from a stock, rent from real estate, and so on. The total return on an asset is the yield plus capital gains.

zero lower bound A practical limit on monetary policy: conventional tools like open market operations cannot move the policy rate below zero.