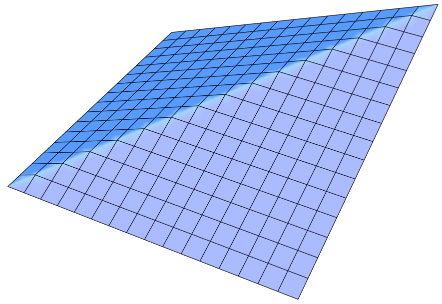

ECON 202: Macroeconomic Theory I

Prof. Fiona Rahman • Winter 2016 • University of Waterloo
TR 5:00–6:20PM in RCH 211

Prepared by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is `https://rob-zimmerman.github.io/coursenotes`.

Contents

1. National Accounting, Production, and Saving

Macroeconomic Measurement

Gross domestic product, or **GDP**, is the market value of final production within an economy's borders during a specified period. The same economic activity can also be measured as total expenditure on that production or as the total income generated by it. Total production, total expenditure, and total income are therefore three ways of measuring the same thing.

1. Under the **production approach**, GDP is the market value of all final goods and services newly produced within a fixed period of time.

   The use of **market value** means that production is valued at market prices. In principle, GDP is a sum of values of the form
   $$\sum_{i=1}^{n} p_i q_i.$$
   Consequently, nonmarket goods and services are generally excluded. For example, unpaid household work and activity in the underground economy are not fully captured in measured GDP. There are important exceptions in which statisticians impute a value instead of observing a market transaction directly. A standard example is owner occupied housing: homeowners are treated as though they pay rent to themselves for the housing services they consume, and those services are included at an estimated value.

   Only **final goods and services** are included. Intermediate goods and services are used up in the production of something else, so counting them separately would count the same production twice. For example, the processor installed in a computer and the transportation services used to deliver it are not counted again after their values have been incorporated into the computer's retail price. The **value added approach** avoids double counting by adding the value created at each stage of production; the resulting total equals the final selling price. Newly produced capital goods are final goods, and inventory investment is also included. Because inventory investment is the change in inventories during the period, an increase in inventories counts as an increase in economic activity.

   The requirement that goods and services be **newly produced** excludes the resale of existing goods. The sale of a used car does not represent new automobile production and therefore does not enter GDP. Services provided by a dealership in arranging the sale do count because those services are newly produced.

   A **fixed period of time** makes GDP a flow rather than a stock. A flow is measured per unit of time, whereas a stock is measured at a point in time. GDP is usually reported for a quarter or a year.

2. The **expenditure approach** measures GDP by adding spending on domestically produced final goods and services. It gives the national income identity
   $$Y = C + I + G + NX, \tag{0.1}$$
   where C is consumption, I is investment, G is government purchases, and NX is net exports.

   Consumption expenditure is the largest component of GDP, accounting for roughly 70% of total expenditure. It consists of household spending on final goods and services in three categories:

(a) consumer durables, such as cars, appliances, and computers;

(b) nondurable goods, such as food and gasoline, which are used up more quickly; and

(c) services, including housing services.

The purchase of a house is not consumption. Instead, the flow of housing services provided by the house is included in consumption.

In macroeconomics, **investment** means spending on currently produced capital goods that will help produce other goods and services in the future. It accounts for roughly 15% of GDP and includes

(a) business fixed investment, such as equipment and structures;

(b) inventory investment; and

(c) residential investment, which includes only newly built houses and apartments.

> **Note:** Economic investment is the purchase of new physical capital, not the purchase of financial assets such as stocks or bonds.

Government purchases, denoted by G, account for roughly 20% of GDP. They include spending by all levels of government. In practice, state and local governments together often spend more than the federal government. Government purchases include government consumption, such as health care and policing, and government investment in capital goods. Transfer payments, such as social insurance, employment insurance, and Medicare, are excluded because they are not payments for newly produced goods and services. Interest payments on government debt are excluded for the same reason.

Net exports are exports minus imports:

$$NX = EX - IM. \tag{0.2}$$

Imports are subtracted because C, I, and G include spending on domestic and foreign goods, whereas GDP measures only domestic production. Thus,

$$\begin{aligned} Y = {} & (C_{\text{dom}} + C_{\text{im}}) + (I_{\text{dom}} + I_{\text{im}}) + (G_{\text{dom}} + G_{\text{im}}) + EX \\ & - \underbrace{(C_{\text{im}} + I_{\text{im}} + G_{\text{im}})}_{IM}. \end{aligned}$$

Net exports are also called the **trade balance**. A country has a **trade deficit** when $NX < 0$.

The relative importance of the expenditure components differs across countries, but several patterns are common. In the United States, consumption has grown steadily as a share of GDP. Investment is smaller but much more volatile, so it plays a disproportionately important role in business cycle fluctuations. Government purchases are fairly stable as a share of GDP, and net exports are usually negative. China has historically had a lower consumption share and a higher investment share than the United States, while Canada has tended to resemble the United States more closely.

3. The **income approach** measures GDP by adding the incomes generated in domestic production. The first four items below are domestic income categories; the final item is a separate adjustment used to move from GDP to GNP.

   (a) Employee compensation, including wages, salaries, and benefits, is the largest category and accounts for more than half of total income.

   (b) Other income includes the income of self-employed workers, rents and royalties, net interest, and sales taxes added to business income. This category is roughly 20% of the total.

   (c) Corporate profits account for about 12%.

   (d) Depreciation reflects wear, tear, and obsolescence of capital. Because firms deduct depreciation when calculating profits, it must be added back when moving to a gross measure of income. It accounts for about 15% of the total.

   (e) Net factor income from abroad is the income that domestic residents earn abroad minus the income that foreign residents earn domestically. It is not part of GDP, which follows the location of production rather than the residence of the income recipient. It is generally small relative to GDP and may be negative.

   Employee compensation, other income, and corporate profits are net income components. Adding depreciation converts a net measure into a gross measure. Gross national product, or GNP, then adjusts GDP for net factor income from abroad:

$$\text{GNP} = \text{GDP} + \text{net factor income from abroad.}$$

GDP is a measure of production, not a complete measure of human development or social welfare. A country can have high GDP and still perform poorly in literacy, life expectancy, political freedom, or physical safety.

The distinction between stocks and flows is important throughout macroeconomics. A bathtub provides a useful analogy: the amount of water currently in the tub is a stock, while the water entering through the tap or leaving through the drain is a flow. Many economic stocks are accumulated through flows over time.

Private disposable income is the income available to the private sector for spending or saving.

It is

$$\text{private disposable income} = \text{GDP} + \text{net factor income from abroad} + \text{government transfers} \\ + \text{interest on government debt} - \text{taxes}.$$

Interest on government debt is included because it is paid to individuals and other private holders of government bonds.

Net government income is

$$\text{net government income} = \text{taxes} - \text{transfers} - \text{interest payments on debt}.$$

It follows that

$$\text{private disposable income} + \text{net government income} = \text{GNP}.$$

A central measurement problem is to separate changes in economic activity from changes in the price level. **Nominal GDP** values output at current market prices, so it rises when production increases or when prices rise. **Real GDP** removes the effect of price changes by valuing quantities at prices from a base year. In shorthand, nominal GDP is the price level times real GDP, while real GDP is constructed from prices in the base year and current quantities.

Economic data are often **seasonally adjusted** because production varies predictably with the season. For example, some forms of production decline during colder months. In **chain-weighted** measures, the reference prices are updated continuously rather than fixed forever. We can, for example, use average prices from 2014 and 2015 to calculate real GDP growth across that interval. This method reduces distortions that arise when the prices of particular goods move very differently from the general price level.

Remark 0.1 The real expenditure components in chain-weighted data do not necessarily add exactly to real GDP. Nominal data are therefore safer when the objective is to study expenditure shares.

The distinction between nominal and real GDP gives several measures of the price level and inflation.

1. The **GDP deflator**, also called the implicit price deflator, measures the price level for

domestically produced final goods and services. In year t,

$$\text{GDP deflator}_t = 100 \times \frac{\text{nominal GDP in year } t}{\text{real GDP in year } t\text{, measured in dollars from the base year}}.$$

The GDP deflator equals 100 in the base year.

Example 0.2 Suppose nominal GDP in 2015 is \$15 trillion and real GDP in 2015, measured in 2005 dollars, is \$12 trillion. Calculate the GDP deflator and interpret it.

Solution. The GDP deflator is

$$100 \times \frac{15}{12} = 125.$$

Thus, the overall price level is 25% higher than in the base year. □

2. The **personal consumption expenditure deflator** is defined similarly, but applies only to the consumption component C of GDP.

3. The **consumer price index**, or CPI, is often described as a cost of living index. It measures the average price of a basket of goods and services purchased by a typical urban consumer. If the basket from the base year contains 2 apples and 10 litres of gasoline, then its value in 2015 relative to its value in 2005 is

$$\text{CPI}_{2015} = 100 \times \frac{10(\text{gas price in 2015}) + 2(\text{apple price in 2015})}{10(\text{gas price in 2005}) + 2(\text{apple price in 2005})}.$$

This normalization sets the CPI equal to 100 in the base year.

The CPI can overstate the true increase in the cost of living, which matters for indexed payments and for monetary and fiscal policy, for three reasons:

(a) Consumers substitute away from goods that become relatively expensive.

(b) A higher price may partly reflect improved quality rather than pure inflation.

(c) New goods may reduce the true cost of attaining a given standard of living before they are fully incorporated into the basket.

The **inflation rate** is the percentage change in the price level:

$$\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}}. \tag{0.3}$$

Different price indices need not produce exactly the same inflation rate, although their measures are usually close in practice.

For small changes, the percentage change in a product is approximately the sum of the percentage changes in its factors:

$$\%\Delta(xy) \approx \%\Delta x + \%\Delta y.$$

Consequently, nominal GDP growth is approximately real GDP growth plus inflation.

The civilian population is divided into three labour market categories:

1. The **employed**, denoted by E, are working full time or part time, or are temporarily absent from work because of illness or vacation.
2. The **unemployed**, denoted by U, did not work during the preceding week and either looked for work during the preceding four weeks or are waiting to return from a layoff.
3. People **not in the labour force** did not work during the preceding week and have not recently looked for work. This group includes discouraged workers, who would like to work but have stopped searching, as well as people who are out of the labour force voluntarily, such as retirees, students, and homemakers.

The labour force is $U + E$, so the **unemployment rate** is

$$\frac{U}{U+E}.$$

This standard measure excludes discouraged workers and does not fully capture underemployment, such as part-time workers who would prefer full-time work. Broader measures, such as U-6 in the United States, are therefore often substantially higher.

The **labour force participation rate** is the labour force divided by the adult population, while the **employment ratio** is employment divided by the adult population. These measures are typically based on a **household survey** of about 60,000 households. The household survey includes self-employed workers and counts workers rather than jobs. By contrast, the **establishment survey** samples about 150,000 businesses, excludes the self-employed, and counts jobs rather than workers.

An **interest rate** is the cost of borrowing, or equivalently the rental price of funds. Macroeconomists often speak of "the" interest rate because many market rates move together. Important examples include the prime rate, the overnight lending rate or federal funds rate, Treasury bill rates, ten-year Treasury bond rates, and mortgage rates; the London Interbank Offered Rate was another important benchmark when the course was taught.

The **ex ante real interest rate** is the increase in purchasing power that a lender expects to receive after accounting for expected inflation. The approximate **Fisher equation** is

$$i = r + \pi^e, \tag{0.4}$$

where i is the nominal interest rate, r is the real interest rate, and π^e is expected inflation. The exact relationship is

$$1 + i = (1 + r)(1 + \pi^e). \tag{0.5}$$

A low real interest rate strengthens the incentive to borrow and invest while weakening the incentive to lend. Nominal and real rates need not move together. In the United States during the 1970s, for example, nominal rates were high but real rates were often negative because inflation was even higher.

Aggregate Production and Productivity

Real GDP is determined by the inputs available to the economy and by the technology used to transform those inputs into output. A **production function** summarizes this relationship:

$$Y = F(K, L), \tag{0.6}$$

where Y is real GDP, K is capital, and L is labour. Labour is measured by the number of workers, with hours per worker treated as fixed for now, while capital is the real value of the economy's capital stock. In this long-run framework, the supplies of both factors are fixed at $\bar{K}$ and $\bar{L}$ and are assumed to be fully employed. Over long periods, labour and capital have often received approximately 70% and 30% of national income, respectively.

A common specification is the **Cobb–Douglas production function**

$$Y = AK^{0.3}L^{0.7}, \tag{0.7}$$

where A is **total factor productivity**, sometimes called the **Solow residual**. Total factor productivity measures how efficiently capital and labour are used together. Once Y, K, and L have been measured, A can be inferred from the data.

Public discussion often focuses on labour productivity alone, but output per worker depends on both productive efficiency and the capital available to each worker. If $y = Y/L$ and $k = K/L$, then dividing (0.7) by L gives

$$y = A\left(\frac{K}{L}\right)^{0.3} = Ak^{0.3}.$$

Thus, output per worker depends on productivity A and capital per worker k. Sustained increases in living standards usually depend more on rising productivity than on continually raising capital per worker.

The Cobb–Douglas specification has two important properties:

1. It has **constant returns to scale**. If capital and labour both rise by the same percentage, output rises by that percentage. In particular,

$$F(2K, 2L) = A(2K)^{0.3}(2L)^{0.7} = 2F(K, L).$$

2. It has **diminishing marginal products**. When one factor increases while the other is fixed, the additional output produced by another unit of that factor declines. The marginal products are

$$\text{MPK} = 0.3\frac{Y}{K} \qquad \text{and} \qquad \text{MPL} = 0.7\frac{Y}{L}.$$

The concave curve in Figure 1 illustrates the diminishing marginal product of capital when labour and productivity are fixed.

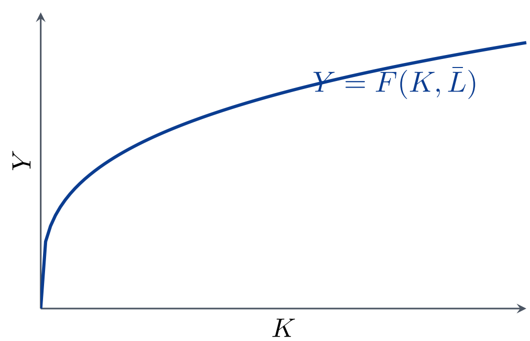

FIGURE 1

A **supply shock** changes productivity A while the quantities of capital and labour remain fixed. Improvements such as faster computers or better software are positive technology shocks. Natural events such as blizzards and earthquakes are negative shocks. Because this reduced-form production function omits energy as a separate input, changes in its price or availability can also appear as supply shocks.

A negative supply shock shifts the production function downward, as shown in Figure 2. Output falls at every level of capital, and both the marginal product of labour and the marginal product of capital decline. A positive supply shock has the opposite effects.

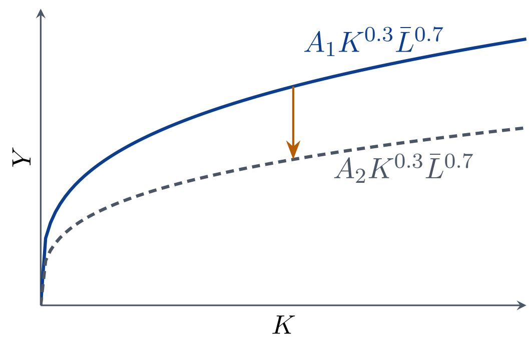

FIGURE 2

To determine factor prices, assume that firms are perfectly competitive and take prices as given. Workers receive the market wage, capital is rented at the market rental price, everyone who wants to work at the prevailing wage can find work, and all available capital is used. A firm that employs more labour or capital can produce more output, but its production costs also rise.

If P is the price of output, R is the nominal rental price of capital, and W is the nominal wage, nominal economic profit is

$$P \underbrace{F(K,L)}_{Y} - RK - WL.$$

Dividing by the price level gives real profit:

$$\Pi = F(K,L) - \frac{R}{P}K - \frac{W}{P}L.$$

Define the real rental price of capital by $r_c = R/P$ and the real wage by $w = W/P$. Then

$$\Pi = F(K,L) - r_c K - wL. \tag{0.8}$$

Notation 0.3 The symbol r_c denotes the real rental price of capital, while r denotes the real interest rate.

A profit-maximizing firm with a differentiable, concave production function employs each factor until its marginal product equals its real price, provided the optimum is interior:

$$\text{MPK} = r_c \qquad \text{and} \qquad \text{MPL} = w. \tag{0.9}$$

The firm adds capital as long as the extra output produced by another unit is worth at least its real rental cost. The same reasoning applies to labour.

Figure 3 draws this factor-market argument for capital. The marginal product of capital is the demand curve D^K, while the fixed long-run capital stock $\bar{K}$ gives a vertical supply curve S^K.

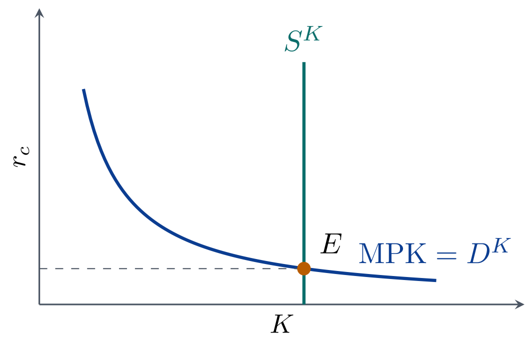

FIGURE 3

The equilibrium rental price is r_c^*. At any higher rental price, the fixed capital stock exceeds the quantity firms want to rent, so excess supply pushes the rental price downward. The labour market works in the same way: the marginal product of labour is the demand curve for labour, and the real wage moves toward w^*. The supply curves for labour and capital are vertical because both factor supplies are fixed in this long-run model.

The Cobb–Douglas exponents also determine the distribution of income. Real labour income is

$$wL = \text{MPL} \cdot L = \left(0.7\frac{Y}{L}\right) L = 0.7Y,$$

while real capital income is

$$r_c K = \text{MPK} \cdot K = \left(0.3\frac{Y}{K}\right) K = 0.3Y.$$

Together, labour income and capital income exhaust national income. Total output can therefore change while the shares received by labour and capital remain approximately constant, a prediction broadly consistent with long-run data.

Real wages are closely tied to labour productivity because

$$w = \text{MPL} = 0.7\frac{Y}{L}.$$

Labour productivity growth slowed during the 1970s, possibly because energy prices rose, and accelerated after about 1995, possibly because of computers and information technology. A sudden increase in oil prices is a negative supply shock: it lowers both marginal products, shifts factor demand curves downward, and reduces the real wage and the real rental price of capital. If common stock is a claim on capital income, stock prices may fall as well.

Example 0.4 The production function is $Y = AK^{0.3}L^{0.7}$. Mexico has output $Y = 1 \times 10^{12}$, capital $K = 0.18 \times 10^{12}$, and labour $L = 105 \times 10^6$. Spain has output $Y = 2 \times 10^{12}$, capital $K = 0.76 \times 10^{12}$, and labour $L = 50 \times 10^6$. Calculate total factor productivity in each country.

Solution. Solving (0.7) for total factor productivity gives

$$A = \frac{Y}{K^{0.3}L^{0.7}}.$$

For Mexico,

$$\begin{aligned}
A_{\text{Mexico}} &= \frac{1 \times 10^{12}}{(0.18 \times 10^{12})^{0.3}(105 \times 10^6)^{0.7}} \\
&= \frac{1}{0.18^{0.3}105^{0.7}} \frac{10^{12}}{10^{3.6+4.2}} \\
&= \frac{1}{(0.59783)(25.99157)} 10^{4.2} \\
&= 0.064356 \times 10^{4.2} \\
&\approx 1019.966.
\end{aligned}$$

For Spain,

$$\begin{aligned}
A_{\text{Spain}} &= \frac{2 \times 10^{12}}{(0.76 \times 10^{12})^{0.3}(50 \times 10^6)^{0.7}} \\
&= \frac{2}{0.76^{0.3}50^{0.7}} 10^{4.2} \\
&\approx 0.14045 \times 10^{4.2} \\
&\approx 2225.903.
\end{aligned}$$

Spain's measured total factor productivity is therefore a little more than twice Mexico's in this calculation. □

Saving and Investment

Wealth is the value of assets minus liabilities at a point in time. Because wealth is measured at a point in time, it is a stock. Saving, by contrast, is a flow that changes wealth over time.

1. **Private saving** is private disposable income minus consumption. Define net taxes by

   $$T = \text{taxes} - \text{transfers to individuals} - \text{interest payments on government debt}.$$

   Interest payments on government bonds are subtracted because they are paid to the private sector. Private disposable income is therefore $Y_D = Y - T$, and private saving is

   $$S_P = Y_D - C = Y - T - C. \tag{0.10}$$

   The ratio S_P/Y_D is the private saving rate. Investment is not subtracted from disposable income because it adds to future productive capacity, income, and wealth.

2. **Government saving** is

   $$S_G = T - G. \tag{0.11}$$

   This quantity is the government budget surplus; when $T - G < 0$, the government runs a budget deficit. Government purchases can be decomposed as $G = C_G + I_G$, where C_G is government consumption and I_G is government investment. The distinction is not always sharp in practice, so it is common to work directly with total government purchases G.

3. **National saving** is the sum of private and government saving:

   $$S = S_P + S_G = Y - C - G. \tag{0.12}$$

   It is the portion of GDP that is not used for current consumption by households or government. The ratio S/Y is the national saving rate.

Substituting the expenditure identity (0.1) into (0.12) gives

$$S = \underbrace{(C + I + G + NX)}_{Y} - C - G = I + NX. \tag{0.13}$$

This simplified section abstracts from cross-border factor income and transfers. Under that convention, the current account reduces to net exports, and national saving can finance domestic investment I or net purchases of foreign assets. Rearranging (0.13) yields the **net capital out-**

flow identity

$$S - I = NX. \tag{0.14}$$

The left side is **net capital outflow**, also called net foreign investment, while the right side is the trade balance. The equality reflects the fact that international flows of funds and international flows of goods and services are two sides of the same transaction.

1. If $NX > 0$, the country has a trade surplus and sends capital abroad to finance foreigners' net purchases of its goods.
2. If $NX < 0$, the country has a trade deficit and borrows from abroad to finance part of its purchases. Equivalently, $S < I$, so domestic investment exceeds national saving and net capital outflow is negative.
3. Balance of payments accounts record flows of goods, services, income, transfers, and assets between countries. In complete accounts, the current account includes net factor income and transfers as well as the trade balance; it equals NX only under the simplification used here.

Example 0.5 Apple sells an iPod in the United Kingdom for £100, equivalent to US$200, and receives the proceeds in a British bank account. Explain how the sale affects U.S. net exports and net capital outflow, and describe the ways Apple can use the proceeds.

Solution. U.S. net exports and the U.S. trade balance rise by US$200. By (0.14), this increase is matched by a US$200 increase in net capital outflow. Apple can use the £100 in any of the following ways:

1. Keep the £100 in the bank, which is equivalent to lending the funds.
2. Buy shares in Rolls Royce, thereby financing a foreign firm.
3. Build an Apple store in the U.K., thereby acquiring a foreign real asset.
4. Deposit the £100 in a U.K. bank, which can lend the funds onward.

Each option leaves the United States with an additional foreign asset and therefore represents net capital outflow. □

Remark 0.6 Net foreign assets equal domestically owned foreign assets minus foreign-owned domestic assets. Wealth can change even when saving is zero because asset prices, including housing and stock market prices, can change. Saving affects wealth, but wealth changes can also affect saving by changing consumption behaviour.

Several policies can stimulate saving:

1. Tax consumption rather than income.
2. Offer saving incentives such as tax-advantaged accounts.
3. Raise the after-tax return on saving, for example by lowering capital gains taxes.
4. Reduce the government budget deficit.

International capital flows take many forms, including purchases of foreign stocks and bonds, purchases of foreign physical assets such as buildings, and deposits in foreign banks. Regardless of their form, (0.14) connects these financial transactions to the accompanying flow of goods and services.

Goods Market Equilibrium

In a **closed economy**, $NX = 0$, so the national income identity becomes

$$Y = C + I + G.$$

It follows from (0.12) that national saving equals investment:

$$S = Y - C - G = I. \tag{0.15}$$

Goods market equilibrium requires production to equal the total quantity of output demanded. The real interest rate adjusts until desired saving equals desired investment, as shown in Figure 4.

Desired consumption depends on disposable income $Y - T$, the real interest rate r, and autonomous consumption $\bar{C}$:

$$C^d = \bar{C} + C(Y - T, r).$$

In the long run, government purchases $\bar{G}$, net taxes $\bar{T}$, capital $\bar{K}$, and labour $\bar{L}$ are exogenous. Output is consequently fixed at

$$Y = F(\bar{K}, \bar{L}) = \bar{Y}.$$

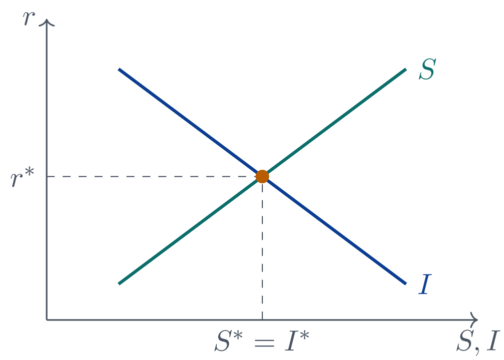

Figure 4

Combining these assumptions with (0.12) gives desired saving as a function of the real interest rate:

$$S^d(r) = \bar{Y} - \bar{C} - C(\bar{Y} - \bar{T}, r) - \bar{G}. \tag{0.16}$$

When r rises, the substitution effect gives households a stronger incentive to save and a weaker incentive to consume. The income effect can work in the opposite direction, so the upward-sloping saving curve used here is an assumption about the net response.

Desired investment can be written as $I^d = \bar{I} + I(r)$. A higher real interest rate reduces the net present value of investment projects and makes them less attractive, so the investment curve slopes downward. This opportunity cost applies even to a firm using its own funds, since those funds could instead earn the market return. A change in r causes movements along the saving and investment curves; changes in their other determinants shift the curves.

The saving curve shifts in Figure 5 illustrate several such changes.

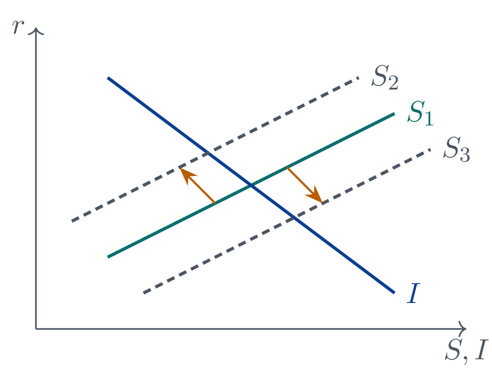

Figure 5

1. If households become more optimistic, autonomous consumption $\bar{C}$ rises. Desired saving

falls, the saving curve shifts left from S_1 to S_2, and the long-run real interest rate rises.

2. Higher taxes reduce disposable income and consumption. If consumption rises with disposable income by less than one-for-one, government saving rises by more than private saving falls. National saving then rises in the long run, the saving curve shifts right from S_1 to S_3, the real interest rate falls, and equilibrium saving and investment both rise.
3. Higher government purchases reduce national saving directly through (0.12). The saving curve shifts left from S_1 to S_2, the real interest rate rises, and saving and investment fall. The reduction in private investment is called **crowding out**.
4. A larger government budget surplus raises $S_G = T - G$. The saving curve shifts right from S_1 to S_3, saving and investment rise, and the real interest rate falls.

Remark 0.7 An increase in aggregate saving does not mean that every individual's bank balance becomes larger.

Intuition: Why can higher taxes raise national saving? Higher taxes raise government saving one-for-one through (0.11). If the marginal propensity to consume out of disposable income lies strictly between zero and one, the accompanying fall in disposable income reduces private saving by less than one-for-one because consumption also falls. Total national saving $S = S_P + S_G$ therefore rises under that assumption.

Changes in desired investment that are not caused by r shift the entire investment curve. Greater business optimism and tax changes that make investment more attractive are two examples. In Figure 6, the shift from I_1 to I_2 raises the real interest rate and increases equilibrium saving and investment. The response combines a shift of desired investment with a movement along the saving curve.

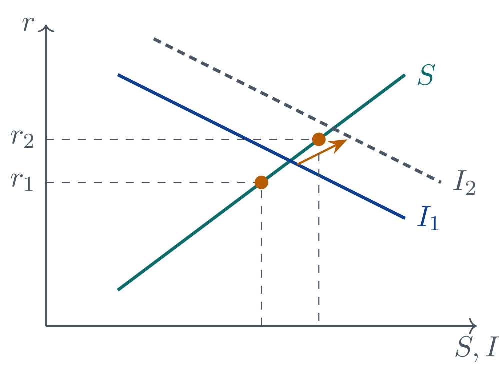

Figure 6

Small Open Economy

A **small open economy** takes the world real interest rate as given and cannot influence it. Under perfect international capital mobility, its domestic real interest rate equals the world rate:

$$r_{\text{domestic}} = r^W. \tag{0.17}$$

Goods market and financial market equilibrium are described by (0.13), namely $S = I + NX$. Let r_E denote the real interest rate that would prevail if the economy were closed. The two possible relationships between r^W and r_E are illustrated in Figure 7.

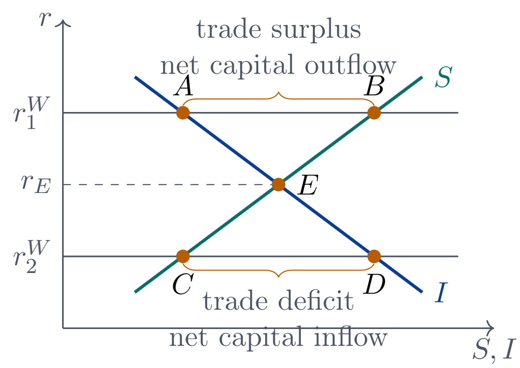

FIGURE 7

1. If $r_1^W > r_E$, desired saving exceeds desired investment. The excess saving produces $NX > 0$, so the country has a trade surplus and a net capital outflow.
2. If $r_2^W < r_E$, desired investment exceeds desired saving. Borrowing from abroad finances the shortfall, producing $NX < 0$, a trade deficit, and a net capital inflow.

The world economy as a whole behaves like a large closed economy that determines r^W. If world saving rises or world investment falls, the world real interest rate declines. A small open economy must accept that lower rate; its domestic investment rises and its net exports fall.

Now consider an increase in domestic saving caused by lower consumption, higher taxes, or lower government purchases. Because the economy is small and open, the domestic interest rate remains fixed at r^W. In Figure 8, saving rises from point B to point C, while investment remains at point A. Net exports consequently increase from the distance AB to the distance AC.

If the economy were closed, both the real interest rate and investment would change. Here the interest rate cannot change, so investment remains fixed and the additional saving becomes net

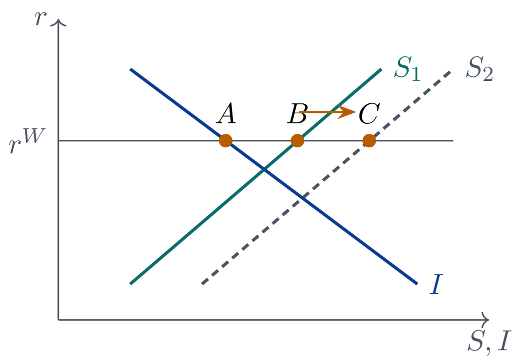

FIGURE 8

exports. Higher saving increases wealth in either model: a closed economy accumulates more domestic capital, while a small open economy accumulates more net foreign assets.

An autonomous increase in desired domestic investment has the complementary effect. At the fixed world interest rate in Figure 9, investment rises from point A to point B, while saving remains at point C. Net exports narrow from the distance AC to the distance BC.

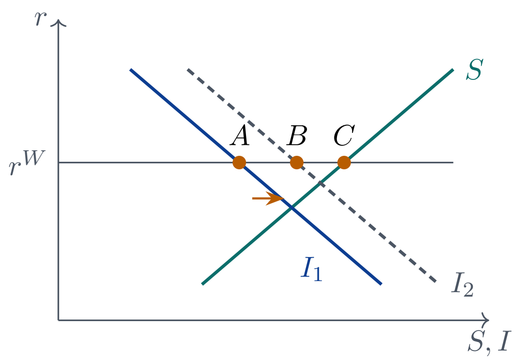

FIGURE 9

Total wealth does not change immediately, but its composition does: the economy holds fewer foreign assets and more domestic capital.

A **large open economy**, such as the United States or the euro area, can influence its domestic real interest rate. An increase in desired saving tends to lower that rate, while an increase in desired investment tends to raise it, as in the model of a closed economy. At the same time, both changes affect net exports, as in the model of a small open economy. A large open economy therefore combines features of the two benchmark cases.

2. Money, Inflation, and Growth

Money and Inflation

Currency consists of bills and coins. **Money** is broader: it is any asset generally accepted as payment for goods and services or for the repayment of debts. Deposits in chequing accounts therefore count as money. Wealth is broader still, with money forming only one component of total wealth.

Money serves three basic functions:

1. As a **medium of exchange**, money makes trade easier and reduces transaction costs. Barter instead requires a double coincidence of wants.
2. As a **unit of account**, money permits prices to be quoted in dollars rather than in complicated barter ratios.
3. As a **store of value**, money preserves purchasing power over time and provides the most liquid form of wealth.

The U.S. Federal Reserve System consists of twelve Federal Reserve Banks and the Board of Governors. Its principal functions are to conduct monetary policy, issue currency, clear cheques, and oversee the financial system. The Federal Open Market Committee directs open-market operations, while the European Central Bank serves as the corresponding institution for the euro area.

In an **open-market purchase**, a central bank buys bonds from banks or the public. The sellers receive money, bank reserves rise, lending can expand, and deposits can increase. The monetary base therefore rises immediately, while broader measures of money can rise as banks and their customers respond. An open-market sale reverses this process.

Two monetary aggregates provide standard measures of money. Under the definitions used when the course was taught, $M1$ consists of currency, travellers' cheques, demand deposits, and other chequable deposits, while $M2$ consists of $M1$ plus money market balances, time deposits, and savings accounts. Statistical definitions evolve, but the distinction between a narrower transactions measure and a broader measure remains useful. Currency in these measures is cash held by the nonbank public, not cash stored in bank machines. A surprisingly large share is held outside ordinary daily transactions, including by foreigners and participants in the underground economy. Although $M1$ is much smaller than $M2$ and the two often move together, they diverge often enough that the Federal Reserve places more emphasis on interest rates than on any one monetary aggregate.

The Quantity Theory of Money

The quantity theory of money is a classical long-run theory. It assumes that wages and prices adjust sufficiently to bring the economy to long-run equilibrium. In Irving Fisher's framework, M is the money supply, P is the price level, Y is real aggregate income, and PY is nominal income or total expenditure.

The **velocity of money** is

$$V = \frac{PY}{M}. \tag{0.18}$$

It is the average number of times per year that a dollar is used to purchase goods and services. Rearranging (0.18) gives the **equation of exchange**

$$MV = PY. \tag{0.19}$$

Velocity depends on institutions and technology. Credit cards and electronic payments, for example, reduce the money balances needed to support a given volume of expenditure and thereby raise velocity. Fisher argued that velocity was reasonably stable over long periods.

The **demand for money** is the amount of money that people want to hold. In equilibrium, money supply equals money demand. If $k = 1/V$ is treated as constant, the demand for purchasing power, or real money balances, is

$$\frac{M^d}{P} = kY. \tag{0.20}$$

Equivalently, if velocity is constant at $\bar{V}$, then

$$PY = M\bar{V}. \tag{0.21}$$

Under the quantity theory, the money supply determines nominal income.

The **classical dichotomy** separates the long-run determination of real and nominal variables. Real output depends on the production function and the available quantities of capital and labour, not on the price level. If $Y = \bar{Y}$, then (0.21) gives

$$P = \frac{M\bar{V}}{\bar{Y}}.$$

The price level is proportional to the money supply. A change in money therefore changes nominal variables such as the price level, but not real variables such as real GDP in the long run. This result is the **neutrality of money**. Milton Friedman summarized the implication by describing inflation as a monetary phenomenon.

Using the usual small-change approximation in (0.19) gives

$$\%\Delta M + \%\Delta V \approx \%\Delta P + \%\Delta Y.$$

If velocity is constant, $\%\Delta V = 0$, and $\%\Delta P$ is the inflation rate π. Hence,

$$\pi \approx \%\Delta M - \%\Delta Y. \tag{0.22}$$

Inflation is therefore approximately determined by money growth in excess of real output growth.

The empirical performance of this conclusion differs by time horizon:

1. In the long run, inflation is strongly related to money growth. Countries with very high rates of money growth typically experience very high inflation.
2. In the short run, the relationship is much weaker. In U.S. data from periods such as 1965–2013, the correlation was limited, probably because wages and prices are not fully flexible over short horizons.

Hyperinflation is conventionally defined as inflation of at least 50% per month. Sustained for a year, that pace raises the price level by roughly 130 times because the monthly changes compound. Hyperinflation usually follows an extreme fiscal imbalance. A government that cannot raise enough taxes or borrow because lenders no longer trust it may finance spending by creating money. The resulting rapid money growth produces massive inflation through (0.22).

The approximate Fisher equation (0.4) connects inflation to interest rates. If the classical dichotomy holds, inflation does not affect the long-run real interest rate. A rise in expected inflation should therefore raise the nominal interest rate one-for-one. This **Fisher effect** works reasonably well over long periods and across countries, but less well in the short run.

Inflation imposes several costs:

1. With **anticipated inflation**, wages and prices may rise together, but the economy still incurs the following costs:

   (a) **Shoe-leather costs** arise because people reduce cash holdings and move funds more frequently.

(b) **Menu costs** arise because firms must change prices in catalogues, labels, websites, and internal systems.

(c) **Tax distortions** arise when tax systems are not fully indexed, so inflation changes effective tax burdens.

(d) **Relative price distortions** arise because firms change prices at different times.

(e) The unit of account becomes a less reliable **yardstick**, which makes price comparisons less informative.

2. **Unanticipated inflation** creates uncertainty, produces poor decisions, and redistributes wealth. If inflation is lower than expected, the realized real burden of debt is higher than borrowers planned. If inflation is higher than expected, lenders receive lower real returns than anticipated. Businesses can also make costly errors in wage contracts, pricing, and investment plans.

3. **Relative price variability** can persist even when firms try to adjust prices for expected inflation. If a firm raises its price by too little, its good becomes temporarily cheap, demand rises, and the firm may overproduce. This problem becomes more severe as average inflation rises.

Deposit creation, the determinants of the money supply, and the money multiplier together form the **money supply process** underlying these monetary aggregates.

The Solow Growth Model

Differences in real GDP per capita across countries are the cumulative result of differences in long-run growth. Nearly all countries grew over the second half of the twentieth century and the beginning of the twenty-first century, but their growth rates varied enormously. Japan, China, South Korea, and India experienced especially rapid growth over important periods, whereas some other countries stagnated or became poorer.

The **Solow growth model** explains long-run income differences by connecting saving and population growth to capital accumulation. Begin with a closed economy that has no government sector and a constant labour force. The model has four building blocks.

1. The **production function** is Cobb–Douglas:

$$Y_t = AK_t^{0.3}L_t^{0.7}. \tag{0.23}$$

Dividing (0.23) by labour gives

$$y_t = \frac{Y_t}{L_t} = A\left(\frac{K_t}{L_t}\right)^{0.3} = Ak_t^{0.3},$$

where $y_t = Y_t/L_t$ is output per worker and $k_t = K_t/L_t$ is capital per worker. Similarly, $c_t = C_t/L_t$ and $i_t = I_t/L_t$ are consumption and investment per worker. Productivity A is exogenous: the basic model takes technology as given rather than explaining how it changes. The production function is increasing and concave because capital has a positive but diminishing marginal product.

2. The **investment function** follows from the resource constraint $y_t = c_t + i_t$. If households save a constant fraction s of income, then consumption is $(1-s)y_t$ and investment is

$$i_t = sy_t = sAk_t^{0.3}. \tag{0.24}$$

As Figure 10 shows, investment has the same shape as production but lies below it when $0 < s < 1$.

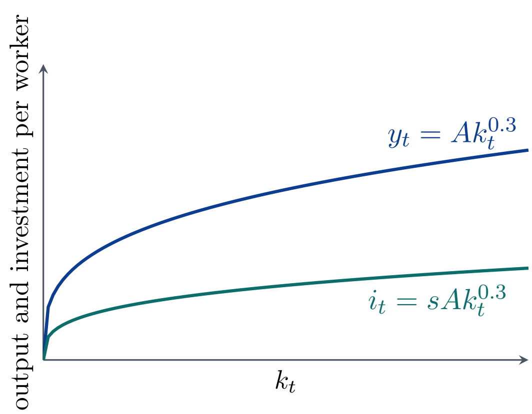

Figure 10

3. The **capital accumulation equation** compares new investment with depreciation. If the constant fraction δ of the capital stock wears out in each period, then

$$\Delta k_t = i_t - \delta k_t = sAk_t^{0.3} - \delta k_t. \tag{0.25}$$

Investment increases the capital stock, whereas depreciation reduces it.

4. A **steady state** is a long-run equilibrium in which capital per worker does not change. Setting $\Delta k_t = 0$ in (0.25) gives the steady-state condition

$$sA(k^*)^{0.3} = \delta k^*. \tag{0.26}$$

Because $k^* > 0$, dividing by $(k^*)^{0.3}$ and solving gives

$$k^* = \left(\frac{sA}{\delta}\right)^{1/0.7}, \qquad y^* = A(k^*)^{0.3}, \qquad \text{and} \qquad c^* = (1-s)y^*.$$

The adjustment toward the steady state is illustrated in Figure 11. When $k_t < k^*$, investment exceeds depreciation, so $\Delta k_t > 0$ and capital per worker rises. When $k_t > k^*$, depreciation exceeds investment, so $\Delta k_t < 0$ and capital per worker falls. The economy therefore converges to k^*, and output and consumption per worker converge to $y^* = A(k^*)^{0.3}$ and $c^* = (1-s)y^*$, respectively.

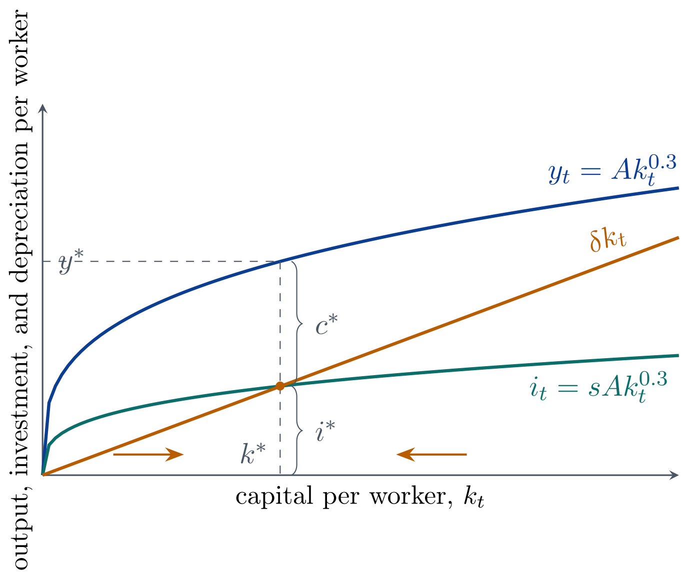

FIGURE 11

The model predicts **conditional convergence**: economies with similar production functions, saving rates, depreciation rates, and labour force growth should approach similar capital stocks and incomes per worker. This prediction fits broadly comparable rich countries better than it fits the entire world because productivity and other structural conditions differ substantially across countries. The rapid postwar recoveries of Germany and Japan are prominent examples of convergence from unusually low capital stocks toward their long-run paths.

Suppose now that the saving rate rises from s_1 to s_2. The adjustment has three parts:

1. The investment curve shifts upward from $s_1 A k_t^{0.3}$ to $s_2 A k_t^{0.3}$.
2. At the original steady state, investment exceeds depreciation, so capital and output per worker rise during the transition.

3. The economy converges to a new steady state with higher k^* and y^*.

Both the curve shift and the transition paths appear in Figure 12.

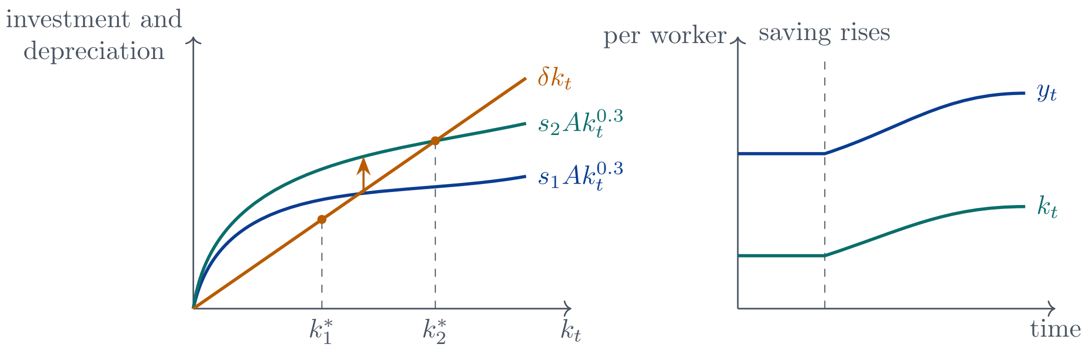

FIGURE 12

The higher saving rate raises the growth rate temporarily while the economy moves toward the new steady state. Once the transition is complete, growth per worker returns to zero. Saving therefore has a **level effect**, not a permanent growth effect. When the labour force is approximately constant, a higher national saving rate thus supports more investment relative to income and a higher long-run level of income per person.

Population growth changes this conclusion by continually increasing the number of workers among whom the capital stock must be divided.

Allow the labour force and population to grow at the common rate n, which requires the worker-to-population ratio to remain constant. Population growth creates **capital dilution**: the existing capital stock must be divided among more workers. Maintaining a given value of k_t requires nk_t units of new capital per current worker in addition to the δk_t units required to replace depreciated capital. More precisely, if $L_{t+1} = (1+n)L_t$ and $K_{t+1} = (1-\delta)K_t + I_t$, then the exact discrete-time law of motion is

$$k_{t+1} - k_t = \frac{sAk_t^{0.3} - (\delta + n)k_t}{1+n}. \tag{0.27}$$

The term $(\delta + n)k_t$ is **break-even investment**: the amount of investment needed to keep capital per worker unchanged.

At the positive steady state, the numerator in (0.27) is zero, so

$$k^* = \left(\frac{sA}{\delta + n}\right)^{1/0.7}.$$

In a steady state with constant productivity, output, capital, and labour all grow at rate n, but output and capital per worker are constant. Population growth can therefore explain sustained growth in total output, but it cannot explain sustained growth in living standards.

Suppose population growth rises from n_1 to n_2. As the left panel of Figure 13 shows, the break-even-investment line becomes steeper and the steady-state capital stock falls from k_1^* to k_2^*. The right panel traces the associated declines in capital and output per worker.

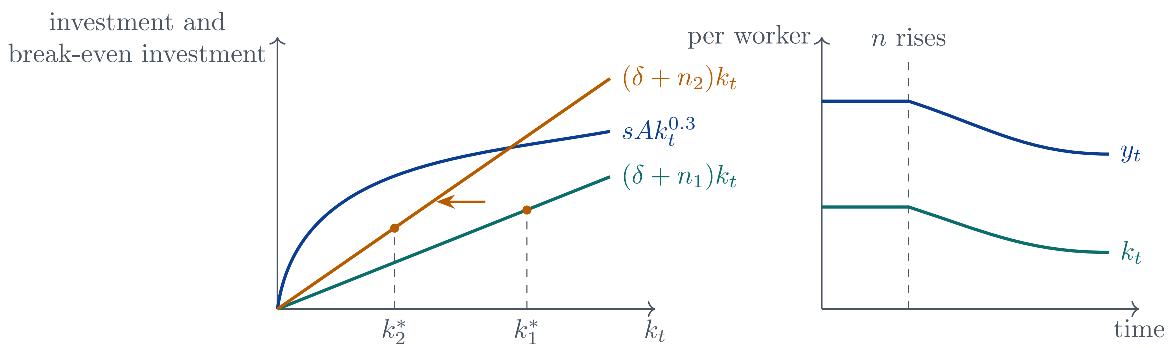

FIGURE 13

The model thus predicts that higher population growth reduces long-run income per person. Comparisons across countries broadly support this association, and the result has sometimes been used to justify controversial population policies. The direction of causation is not settled, however, because higher income can itself reduce fertility.

Productivity provides another source of change. A one-time rise from A_1 to A_2 shifts both the production and investment curves upward. The response, shown in Figure 14, has two stages:

1. Output per worker rises immediately because the existing capital stock becomes more productive.
2. The higher investment generated by the productivity increase gradually raises capital and output per worker further.

The new steady state satisfies

$$(\delta + n)k_2^* = sA_2(k_2^*)^{0.3}.$$

If productivity grows at the ordinary discrete rate g_A, capital and output per worker also grow along a balanced growth path. In the Cobb–Douglas specification $y_t = A_t k_t^{0.3}$, the exact growth-

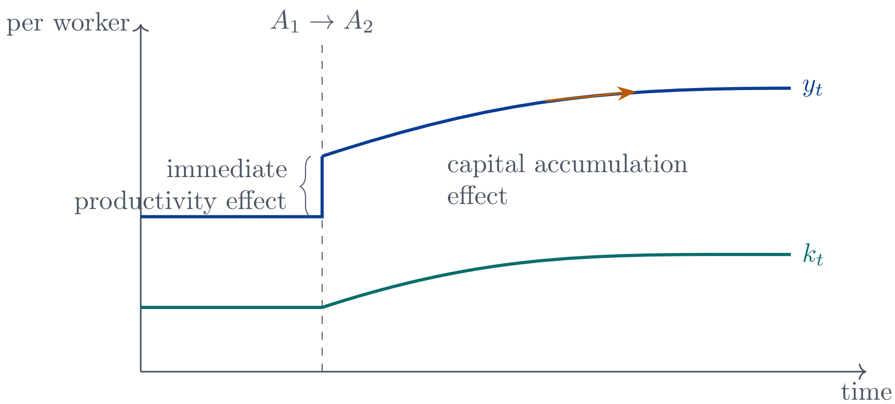

Figure 14

factor relation is

$$1 + g_y = (1 + g_A)(1 + g_k)^{0.3} \qquad \text{and} \qquad g_y = g_k.$$

Consequently,

$$g_y = g_k = (1 + g_A)^{1/0.7} - 1 \approx \frac{g_A}{0.7} \approx 1.43 g_A,$$

where the approximation is accurate for small growth rates. Living standards therefore grow more than one-for-one with the direct growth of the productivity term because higher productivity also induces capital accumulation.

The principal conclusions of the Solow model are as follows:

1. Among economies with similar structural parameters, those with lower initial income tend to grow faster as they converge toward steady state.
2. A higher saving rate raises the long-run levels of capital and output per worker, but not their permanent growth rates.
3. Higher population growth lowers steady-state output per person through capital dilution.
4. Higher productivity raises output directly and indirectly through additional capital accumulation.

The model's principal limitation is that it treats the ultimate source of sustained productivity growth as exogenous.

Growth Accounting

Growth accounting decomposes output growth into contributions from productivity, capital, and labour. Let $g_X = (X_t - X_{t-1})/X_{t-1}$ denote the ordinary discrete growth rate of variable X. Taking growth factors of the aggregate production function (0.23) gives the exact identity

$$\log(1+g_Y) = \log(1+g_A) + 0.3\log(1+g_K) + 0.7\log(1+g_L). \tag{0.28}$$

For small growth rates, $\log(1+g_X) \approx g_X$, so (0.28) becomes the familiar first-order approximation

$$g_Y \approx g_A + 0.3g_K + 0.7g_L.$$

The coefficients 0.3 and 0.7 are the output elasticities of capital and labour. Thus, to first order and holding the other inputs constant, capital growth of 5% contributes about $0.3(5\%) = 1.5$ percentage points to output growth, labour growth of 5% contributes about $0.7(5\%) = 3.5$ percentage points, and productivity growth of 5% contributes about the full 5 percentage points.

Productivity is not observed directly. Rearranging the production function gives the measured productivity level

$$A_t = \frac{Y_t}{K_t^{0.3}L_t^{0.7}}, \tag{0.29}$$

and the first-order growth-accounting approximation gives

$$g_A \approx g_Y - 0.3g_K - 0.7g_L.$$

This inferred productivity growth is the **Solow residual**. It includes genuine technological progress as well as any omitted influences or measurement error that change output without being recorded as capital or labour growth.

Postwar U.S. growth accounting attributes the slowdown in output growth primarily to slower productivity growth rather than to slower capital or labour growth. Proposed explanations include oil price shocks, heavier regulation, and a shift from manufacturing toward services. In later periods, slower labour force growth also became more important. Internationally, some of the Asian Tigers relied heavily on capital accumulation, whereas others relied more on rapid productivity growth. Taken together, these comparisons suggest that productivity growth is more important than factor accumulation alone as a sustained source of higher living standards.

Technology, Policy, and Institutions

The large differences across countries in capital accumulation and productivity raise a deeper question: why do some economies create productive resources and ideas much more rapidly than others? In the growth model, **technology** is represented by A because it raises the efficiency of both capital and labour.

Ideas differ from ordinary physical inputs in two important ways. An idea is **nonrival** because one person's use does not prevent another person from using it, and it is reusable because the same knowledge can contribute to production repeatedly. Ideas may also have low **excludability**: inventors can find it difficult to prevent others from benefiting from their discoveries. These knowledge spillovers create a gap between the social return and the private return to innovation, so private firms may invest less in research than is socially desirable.

Public policy can promote productivity through several channels:

1. **Infrastructure** consists of public capital such as roads, ports, and bridges. It can lower transportation costs and support trade, tourism, and production. The U.S. interstate highway system is a standard example. Measuring the causal effect is difficult because fast-growing governments may spend more, while poorly chosen projects may be wasteful or vulnerable to corruption.
2. **Human capital** is the knowledge and skill embodied in workers. Greater educational attainment is associated with substantially higher lifetime earnings and national income, although education by itself cannot make every poor country rich. Better health care can also increase human capital by making workers healthier and more productive.
3. Governments can encourage **research and development**. Although private firms devote substantial resources to research, knowledge spillovers may still cause underinvestment. Governments can respond in three ways:
   (a) Governments can fund research directly through agencies, grants, and universities.
   (b) Governments can provide tax incentives such as research and development tax credits.
   (c) Governments can protect patents to make innovation more excludable, although patent terms must be balanced carefully.

Growth also depends on **institutions**, the rules, organizations, and customs that govern behaviour in an economy. Secure **property rights** give people confidence that they will retain the returns to work, saving, and innovation. Fear of theft, arbitrary seizure, patent violation, or weak contract enforcement discourages investment and slows capital accumulation.

An effective legal system therefore requires

1. enforceable contracts;
2. adequate courts, judges, and enforcement resources; and
3. meaningful access to legal representation.

Differences in colonial history and legal institutions are often used to explain why some former colonies have performed far better than others. Corruption, excessive bureaucracy, arbitrary expropriation, war, rebellion, and kleptocratic government all weaken property rights. Indicators that compare countries, including the World Bank's historical *Doing Business* measures, have been used to study these barriers and encourage reform.

Foreign aid remains controversial in this institutional setting. Critics argue that aid can strengthen harmful political incentives, whereas supporters argue that carefully designed aid can improve public health, infrastructure, and economic development.

Solow's model leaves productivity growth unexplained. **Endogenous growth theory**, associated here with Paul Romer, instead models the creation of new ideas explicitly.

In the Romer model, labour is divided between production and research. Let L_p be labour used to produce goods and services, let L_A be labour used to create ideas, and let $\bar{N}$ be the total population. If the constant fraction α of the population conducts research, then

$$\bar{N} = L_p + L_A, \qquad L_A = \alpha \bar{N}, \qquad \text{and} \qquad L_p = (1-\alpha)\bar{N}. \tag{0.30}$$

Output per production worker retains the Cobb–Douglas form

$$y_{p,t} = A_t k_{p,t}^{0.3},$$

where $k_{p,t} = K_t / L_p$.

New ideas are produced using both research labour and the existing stock of knowledge:

$$A_{t+1} - A_t = \Delta A_t = \chi A_t L_A, \tag{0.31}$$

where χ measures the productivity of research labour in creating ideas. Existing knowledge makes it easier to create additional knowledge and simultaneously raises current production. Because ideas are nonrival, this cumulative process can sustain growth rather than merely produce convergence to a fixed steady state.

Dividing (0.31) by A_t and using (0.30) yields

$$g_A = \frac{\Delta A_t}{A_t} = \chi \alpha \bar{N}. \tag{0.32}$$

Technology growth therefore depends on the productivity of researchers, the fraction of the population engaged in research, and the size of the population.

Because $L_p/\bar{N} = 1 - \alpha$, output per person is

$$y_t = \frac{Y_t}{\bar{N}} = \frac{Y_t}{L_p}\frac{L_p}{\bar{N}} = (1-\alpha)y_{p,t} = (1-\alpha)A_t k_{p,t}^{0.3}. \tag{0.33}$$

When technology grows at a constant rate, output per person follows a balanced growth path rather than converging to a fixed level. With the same production and capital accumulation assumptions used above, the exact discrete long-run growth rate is $g_y = (1+g_A)^{1/0.7} - 1$, which is approximately $g_A/0.7 \approx 1.43 g_A$ for small growth rates.

Equation (0.32) and the capital accumulation mechanism identify four influences on endogenous growth.

1. A rise in α, the fraction of the population conducting research, creates the transition in Figure 15. Moving labour out of production can reduce output per person immediately, even though capital per production worker rises mechanically when fewer workers remain in production. Capital then adjusts toward its new path, and the larger research sector produces a permanently higher rate of idea creation. The economy therefore faces a short-run versus long-run tradeoff: less output initially, but faster growth thereafter.
2. A rise in χ, the productivity of research labour, raises the growth rate of technology immediately without requiring labour to leave current production. In Figure 16, output remains continuous when χ changes, but its growth path becomes steeper. Policies that improve education, research quality, or innovation incentives can all raise χ.
3. A rise in total population $\bar{N}$ illustrates an important contrast between the two growth models. Solow emphasizes capital dilution, whereas Romer emphasizes that a larger population also provides more potential researchers. The complete response is shown in Figure 17.
   (a) A sudden population increase initially lowers capital and output per person through capital dilution.

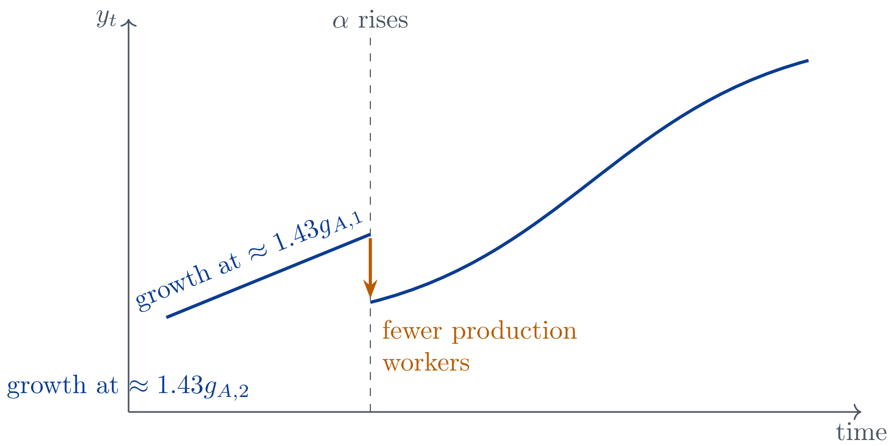

FIGURE 15

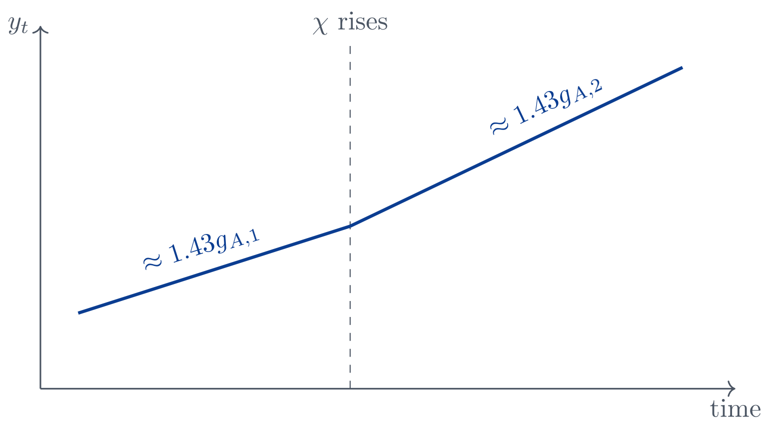

FIGURE 16

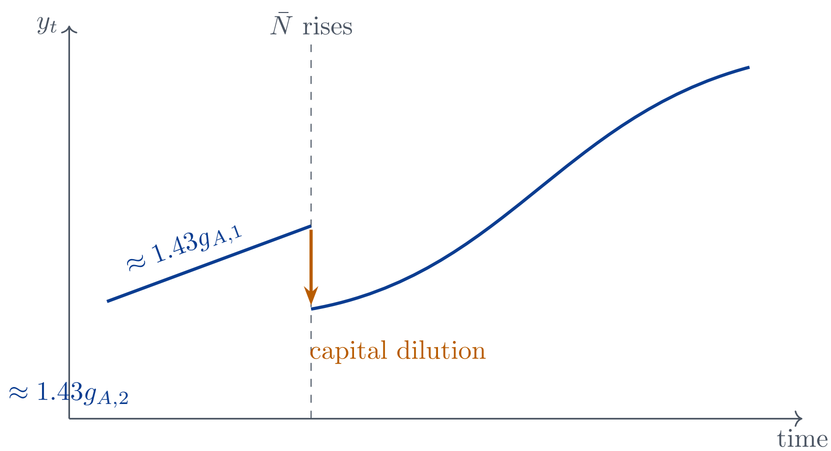

FIGURE 17

(b) Capital per production worker then recovers toward its balanced growth path.

(c) Because $L_A = \alpha\bar{N}$, the larger population ultimately supports a permanently higher rate of technological growth.

The Solow dilution effect often appears important for individual countries. For the world economy as a whole, the Romer effect may be more relevant because more people can create more ideas, and technological knowledge can spill across national borders.

4. In both the Solow model and this basic Romer model, a higher saving rate raises growth only temporarily. It increases capital accumulation and moves output to a permanently higher path, as in Figure 18, but it does not affect $g_A = \chi\alpha\bar{N}$. The long-run growth rate is therefore unchanged.

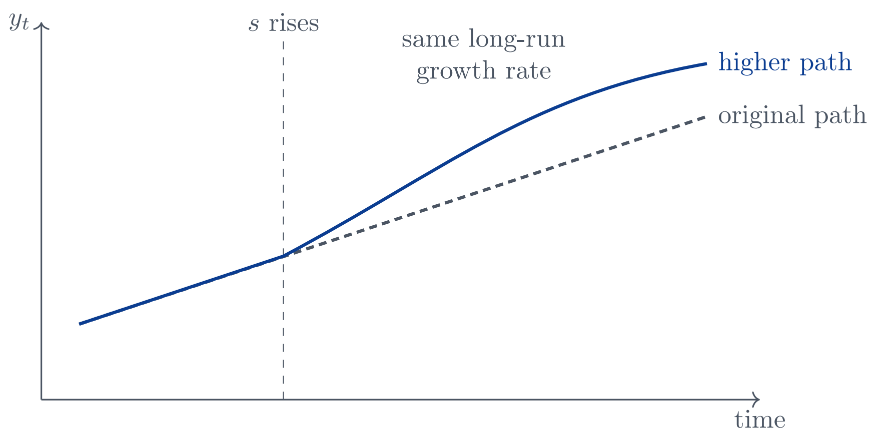

FIGURE 18

> **Remark 0.8** More elaborate endogenous growth models can allow saving to raise productivity indirectly by financing broader forms of capital.

3. Business Cycles and Aggregate Demand

Business Cycles

Paul Samuelson, 1966: "Wall Street indexes predicted nine of the last five recessions."

Definition 0.9 A **business cycle** is a recurring but not strictly periodic fluctuation in aggregate economic activity during which many sectors expand or contract together.

The path of actual output Y fluctuates around potential output Y^P, as illustrated in Figure 19. **Potential output** is the level of output produced when prices have adjusted and resources are employed at normal sustainable rates. A peak marks the transition from expansion to contraction, whereas a trough marks the transition from contraction to expansion. A particularly severe and prolonged recession is called a depression.

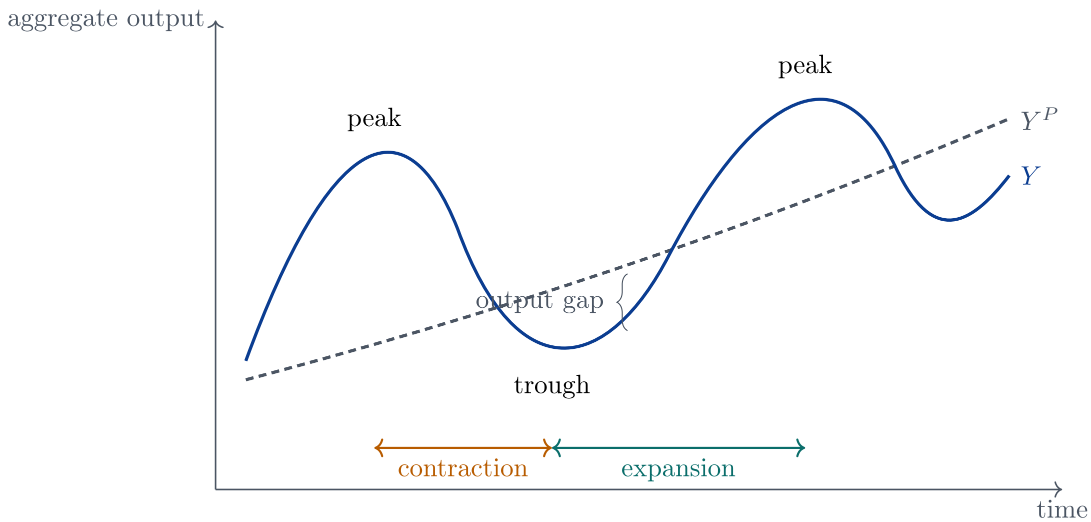

Figure 19

The **output gap** is $Y - Y^P$. It is negative when actual output is below potential and positive when actual output is above potential. In the United States, the National Bureau of Economic Research identifies the dates of business cycle peaks and troughs.

Macroeconomic indicators can be classified both by their direction and by their timing. A **procyclical** variable rises during expansions and falls during contractions; a **countercyclical** variable moves in the opposite direction; and an **acyclical** variable has no systematic relationship with the cycle. A leading indicator tends to move before aggregate activity, a coincident indicator moves at approximately the same time, and a lagging indicator moves afterward.

1. Real GDP is the broadest measure of aggregate activity and is commonly used to track the business cycle. The most severe U.S. contractions include 1929–1933 and 1937–1938, with major postwar recessions in 1973–1975, 1981–1982, and 2007–2009.
2. Real consumer spending is usually procyclical and coincident.

3. Investment spending is also procyclical and coincident, but it is much more volatile than consumption.
4. Unemployment is strongly countercyclical.
5. Inflation is generally procyclical and tends to lag the cycle.
6. Financial variables provide additional signals:
   (a) Stock prices are usually procyclical and often leading.
   (b) Treasury bill rates, term spreads, and spreads between corporate and government bond yields can also be useful because credit conditions often worsen before or during downturns. A Treasury bill is a short-term U.S. government bond.

International business cycles are often correlated with the U.S. cycle because the U.S. economy is large and because global financial markets are highly integrated.

Macroeconomics distinguishes sharply between long-run and short-run analysis. The classical view stresses that economies adjust toward long-run equilibrium relatively quickly, so policy should concentrate on economic growth. Keynes argued that short-run fluctuations matter greatly because an economy may remain away from long-run equilibrium for an extended period. The two perspectives differ in four related ways:

1. Long-run classical models treat prices and wages as flexible, whereas short-run Keynesian models allow them to adjust slowly.
2. The classical dichotomy separates real variables from nominal variables in the long run. In the short run, changes in money and inflation can affect real output, real interest rates, saving, and investment.
3. Long-run output is determined by productivity and the available quantities of capital and labour, while the long-run real interest rate equates saving and investment. Short-run output can instead be constrained by aggregate demand.
4. Long-run policy emphasizes economic growth, whereas short-run policy emphasizes stabilizing fluctuations in output and unemployment.

Classical models usually assume perfect competition and flexible prices.

In reality, many goods and services are differentiated, so firms possess some market power. This market structure is called **monopolistic competition**. Prices can then be sticky rather than adjusting immediately to clear every market.

Sources of price stickiness include the following:

1. **Menu costs** are the time and resources required to change prices, including the costs of updating catalogues, advertising, labels, and pricing systems. Some empirical studies find that these costs are large enough to affect firms' behaviour materially.
2. Firms may avoid frequent price changes because customers dislike them. Limited attention can also lead firms to ignore some aggregate shocks.
3. Under **staggered price setting**, firms reset their prices at different times. Even if each individual price eventually adjusts, this staggering slows adjustment of the aggregate price level.

The empirical evidence is mixed. Some surveys find that many firms change prices only once or twice a year, whereas retail price data often show more frequent changes. Firms may also react quickly to shocks specific to their own products but more slowly to economy-wide shocks such as changes in monetary policy.

The aggregate demand–aggregate supply model incorporates these ideas by treating prices and wages as sticky in the short run. Wage stickiness belongs to the same general problem because a wage is the price of labour.

The IS Curve

The Great Depression challenged the classical view because aggregate output collapsed even though capital and labour did not change much. Keynes responded by emphasizing aggregate demand.

The **IS curve** shows the combinations of the real interest rate and aggregate output for which the market for goods and services is in equilibrium. Its derivation begins with **planned expenditure**, the amount that households, firms, governments, and foreigners intend to spend on domestically produced goods and services. Planned expenditure need not equal actual expenditure because firms may experience unplanned changes in inventories. In the Keynesian framework, planned expenditure is aggregate demand:

$$PE = C + I + G + NX. \tag{0.34}$$

The consumption function is an aggregate relationship. Disposable income is $Y_D = Y - T$, and consumption is

$$C = \bar{C} + m(Y - T) - cr, \tag{0.35}$$

where $\bar{C}$ is autonomous consumption, $m \in (0,1)$ is the marginal propensity to consume, and $c > 0$ measures the sensitivity of consumption to the real interest rate. Autonomous consumption does not depend on current income, taxes, or the real interest rate within the model; it can reflect household expectations, wealth, or optimism about future income. The marginal propensity to consume is the fraction of an additional dollar of disposable income that households spend. Keynes's simplest consumption function omitted the term $-cr$, but interest-sensitive durable purchases make this extension useful.

Investment has two components:

1. **Fixed investment** is business spending on equipment and structures together with residential construction.
2. **Inventory investment** is the change in firms' holdings of raw materials, parts, and finished goods. It is much smaller than fixed investment, but it is important for fluctuations in aggregate output. Inventory investment can be unplanned, in which case firms adjust subsequent production to eliminate the unwanted inventory change.

Planned investment is fixed investment plus planned inventory investment. Firms invest when the expected payoff exceeds the real cost of borrowed funds or the opportunity cost of using internal funds. Expectations therefore matter alongside interest rates; Keynes used the phrase "animal spirits" for sharp changes in business confidence. Let the relevant real borrowing rate be $r_i = r + \bar{f}$, where r is the risk-free real interest rate and $\bar{f}$ captures financial frictions such as credit spreads. Planned investment is

$$I = \bar{I} - dr_i = \bar{I} - d(r + \bar{f}), \tag{0.36}$$

where $\bar{I}$ is autonomous planned investment and $d > 0$ is its sensitivity to the borrowing rate. The real rate, rather than the nominal rate, governs this investment decision.

Net exports are

$$NX = \overline{NX} - xr, \tag{0.37}$$

where $\overline{NX}$ represents exogenous influences such as foreign tastes and foreign income, and $x > 0$ measures interest sensitivity. A rise in the U.S. real interest rate makes dollar assets more attractive, appreciates the dollar, makes U.S. goods more expensive to foreigners, and makes foreign goods cheaper to Americans. Net exports consequently fall.

Remark 0.10 Under a fixed exchange rate, the interest rate channel through currency appreciation is much weaker, so net exports remain closer to $\overline{NX}$.

Finally, government purchases and taxes are exogenous: $G = \bar{G}$ and $T = \bar{T}$. A more realistic model could allow taxes to rise with income, but holding them fixed isolates the basic expenditure mechanism.

Goods market equilibrium requires actual output to equal planned expenditure. Substituting (0.35), (0.36), and (0.37) into (0.34) gives

$$\begin{aligned} Y &= \bar{C} + m(Y - \bar{T}) - cr + \bar{I} - d(r + \bar{f}) + \bar{G} + \overline{NX} - xr \\ &= \underbrace{\bar{C} + \bar{I} - d\bar{f} + \bar{G} + \overline{NX} - m\bar{T}}_{\bar{A}} + mY - \underbrace{(c + d + x)}_{b} r, \end{aligned} \tag{0.38}$$

where $\bar{A}$ is autonomous expenditure and $b = c + d + x > 0$ is the total interest sensitivity of planned expenditure. Collecting the output terms yields $(1 - m)Y = \bar{A} - br$, so the IS curve is

$$\boxed{Y = \frac{\bar{A}}{1 - m} - \frac{b}{1 - m} r.} \tag{0.39}$$

Equivalently, substituting the definitions of $\bar{A}$ and b into (0.39) gives

$$Y = \frac{\bar{C} + \bar{I} - d\bar{f} + \bar{G} + \overline{NX} - m\bar{T}}{1 - m} - \frac{c + d + x}{1 - m} r.$$

Because $0 < m < 1$, the multiplier $1/(1 - m)$ is positive. Increases in $\bar{C}$, $\bar{I}$, $\bar{G}$, or $\overline{NX}$, and decreases in $\bar{T}$ or $\bar{f}$, raise equilibrium output at every real interest rate and shift the IS curve. In contrast, a change in r produces a movement along the curve. Its negative slope reflects the fall in consumption, investment, and net exports caused by a higher real interest rate. A simple numerical example is $Y = 12 - r$.

Points away from the IS curve generate unplanned inventory changes. In Figure 20, point A has excessive production at the given interest rate. Inventories accumulate unexpectedly, so firms cut production until the economy reaches point B. At point C, planned expenditure exceeds production, inventories fall unexpectedly, and firms raise production until the economy reaches point D.

The name IS comes from the saving–investment condition developed earlier in (0.15). If $G = 0$

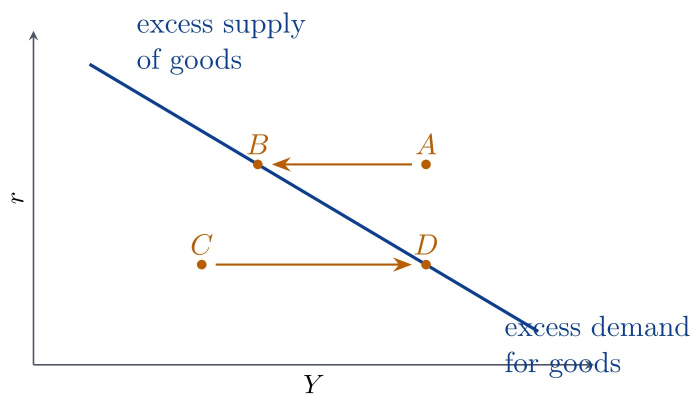

FIGURE 20

and $NX = 0$, goods market equilibrium gives

$$Y = C + I \qquad \text{and} \qquad S = Y - C = I.$$

Thus, every point on the IS curve also represents equality between desired saving and desired investment in this special case. The mechanism is displayed in Figure 21: higher output raises desired saving, which permits saving and investment to be equal at a lower real interest rate.

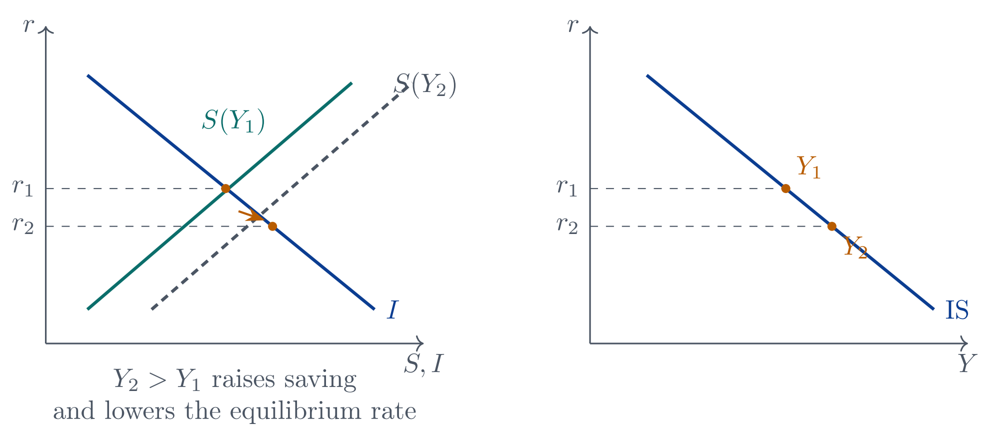

FIGURE 21

The IS curve shifts only when an autonomous determinant changes; a change in the real interest rate itself causes a movement along the curve. The directions of the shifts are summarized in Figure 22.

Equation (0.39) gives the comparative statics:

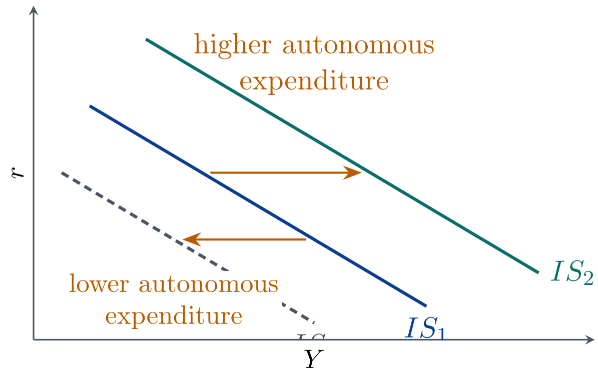

FIGURE 22

1. An increase in $\bar{G}$, $\bar{C}$, $\bar{I}$, or $\overline{NX}$ raises autonomous expenditure and shifts the IS curve to the right.
2. An increase in the financial friction $\bar{f}$ or in taxes $\bar{T}$ reduces autonomous expenditure and shifts the IS curve to the left.

For example, when government purchases rise, firms must produce more output to meet the additional demand at any given real interest rate. Historical episodes such as the buildup associated with the Vietnam War are standard examples of rightward IS shifts. Fiscal policy works through these channels by changing government purchases or taxes.

Monetary Policy and Aggregate Demand

The **monetary policy curve**, or MP curve, describes how the central bank chooses the real interest rate in response to inflation. Combining this curve with the IS curve produces the aggregate demand curve.

The Federal Reserve implements monetary policy by influencing the nominal federal funds rate, principally through its management of financial system liquidity. Expenditure decisions depend on the real interest rate. By the approximate Fisher equation (0.4), $r \approx i - \pi^e$. Expected inflation usually changes slowly in the short run, so changing the nominal policy rate changes the real rate as well. In the long run, however, saving and investment determine the real interest rate.

The MP curve is

$$r = \bar{r} + \lambda\pi, \tag{0.40}$$

where $\bar{r}$ is the autonomous component of the policy stance and $\lambda > 0$ measures the central bank's response to inflation. The positive slope appears in Figure 23.

A central bank seeking price stability normally follows the **Taylor principle**: when inflation rises by one percentage point, it raises the nominal interest rate by more than one percentage point, causing the real interest rate to rise. Without this response, higher inflation would reduce the real rate, stimulate planned expenditure, and reinforce inflation.

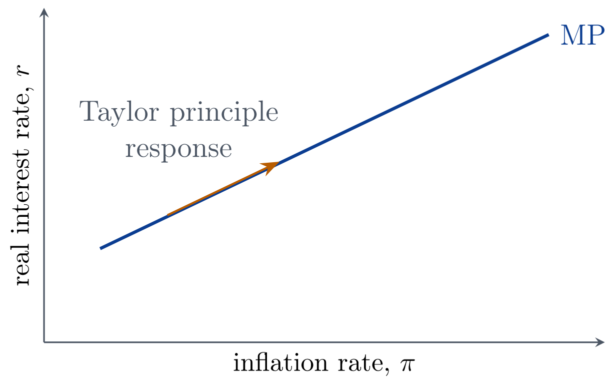

FIGURE 23

The Taylor principle is distinct from a Taylor rule. The principle states only that nominal rates must rise more than one-for-one with inflation; a Taylor rule, developed later, specifies a complete policy formula involving inflation and the output gap. The 1970s are often interpreted as a period in which inflation rose without a sufficiently strong policy response.

An increase in the real policy rate is a **monetary tightening**, and a decrease is a **monetary easing**. An automatic response to inflation is a movement along the MP curve. An autonomous change in $\bar{r}$ shifts the entire curve: raising $\bar{r}$ shifts it upward, while lowering $\bar{r}$ shifts it downward. These shifts appear in Figure 24.

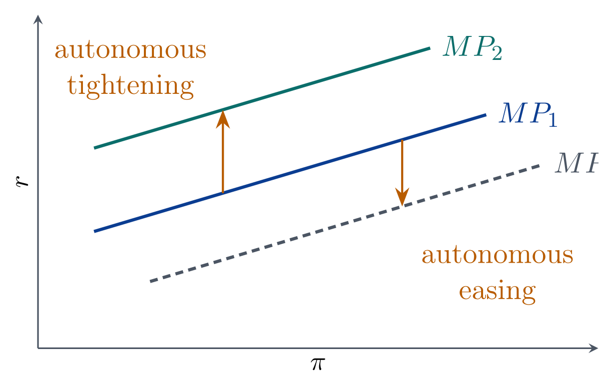

FIGURE 24

In normal times, the central bank often moves along the MP curve by raising the real interest rate when inflation rises. The 2004–2006 tightening cycle is an example of this response. During the 2007–2009 financial crisis, by contrast, the Federal Reserve shifted the policy stance itself: it lowered rates before inflation had collapsed because it anticipated severe economic weakness.

The Aggregate Demand Curve

The MP curve links inflation to the real interest rate, while the IS curve links the real interest rate to output. Combining them gives the **aggregate demand curve**, or AD curve, which relates inflation to the quantity of aggregate output demanded.

Example 0.11 Suppose the MP curve is $r = 1 + 0.5\pi$ and the IS curve is $Y = 12 - r$. Derive the aggregate demand curve.

Solution. Substituting the policy rule into the IS curve gives

$$Y = 12 - (1 + 0.5\pi) = 11 - 0.5\pi.$$

Thus, higher inflation induces a higher real policy rate and reduces aggregate output demanded. □

More generally, substitute the MP curve (0.40) into the IS curve (0.39):

$$\begin{aligned} Y &= \frac{\bar{A}}{1-m} - \frac{b}{1-m}(\bar{r} + \lambda\pi) \\ &= \frac{\bar{A} - b\bar{r}}{1-m} - \frac{b\lambda}{1-m}\pi. \end{aligned} \tag{0.41}$$

In terms of the underlying expenditure parameters, (0.41) is

$$Y = \frac{\bar{C} + \bar{I} - d\bar{f} + \bar{G} + \overline{NX} - m\bar{T} - (c + d + x)\bar{r}}{1-m} - \frac{(c+d+x)\lambda}{1-m}\pi.$$

The aggregate demand curve slopes downward because higher inflation causes the central bank to raise the real interest rate, which reduces planned consumption, investment, and net exports.

The autonomous factors that shift the IS curve also shift the aggregate demand curve in the same direction. An autonomous monetary tightening raises $\bar{r}$ and shifts aggregate demand to the left; an autonomous easing lowers $\bar{r}$ and shifts aggregate demand to the right. Figure 25 shows how an upward shift of the MP curve becomes a leftward shift of the AD curve.

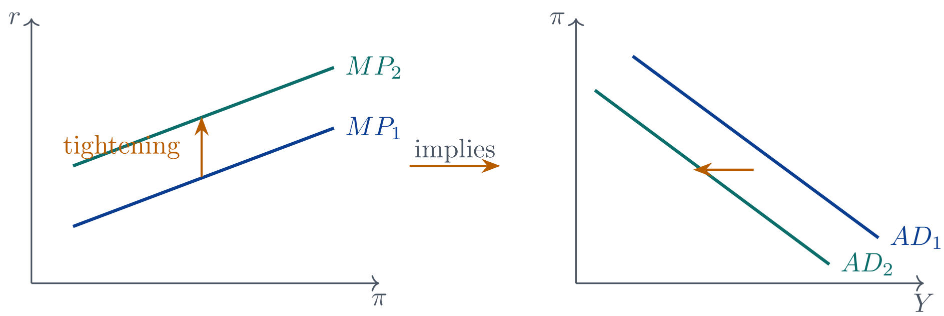

FIGURE 25

The Money Market and Interest Rates

The money market determines the equilibrium nominal interest rate at which the supply of money equals the demand for money. Because people care about purchasing power, the relevant object is the demand for **real money balances**. The **liquidity preference function** is

$$\frac{M^d}{P} = L(i, Y), \qquad \frac{\partial L}{\partial i} < 0, \qquad \text{and} \qquad \frac{\partial L}{\partial Y} > 0. \tag{0.42}$$

Here M^d/P is real money demand, i is the nominal interest rate, and Y is real income. A lower nominal interest rate reduces the opportunity cost of holding money rather than interest-bearing assets, so money demand rises. Higher income raises money demand because

1. a larger volume of transactions requires more money; and
2. wealthier households tend to hold more of every normal asset, including money.

Money demand is related to the nominal interest rate because the nominal return forgone by holding currency is the relevant opportunity cost. Planned expenditure, by contrast, depends more directly on the real interest rate.

In the short run, the price level is sticky at $\bar{P}$. For a given income level, (0.42) therefore produces a downward-sloping demand curve for real balances. In this simple model, the Federal Reserve fixes the nominal money supply at $\bar{M}$, so the supply of real balances is vertical at $\bar{M}/\bar{P}$. Their intersection in Figure 26 determines i^*.

If $i > i^*$, money supply exceeds money demand. People try to exchange excess money for bonds and other interest-bearing assets, which pushes the nominal interest rate down toward equilibrium.

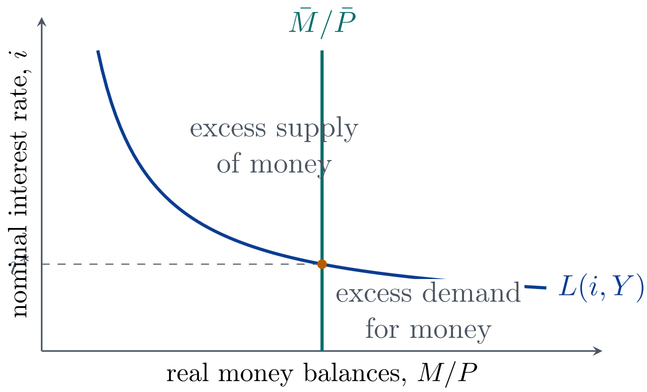

FIGURE 26

If $i < i^*$, money demand exceeds money supply, and the nominal interest rate rises toward equilibrium.

The equilibrium interest rate also changes when either curve shifts:

1. Higher income increases transaction demand for money. As Figure 27 shows, the money demand curve shifts to the right and the equilibrium nominal interest rate rises.

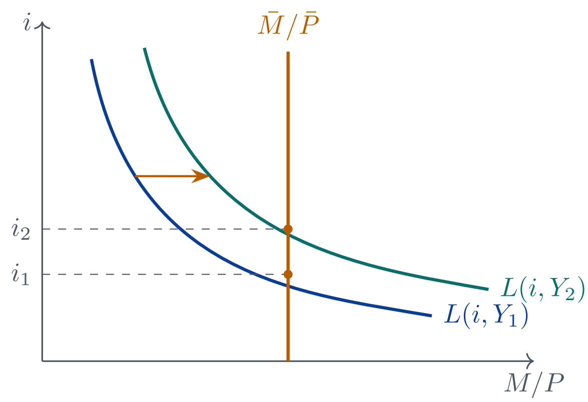

FIGURE 27

2. The supply of real balances changes when either nominal money or the price level changes. If the Federal Reserve raises nominal money, the real money supply curve shifts right and the nominal interest rate falls, as in Figure 28. If the price level rises while nominal money is fixed, real balances fall. The real money supply curve shifts left and the nominal interest rate rises, as shown in Figure 29.

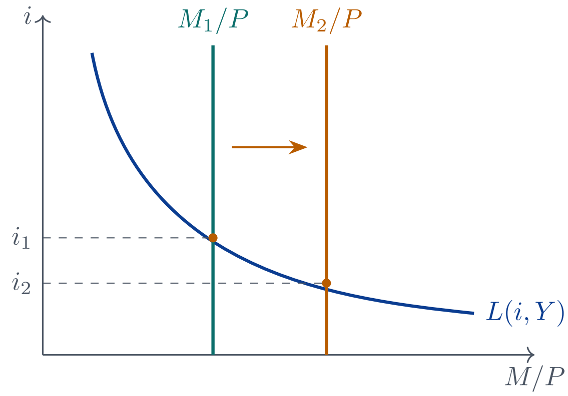

FIGURE 28

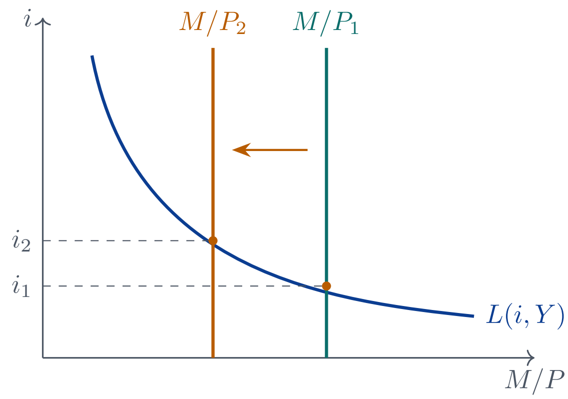

FIGURE 29

Keynes rejected the assumption of constant velocity and emphasized the interest sensitivity of money demand. His liquidity preference theory distinguishes three motives for holding money:

1. The **transaction motive** arises from the need to make routine purchases.
2. The **precautionary motive** arises from the desire to meet unexpected expenses.
3. The **speculative motive** arises from choosing between money and interest-bearing assets as expected returns change.

At money market equilibrium, $M/P = L(i, Y)$. Substituting this relation into the velocity identity (0.18) gives

$$V = \frac{PY}{M} = \frac{Y}{L(i, Y)}. \tag{0.43}$$

Velocity therefore varies with the nominal interest rate and income rather than remaining constant.

Portfolio theories of money demand treat money as one asset among many. The demand for any asset depends on four characteristics relative to the available alternatives:

1. Greater wealth tends to increase the quantity held of every normal asset.
2. A higher expected return makes an asset more attractive.
3. Greater risk makes an asset less attractive to risk-averse investors.
4. Greater liquidity makes an asset more useful for transactions.

These considerations connect naturally to liquidity preference. Income and wealth tend to move together; the interest rate changes the return on money relative to bonds; money is nominally safe but exposed to inflation risk; and the availability of other liquid assets reduces the need to hold money itself. Real estate and gold, for example, may be attractive as inflation hedges even though they are less liquid than money.

Seven variables summarize the principal determinants of money demand:

1. A higher nominal interest rate reduces money demand by raising the opportunity cost of holding it.
2. Higher income increases transaction demand.
3. Improvements in payment technology can reduce the balances needed to conduct a given volume of transactions.

4. Greater wealth tends to increase money demand when money is a normal asset.
5. Greater risk among alternative assets tends to increase demand for relatively safe money balances.
6. Greater inflation risk reduces the attractiveness of money's real return and may increase demand for inflation hedges.
7. Greater liquidity among alternative assets reduces the special liquidity advantage of money.

Two empirical questions determine whether monetary aggregates are useful policy targets.

1. Money demand is sensitive to interest rates, although its sensitivity is normally finite. Greater sensitivity makes velocity more variable and weakens the link between the money supply and aggregate expenditure. At very low nominal interest rates, the economy may approach a **liquidity trap**, represented in Figure 30.

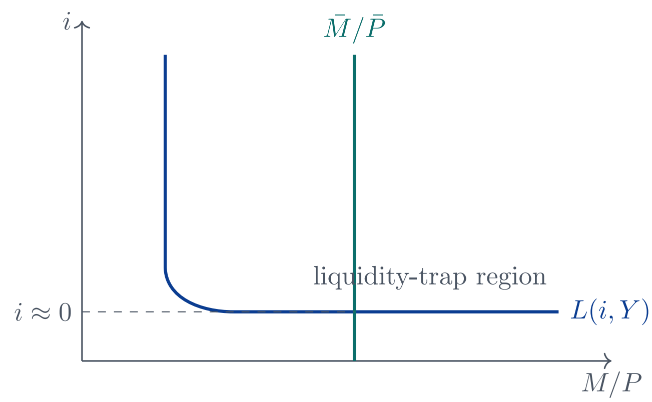

FIGURE 30

   In this region, the money demand curve is extremely flat: the public willingly absorbs additional money with almost no further decline in the nominal interest rate. Conventional monetary expansion therefore becomes much less effective.

2. The stability of money demand has changed over time. It appeared reasonably stable until the 1970s but became less stable thereafter. When the function $L(i, Y)$ shifts unpredictably, a fixed money supply target produces unstable interest rates and expenditure. This instability is one reason modern central banks emphasize interest rate targets rather than monetary aggregate targets.

4. Aggregate Supply and Stabilization Policy

Aggregate Supply and the Phillips Curve

The Phillips curve provides the foundation for aggregate supply. A. W. Phillips (1958) documented an inverse relationship between unemployment and wage growth in the United Kingdom. Wages tended to rise rapidly when unemployment was low and much more slowly when unemployment was high. Economists soon generalized this relationship from wage inflation to price inflation: in a tight labour market, firms compete more aggressively for workers, nominal wages rise more quickly, and firms pass at least some of those labour costs into prices. During the 1960s, many policymakers interpreted the apparently stable relationship as a permanent tradeoff between unemployment and inflation. That interpretation proved mistaken.

The analyses of Friedman (1968) and Phelps (1967) corrected the earlier view by emphasizing that workers and firms care about **real wages**, not only nominal wages. When people expect inflation, they demand higher nominal wages to preserve purchasing power. Wage growth and price inflation therefore depend on expected inflation as well as labour market tightness. In the long run, when wages and prices are flexible, the economy tends toward a **natural rate of unemployment**. Full employment does not mean zero unemployment because frictional and structural unemployment remain even in a healthy economy.

The resulting **expectations-augmented Phillips curve** is

$$\pi = \pi^e - \omega(U - U_n), \tag{0.44}$$

where π is inflation, π^e is expected inflation, U is the unemployment rate, U_n is the natural rate of unemployment, and $\omega > 0$ measures the sensitivity of inflation to the unemployment gap $U - U_n$.

When unemployment is below its natural rate, the unemployment gap is negative, so inflation exceeds expected inflation. When unemployment is above its natural rate, inflation is below expected inflation. In the long run, expected inflation adjusts to actual inflation and unemployment returns to U_n. There is therefore no permanent tradeoff between inflation and unemployment: nominal changes do not permanently alter real economic activity.

Figure 31 distinguishes the temporary tradeoff along a short-run Phillips curve from the vertical long-run Phillips curve.

Suppose expected inflation is 2% and the natural rate of unemployment is 5%. The economy

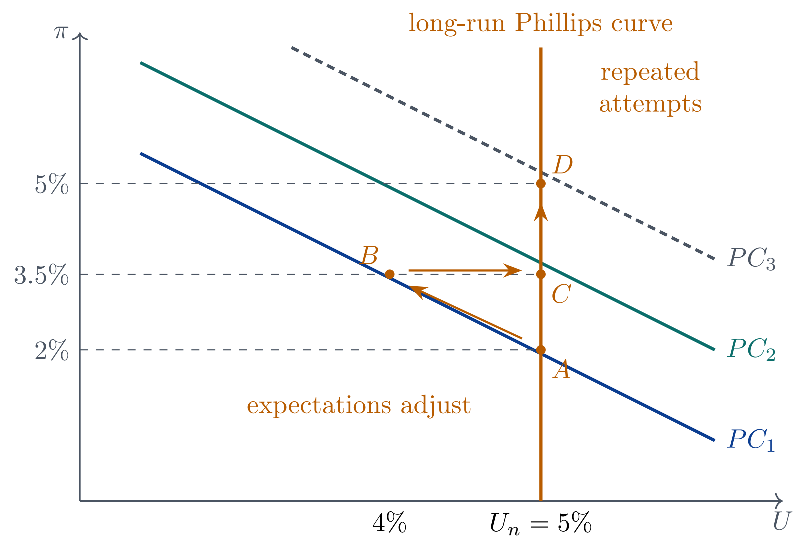

FIGURE 31

begins at point A on PC_1. If expansionary policy pushes unemployment to 4%, the economy moves along PC_1 to point B. In the short run, lower unemployment is obtained at the cost of higher inflation.

That improvement does not last. Once people observe higher inflation, they revise expected inflation upward. The short-run Phillips curve shifts from PC_1 to PC_2, and unemployment returns to its natural rate at point C. Repeated attempts to hold unemployment below U_n raise expected and actual inflation further, eventually moving the economy to a point such as D on PC_3. At both C and D, actual inflation equals expected inflation and unemployment equals its natural rate.

The diagram therefore establishes three conclusions:

1. The long-run Phillips curve is vertical, so there is no long-run tradeoff between inflation and unemployment.
2. There is a short-run tradeoff along any given short-run Phillips curve, as the movement from A to B demonstrates.
3. Each expected inflation rate generates a different short-run Phillips curve; PC_1, PC_2, and PC_3 must therefore be distinguished from the long-run curve.

By the 1970s, the simple Phillips curve relationship from the 1960s had broken down, as Friedman and Phelps had predicted. After a period of very low unemployment, expected inflation rose and the apparent tradeoff disappeared. Large oil shocks simultaneously raised inflation. Such a **price**

shock is a change in production costs or import prices that affects inflation independently of expected inflation and the unemployment gap. Cost-push wage shocks, in which nominal wage growth exceeds the rate justified by productivity growth, belong to the same category.

Adding the price shock term ρ gives the modern Phillips curve

$$\pi = \pi^e - \omega(U - U_n) + \rho. \tag{0.45}$$

This modern version of the Phillips curve assumes that wages and prices are sticky in the short run. If wages and prices were perfectly flexible, inflation would respond much more strongly to the unemployment gap, ω would be very large, and the short-run Phillips curve would become very steep. In the limiting case, it would become vertical and collapse into the long-run Phillips curve.

Under **adaptive expectations**, also called backward-looking expectations, people use the previous period's inflation to predict the current period, so $\pi^e = \pi_{-1}$. Equation (0.45) then has the two equivalent forms

$$\pi = \pi_{-1} - \omega(U - U_n) + \rho \qquad \text{and} \qquad \Delta\pi = \pi - \pi_{-1} = -\omega(U - U_n) + \rho. \tag{0.46}$$

This is the **accelerationist Phillips curve**. In the absence of a price shock, inflation accelerates when unemployment is below its natural rate and decelerates when unemployment is above it. Expectations adjust gradually, and many wage and price contracts are tied to past inflation, so both mechanisms contribute to short-run stickiness. The natural rate is consequently also called the **nonaccelerating inflation rate of unemployment**, or NAIRU.

Aggregate supply is the output counterpart of the Phillips curve. Figure 32 previews the transformation: Okun's law maps a low unemployment gap into a positive output gap, turning the downward-sloping short-run Phillips curve into an upward-sloping short-run aggregate supply curve.

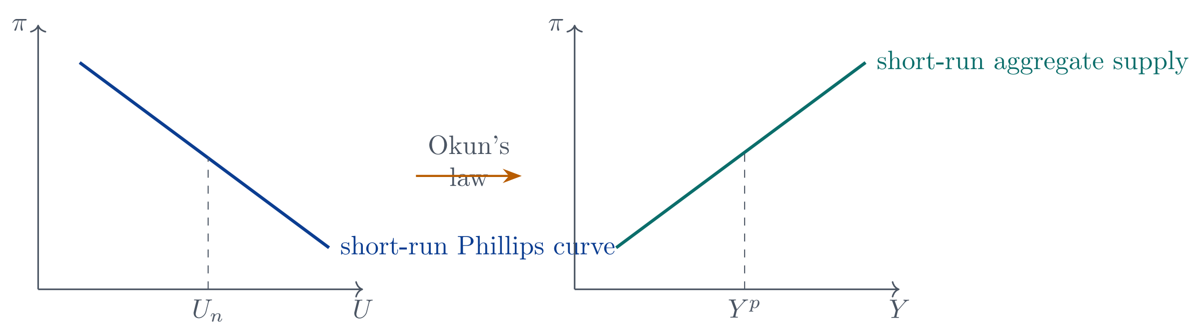

Figure 32

An aggregate supply curve relates output Y to inflation π. In the long run, technology, capital, and labour determine output. Output produced when unemployment equals its natural rate is **potential output**, denoted Y^p. Because inflation does not determine these real productive resources, the long-run aggregate supply curve is vertical at Y^p. It is the output counterpart of the vertical long-run Phillips curve.

To derive short-run aggregate supply, use **Okun's law** to replace the unemployment gap with the output gap. In a linear approximation,

$$U - U_n = -\kappa(Y - Y^p), \qquad \text{and} \qquad \kappa > 0. \tag{0.47}$$

The units of κ depend on how output is measured. When the output gap is a percentage deviation from potential, a common empirical rule of thumb is $\kappa \approx 0.5$: a one-percentage-point increase in output relative to potential is associated with unemployment roughly one-half of a percentage point below its natural rate.

Substituting (0.47) into the modern Phillips curve (0.45) gives the **short-run aggregate supply curve**:

$$\boxed{\pi = \pi^e + \gamma(Y - Y^p) + \rho,} \qquad \text{and} \qquad \gamma = \kappa\omega > 0. \tag{0.48}$$

Here $Y - Y^p$ is the output gap, ρ is the price shock, and γ measures the sensitivity of inflation to the output gap. Equation (0.48) is the Phillips curve written in terms of output rather than unemployment; the two formulations describe the same economic mechanism.

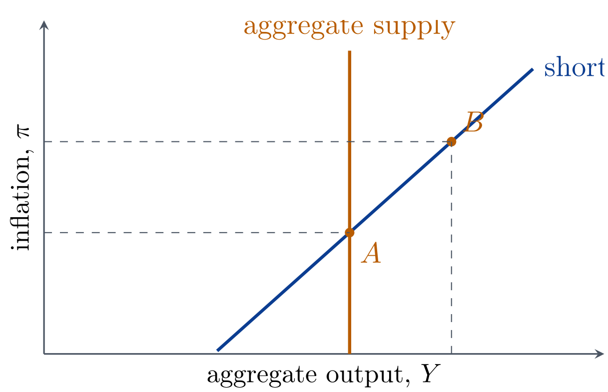

FIGURE 33

Example 0.12 Suppose last year's inflation rate was $\pi_{-1} = 2\%$, there is no supply shock, potential output is $Y^p = \$10$ trillion, and $\gamma = 1.5$ percentage points per trillion dollars. Find inflation when output is \$10 trillion and when it is \$11 trillion.

Solution. Under adaptive expectations, the short-run aggregate supply equation is

$$\pi = 2 + 1.5(Y - 10).$$

At potential output, $Y = 10$, so $\pi = 2\%$; this is point A in Figure 33. If output rises to $Y = 11$, then

$$\pi = 2 + 1.5(11 - 10) = 3.5\%,$$

which is point B. Higher output pushes unemployment below its natural rate, accelerates wage growth, and feeds into faster price growth. A larger γ makes inflation more sensitive to the output gap. With perfectly flexible wages and prices, γ would become extremely large and the short-run aggregate supply curve would approach the vertical long-run curve. □

Both long-run and short-run aggregate supply can shift.

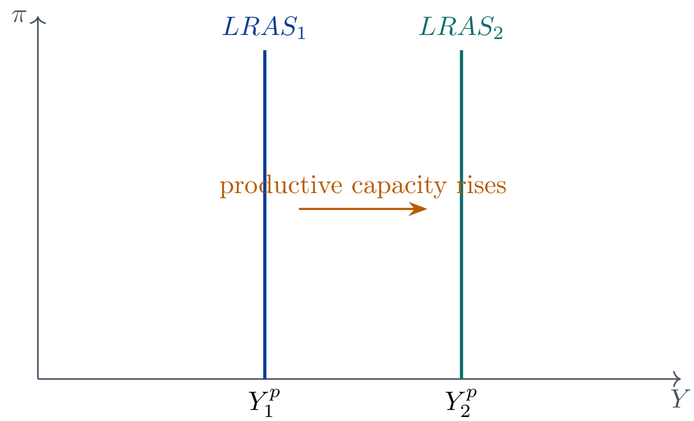

FIGURE 34

In the long run, $Y = F(A, K, L)$, so growth in technology, capital, or labour shifts long-run aggregate supply to the right, as in Figure 34. A change in the natural rate of unemployment also changes potential output. If the natural rate falls, the economy can sustain more employment and output, so long-run aggregate supply shifts right.

Short-run aggregate supply shifts for three main reasons:

1. Higher expected inflation raises inflation at every level of output, shifting short-run aggregate supply upward and to the left.
2. An adverse price shock, represented by a rise in ρ, also shifts short-run aggregate supply upward and to the left.

3. A persistent output gap changes inflation expectations over time. If output remains above potential, current inflation rises, expected inflation rises in the next period, and short-run aggregate supply shifts upward. If output remains below potential, inflation and expected inflation gradually fall, shifting short-run aggregate supply downward and to the right.

These expectation-driven shifts continue until output returns to potential.

The AD–AS Model

Because the AD–AS model contains both short-run and long-run aggregate supply, it distinguishes short-run from long-run equilibrium. **Short-run equilibrium** occurs where aggregate demand intersects short-run aggregate supply, as in Figure 35. At that point, planned expenditure equals output and inflation is consistent with current expectations, the output gap, and price shocks.

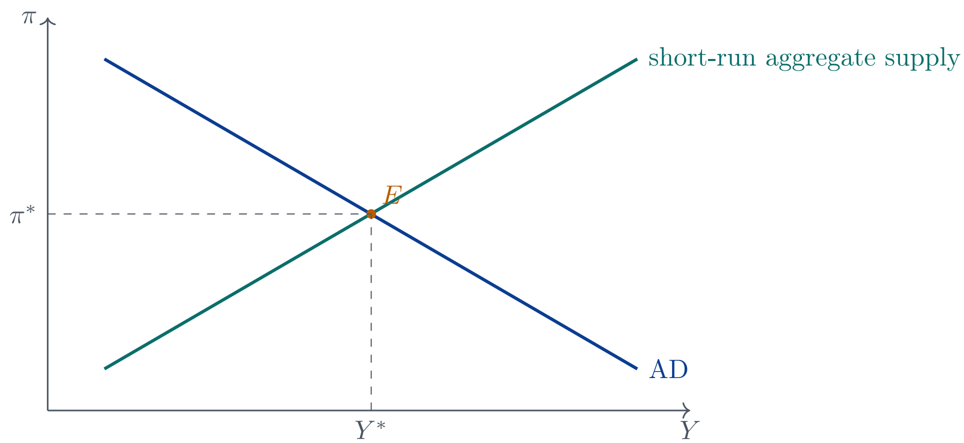

FIGURE 35

Example 0.13 Suppose aggregate demand is $Y = 11 - 0.5\pi$ and short-run aggregate supply is $\pi = 2 + 1.5(Y - 10)$. Find the short-run equilibrium.

Solution. Substituting the supply equation into aggregate demand gives

$$Y = 11 - 0.5\big[2 + 1.5(Y - 10)\big] = 17.5 - 0.75Y.$$

Thus, $1.75Y = 17.5$, so $Y^* = 10$. Substitution into either original equation gives $\pi^* = 2\%$. In this numerical case, short-run equilibrium also happens to be long-run equilibrium because $Y^* = Y^p = 10$. $\square$

In general, long-run equilibrium is more restrictive: it requires $Y^* = Y^p$ in addition to equality between aggregate demand and short-run aggregate supply. If short-run output differs from potential, wages, prices, and inflation expectations adjust until short-run aggregate supply shifts enough to close the output gap.

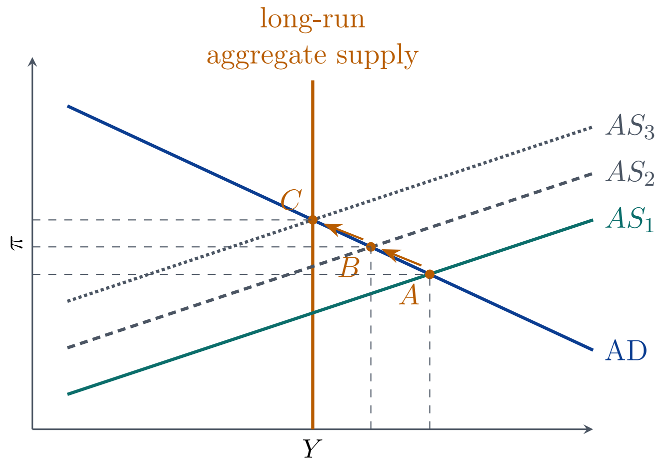

FIGURE 36

The economy begins at point A, where AD intersects AS_1 and $Y_1 > Y^p$. Unemployment is below its natural rate, so tight labour markets increase wage and price inflation. Expected inflation then rises, shifting short-run aggregate supply upward to AS_2 and moving the economy to point B. If output still exceeds potential, the process continues until the economy reaches point C, where AD, AS_3, and long-run aggregate supply intersect at Y^p. The reverse process operates when output begins below potential. This gradual movement back to potential is the economy's **self-correcting mechanism**, traced in Figure 36.

Demand shocks follow directly from the aggregate demand equation (0.41). A fall in the autonomous real policy rate $\bar{r}$ or taxes $\bar{T}$, a rise in autonomous consumption $\bar{C}$, or any other increase in autonomous expenditure shifts aggregate demand to the right.

In Figure 37, the economy initially rests at point A, where AD_1, AS_1, and long-run aggregate supply intersect. A positive demand shock shifts aggregate demand to AD_2 and moves the economy to point B in the short run. Output and inflation both rise, and output exceeds potential. Expected inflation subsequently rises, shifting short-run aggregate supply upward to AS_2 until the economy reaches point C. In the long run, output returns to potential and only inflation remains higher.

This experiment holds all other policies constant; in particular, it assumes that the central bank does not offset the demand shock. When policy responds, the adjustment path can differ.

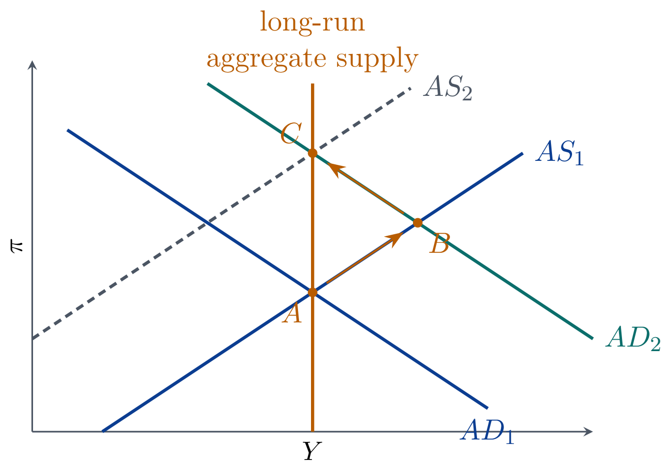

Figure 37

The Volcker disinflation is a clear example of a negative demand shock created by policy. In 1979, inflation was extremely high, so the Federal Reserve sharply tightened policy by raising the federal funds rate. Aggregate demand shifted left. In the short run, output and inflation both fell. Over roughly six years, inflation fell from about 13.5% to 1.9%, while unemployment rose from about 7.1% to 9.6% before falling back toward 7.0%.

Another example occurred in the early 2000s. The bursting of the technology bubble, the aftermath of the September 11 attacks, the Enron collapse, and related accounting scandals raised corporate borrowing rates and made investment more costly. Aggregate demand weakened. Unemployment rose from about 4.0% to 6.0% before easing to 5.5%, while inflation fell from about 3.4% to 1.6% and later recovered to approximately 2.7%.

Figure 38 summarizes a negative demand shock. The economy moves from A to B in the short run; with output below potential, expected inflation then falls and short-run aggregate supply shifts downward until the economy reaches C.

Supply shocks may be temporary or permanent, depending on whether they change potential output.

1. A **temporary adverse supply shock** raises production costs without changing potential output. Examples include a poor harvest, a temporary cost-push wage shock, or a temporary surge in import prices. In Figure 39, short-run aggregate supply shifts from AS_1 to AS_2 and the economy moves from A to B. Output falls, unemployment rises, and inflation rises. This combination of stagnation and inflation is **stagflation**. Because the shock is temporary, long-

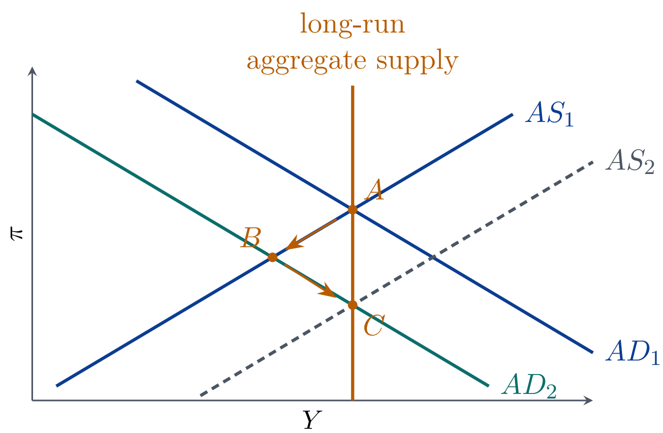

FIGURE 38

run aggregate supply does not move. If it dissipates before expected inflation has materially changed, short-run aggregate supply returns to AS_1 and the economy returns to A; otherwise, the expectation component of aggregate supply must adjust as well. A favourable temporary supply shock has the opposite effects.

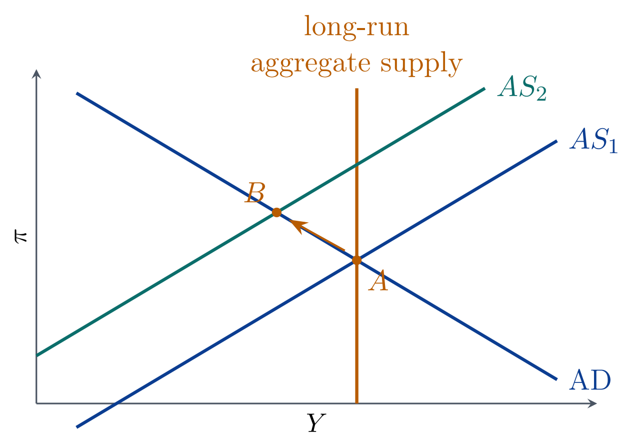

FIGURE 39

2. A **permanent adverse supply shock** reduces productive capacity. Burdensome new regulations, permanent losses of productive resources, or persistent increases in production costs can have this effect.

   In Figure 40, the economy initially moves from A to B, with lower output and higher inflation. Unlike a temporary shock, the permanent shock also shifts long-run aggregate supply from $LRAS_1$ to $LRAS_2$. Further adjustment of expected inflation shifts short-run aggregate supply to AS_3, producing the new long-run equilibrium C. Output is permanently lower and inflation is higher. The experiment holds monetary policy fixed, so aggregate demand does not move.

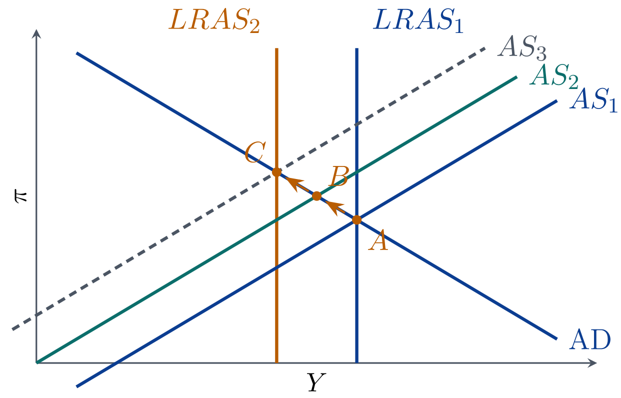

FIGURE 40

Favourable permanent supply shocks are also possible. In the late 1990s, lower health care costs, the computer revolution, and favourable demographic changes appear to have reduced the NAIRU in the United States and raised productive capacity. Long-run aggregate supply shifted right, short-run aggregate supply shifted downward and to the right, output rose, and inflation fell. This adjustment is illustrated in Figure 41.

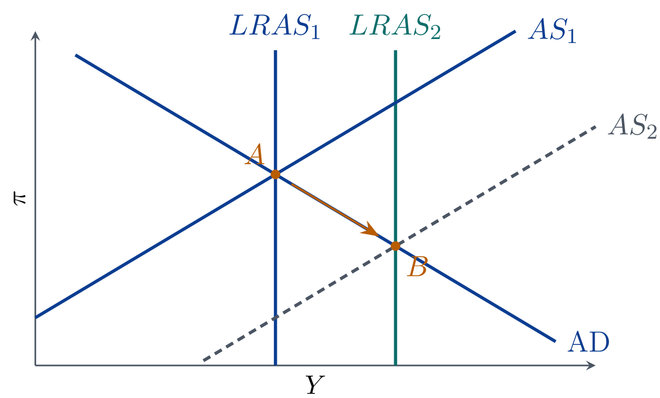

FIGURE 41

The AD–AS model yields four general conclusions:

1. The self-correcting mechanism tends to return the economy to potential output and the NAIRU over time.
2. A shift in aggregate demand affects output in the short run but not in the long run. It affects inflation in the short run and usually even more in the long run.

3. A temporary supply shock affects output and inflation only in the short run when aggregate demand remains fixed.
4. A permanent supply shock affects output and inflation in both the short run and the long run.

The 2007–2009 financial crisis combined adverse supply and demand shocks in a "perfect storm." Strong global oil demand and slowing production pushed oil prices toward $140 per barrel, creating an adverse supply shock. At the same time, the collapse of subprime credit produced a severe financial crisis: credit conditions tightened, and household expenditure and business investment fell sharply.

In Figure 42, the adverse supply shock shifts short-run aggregate supply from AS_1 to AS_2, while the demand shock shifts aggregate demand from AD_1 to AD_2. The economy moves from A to B, where output is lower, unemployment is higher, and inflation is higher. After 2008, oil prices fell sharply, so the supply shock largely reversed and aggregate supply moved back toward AS_1. The financial crisis nevertheless deepened, shifting aggregate demand farther left to AD_3. The economy moved to C, with even lower output and higher unemployment, but now with falling rather than rising inflation.

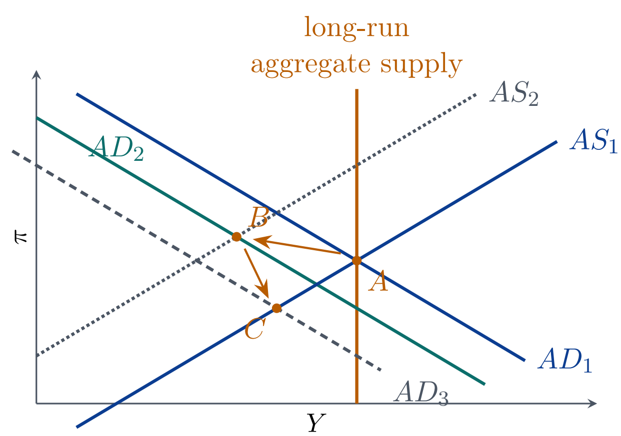

Figure 42

China's experience during 2007–2009 differed because export demand was central to its expansion.

Chinese growth had depended heavily on exports. When foreign demand collapsed in October 2008, China experienced a large negative demand shock. In Figure 43, aggregate demand shifts from AD_1 to AD_2 and the economy moves from A to B, with lower output and lower inflation. A large fiscal stimulus and easier monetary policy then moved aggregate demand rapidly back toward its original position.

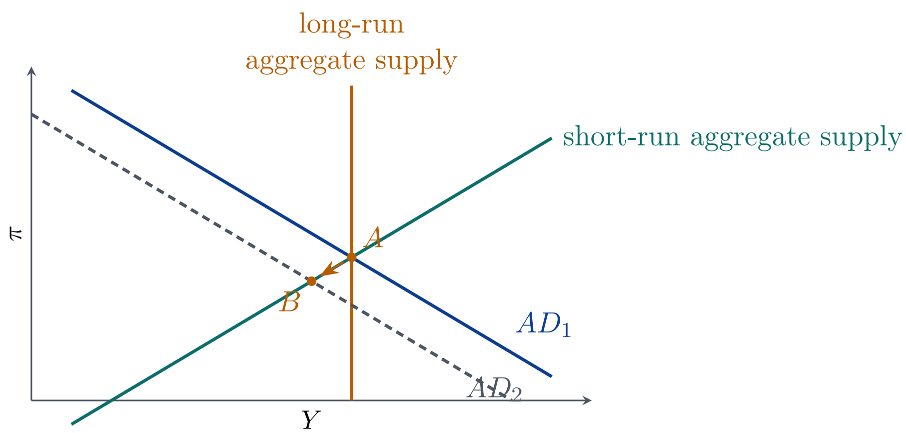

Figure 43

Macroeconomic Policy and AD–AS Analysis

Stabilization policy has two primary objectives:

1. Policymakers seek to stabilize economic activity by keeping output near potential and unemployment near its sustainable level.
2. Policymakers seek to stabilize inflation at a low positive rate.

Financial stability supports both objectives because a functioning financial system transmits policy and credit to the real economy. High unemployment causes hardship and represents lost output, but full employment does not mean zero unemployment. Frictional unemployment includes workers changing jobs and recent graduates entering the labour market, while structural unemployment reflects mismatches between workers' skills and employers' needs. Monetary policy can influence cyclical unemployment but has much less effect on frictional or structural unemployment. Policymakers therefore focus on the NAIRU. When unemployment is near that rate, output is ordinarily near potential and the output gap is close to zero. The NAIRU is difficult to measure and can change when, for example, labour market institutions or retraining programs reduce structural unemployment.

Price stability is also symmetric in practice. High and volatile inflation distorts price signals, complicates planning, and redistributes wealth unpredictably. Inflation that is too low raises the risk of deflation and leaves less room to reduce real interest rates when nominal rates approach zero. Central banks therefore aim for a low positive target, often between 1% and 3%, and seek to keep the inflation gap $\pi - \pi^T$ small.

Definition 0.14 Let

$$\widetilde{Y} = 100\left(\frac{Y - Y^p}{Y^p}\right)$$

denote the output gap in percentage points. This normalization puts the two stabilization gaps on comparable scales. A simple **policy loss function** is

$$\mathcal{L} = \alpha(\pi - \pi^T)^2 + (1-\alpha)\widetilde{Y}^2, \quad 0 \le \alpha \le 1. \tag{0.49}$$

The parameter α is the relative weight on inflation stabilization, while $1-\alpha$ is the weight on output stabilization.

A **hierarchical mandate** gives priority to inflation stability. The European Central Bank, the Bank of England, and the Bank of Canada are commonly described this way. A **dual mandate**, such as the one used by the U.S. Federal Reserve, gives sustainable employment and stable prices equal standing. The practical question is whether these goals usually reinforce each other or whether policymakers often have to trade one off against the other.

The monetary benchmark connecting these objectives is the **natural real interest rate**, denoted r^*. It is the equilibrium real interest rate consistent with $Y = Y^p$ and $\pi = \pi^T$. On the MP curve, r^* is associated with target inflation; in the AD–AS diagram, the same target is associated with potential output, as shown in Figure 44. If the actual real policy rate is below r^*, aggregate demand is too strong, output tends to exceed potential, and inflation tends to rise. If it is above r^*, demand is too weak and output tends to fall below potential. Central banks devote considerable effort to estimating this unobservable benchmark.

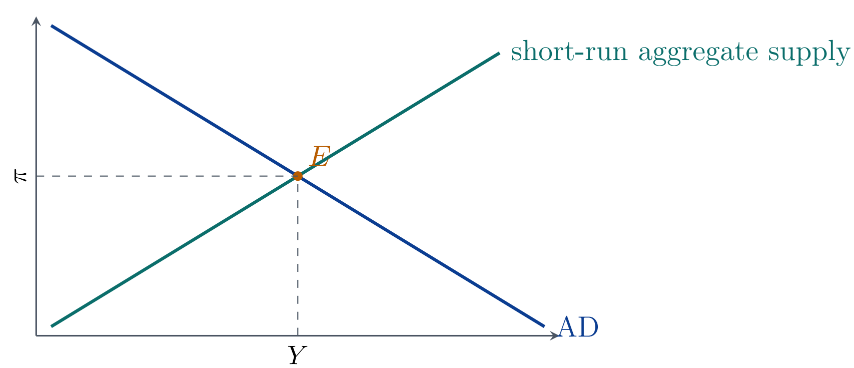

Figure 44

Consider first a negative aggregate demand shock.

A negative demand shock shifts aggregate demand from AD_1 to AD_2 in Figure 45, moving the economy from A to B. Output falls below potential and inflation falls below target. The monetary

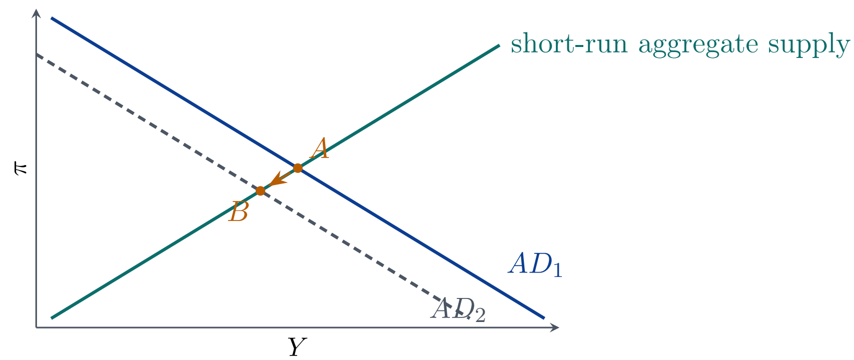

FIGURE 45

policy diagrams describe two possible responses.

If the central bank makes no autonomous policy change, both output and inflation fall. Economic slack gradually causes wages, prices, and expected inflation to adjust; short-run aggregate supply shifts downward, and output eventually returns to potential at a lower inflation rate. The real policy rate moves downward along the unchanged MP curve as inflation falls, as shown in Figure 46. This self-correction can be slow, and the persistent recession and inflation shortfall are costly. Fiscal policy could support demand, but it ordinarily operates more slowly.

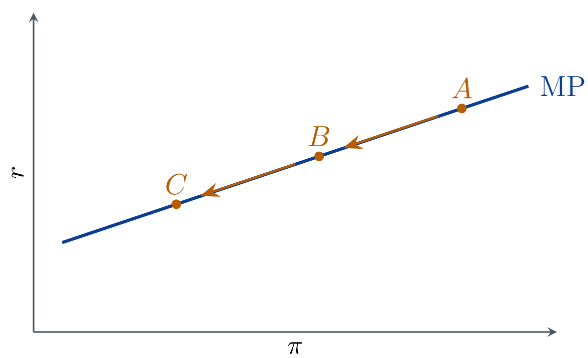

FIGURE 46

Alternatively, an autonomous monetary easing shifts the MP curve downward from MP_1 to MP_2 in Figure 47. The lower real rate stimulates expenditure and restores aggregate demand to its original position. Output returns to potential and inflation returns to target without waiting for slow wage and price adjustment.

The natural real interest rate is the same r^* that equates desired saving and investment in (0.15). Monetary policy can move the actual real rate around this benchmark but cannot determine its long-run value. A pure demand shock therefore creates a **divine coincidence**: policy that

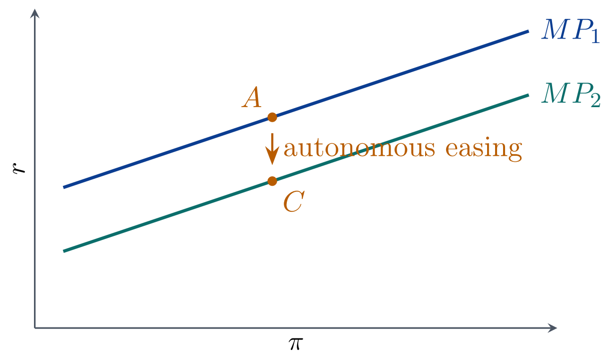

Figure 47

stabilizes inflation also stabilizes output.

An adverse permanent supply shock requires a different adjustment because potential output itself falls. With no autonomous monetary response, the economy moves from A to B in the short run and then to C as short-run aggregate supply continues to adjust. Output is permanently lower, inflation is higher, and the natural real interest rate is assumed to be higher. The new long-run equilibrium nevertheless has a zero output gap because actual output equals the new, lower potential output. The associated movement along the MP curve appears in Figure 48.

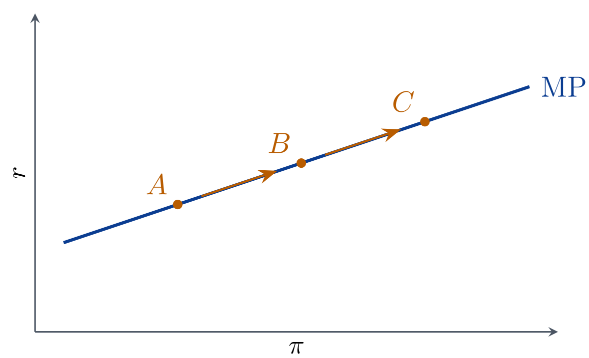

Figure 48

The central bank can instead tighten autonomously to stabilize inflation: The higher real rate shifts aggregate demand left and returns inflation to target at point C, while output equals the new, lower potential level, as shown in Figure 49. For a permanent supply shock, stabilizing inflation is therefore compatible with eliminating the output gap. Once again, the two objectives do not truly conflict.

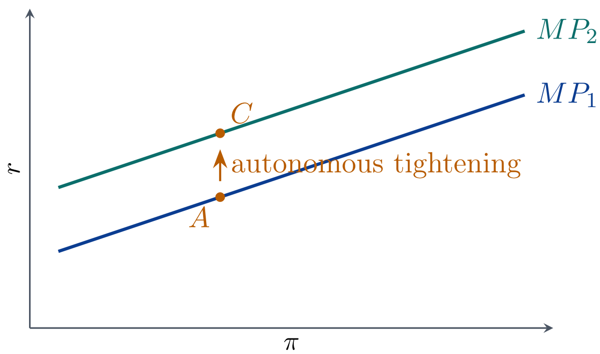

Figure 49

Responses to Temporary Supply Shocks

An adverse temporary supply shock raises inflation while reducing output, so policymakers cannot simultaneously eliminate both the inflation gap and the output gap. They have three broad options.

The first option is to make no autonomous policy change. The MP curve and long-run aggregate supply remain fixed. Short-run aggregate supply shifts upward, moving the economy from A to B, where inflation is above target and output is below potential. The policy rate rises only by moving along the unchanged MP curve, as indicated in Figure 50. If the shock disappears, short-run aggregate supply returns to its original position and the economy returns to A. This strategy avoids an additional policy disturbance, but the short run may be painful and its duration is unknown.

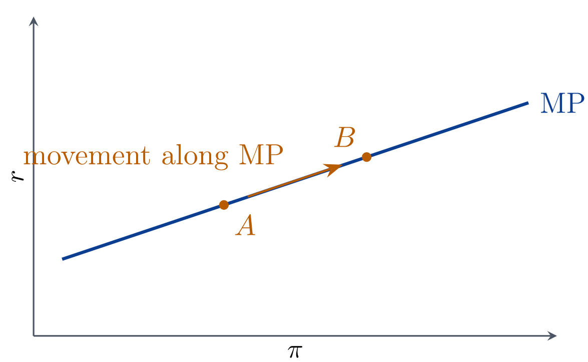

Figure 50

The second option is to give priority to short-run inflation stability by tightening monetary policy.

The central bank raises the real rate and shifts the MP curve upward, as in Figure 51. Lower

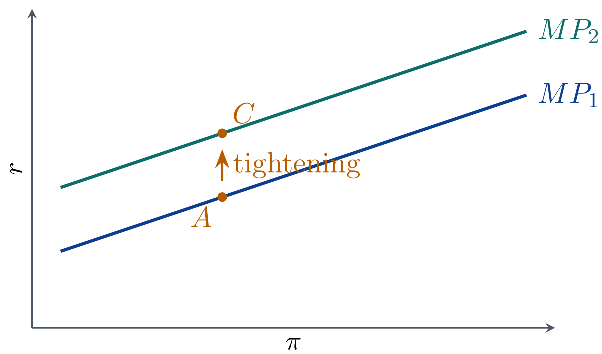

Figure 51

expenditure shifts aggregate demand left, holding inflation at target but pushing output even farther below potential. The additional slack eventually shifts short-run aggregate supply downward. Once the temporary shock disappears, the central bank must reverse the tightening. Inflation is stabilized at the cost of a larger output gap and a deeper short-run downturn.

The third option is to give priority to short-run economic activity by easing monetary policy.

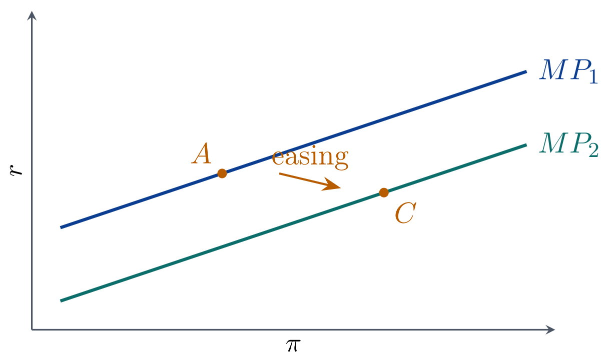

Figure 52

The central bank shifts the MP curve downward in Figure 52, which raises aggregate demand. Output remains closer to potential despite the adverse supply shock, but inflation rises further. The economy moves from the initial point A, through the shock point B in the AD–AS diagram, to point C, where activity is stabilized at the cost of a larger inflation gap.

Figure 53 compares the three responses.

Unlike demand shocks and permanent supply shocks, a temporary supply shock creates a genuine short-run policy tradeoff. A central bank that frequently faces such shocks must decide how much weight to place on each term in the loss function (0.49).

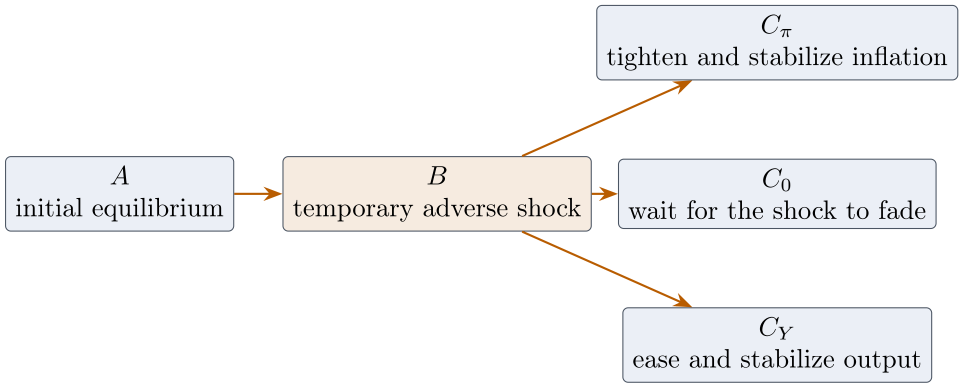

FIGURE 53

Policy Activism and Taylor Rules

Suppose unemployment rises because aggregate demand falls. **Nonactivists**, often associated with the classical tradition, argue that wages and prices are flexible enough for self-correction to work quickly. On this view, short-run aggregate supply will adjust and unemployment will return to its natural rate without much policy intervention. **Activists**, often associated with Keynesian economics, argue that wages and prices are sticky, so self-correction is slow and policy should respond.

The case for activism is weakened by policy lags. Policymakers face a *data lag*, since good macroeconomic data arrive with delay. They face a *recognition lag*, because it takes time to determine whether incoming data really signal a recession. Fiscal policy also faces a *legislative lag*, since tax and spending measures must pass through the political process, and an *implementation lag*, since approved measures still take time to administer. Finally, both fiscal and monetary policy face an *effectiveness lag*: even after a policy is in place, its effects on the economy may take many months to appear and can vary in size. Because of these lags, policy intervention may arrive after self-correction has already begun and may therefore overshoot in the wrong direction.

The 2009 policy debate illustrates the issue. The United States entered a deep recession with unemployment already around 7% and rising. Monetary policy was aggressive, but the federal funds rate was already near zero. Activists argued that this made large fiscal stimulus necessary to raise aggregate demand. Nonactivists argued that fiscal policy would act too slowly and might destabilize output and inflation if the economy had already started to recover by the time the stimulus took effect. The effectiveness of that stimulus remains debated.

A more explicit version of the Taylor principle is the **Taylor rule**. Use the percentage output gap $\widetilde{Y}$ defined above. If inflation and interest rates are measured in percentage points, a benchmark

Taylor rule is

$$\underbrace{i}_{\substack{\text{nominal policy}\\ \text{interest rate}}} = \underbrace{\pi}_{\text{inflation}} + \underbrace{2}_{\substack{\text{equilibrium real}\\ \text{interest rate}}} + \underbrace{\frac{1}{2}(\pi - \pi^T)}_{\text{inflation gap}} + \underbrace{\frac{1}{2}\widetilde{Y}}_{\text{output gap}} . \tag{0.50}$$

The value 2 represents a historical benchmark for the equilibrium real interest rate, while the two coefficients of 1/2 give equal weight to the inflation and output gaps. When inflation rises by one percentage point while the target and output gap remain unchanged, (0.50) raises the nominal policy rate by 1.5 percentage points. The real rate therefore rises by 0.5 percentage points, as required by the Taylor principle.

The rule is more specific than the principle because it responds to both inflation and economic activity and assigns an exact policy rate to every observed pair of gaps. Although the rule has sometimes tracked actual policy decisions reasonably well, central banks do not follow it mechanically. Inflation and potential output are measured imperfectly and are often revised; suitable coefficients may change over time; and policymakers respond to forecasts rather than only to current observations. The Taylor rule is therefore best interpreted as a benchmark against which a policy decision can be evaluated.

The Taylor rule and the monetary policy curve express the same central idea: the central bank raises the real interest rate when inflation rises. Both frameworks also allow an autonomous adjustment in response to economic conditions. For example, after a temporary adverse supply shock, the bank can lower the real rate to support output, but only by accepting a larger inflation gap.

Remark 0.15 The Taylor rule does not eliminate the need for judgment. A formula cannot by itself resolve the tradeoff between inflation and output stability created by a temporary supply shock.

The choice of a long-run inflation target is ultimately a monetary policy decision. To see how a higher target is implemented, suppose that the central bank raises its target from π_1^T to π_2^T.

The central bank first eases policy by shifting the monetary policy curve from MP_1 to MP_2. The real interest rate falls at the initial inflation rate, expenditure rises, and the aggregate demand curve shifts from AD_1 to AD_2. The economy consequently moves from A to B: output temporarily exceeds potential, and inflation rises from π_1^T to π_2. The positive output gap then causes the short-run aggregate supply curve to shift upward from AS_1 to AS_2. The economy ultimately reaches C, where output has returned to potential, inflation equals the new target π_2^T, and the real interest rate has returned to its equilibrium value r^*. Figure 54 shows both stages of this transition.

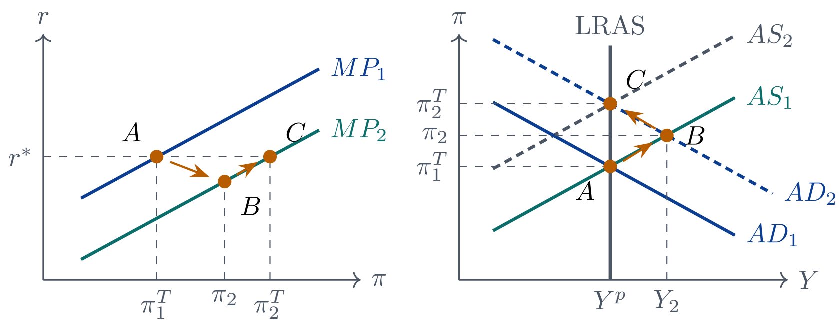

FIGURE 54

> **Remark 0.16** Monetary policy determines inflation in the long run, but it does not determine potential output or the long-run equilibrium real interest rate. This is an application of long-run monetary neutrality.

Persistent inflation can arise through either of two closely related mechanisms.

1. **Cost-push inflation** begins when an adverse supply shock or an increase in nominal wage demands raises firms' production costs. The short-run aggregate supply curve shifts upward, which would ordinarily reduce output below potential. If policymakers accommodate the shock by shifting aggregate demand from AD_1 to AD_2, they can keep output at potential, but they validate a higher inflation rate. Workers may then demand another increase in nominal wages to preserve their purchasing power. Repeated upward shifts of short-run aggregate supply, followed by accommodating shifts of aggregate demand, generate the inflationary spiral in Figure 55.

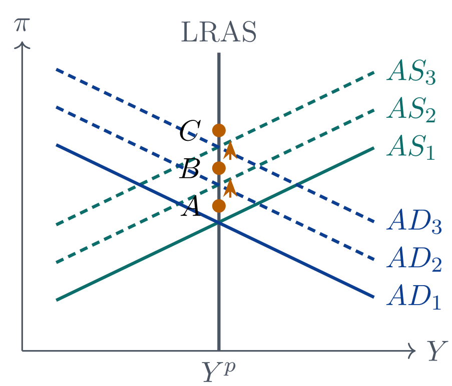

FIGURE 55

2. **Demand-pull inflation** begins when policymakers repeatedly expand aggregate demand in

an attempt to keep unemployment below its natural rate or output above potential. A target $U^T < U_n$ is equivalent to an output target $Y^T > Y^p$. The initial shift from AD_1 to AD_2 moves output temporarily toward Y^T, but it also raises inflation. Because output exceeds the economy's productive capacity, the short-run aggregate supply curve subsequently shifts upward and returns output to potential. Achieving the excessive output target again requires another expansion of aggregate demand, so inflation continues to rise, following the cycle in Figure 56.

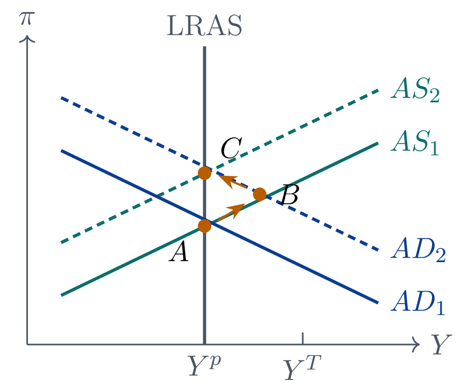

FIGURE 56

In principle, demand-pull inflation is associated with unemployment below the nonaccelerating inflation rate of unemployment, whereas cost-push inflation can occur while unemployment is above that rate. In practice, the natural rate is difficult to estimate and the two mechanisms interact. Demand-pull inflation can induce cost-push inflation when rising expected inflation leads workers to demand higher nominal wages. Actual inflation episodes therefore usually contain elements of both mechanisms.

The **Great Inflation** in the United States from 1965 to 1982 illustrates this interaction. Policymakers appear to have targeted unemployment of approximately 4% when the natural rate was probably closer to 5% or 6%. Expansionary policy pushed aggregate demand beyond a sustainable level, inflation rose, expectations adjusted, nominal wage demands increased, and the process became self-reinforcing. Under Federal Reserve Chair Paul Volcker, the central bank eventually broke the cycle by raising nominal interest rates sharply.

Remark 0.17 Some central banks have at times moved slightly below zero in nominal terms, but the core issue remains that conventional rate cuts become severely constrained near zero.

Monetary Policy at the Zero Bound

The **zero lower bound** constrains conventional monetary policy because nominal interest rates cannot be pushed far below zero. At a sufficiently negative interest rate, depositors can avoid the charge by holding currency, which pays a nominal return of zero. Some central banks have nevertheless implemented slightly negative policy rates, so the more precise modern term is the **effective lower bound**. The following analysis uses zero as the relevant bound.

Suppose that expected inflation moves closely with actual inflation. At point A, inflation and expected inflation are 3%, the real interest rate is 2%, and output equals potential. The approximate Fisher equation implies that the nominal interest rate is about 5%. If inflation falls to 2%, the desired policy response may call for a real interest rate of about -2%, which requires a nominal policy rate of approximately zero. The economy is then at point B, and conventional interest rate policy has reached its lower bound.

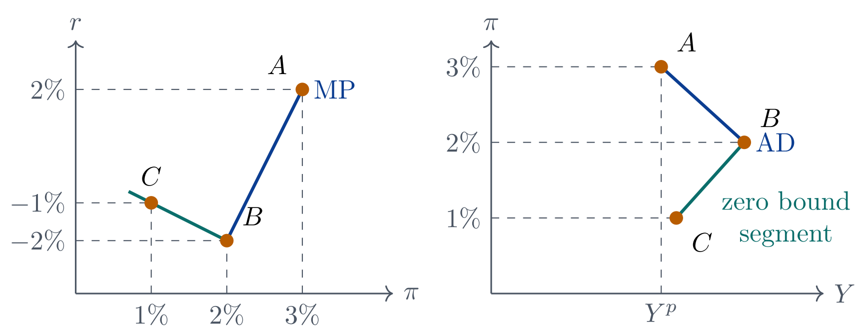

Figure 57

Figure 57 shows how the monetary policy and aggregate demand curves bend at the zero bound. From A to B, lower inflation is accompanied by a sufficiently large reduction in the nominal policy rate, so the real interest rate falls and aggregate demand rises. Once the nominal rate is fixed at zero, however, the approximate Fisher equation becomes

$$r \approx -\pi^e. \tag{0.51}$$

The exact Fisher relation gives $r = -\pi^e/(1+\pi^e)$ when $i = 0$, so (0.51) is accurate for modest inflation rates. If expected inflation falls from 2% to 1%, the actual real interest rate rises from approximately -2% to approximately -1%, even if the central bank would prefer a rate of -2% or less. The monetary policy curve therefore bends backward between B and C: lower inflation now raises the real interest rate. The aggregate demand curve inherits the kink. Between A and B, lower inflation raises output in the usual way; between B and C, lower inflation raises the real interest rate and reduces output.

This creates a dangerous situation when the economy is already weak. An adverse shock can push output below potential, lower inflation, and thereby raise the real interest rate. The higher real interest rate then reduces demand further, which lowers inflation again. The usual self-correcting mechanism no longer works.

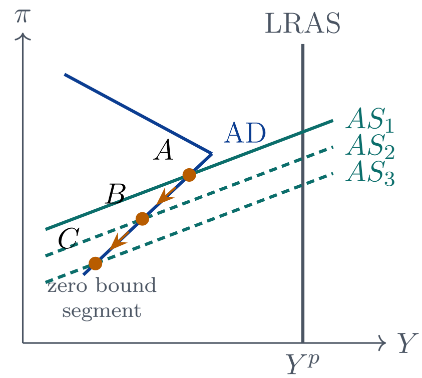

FIGURE 58

At point A in Figure 58, output is already below potential. Ordinarily, the resulting slack would shift short-run aggregate supply downward and help output recover. At the lower bound, a shift from AS_1 to AS_2 instead moves the economy from A to B: both inflation and output fall. Lower inflation raises the real interest rate through (0.51), depresses demand, and creates still more slack. Further short-run aggregate supply adjustment moves the economy toward C. This **deflationary spiral** reverses the usual self-correcting mechanism, so the economy need not return to long-run aggregate supply on its own.

When conventional interest rate policy is exhausted, the central bank can turn to **unconventional monetary policy**. The real borrowing cost faced by a firm or household can be represented as

$$r^{\text{borrow}} = r + f, \tag{0.52}$$

where f is a financial premium. It may include a credit spread that compensates lenders for risk and a term premium that separates long-term rates from the expected path of short-term rates. Even when the central bank cannot reduce r through another cut in its policy rate, it can stimulate expenditure by reducing f.

1. **Liquidity provision** is especially important when credit markets seize up and financial premiums rise sharply. The central bank can lend against collateral or otherwise supply liquid funds to financial institutions. If this intervention restores lending and reduces f, the borrowing cost in (0.52) falls and aggregate demand rises.
2. **Asset purchases** raise the prices and lower the yields of the securities purchased. Purchases

of long-term government bonds can reduce the term premium, while purchases of riskier private securities can compress credit spreads. Either channel can reduce the longer-term borrowing rates relevant to investment even when the short-term policy rate is already at zero.

3. **Managing expectations**, often called forward guidance, means committing to keep the policy rate low for longer than markets would otherwise expect or committing to a higher future inflation path. A credible promise of low future short-term rates lowers long-term rates today. A credible promise of higher inflation lowers the real interest rate. Both channels shift aggregate demand to the right.

Figure 59 illustrates the first two channels. A reduction in financial premiums shifts aggregate demand from AD_1 to AD_2 and moves the economy from A to B. In the favourable case shown, output returns to potential.

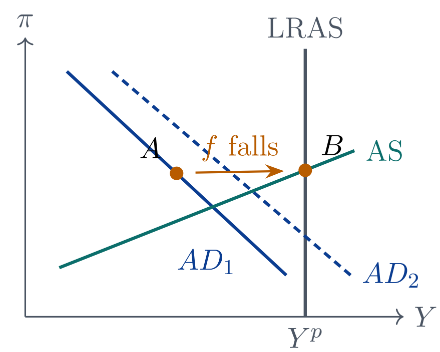

FIGURE 59

Expectations management can also exploit the unusual slope of aggregate demand at the lower bound. Suppose the central bank credibly commits to a higher inflation target. The increase in expected inflation shifts short-run aggregate supply upward from AS_1 to AS_2 in Figure 60. With the nominal interest rate fixed at zero, higher expected inflation also reduces the real interest rate and raises aggregate expenditure. The diagram folds this expenditure response into its upward-sloping zero-bound aggregate demand segment, so the economy moves along that segment from A to B, where output has returned to potential. If expected inflation were instead treated as a separate parameter on the aggregate demand curve, the same real-rate channel would appear as a rightward shift of that curve. This policy requires credibility: an announcement that does not change expectations will not change the real interest rate.

Liquidity provision and asset purchases expand the central bank's balance sheet and usually increase the monetary base. These interventions are often grouped under **quantitative easing**. Their effectiveness depends on whether they reduce the borrowing rates relevant to expenditure.

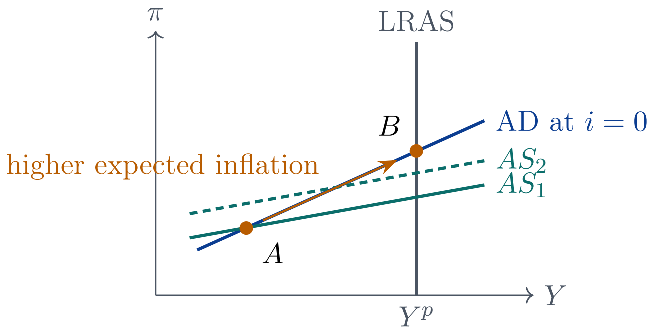

FIGURE 60

Purchasing short-term government securities may have little effect when those securities and reserves are already close substitutes and the purchases do not change credit spreads or long-term rates. For this reason, the Federal Reserve emphasized the term **credit easing**: the objective was not merely to enlarge the balance sheet, but to change its composition in a way that reduced private borrowing costs.

Abenomics provides an important application. When Japan changed its monetary policy strategy in 2013, the economy was experiencing weak growth, deflation, and a policy rate at its lower bound. The Bank of Japan raised its inflation target from 1% to 2% to change expectations. It also began a massive asset purchase program that shifted emphasis away from short-term government securities and toward long-term government bonds and some private assets, including real estate investment trusts. The intended channels were a lower term premium, smaller credit spreads, and higher expected inflation.

By 2015, the results remained disappointing. Inflation was still below target, household consumption had not grown strongly, lower interest rates had not produced a large increase in bank lending, private investment remained sluggish, and export growth was weakening. The episode illustrates that unconventional instruments can alter financial prices without necessarily producing a proportionate increase in borrowing and expenditure.

Remark 0.18 Unconventional policy can help at the effective lower bound, but its effects are uncertain. Success depends on financial conditions, central bank credibility, and the willingness of households and firms to borrow and spend.

5. Fiscal Policy and the Government Budget

The Government Budget

Fiscal policy includes the spending and revenue decisions of every level of government. In the United States, for example, the federal government accounts for roughly half of government spending, while state and local governments account for the remainder. In 2012, total government spending was approximately 45% of gross domestic product in Canada and 40% in the United States, although the exact share depends on which expenditures and levels of government are included.

Government outlays have four main components.

1. **Government purchases** are payments for currently produced goods and services. They can be decomposed as
$$G = G_I + G_C, \tag{0.53}$$
where G_I is public investment in assets such as schools, roads, and other infrastructure, and G_C consists of current purchases such as national defence and routine operating expenditures.
2. **Transfers** are payments to individuals that are not made in exchange for currently produced goods and services. Employment insurance, social assistance, and old-age benefits are examples. Many transfer programs are established in legislation and are therefore described as entitlements.
3. **Grants in aid** are transfers from one level of government to another, such as federal assistance to provincial, state, or local governments. A grant is an expenditure for the government that provides it and revenue for the government that receives it. It cancels when the budgets of all levels of government are consolidated.
4. **Net interest payments** equal the government's interest expenses less the interest income it receives on financial assets and loans.

Government revenue comes primarily from taxes. Its main components are personal taxes on income and property, contributions to social insurance programs, taxes on production and imports such as sales taxes and tariffs, and corporate income taxes on business profits.

Let R_t denote total government revenue, TR_t transfers, and NIP_t net interest payments. For the consolidated government sector, the overall budget deficit is

$$D_t = G_t + TR_t + NIP_t - R_t. \tag{0.54}$$

A positive D_t is a deficit, while a negative D_t is a surplus. The net tax variable used in the saving identities is $T_t = R_t - TR_t - NIP_t$, so (0.54) is equivalently $D_t = G_t - T_t = -S_{G,t}$ by (0.11).

When an overall deficit is financed entirely by issuing bonds, the government's one-period budget constraint is

$$D_t = B_t - B_{t-1} = \Delta B_t, \tag{0.55}$$

where B_t is the nominal stock of government debt at the end of period t. Raising taxes or reducing outlays shrinks the deficit; any remaining deficit must be financed. Bond financing increases outstanding debt according to (0.55). Monetary financing is introduced separately below.

Government Debt

The deficit is a flow over a period of time, whereas government debt is a stock at a point in time. Because the capacity to service debt depends on national income, debt is usually reported as a share of nominal gross domestic product. If P_tY_t is nominal gross domestic product, the debt-to-GDP ratio is

$$b_t = \frac{B_t}{P_tY_t}. \tag{0.56}$$

Both the numerator and the denominator of (0.56) are nominal. Let γ_t be the growth rate of nominal gross domestic product and let $d_t = D_t/(P_tY_t)$ be the deficit as a share of current nominal gross domestic product. Combining (0.55) and (0.56) gives

$$b_t = \frac{b_{t-1}}{1+\gamma_t} + d_t, \tag{0.57}$$

$$b_t - b_{t-1} = d_t - \frac{\gamma_t}{1+\gamma_t} b_{t-1}. \tag{0.58}$$

Thus, holding nominal gross domestic product fixed, a deficit raises the debt ratio; holding nominal debt fixed, real growth and inflation raise nominal gross domestic product and reduce the ratio. A country can run a deficit while its debt ratio falls if nominal income grows sufficiently quickly. Conversely, the ratio can rise sharply during a recession even when the government reduces its deficit.

Debt ratios vary widely across countries and across accounting definitions. Estimates for gross government debt in 2011 were approximately 106% of gross domestic product in Canada, 103% in the United States, and 228% in Japan. Comparisons must distinguish gross from net debt and must use the same definition of the government sector.

Entitlement programs create an additional long-run fiscal challenge. The United States established Social Security in 1935 as a pay-as-you-go pension system: current workers' contributions finance current retirees rather than being saved in personal accounts for those same workers. Public pensions and health programs such as Medicare and Medicaid face similar demographic pressure. As people live longer and the ratio of retirees to workers rises, a given contribution rate finances a smaller benefit per retiree. Proposed reforms generally involve some combination of investing funds for a higher expected return, raising contributions or taxes, and reducing benefits. Each option entails economic risk or political cost.

A **sovereign debt crisis** is a collapse in the market for a government's bonds caused by doubts about repayment. The mechanism can become self-reinforcing, as shown in Figure 61.

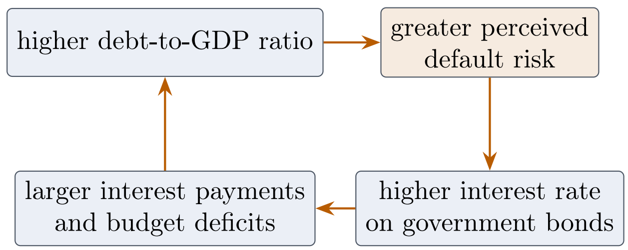

FIGURE 61

As the debt ratio rises, investors may demand a larger default risk premium. The higher interest rate raises net interest payments and future deficits, causing the debt ratio to rise still further. Several European countries experienced this dynamic during the euro area crisis. Membership in the monetary union had initially caused their interest rates to converge toward the low rates of safer borrowers. The subsequent recession reduced output and tax revenue, while financial sector bailouts raised government spending. These forces were especially severe in Greece and other heavily indebted members of the euro area.

Fiscal Policy in the Long Run

There is a long-standing debate about whether public debt burdens future generations. Debt need not impose a net burden when it finances productive public capital or human capital. If the resulting increase in productivity raises future wages and tax revenue by enough, future citizens inherit both the liability and an offsetting productive asset. Moreover, repayment of domestically held debt transfers resources from domestic taxpayers to domestic bondholders rather than sending resources out of the country.

Several considerations nevertheless make high public debt costly.

1. A deficit reduces government saving. From (0.12) and (0.13),

$$S = \underbrace{(Y - T - C)}_{S_P} + \underbrace{(T - G)}_{S_G} = I + NX. \tag{0.59}$$

   If government saving becomes more negative while private saving is unchanged, domestic investment, net exports, or both must fall. The reduction in investment caused by government borrowing is called **crowding out**. It leaves a smaller future capital stock and lowers future output.

2. Government borrowing does not necessarily finance productive investment. When it finances current purchases, transfers, or poorly selected projects, future taxpayers inherit the liability without an offsetting increase in productive capacity.

3. If foreigners hold part of the debt, future interest and principal payments transfer resources abroad. The same decline in national saving can reduce net exports through (0.14), because foreign purchases of domestic bonds provide the capital inflow that finances a current account deficit.

4. Debt finance redistributes income. The taxpayers who finance debt service need not be the households that own government bonds, and bondholders tend to be wealthier than the average taxpayer. Debt service can therefore increase after-tax income inequality.

5. Some countries exhibit **debt intolerance**: investors fear repudiation or default and demand a large risk premium at debt ratios that other governments can sustain. The feedback in Figure 61 can then begin at a comparatively low debt ratio.

6. Servicing a large debt may require high future tax rates. Taxes create wedges between private and social returns, so they can distort decisions to work, save, and invest. These distortions reduce economic efficiency and can slow long-run growth.

The principle of **tax smoothing** qualifies the case against deficits. Because the efficiency cost of taxation generally rises more than proportionately with the tax rate, the government should avoid large, temporary changes in tax rates. A temporary surge in spending can therefore justify a temporary deficit followed by later surpluses. This argument is much weaker when the spending increase is permanent: without an eventual permanent increase in revenue or reduction in other outlays, the debt stock continues to rise.

Fiscal Policy in the Short Run

Expansionary fiscal policy consists of an increase in government purchases, a reduction in taxes, or an increase in transfers. Each measure raises planned expenditure and shifts aggregate demand to the right. Starting from a recessionary equilibrium such as A in Figure 62, an expansion moves the economy to B, raising both output and inflation. Contractionary fiscal policy has the opposite effects.

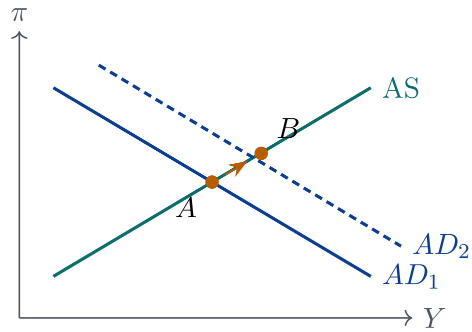

FIGURE 62

The simplest Keynesian multiplier follows from the goods market equilibrium condition

$$Y = \bar{A} + m(Y - T) + G,$$

where $m \in (0, 1)$ is the marginal propensity to consume and $\bar{A}$ collects the remaining autonomous expenditures. Holding $\bar{A}$ fixed and taking changes gives

$$(1 - m)\Delta Y = -m\Delta T + \Delta G.$$

The government purchases and tax multipliers are therefore

$$\frac{\Delta Y}{\Delta G} = \frac{1}{1-m} > 1 \qquad \text{and} \qquad \frac{\Delta Y}{\Delta T} = -\frac{m}{1-m}. \tag{0.60}$$

An increase of one dollar in government purchases enters planned expenditure directly. A tax cut of one dollar affects consumption only through the fraction m that households spend, so the tax multiplier is smaller in absolute value.

Equation (0.60) holds the real interest rate fixed. In the full aggregate demand model, a fiscal expansion raises inflation, and the central bank responds under the Taylor principle by raising the real interest rate. Let $b = c + d + x > 0$ denote the total interest sensitivity of consumption, investment, and net exports. Combining the aggregate demand equation (0.41) with the short-run aggregate supply equation (0.48) yields

$$(1-m)\Delta Y = \Delta G - b\lambda\Delta\pi \qquad \text{and} \qquad \Delta\pi = \gamma\Delta Y,$$

when expected inflation, potential output, and price shocks are fixed. Hence,

$$\frac{\Delta Y}{\Delta G} = \frac{1}{1-m+b\lambda\gamma} < \frac{1}{1-m} \qquad \text{and} \qquad \frac{\Delta Y}{\Delta T} = -\frac{m}{1-m+b\lambda\gamma}. \tag{0.61}$$

The term $b\lambda\gamma$ captures monetary policy crowding out: fiscal expansion raises inflation, the real interest rate rises, and interest-sensitive consumption, investment, and net exports fall. Expectations can reduce the multiplier further. If households and firms anticipate that current deficits will require higher future taxes, they may cut current private expenditure. The resulting multiplier can be smaller than one.

These mechanisms underlie the debate over the composition of fiscal stimulus. Proponents of tax cuts emphasize disposable income, private choice, and favourable long-run incentives. Proponents of government purchases emphasize their direct effect on demand and the opportunity to remedy deficiencies in public infrastructure and human capital. Estimates available when the course was taught ranged widely, often from approximately 0.5 to 1.5, because the result depends on the policy instrument, expectations, monetary policy response, and condition of the economy.

The effective lower bound changes the monetary policy response. On the normal portion of the aggregate demand curve, a fiscal expansion moves the economy from A_1 to B_1 in the left panel of Figure 63. Inflation rises, the central bank raises the real interest rate, and crowding out limits the increase from Y_1 to Y_2. At the lower bound in the right panel, the same expansion moves the economy from A_2 to B_2 and generates the larger increase from Y_1 to Y_3.

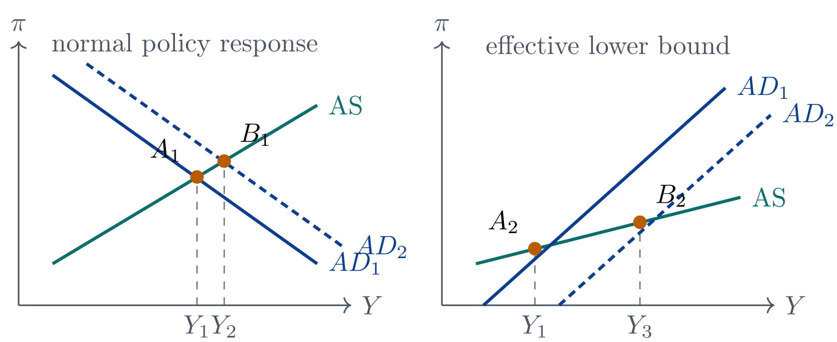

FIGURE 63

At the effective lower bound, the nominal interest rate remains at zero while the fiscal expansion raises expected inflation along with actual inflation. Equation (0.51) then implies that the real interest rate falls. Rather than crowding out private expenditure, the interest rate channel reinforces the expansion. Fiscal multipliers may therefore be substantially larger in a severe slump at the lower bound than in normal conditions.

Fiscal policy can affect aggregate supply as well as aggregate demand. A temporary cut in the employer portion of payroll taxes lowers firms' labour costs and acts as a temporary favourable supply shock, shifting short-run aggregate supply from AS_1 to AS_2. If the policy also raises workers' take-home pay, aggregate demand shifts from AD_1 to AD_2. The two effects move the economy from A to B in Figure 64. Output unambiguously rises, while inflation can rise or fall depending on the relative sizes of the demand and supply shifts; the diagram shows a small increase.

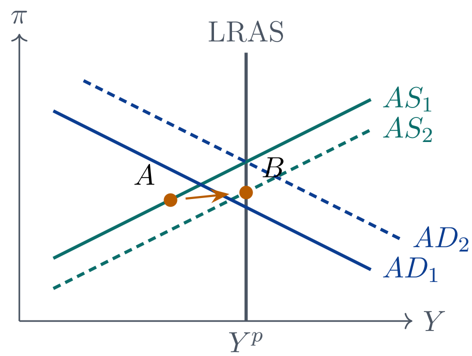

Figure 64

If the payroll tax cut is permanent, it may also raise labour demand, labour supply when workers' marginal tax wedge falls, investment, and potential output. Supply-side economists extend this reasoning to permanent reductions in marginal tax rates more generally. In Figure 65, aggregate demand shifts from AD_1 to AD_2, short-run aggregate supply shifts from AS_1 to AS_2, and long-run aggregate supply shifts from $LRAS_1$ to $LRAS_2$. Potential output rises from Y_1^p to Y_2^p. If the supply response is sufficiently strong, the new long-run inflation rate can be lower despite the increase in aggregate demand.

Tax reductions can partly finance themselves if the resulting increase in economic activity expands the tax base. Let revenue be $R = \tau\mathcal{B}(\tau)$, where τ is the tax rate and $\mathcal{B}$ is the tax base. Then

$$\frac{\Delta R}{R} \approx \frac{\Delta\tau}{\tau} + \frac{\Delta\mathcal{B}}{\mathcal{B}}.$$

A tax cut fully pays for itself only if the percentage increase in the tax base exceeds the percentage reduction in the tax rate. That strong response is controversial, although a partial revenue offset is plausible.

Budget balance alone does not determine the short-run effect of fiscal policy. In the elementary

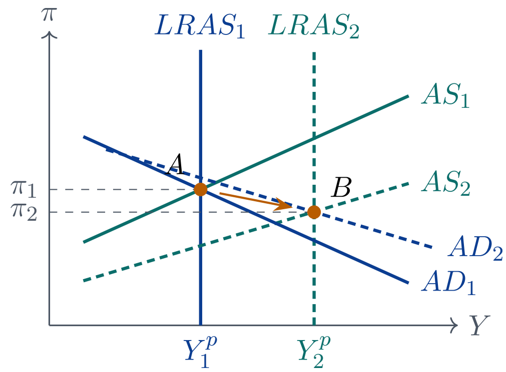

FIGURE 65

Keynesian model, an equal increase in government purchases and taxes has a **balanced-budget multiplier** of

$$\left.\frac{\Delta Y}{\Delta G}\right|_{\Delta T=\Delta G} = \frac{1}{1-m} - \frac{m}{1-m} = 1. \tag{0.62}$$

Thus, a balanced-budget increase in government purchases is expansionary in that model. Conversely, deficit reduction through lower government purchases or higher taxes has a direct contractionary effect. A credible and sustained deficit reduction may nevertheless lower expected future taxes, reduce tax distortions and default risk, lower interest rates, and stimulate consumption or investment. Expectations of stronger future output can offset part of the direct contraction.

Episodes in Denmark and Ireland in which fiscal retrenchment coincided with strong growth are often cited as examples of **expansionary austerity**. The European debt crisis renewed the debate.

1. Supporters argued that credible deficit reduction could lower perceived default risk and interest rates, stop the sovereign debt spiral in Figure 61, and reduce expected future taxes. These channels could increase consumption and investment enough to offset the direct contraction.
2. Critics argued that cuts in government purchases and increases in taxes impose immediate hardship and reduce aggregate demand. The contraction is especially large when fiscal multipliers are high and monetary policy is constrained by the effective lower bound. Lower output also reduces tax revenue, so the deficit may fall by less than planned. Finally, a sufficiently severe decline in nominal gross domestic product can raise the debt-to-GDP ratio through (0.58), even if the nominal deficit becomes smaller.

Whether fiscal consolidation is expansionary therefore depends on the state of the economy, the response of monetary policy, the credibility and composition of the plan, and the sensitivity of private expenditure to expectations.

Deficits, Inflation, and Ricardian Equivalence

Persistent large deficits create inflationary pressure only if they ultimately lead to sustained money creation. In a simplified consolidated government budget, a deficit can be financed by bonds held by private investors or by bonds acquired by the central bank:

$$D_t = \Delta B_t^{\text{private}} + \Delta B_t^{\text{central bank}} \qquad \text{and} \qquad \Delta B_t^{\text{central bank}} = \Delta M_t^B. \tag{0.63}$$

Here M_t^B denotes the monetary base, distinguishing it from the broader monetary aggregate M_t used in the quantity theory. The second equality captures a central bank purchase financed by creating base money and suppresses the other entries on the central bank's balance sheet. This process is called **debt monetization**. It is not the deficit itself that causes inflation: bond-financed deficits need not increase the money supply. Inflation results when persistent base creation is transmitted into persistent growth of the money balances used for transactions beyond the growth of real money demand, as in (0.22).

The resources obtained by issuing money are called **seigniorage**. Its real value is

$$s_t = \frac{\Delta M_t^B}{P_t}. \tag{0.64}$$

To relate this revenue to inflation, suppose real base-money balances $m_t^B = M_t^B / P_t$ are constant. Base money and prices then grow at the same gross rate, so

$$s_t = \left(1 - \frac{1}{1+\pi_t}\right) \frac{M_t^B}{P_t} = \frac{\pi_t}{1+\pi_t} m_t^B \approx \pi_t m_t^B. \tag{0.65}$$

The approximation is accurate when inflation is not too large. Equation (0.65) says that real seigniorage revenue is approximately the inflation rate multiplied by real base-money balances. Seigniorage is therefore called an **inflation tax**: inflation reduces the real purchasing power of the public's base-money balances while transferring resources to the issuer of money.

The government cannot obtain unlimited real revenue by printing money. Higher inflation encourages households and firms to economize on base-money balances, so the tax base m_t^B shrinks. Beyond some point, a higher inflation rate can reduce rather than increase seigniorage revenue. Hyperinflations, including Zimbabwe's, illustrate how large deficits, rapid money creation, and collapsing real money demand can reinforce one another. Concern about this form of fiscal dominance motivates independent central banks and fiscal rules that limit persistent deficits and monetary financing.

Ricardian equivalence offers a different perspective on bond-financed deficits. Suppose the government cuts current lump-sum taxes by $\delta > 0$, leaves purchases unchanged, and issues debt of δ. If the real interest rate is r, repaying the debt next period requires future taxes to rise by $(1+r)\delta$. The present value of the future tax increase is

$$\frac{(1+r)\delta}{1+r} = \delta, \tag{0.66}$$

which exactly equals the current tax cut. A forward-looking household recognizes that the timing of taxes has changed but the present value of its tax burden has not. It saves the entire tax cut, including the interest needed to meet the future tax liability, rather than increasing consumption.

The contrast appears in Figure 66. Under Ricardian equivalence, private saving rises by the amount of the tax cut. If households instead consume part of the tax cut, aggregate demand rises in the short run and national saving falls.

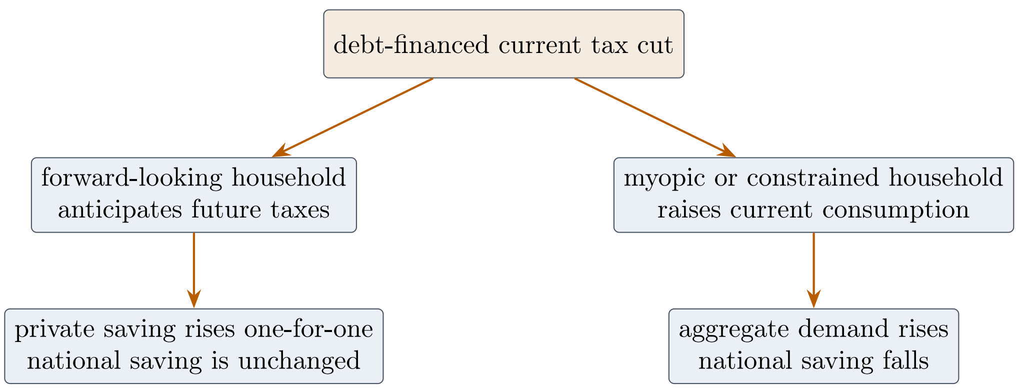

Figure 66

If Ricardian equivalence holds exactly, three conclusions follow.

1. The tax cut does not increase consumption and therefore does not shift aggregate demand, apart from any direct supply-side effect of the lower tax rate.
2. National saving is unchanged. Government saving falls by δ, but private saving rises by δ. There is consequently no crowding out of domestic investment or net exports through (0.59).
3. Households use the additional saving to buy the newly issued government bonds. The government therefore need not obtain financing through money creation, so the deficit need not raise inflation.

Conclusions (1)–(3) require strong assumptions about household behaviour.

There are three principal objections:

1. Households may be myopic. If they do not fully anticipate the future tax liability, they spend part of the tax cut and private saving rises by less than government saving falls.
2. Some households face borrowing constraints. A student with low current income, for example, may be unable to borrow as much as desired against future earnings. A current tax cut relaxes that constraint and raises consumption even if the student understands that future taxes will be higher.
3. Current taxpayers may not care fully about taxes imposed after their lifetimes. If debt repayment falls partly on future generations and households do not leave offsetting bequests, the current tax cut raises the present generation's lifetime resources and consumption.

The empirical evidence is mixed because observed saving also responds to income, wealth, credit conditions, and expectations. The tax cuts enacted under U.S. President George W. Bush were followed by lower private saving, which is inconsistent with exact Ricardian equivalence, but the simultaneous housing and stock market boom raised household wealth and independently reduced saving. The practical importance of Ricardian equivalence is therefore an empirical question about how households respond to a change in the timing of taxes.