Macroeconomics – Luigi Iovino

Chapter 2 – A Tour of the Book

2.1 Aggregate Statistics

In economics, **aggregate data** or data aggregates are high-level data that are composed from a multitude or combination of other more individual data, such as:

- in **macroeconomics**, data such as the overall price level or overall inflation rate; and
- in **microeconomics**, data of an entire sector of an economy composed of many firms, or of all households in a city or region.

Three aggregate statistics

When macroeconomists study an economy they first look at 3 variables:

- **Aggregate Output**
  - **National Income and Product Accounts**
    Accounting System used to measure aggregate economic activity.
  - **Measure of Aggregate Output** in the national income accounts Gross Domestic Product (GDP).
- **Unemployment Rate**
- **Inflation Rate**

Gross Domestic Product (GDP)

There are three ways of defining GDP:

- **GDP** is the **value** of **final goods and services** produced in the economy during a given period.
  - A **final good** is a good that is intended for final consumption.
  - An **intermediate good** is a good used in the production of another good.

- **GDP** is the **sum** of **value added** in the economy during a given period.
  - **Value added** by a firm is the value of its **production minus** the value of the **intermediate goods** used in production.

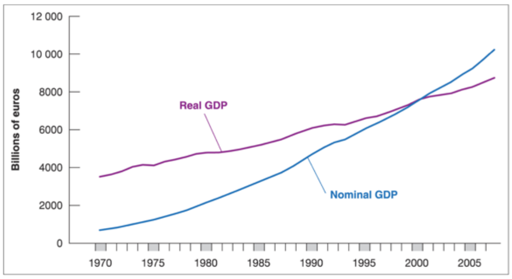
- **GDP** is the **sum** of **incomes** in the economy during a given period. Value added is used to remunerate:
  - Workers -> **Labor Income**
  - Firms -> **Capital Income or Profit Income**
  - Government -> **Indirect Taxes**

Nominal Gross Domestic Product

- **Nominal GDP** is the sum of the quantities of final goods produced multiplied by their **current prices**.
- **Nominal GDP increases** over time because of **2 reasons**:
  - **Production** of most goods **increases** over time.
  - **Prices** of most goods also **increase** over time.

Real Gross Domestic Product

- **Real GDP** is constructed as the sum of the quantities of final goods multiplied by **constant prices** (rather than current).
- **Real GDP increases** over time because **production** of most goods **increases** over time.



Nominal and real GDP in the EU15 since 1970
Since 1970, nominal GDP in the EU15 increased by a factor of 14. Real GDP increased by a factor of 2.5
Source: OECD Economic Outlook database

Measuring Real GDP

- Simple in an economy with one good, say cars:

Year	Quantity of cars	Price of cars	Nominal GDP	Real GDP (in 2005 prices)
2004	10	€20,000	€200,000	€240,000
2005	12	€24,000	€288,000	€288,000
2006	13	€26,000	€338,000	€312,000

  - To construct real GDP, multiply the number of cars in each year by a **common price.**
  - Suppose we use the **price of cars in 2005** as the common price.
  - This approach gives us **real GDP in 2005 prices**.

- Would the **level** and the **rate of change** of real GDP be different had we used the price in a different year, say 2006 rather than 2005?
- Real GDP more difficult to measure in an economy with many final goods.
-

GDP: Level versus Growth Rate

- Real GDP level in period t, $\mathbf{Y_t}$.
- Real GDP growth: $\mathbf{(Y_t - Y_{t-1}) / Y_{t-1}}$.
- Periods of **positive GDP growth** are called **Expansions**.
- Periods of **negative GDP growth** (at least 2 consecutive quarters) are called **Recessions**.

Going back to a previous example of an economy with only one good, cars:

Year	Quantity of cars	Price of cars	Nominal GDP	Real GDP (in 2005 prices)	Real GDP (in 2006 prices)
2004	10	€20,000	€200,000	€240,000	€260,000
2005	12	€24,000	€288,000	€288,000	€312,000
2006	13	€26,000	€338,000	€312,000	€338,000

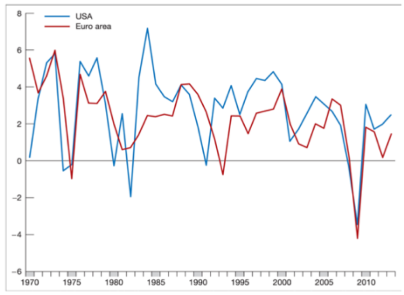
- GDP growth using real GDP at 2005 prices
  - Growth from 2004 to 2005 = (312,000 – 260,000)/260,000 = 0.2 = 20%
  - Growth from 2005 to 2006 = (338,000 – 312,000)/312,000 = 0.083 = 8.3%

- GDP growth using real GDP at 2006 prices
  - Growth from 2004 to 2005 = (312,000 – 260,000)/260,000 = 0.2 = 20%
  - Growth from 2005 to 2006 = (338,000 – 312,000)/312,000 = 0.083 = 8.3%

- With only one good, the growth rate is unaffected by the base year.

GDP Growth Rate

Growth rates of GDP in the EU15 and in the USA since 1970



Since 1970, both the EU-15 and the US economy have gone through a series of expansions, interrupted by short recessions. The recession associated with the current crisis has been particularly deep.
Source: OECD *Economic Outlook* database

2.2 Unemployment

Measuring Unemployment

Parameters to evaluate aggregate statistics:

- **Working Age Population**, number of people who can work, **P**.
- **Employment**, number of people who have a job, **N**.
- **Unemployment**, number of people who don't have a job but are looking for one, **U**.
- **Labor Force**, sum of employment and unemployment, **L**.

L = N + U, that is **Labor Force = Employment + Unemployment**.
P = L + Out of Labor Force, i.e. **Working Age Population = Labor Force + Out of the Labor Force**.

Unemployment Rate

Ratio of unemployed people to the number of people in the labor force, **u = U/L**.

E.g.: In 2008 in the U.S. 144.4 million people were employed, 7.0 million people were unemployed, so the unemployed rate was 7.0/(144.4 +7.0) x 100 = 7.1%.
Question: What would happen to the unemployment rate if all workers without a job gave up looking for one?

Participation Rate

Ratio of people in the Labor Force to the Working Age Population, **Participation Rate = L / P.**

Why do Economists care about Unemployment?

Economists care about unemployment for two reasons:

- Because of its direct effects on the welfare of the unemployed.
- Because it provides a signal that the economy may not be using some of its resources efficiently.

2.3 Inflation Rate

- **General Level of Prices**, average price of goods and services in the economy, P_t.
- **Inflation rate**, rate at which the general price level grows, $\pi_t = (P_t - P_{t-1})/P_{t-1}$.

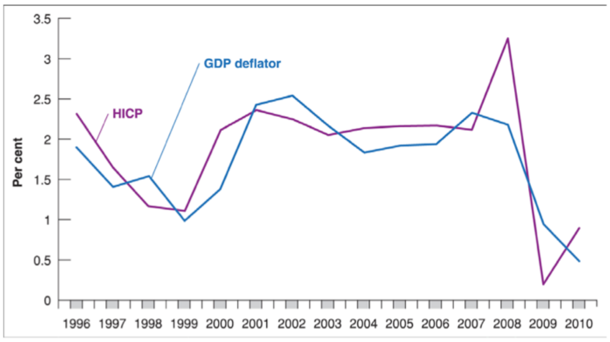
Inflation and Deflation

- **Inflation**, sustained raise in the price level; corresponds to a positive inflation rate.
- **Deflation**, sustained decline in the price level; corresponds to a negative inflation rate.

Price Level

Economists use **2 definitions** of the aggregate price level, **2 price indexes**:

- **GDP Deflator, P_t**
    - GDP deflator in t is defined as the ratio of Nominal GDP to Real GDP in t
    - $\$Y_t = P_t Y_t$, Nominal GDP = GDP deflator x Real GDP.
    - $P_t = \$Y_t / Y_t$, GDP deflator = Nominal GDP / Real GDP.
    - GDP Deflator is an index number, set equal to 1 in the base year.



Inflation rate, using the HICP and the GDP deflator in the euro area since 1996
The inflation rates, computed using either the HICP or the GDP deflator,
are largely similar
Source: Eurostat, OECD Statistics Database

▪ The **GDP Deflator** measures the **average price of output**, i.e. of the final goods produced in the economy.

Year	Cars		Books	
	Q^C	P^C	Q^B	P^B
2000	12	24,000	10	5,000
2010	13	30,000	20	2,000

GDP-2000: 12x24,000+10x5,000 = 338,000

GDP-2010: 13x30,000+20x2,000 = 430,000

Real GDP-2000: 12x24,000+10x5,000 = 338,000

Real GDP-2010: 13x24,000+20x5,000 = 412,000

Deflator-2000: $$P_{2000} = \frac{€Y_{2000}}{Y_{2000}} = \frac{338{,}000}{338{,}000} = 1$$

Deflator-2010: $$P_{2010} = \frac{€Y_{2010}}{Y_{2010}} = \frac{430{,}000}{412{,}000} = 1.044$$

Inflation rate: $$\pi_{2000-2010} = \frac{P_{2010} - P_{2000}}{P_{2000}} = \frac{1.044 - 1}{1} = 0.044$$

- **Consumer Price Index**

▪ The **consumer price index**, or **CPI**, measures the **average price of consumption**, i.e. of the goods consumed by the households.

- ▪ The **CPI** gives the cost in dollars of a specific list of goods and services over time, which attempts to represent the **consumption basket of a typical urban consumer**.
- ▪ The **CPI** measures the **cost of living**.

The set of goods produced in the economy is not the same as the set of goods purchased by consumers, for two reasons:

- Some of the goods are sold to firms, to the government or to foreigners.
- Some of the goods are not produced domestically but are imported from abroad.

The **HICP** *(Harmonized Index of Consumer Prices, i.e. CPI of EU area)* and the **GDP Deflator** move together most of the time. In most years, the two inflation rates differ by less than 0.5%.

There are clear exceptions to the first conclusion: In 1998, 2002 and 2009, the increase in the HICP was slightly smaller than the increase in the GDP Deflator. That means that the price of goods consumed in the euro area (measured by the HICP) was lower than the price of goods produced in the euro area (measured by the GDP Deflator).

Why do Economists care about Inflation?
Economists care about inflation for two reasons:

- During periods of inflation, not all prices and wages rise proportionately, inflation affects relative prices and income distribution.
- Inflation leads to other distortions.

Short Run, Medium Run and Long Run

Movements in **Demand** for goods produce movements in **Output**. However, what matters when it comes to aggregate output is the supply side, how much the economy can produce. How much can be produced depends on how advanced the technology of the country is, how much capital it is using, and the size and the skills of its labor force.
This suggests that the true determinants of output are factors like a country's education system, the saving rate and the quality of the government. If we want to understand what determines the level of output, we must look at these factors.

Level of Aggregate Output in an economy is determined by:
- In the **Short Run**, say a few years, movements in output are primarily driven by movements in **demand**.
- In the **Medium Run**, say a decade, the economy tends to return to the level of output determined by supply factors: the capital stock, the level of technology and the size of the labor force.
- In the **Long Run**, say a few decades or more, we must look at factors like the education system, the saving rate and the role of the government.

Chapter 3 - The Goods Market

3.1 Composition of GDP

When economists think about year-to-year movements in economic activity, they focus on the interactions among **production**, **income** and **demand**:
- Changes in the **demand** for goods lead to changes in **production**.
- Changes in **production** lead to changes in **income**.
- Changes in **income** lead to changes in the **demand** for goods.

If we want to understand what determines the demand for goods, it makes sense to decompose aggregate output (GDP):
- **Consumption (C)**, which are the goods and services purchased by consumers, ranging from food to airline tickets, to new cars, and so on. Consumption is by far the largest component of GDP.
- **Investment (I)**, sometimes called **Fixed Investment** is composed of:
  - **Non-Residential Investment**, the purchase by firms of new plants or new machines.
  - **Residential Investment**, the purchase by people of new houses or apartments.
- **Government Spending (G)**, which are purchases of goods and services by the federal, state and local governments, including services provided by government employees.

To determine the purchases of a region goods and services we also have to take into consideration:
- **Exports (X)**, purchases of region goods and services by foreigners.
- **Imports (IM)**, purchases of foreign goods and services by region's consumers.

The difference between **Exports** and **Imports** returns the **Trade Balance**:
- If **Exports** > **Imports**, then the region has a **Trade Surplus**.
- If **Imports** > **Exports**, then the region has a **Trade Deficit**.

The last parameter that has to be taken into consideration is **Inventory Investment** that is the difference between **production** and **sales**.

3.2 The Demand for Goods

We denote the total demand for goods by **Z**. Using the decomposition of GDP, we saw:
$$\mathbf{Z = C + I + G + X - IM}$$
However, we are going to make some simplifications on this formula:

- Assume that all firms produce the same good, which can then be used by consumers for consumption, by firms for investment, or by the government. $Z = c_0 + c_1(Y - T) + \bar{I} + G$
- Assume that firms are willing to supply any amount of the good at a given price level P.
- Assume that the economy is closed, that it does not trade with the rest of the world, which is, both exports and imports are zero. (X = IM = 0)

Hence, the resulting equation is: $Y = c_0 + c_1(Y - T) + \bar{I} + G$

Z = C + I + G

We will now deeper discuss the three components of this equation:

- **Consumption (C)**, consumption decisions depend on many factors. But the main one is surely income, or, more precisely, **disposable income** (Y_D), the income that remains once consumers have received transfers from the government and paid their taxes. When their disposable income goes up, people buy more goods. That is:

  The function $C(Y_D)$ is called the **consumption function**. The positive sign below Y_D reflects the fact that when disposable income increases, so does consumption.

  $\longrightarrow$

  According to this equation, we have that consumption is composed of two parameters:
  - **C_0**, that is what people would consume if their disposable income in the current year were equal to zero.
  - **$C_1(Y_D)$**, is called **marginal propensity to consume**, $\underset{(+)}{C = C(Y_D)}$ and gives the effect an additional euro of disposable income has on consumption.
  - **$Y_D = Y - T$**, that is disposable income is given by Income – Taxes.

- **Investment (I)**, for will consider **exogenous** $C = c_0 + c_1(Y_D)$ $\xrightarrow{Y_D \equiv Y - T}$ $C = c_0 + c_1(Y - T)$ the sake of simplicity, we Investment to be an **variable**, that is, as fixed. Indeed, we will denote this by:

- **Government Spending (G)**, it describes the **fiscal policy**, that is, the choice of taxes and spending by the government. Just as we did for investment, we will take G and T as exogenous, for 2 main reasons:
  - Governments do not behave with the same regularity as consumers or firms.
  - One of the tasks of macroeconomics is to think about the implications of alternative spending and tax decisions. We want to be able to say, "If the government was to choose these values for G and T, this is what would happen".

$I = \bar{I}$

3.3 The Determination of Equilibrium Output

We have now reached an equation for demand of the following shape:
We now introduce another strong assumption, that is, firms do not hold inventories. In this case, inventory is always equal to zero, and **equilibrium in the goods market** requires that production Y be equal to the demand for goods Z: **Y = Z**. This equation is called an **equilibrium condition**.
Hence, we can derive the following equation:

In equilibrium, production Y (the left of the equation), is equal to demand (the right side). Demand in turn depends on income, Y, which is itself equal to production.

Note that we are using the same symbol Y for production and income. This is no accident. As we saw earlier, we can look at GDP either from the production side or from the income side. Production and Income are identically equal.

We have to note that by rearranging the last equation, we can derive:
Where $1/(1-c_1)$ is called **Multiplier** and is called **Autonomous Spending**, since it does not depend on output. The multiplier tells us how much the output will increase given an increase in c_0 consumption.

$$Y = C + \bar{I} + G$$

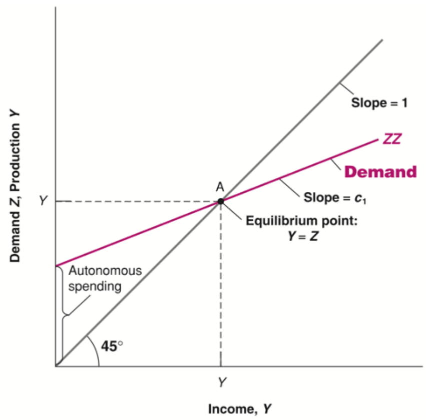
$$Y - T - C = \bar{I} + G - T$$

$$S = \bar{I} + G - T \qquad \bar{I} = S + (T - G)$$

Graphical Representation

$$Y = \frac{1}{1-c_1}[c_0 + \bar{I} + G - c_1 T]$$

$$[c_0 + \bar{I} + G - c_1 T]$$



In the short run, demand determines production. Production is equal to income. Income in turn affects demand.
The consumption function shows how consumption depends on disposable income. The propensity to consume describes how much consumption increases for a given increase in disposable income.
Equilibrium output is the level of output at which production equals demand. In equilibrium, output equals autonomous spending times the multiplier. Autonomous spending is that part of demand that does not depend on income. The multiplier is equal to $1/(1-c_1)$, where c_1 is the propensity to consume.

3.4 Investment equals Savings: An alternative way of thinking about the Goods-Market Equilibrium

Thus far, we have been thinking of equilibrium in the goods market in terms of the equality of the production and the demand for goods. An alternative, but equivalent, way of thinking about equilibrium focuses instead on investment and saving.

Saving is the sum of private saving and public saving:
- **Private Savings (S)**, saving by consumers, is equal to their disposable income minus their consumption: $\mathbf{S = Y_D - C}$. Which we can rewrite as: $\mathbf{S = Y - T - C}$.
- **Public Savings** is equal to taxes minus governments spending, T - G.
    - If taxes exceed government spending, the government is running a **Budget Surplus**.
    - If taxes are less than government spending, the government is running a **Budget Deficit**.

Now returning to the equation for equilibrium in the goods market that we derived earlier. Production must be equal to demand, which in turn is the sum of consumption, investment and government

spending: Subtracting taxes from both sides and moving consumption to the left side: We can now notice that the left side of this equation is simply private saving, (S), so or equivalently:
On the left of the last equation is Investment. On the right is saving, the sum of private saving and public saving. This equation gives us another way of thinking about equilibrium in the goods market. It explains why the equilibrium condition for the goods market is called the **IS Relation**, which stands for "*Investment equals Savings*", that is, what firms want to invest must be equal to what people and the government want to save.

In a modern economy, investment decisions are made by firms, whereas saving decisions are made by consumers and the government. In equilibrium, the last equation tells us all these decisions have to be consistent: investment must equal saving.
To summarize, there are two equivalent ways of stating the condition for equilibrium in the goods market:

- **Production = Demand**
- **Investment = Saving**

Hence, we can derive an equation that characterizes the equilibrium using the second condition:
Where **($1-c_1$)** is called the **marginal propensity to save**.
Or, alternatively: .

3.5 Is the Government Omnipotent? A Warning

According to the equations derived seems that the government, by choosing the level of spending, G, or the level of taxes T, can choose the level of output it wants. Can governments really choose the level of output they want? Obviously not.

There are many aspects of reality that we have not yet incorporated into our model, and all these complicate the government's task. We shall incorporate them in due time. But it is useful to lit them briefly here:

$S = -c_0 + (1 - c_1)(Y - T)$

$\bar{I} = -c_0 + (1 - c_1)(Y - T) + (T - G)$

- Changing government spending or taxes is not always easy. Getting national parliaments to pass bills always takes time and often becomes a premier's nightmare.
- We have assumed that investment remained constant, but investment is also likely to respond, and so too are imports: some of the increased demand by consumers and firms will not be for domestic goods but for foreign goods. All these responses are likely to be associated with complex, dynamic effects, making it hard for governments to assess them with much certainty.
- Anticipation is likely to matter. For example, the reaction of consumers to a tax cut is likely to depend very much on whether they think of the tax cut as transitory or permanent. The more they perceive the tax cut as permanent, the larger will be their consumption response.
- Achieving a given level of output can come with unpleasant side effects. Trying to achieve too high a level of output can, for example, lead to increasing inflation and, for that reason, be unsustainable in the medium run.
- Cutting taxes or increasing government spending can lead to large budget deficits and an accumulation of public debt. A large debt can have adverse effects in the long run. This is a hot issue in almost every advanced country in the world.

In short, the proposition that, by using fiscal policy, the government can affect demand and output in the short run is an important and correct proposition.

Money in Modern Economies

Money & Banking

The simplest way of defining money is through the functions it performs. Money has **three important functions**:

- A **Medium of Exchange**: it is accepted in exchange for goods for the sole reason that it can in turn be traded for other goods;
- A **Store of Value**: it is an asset (the most liquid one), since it allows to trade current goods for future goods;
- A **Unit of Account**: all contracts are denominated in terms of money.

The distinguishing economic feature of money is its medium-of-exchange role: money eliminates the need of a **double coincidence of wants**, and allows trades when credit transactions are impossible. Adam Smith saw it as an essential part of moving from a **subsistence economy**, typically based on barter, to an **exchange economy**. Money can be seen as a social institution that provides a solution to the problem of a lack of trust.

Money has circulated in many forms:

- **Commodity Money**, objects that value in themselves as well as value in their use as money (gold coins).
- **Representative Money**, any type of money that has face value greater than its value as material substance.
  - **Commodity-backed Money**, a claim on a commodity, for example gold or silver certificates.
  - **Fiat Money**, money that is "... *created and issued by the State, but is not convertible by law into anything other than itself, and has no fixed value in terms of an objective standard". (J.M. Keynes)* .

The Financial System

Central Bank

The **Central Bank (CB)** is the authority responsible for policies that affect a country's supply of money and credit. In modern economies, CBs are concerned with **monetary policy** and **financial stability**. In the words of **Jeffers**: "*The central bank, as* ***lender of last resort****, helps ensure systematic stability. In general, what is meant by that is stability of both the banking sector and the financial markets, not only the stock market, but also and specifically the payment systems. In order to carry out this mission, the CB must use preventive measures as well as measures aimed at helping to overcome a crisis when it occurs. In such case, the CB must above all insure that liquidity is available or very quickly restored, because that is what stabilizes the financial system".*

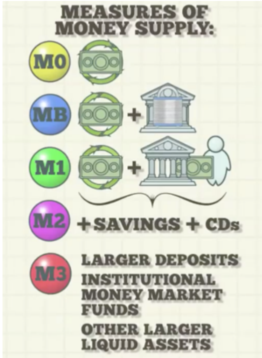
The **European Central Bank (ECB)** claims that "monetary policy concerns the decisions taken by central banks to influence the cost and availability of money in an economy". Furthermore, the ECB believes that "*financial stability is a state whereby the build-up of* ***systemic risk*** *is prevented. Systemic risk can best be described as the risk that the provision of necessary financial products and services by the financial system will be impaired to a point where economic growth and welfare may be materially affected*".

Fiat money is typically issued by the CB, and is recorded as a liability in the CB's **balance sheet**. Although the CB is formally in debt to the holder of its money, that debt can only be repaid in more fiat money. Hence, fiat money is just a special form of an **I-Owe-yoU** (***IOU***), a **financial asset**, which is useful in exchange because it is an IOU that everyone trusts. As any other financial asset, the net aggregate supply of money is zero. In other words, money is a financial asset for the public, and a liability for the CB: at the aggregate level the two positions cancel out.

In modern economies, not only the CB, but also the **commercial banking system** creates fiat money; more precisely:

- The CB is the only issuer of **Base Money** (or **narrow money**). Base money, or "central bank money", comprises IOUs from the CB; i.e. fully liquid claims on the central bank held by the private sector:
    - **Currency** (an IOU to the public).
    - **Central Bank Reserves** (IOUs from the CB to commercial banks).

- The banking system is the main creator of **Broad Money**. Broad Money can be thought of as the money that the private sector has available for transactions, and comprises:
    - **Currency** in the hands of the public, available for transaction purposes.
    - **Bank Deposits** (IOUs from commercial banks to the public).

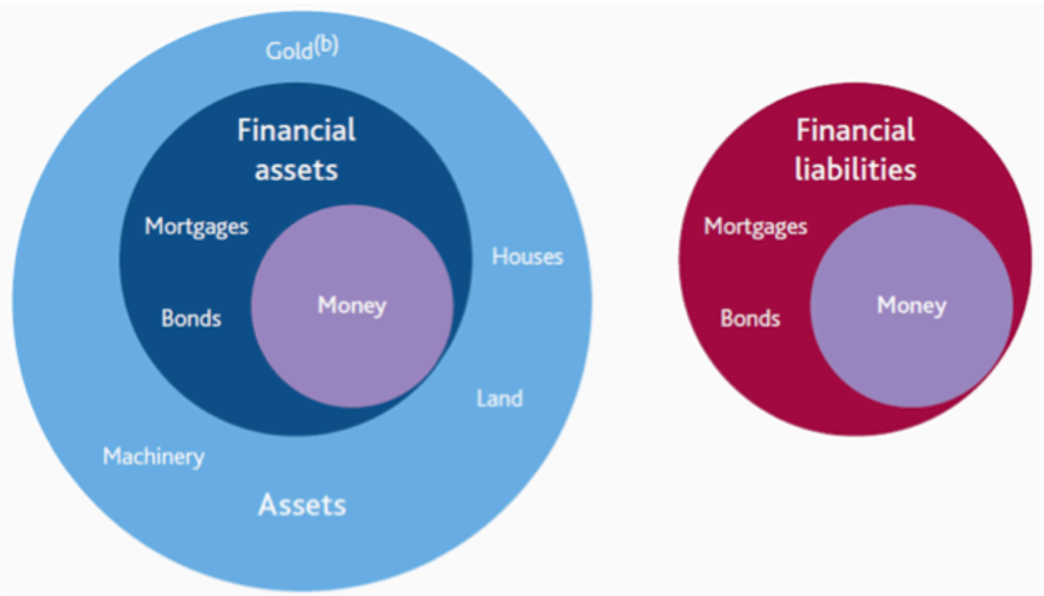


The empirical counterparts to the theoretical concepts of narrow and broad money are the respectively **monetary aggregates** known as **M0** and **M1**, **M2** and **M3**.

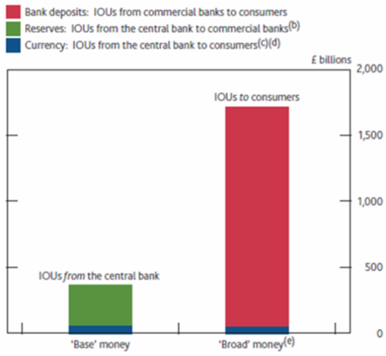
Base money is important because it is by virtue of their position as the only issuer of base money that CBs can implement **monetary policy**. Furthermore, base money is **legal tender** (i.e. the medium of payment recognized by the legal system to be valid for meeting a financial obligation), while bank deposits aren't. During **financial crises**, when doubts arise about the liquidity of commercial bank money, this distinction may become quite relevant.

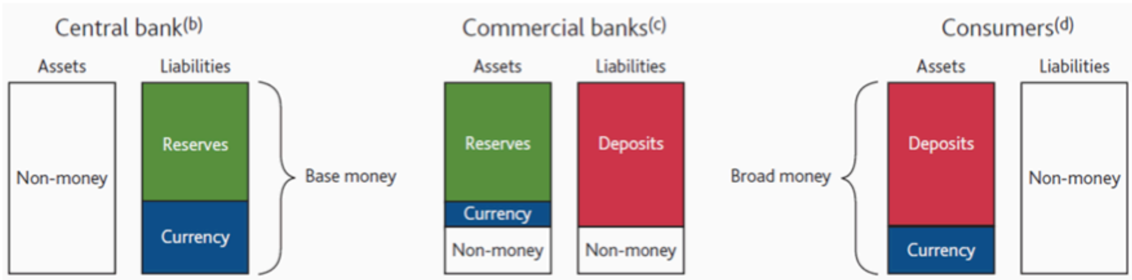
Commercial Banks

Commercial Banks need to hold some currency to meet frequent deposit withdrawals and other outflows. But to use banknotes to carry out this large volume of transactions would be extremely cumbersome! So, banks are allowed to hold a different type of IOY from the CB, known as **Central Bank Reserves**. Reserves are a useful medium of exchange for banks: if one bank wants to make a



payment to another, the CB will adjust their reserves balances accordingly.





In many jurisdictions, CBs implement their monetary policy and defend the stability of the financial system by setting the:

- **Required Reserve Ratio**, a minimum **(reserve + vault cash) / deposits** ratio for some categories of deposits.
- **Interest Rate on Reserves**, in recent years, many CBs began paying interest on excess reserve balances as well as required reserves. Some CBs, as the ECB, have paid negative interest, in effect taxing banks for exceeding their reserve requirements, as an expansionary monetary policy measure.
- **Capital Adequacy Ratio**, also known as **Capital to Risk (Weighted) Assets Ratio**: a minimum **equity/risk-weighted credit exposure** ratio, used to protect depositors and promote stability in the financial system.

The image shows the composition of money supply in the U.K.

Chapter 4 – Financial Markets I

4.1 The Demand for Money

We will now examine the determinants of the **demand for money** (in broad sense) expressed by the private sector (households + firms). Suppose that your **financial wealth** can be hold in two financial assets:

- **Money**, which you can use for transactions, pays no interest. In the real world, as we already mentioned, there are two types of money:
  - **Currency**
  - **Deposits Accounts**
- **Bonds** pay a positive nominal interest rate, $i > 0$, but cannot be used for transactions. In the real world, there are many types of bonds and other financial assets, each associated with a specific interest rate. For the time being, we also ignore this aspect of reality and assume that there is just one type of bond and that it pays, i, the rate of interest.

Assume furthermore that trading in bonds involves **transaction costs**. How should your financial wealth be allocated between money and bonds? This is a **portfolio allocation problem**! If transaction costs are high enough, a diversified portfolio is the natural outcome, i.e. optimal portfolios contain both money and bonds.

At the individual level, the optimal portfolio depends upon:
- **Your level of transaction**: you will want to have enough money on hand to avoid having to sell bonds and incur in transaction costs.
- **The interest rate on bonds**: the higher the interest rate, the higher the incentive to hold bonds.

Money Market Funds (*Money Market Mutual Funds*) pool together the funds of many people. The funds are then used to buy bonds, typically government bonds. Money market funds pay an interest rate close to but slightly below the interest rate on the bonds they hold, the difference coming from the administrative costs of running the funds and from their profit margins.

The **aggregate demand for money** is the sum of all the individual demands, and is denoted by M^d. Therefore, it depends on the overall level of transactions in the economy and on the interest rate. The overall level of transactions in the economy is hard to measure, but it is likely to be roughly proportional to nominal income (€Y). Hence:

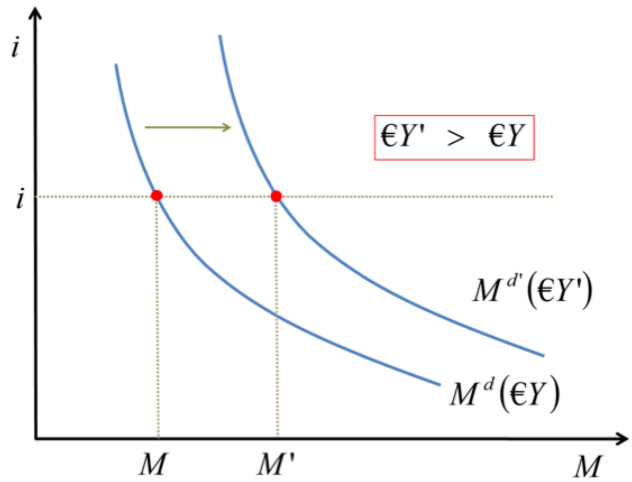
The demand for money, M^d , depends negatively on the interest rate, since the latter represents the opportunity cost of holding money.

The following figure shows the effect on the demand for money of an increase in nominal GDP, for given interest rate: the money demand schedule simply shifts to the right, so that, for each interest rate, the demand for money is now higher.

$$M^d\Big(\underset{(+)}{€Y}, \underset{(-)}{i}\Big) = €Y \cdot L(i).$$

The Supply of Base Money

The supply of base money is firmly in the hands of the CB. To better understand how the CB is able to easily change the amount of money in the hands of the public, let's consider the **CB's Balance Sheet**. The assets of the CB are the bonds it holds in its portfolio, while its liabilities are the stock of currency and reserves.



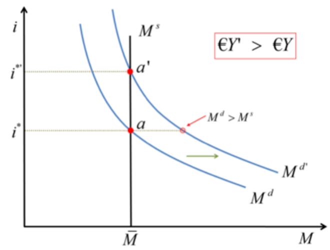
The CB can change the supply of money by buying or selling securities in the bond market: those actions are called **Open Market Operations** (OMOs).

For the sake of simplicity, let's ignore the commercial banking system, set the stock of reserves to zero, and assume that the CB engages in trade directly with households and firms. In general, OMOs lead to equal changes in the CB's assets and liabilities.

- **Expansionary Open Market Operations**, in which the central bank increases the money supply by buying bonds, lead to an increase in the price of bonds and a decrease in the interest rate.
- **Contractionary Open Market Operations**, in which the central bank decreases the money supply by selling bonds, lead to a decrease in the price of bonds and an increase in the interest rate.

The Equilibrium Interest Rate (I)
The supply of currency is firmly in the hand of the CB, and is controlled via OMOs. Suppose the CB decides to supply a **fixed amount of currency**, so that: $M^s = \overline{M} > 0$.
Equilibrium in financial markets requires that **money supply equals money demand**. This equilibrium condition, labeled the **LM Curve**, $\overline{M} = €Y \cdot L(i)$. boils down to:

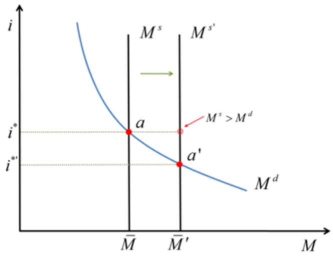
Effects of an increase in Nominal Income



An increase in €Y increases in turn the level of transactions, and therefore the demand for money, for given i. At the initial interest rate, there is an excess demand for money, $M^d > M^s$; in order to decrease the amount of money people want to hold, an increase in the interest rate becomes necessary:
- People try to sell bonds in exchange for currency, but this drives down the price of bonds.
- The drop in the price of bonds drives the interest rate up.

Hence, an increase in €Y implies an increase in the equilibrium interest rate, while M remains unchanged, being determined by the CB.

Effects of an increase in the Supply of Money



An increase in the stock of currency implies that $M^d < M^s$ at the initial interest rate i. In order to increase the amount of money people want to hold, a decrease in the interest rate becomes necessary:
- People try to buy bonds in exchange for currency, but this drives up the price of bonds.
- The increase in the price of bonds drives the interest rate down.

Hence, the increase in M^s implies a decrease in the equilibrium interest rate and a contemporaneous increase in the stock of currency.

Banking and Money Supply
It is now time to focus on the role of the commercial banking system as a key player in the market for money. We already anticipated that the banking system is the main creator of broad money, in the form of short-term checkable deposits. Furthermore, we noted that banks need to hold some currency and central bank reserves to address the liquidity risk associated with bank deposits. Hence, the CB controls not only the amount of currency in the economy, but also the stock of reserves. In the following, we will emphasize the role of banks as financial intermediaries in the process of money creation.

The supply of broad money
In modern economies (i.e. in the presence of fiat money), the "creation" of broad money is essentially in the hands of the banking sector.

Being financial intermediaries, **banks supply loans on demand**, **creating in the process new deposits**, and finance those loans with existing deposits or by short-term borrowing on the interbank market. Intuitively, the idea is that "*lending creates deposits*", and therefore broad money:

- Banks create money, in the form of deposits, by making new loans (or by purchasing securities from the non-bank private sector).
- The amount of deposits influences the amount of desired reserves, which is then supplied on demand by the CB (at a price!).

Summary of Banco Bocconiano Adventures

- Deposits in a Bank do not affect money supply as money is simply "changing hands".
- When banks buy/sell securities from the non-bank private sector they affect the supply of money.
- Bank lending increases the supply of money.
- At the same time, bank lending increases deposits, and the connected risk of unexpected withdrawals, raising the demand for CB reserves and borrowings on the unsecured interbank market.

Hence, the need to borrow short-term in order to reach the desired reserve ratio **makes lending and purchasing securities a costly activity**, over and above the pure **credit risk**, and actually **limits the supply of broad money**!

Hence, **there are strong forces that limit money creation through lending**:

- Market forces constrain lending because individual banks have to lend profitably, and therefore the **demand for lending** plays a key role.
- **Additional loans increase liquidity and credit risks**, and therefore increase respectively the need for reserves and the size of the default premium.

Monetary Policy can be considered the ultimate constraint on money creation:

- Being the monopoly provider of base money, the CB controls the **price of reserves** (directly and indirectly by controlling the interest rate on short-term money markets).
- By controlling the price of reserves, and setting the required reserve and capital adequacy ratios, the **CB indirectly controls the supply of broad money**.

The Equilibrium Interest Rate (II)

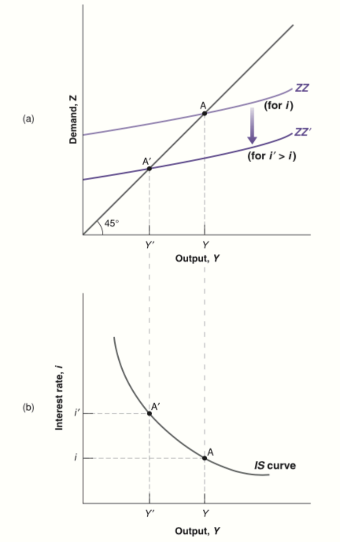
The previous analysis suggests that, while the CB has full control on base money, the control on broad money is only indirect, and it exerted by controlling the price of reserves.
Hence, in a world without commercial banks, the CB would control the supply of currency, and let the interest rate be determined in equilibrium. In a world characterized by a well-functioning commercial banking system, instead, the CB would fine it much easier to control the interest rate (i.e. the price of additional reserves) and let the **money supply equate endogenously the money demand**.

The **LM curve**, in this case, boils down to the following relationship: $\boldsymbol{i = \bar{\imath}}$**.**

At this point, you may wonder whether the price of reserves, i.e. the short-term interest rate on interbank markets, was somehow related to the interest rate on riskless bonds or not. After all, the demand of money depends on the return of riskless bonds, not on the price of reserves.

Chapter 5 – Financial Markets: The IS-LM Model

In the preceding chapters we looked at the goods market and at financial markets. We now look at goods and financial markets together. By the end of this chapter you will have a framework to think about how output and the interest rate are determined in the short run.



5.1 The Goods Market and the IS Relation

The IS – LM model characterizes the implications of equilibrium in both the goods and the financial markets. The IS relation and the IS curve show the combinations of the interest rate and the level of output that are consistent with equilibrium in the goods market. An increase in the interest rate leads to a decline in output. Consequently, the IS curve is downward sloping. Changes in factors that decrease the demand for goods given the interest rate shift the IS curve to the left. Changes in factors that increase the demand for goods given the interest rate shift the IS curve to the right.

5.2 Financial Markets and the LM Relation

The LM relation and the LM curve show the combinations of the interest a rate and the level of output consistent with equilibrium in financial markets. Under the assumption that the central bank chooses the interest rate, the LM curve is a horizontal line at the interest rate chosen by the central bank. (Insert Image)

5.3 Putting the IS and the LM Relations together

The IS relation follows from goods-market equilibrium. The LM relation follows from financial market equilibrium. They must both hold.

- IS Relation: $Y = C(Y - T) + I(Y, i) + G$
- LM Relation: $i = \bar{i}$

Together they determine output. Output, equivalently, production or income, is measured on the horizontal axis. The interest rate is measured on the vertical axis.

- Any point on the downward-sloping IS curve corresponds to equilibrium in the goods market.
- Any point on the horizontal LM curve corresponds to equilibrium in financial markes.
- Only at point A are both equilibrium conditions satisfied. That means point A, with the associated level of output Y and interest rate $\bar{i}$ is the overall equilibrium, the point at which there is equilibrium in both the goods market and the financial markets.

Let's now see what the IS – LM model tells us, by looking separately at the effects of fiscal and monetary policy.

- **Fiscal Policy**, suppose the government decides to reduce/increase the budget deficit:
  - **Fiscal Contraction** (consolidation), shifts the IS curve to the left, leading to a decrease in output.
  - **Fiscal Expansion**, shifts the IS curve to the right, leading to an increase in output.
- **Monetary Policy**, suppose the central bank increases/decreases the interest rate:
  - **Monetary Contraction**, shifts the LM curve up, leading to an increase in the interest rate and a decrease in output.
  - **Monetary Expansion**, shift the LM curve down, leading to a decrease in the interest rate and an increase in output.

How to answer a possible question?

- Ask how the change affects equilibrium in the goods market and how it affects equilibrium in the financial markets. Put another way, does it shift the IS curve and/or the LM curve, and, if so, how? E.g. : *The first question is how the increase in taxes affects equilibrium in the goods market, that is, how it affects the relation between output and the interest rate captured by the IS curve. We derive that at a given interest rate, the increase in taxes decreases output. The IS curve shifts to the left. At the same time, the CB does not change the interest rate. Thus, the LM curve, namely the horizontal line remains unchanged, the LM curve does not shifts.*

- Characterize the effects of these shifts on the intersection of the IS and the LM curves. What does this do to equilibrium output and the equilibrium interest rate? E.g: *After the increase in taxes and the shift to the left of the IS curve, the new equilibrium sees an output decrease. BY assumption, the interest rate does not change. Thus, as the IS curve shifts, the economy moves along the LM curve.*

- Describe the effects in words. E.g.: *The increase in taxes leads to lower disposable income, which causes people to decrease their consumption. This decrease in demand leads, in turn through a multiplier, to a decrease in output and income. At a given interest rate, the increase in taxes leads therefore to a decrease in output. Looking at the components of output, the decrease in income and the increase in taxes both contribute to the decrease of disposable income and, in turn, a decrease in consumption. The decrease in output leads to a decrease in investment. Thus, both consumption and investment decrease.*

5.4 Using a Policy Mix

We have looked so far at fiscal policy and monetary policy in isolation. In practice, the two are often used together. The combination of monetary and fiscal policies is known as the **monetary-fiscal policy mix**.

- Expansionary fiscal policy, say through a decrease in taxes, shifts the Is curve to the right. Expansionary monetary policy shifts the LM curve down. Both fiscal and monetary policies contribute to the increase in output.

There are a number of reasons why policy makers may want to use a policy mix:

- A fiscal expansion means either an increase in government spending or a decrease in taxes, or both. This means an increase in the budget deficit, however running a large deficit and increasing government debt may be dangerous.
- A monetary expansion means a decrease in the interest rate. If the interest rate is very low, then the room for using monetary policy may be limited. In this case, fiscal policy has to do more of the job. If the interest rate is already equal to zero, the case of the zero-lower bound, then fiscal policy has to do all the job.
- Fiscal and monetary policies have different effects on the composition of output. A decrease in income taxes for example will tend to increase consumption relative to investment. A decrease in the interest rate will affect investment more than consumption. Thus, depending on the initial composition of the output, policy makers may want to rely more on fiscal or more on monetary policy.
- Finally, neither fiscal policy nor monetary policy work perfectly. A decrease in taxes may fail to increase consumption. A decrease in the interest rate may fail to increase investment. Thus, in case one policy does not work as well as hoped for, it is better t use both.

Sometimes, the right policy mix is instead to use the two policies in opposite directions, for example, by combining a fiscal consolidation with a monetary expansion. Suppose that the government is running a large budget deficit and would like to reduce it but does not want to trigger a recession. If the government reduces the deficit, say by increasing T or by decreasing G (or both), the IS curve would shift to the left and the output would decrease. However, the recession can be avoided, if monetary policy is also used. If the central bank reduces the interest rate such that the output is kept constant. What happens to consumption and investment in this case? What happens to consumption depends on how the deficit is reduced. If the reduction takes the form of a decrease in government spending rather than an increase in taxes, income is unchanged, disposable income is unchanged, and so consumption is unchanged. If the reduction takes the form of an increase in income taxes, then disposable income is lower and so is consumption. What happens to investment is unambiguous: unchanged output and lower interest rate imply higher investment.

5.5 How Does the IS – LM Model Fit the Facts?

We have so far ignored dynamics:

- Consumers are likely to take some time to adjust their consumption following a change in disposable income.
- Firms are likely to take some time to adjust investment spending following a change in their sales.
- Firms are likely to take some time to adjust investment spending following a change in the interest rate.
- Firms are likely to take some time to adjust production following a change in their sales.

So, in response to an increase in taxes, it takes some time for consumption spending to respond to the decrease in disposable income, some more time for production to decrease in response to the decrease in consumption spending, yet more time for investment to decrease in response to lower sales, for consumption to decrease in response to the decrease in income, and so on.

In response to a decrease in the interest rate, it takes some time for investment spending to respond to the decrease in the interest rate, some more time for production to increase in response to the increase in demand, yet more time for consumption and investment to increase in response to the induced change in output, and so on.

The IS – LM model appears to describe well the behaviour of the economy in the short run. In particular, the effects of monetary policy appear to be similar to those implied by the IS – LM model once dynamics are introduced in the model. An increase in the interest rate due to a monetary contraction leads to a steady decrease in output, with the maximum effect taking place after about eight quarters.

Chapter 6 – Financial Markets II

Until now, we have assumed that there were only two financial assets, money and bonds, ad just one interest rate, the rate on bonds, determined by monetary policy. As you well know, the financial system is vastly more complex than that. There are many interest rates and many financial institutions. And the financial system plays a major role in the economy.

6.1 Nominal versus Real Interest Rates

When we borrow, what we really want to know is how many goods we will have to give up in the future in exchange for the goods we get today. Likewise, when we lend, we want to know how many goods (not how many pounds or euros) we will get in the future for the goods we give up today. The presence of inflation makes this distinction important. What is the point of receiving high interest payments in the future if inflation between now and then is so high that, with what we shall receive the, we shall be unable to buy more goods?

This is where the distinction between nominal interest rates and real interest rates comes in:

- Interest rates expressed in terms of units of national currency are called **nominal interest rates**. If the nominal interest rate for year t is i_t, borrowing one unit of national currency this year requires you to pay $1 + i_t$ units of national currency next year.

- Interest rates expressed in terms of a basket of goods are called **real interest rates**. If we denote the real interest rate for year t by r_t , then, by definition, borrowing the equivalent of one basket of goods this year requires you to pay the equivalent of 1 + r_t baskets of goods next year.

What is the relation between nominal and real interest rates? How do we go from nominal interest rates, which we do observe, to real interest rates, which we typically do not observe? The intuitive answer is: We must adjust nominal interest rate to take into account expected inflation.

After some algebraic passages we can reach the following equation: $r_t \approx i_t - \pi_{t+1}^e$
Where π_{t+1}^e is the expected inflation for $year_{t+1}$.
The equation states that the real interest rate is (approximately) equal to the nominal interest rate minus expected inflation.
- When expected inflation equals zero, the nominal and the real interest rates are equal.
- Because expected inflation is typically positive, the real interest rate is typically lower than the nominal interest rate.
- For a given nominal interest rate, the higher the expected rate of inflation, the lower the real interest rate.

The case where expected inflation happens to be equal to the nominal interest rate is worth looking at more closely. Suppose the nominal interest rate and expected inflation both equal 10%, and you are the borrower. For every euro you borrow this year, you will have to repay 1.10 euros next year. This looks expensive. But euros will be worth 10% less in terms of bread next year. So, if you borrow the equivalent of one pound of bead, you will have to repay the equivalent of one pound of bread next year. The real cost of borrowing, that is, the real interest rate is equal to zero.

Now suppose you are the lender. For every euro you lend this yar, you will receive 1.10 next year. This looks attractive, but euros next year will be worth 10% less in terms of bread. If you lend the equivalent of one pound of bread this year, you will get the equivalent of one pound of bread next year. Despite the 10% nominal interest rate, the real interest rate is equal to zero.

We have assumed so far that there is only one good, bread. But what we have done generalizes easily to many goods. All we need to do is to substitute the **price level**, the price of a basket of goods, for the price of bread in the equation. If we use the **consumer price index (CPI)** to measure the price level, the real interest rate tells us how much consumption we must give up next year to consume more today.

Which interest rate should enter the IS relation? Real Interest, because it is the one that really affects the consumption of people. Although the CB chooses the nominal rate, it cares about the real interest rate because this is the rate that affects spending decisions. To set the interest rate it wants, it thus has to take into consideration expected inflation. If, for example, it wants to set the real interest rate equal to r, it must choose the nominal rate i so that, given expected inflation π^e, the real interest rate, r = i - π^e, is at the level it desires. So, we can think of the CB as choosing the real interest rate. This conclusion comes, however with an important warning. The zero lower bound implies that the nominal interest rate cannot be negative, otherwise people would not want to hold bonds. This implies that the real interest rate cannot be lower than the negative inflation. So, if expected inflation is 2%, for example, then the lowest the real rate can be is 0% - 2% = -2%. So, as long as expected inflation is positive, this allows for negative real interest rates. But if expected inflation turns negative, if people anticipate deflation, then the lower bound on the real rate is positive and can turn out to be high. If, for example, expected deflation is2%. The real rate cannot be less than 2%. This may not be low enough to increase demand for goods by much, and the economy may remain in recession.

6.2 Risk and Risk Premiums

Neither you nor we can borrow at the rate set by the CB. Nor can we borrow at the same rate as the Government. There is a good reason for this. Whoever might be lending to us knows that there is a chance that we may not be able to repay. The same is true for firm that issue bonds. Some firms present little risk and others more. To compensate for the risk, bond holders require a **risk premium**. What determines this risk premium?

- The first factor is the **probability of default** itself. The higher this probability, the higher the interest rate investors will ask for.

  $$(1+i) = (1-p)(1+i+x) + (p)(0)$$

  The left-hand side gives the return on the riskless bond. The right-hand side gives the expected return on the risky bond. With probability (1-p), there is no default and the bon will pay (1+i+x). With probability p, there is default, and the bond will pay nothing. Rearranging gives:

  $$x = (1+i)p/(1-p)$$

- The second factor is the degree of risk aversion of the bond holders. Even if the expected return on the risky bond was the same as on a riskless bond, the risk itself will make them reluctant to hold the risky bond. Thus, they will ask for an even higher premium to compensate for the risk. How much more will depend on their degree of risk aversion. And, if they become more **risk averse**, the risk premium will go up even if the probability of default itself has not changed.

To summarise, in the last two sections we have introduced the concepts of real versus nominal rates and the concepts of a risk premium. In section 6.4, we shall extend the IS-LM model to take both concepts into account. Before we do, let's turn to the role of financial intermediaries. 3 chilometri di distanza

6.3 The Role of Financial Intermediaries

Until now, we have looked at **direct finance**: that is, borrowing directly by the ultimate borrowers from the ultimate lenders. In fact, much of the borrowing and lending takes place through financial intermediaries, which are financial institutions that receive funds from some investors and then lend these funds to others.
Financial intermediaries perform an important function. They develop expertise about specific borrowers and can tailor lending to their specific needs. In normal times, they function smoothly. They borrow and lend, charging a slightly higher interest rate than the rate at which they borrow so as to make a profit.

We start with two definitions:

- **Capital Ratio** of a bank is defined as the ratio of its capital to its assets.
- **Leverage Ratio** of a bank is defined as the ratio of assets to capital, so as the inverse of the capital ratio.

It is traditional to think in terms of leverage and to focus on the leverage ratio. We shall follow tradition. But given the simple relation between the two, the discussion could equivalently be in terms of the capital ratio.
In thinking what leverage ratio it should choose, the bank has to balance two factors:

- A higher leverage ratio implies a higher expected profit rate.
- But a higher leverage ratio also implies a higher risk of bankruptcy (**insolvency**).

Thus, the bank must choose a leverage ratio that takes into account both factors. Too low a leverage ratio means less profit. Too high a leverage ratio means too high a risk of bankruptcy.

Suppose a bank has chosen its preferred leverage ratio, and suppose that the value of its assets declines, then its leverage ratio increases. The bank may be still solvent, but it is clearly more at risk than it was before. What will it want to do? It may want to increase capital, for example by asking other investors to provide funds. But it is also likely to want to decrease the size of its balance sheet. For example, if it can call back some loans, hence reducing assets and then use that to decrease its liabilities.

Suppose that, the decline in the value of the assets is large and the bank will become insolvent and go bankrupt. The borrowers that depended on the bank may have a hard time finding another lender. Why is this relevant to us? Because whether banks remain solvent but cut lending or become insolvent, the decrease in lending that this triggers may well have major adverse macroeconomics effects.

We looked at the case where bank assets declined in value and saw that this led banks to reduce lending. Now consider a case in which investors are unsure of the value of the assets of the bank, and believe, right or wrong, that the value of the assets may have come down. Then, leverage can have disastrous effects:

- If investors have doubt about the value of the bank assets, the safe thing for them is to take their funds out of the bank. But this creates serious problems for the bank, which needs to find the funds to repay the investors. The loans it has made cannot easily be called back. Typically, the borrowers no longer have the funds at the ready. Selling the loans to another bank is likely be difficult as well. In general, the harder it I for others to assess the value of the assets of the bank, the more likely the bank is either to be simply unable to sell them or to have to do it at **fire sale prices**, which are prices far below the true value of the assets of the loans. Such sales, however, only make matters worse for the bank. As the value of the assets decreases, the bank may well become insolvent and go bankrupt. In turn, as investors realise this may happen, this gives them even more reason to want to get their funds out, forcing more fire sales and making the problem worse. Note that this could happen even if the initial doubts of investors were totally unfounded, even if the value of the bank assets had not decreased in the first place. The decision by investors to ask for their funds, and the fire sales tis triggers, can make the bank insolvent even if it was fully solvent to start with.
- Note also that the problem is worse if investors can ask for their funds at short notice. This is clearly the case for **demand deposits**, namely deposits at banks against which cheques can be written. Deposit accounts are so called because people can ask for their funds on demand. The fact that banks' assets are largely composed of loans and their liabilities are largely composed of demand deposits makes them particularly exposed t the risk of runs, and the history of the financial system is full of examples of **bank runs**.

We can summarise what we have just learned in terms of the **liquidity** of assets and liabilities. The lower the liquidity of the assets (i.e. the more difficult they are to sell), the higher the risk of fire sales and the risk that the bank becomes insolvent and goes bankrupt. The higher the liquidity of the liabilities (i.e. the easier it is for investors to get their funds short notice), the higher the risk of fire sales as well, and the risk that the bank becomes insolvent and goes bankrupt. Again, the reason this is relevant for us is that such bankruptcies, if they occur, may well have major macroeconomic consequences.

6.3 Extending the IS-LM Model

IS Relation: $\boldsymbol{Y = C(Y - T) + I(Y, i - \pi^e + x) + G}$
LM Relation: $\boldsymbol{i = \bar{\imath}}$

The LM relation remains the same. The CB still controls the nominal interest rate. But there are two changes to the IS relation, the presence of expected inflation, π^e, and a new term that we shall call the risk premium and denote by x. The two equations make clear that the interest rate entering the LM equation, i , is no longer the same as the interest rate entering the IS relation, $r + x$. Let's call the rate entering the LM equation the (nominal) **policy rate** (because it is determined by monetary policy), and the rate entering the IS equation the (real) **borrowing rate** (because it is the rate at which consumers and firms can borrow).
One simplification: as we discussed in Section 6.2, although the CB formally chooses the nominal interest rate, it can choose it in such a way as to achieve the real interest rate it wants.

Thus, we can think of the CB as choosing the real policy rate directly and rewrite the two equations as:

IS Relation: $\boldsymbol{Y = C(Y - T) + I(Y, r + x) + G}$
LM Relation: $\boldsymbol{r = \bar{r}}$

The CB chooses the real policy rate, r. But the real interest rate relevant for spending decisions is the borrowing rate, $r + x$, which depends not only on the policy rate, but also on the risk premium.

Suppose that, for some reason, x increases. There are many potential scenarios here. This may be for example because investors have become more risk averse and require a higher risk premium, or it may be because one financial institution has gone bankrupt and investors have become worried about the health of other banks, starting a run, forcing these other banks to reduce lending. The IS curve shifts to the left. At the same policy rate r, the borrowing rate, $r + x$, increases, leading to a decrease in demand and a decrease in output. Problems in the financial system lead to a recession. Put another way, a financial crisis becomes a macroeconomic crisis.

What can policy do? Just as before, fiscal policy, be it an increase in G or a decrease in T, can shift the IS curve to the right and increase output. But a large increase in spending or a cut I taxes may imply a large increase in the budget deficit and the government may be reluctant to do so.
Given that the cause of the low output is that the interest rate facing borrowers is too high, monetary policy appears to be a better tool. Indeed, a sufficient decrease in the policy rate can in principle be enough to return output to its initial level. In effect, in the face of the increase in x, the CB must decrease r so as to keep $r + x$, the rate relevant to spending decisions, unchanged.

Given the zero lower bound on the nominal rate, the lowest real rate the CB can achieve is by $r = i - \pi^e = 0 - \pi^e = -\pi^e$. In other words, the lowest real policy rate the CB can achieve is the negative of inflation. If inflation is high enough, say for example 5%, then a zero nominal rate implies a real rate of -5%, which is likely to be low enough to offset the increase in x. But if inflation is low or even negative, then the lowest real rate the CB can achieve may not be enough to offset the increase in x. It may not be enough to return the economy to its initial equilibrium. As we shall see, two characteristics of the recent crisis were indeed a large increase in x and low actual and expected inflation, limiting how much CB could use monetary policy to offset the increase in x.

Chapter 7 – The Labour Market

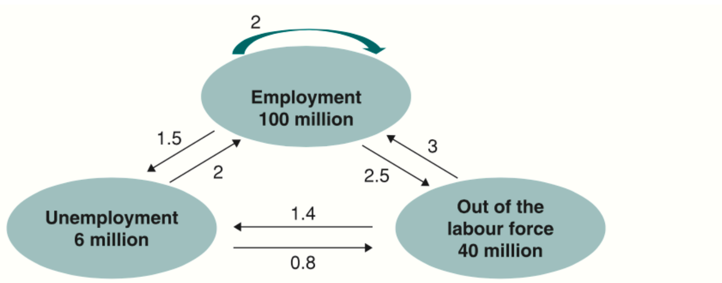
In the medium run, the economy returns to a level of output associated with the natural rate of unemployment. Chapter 7 looks at equilibrium in the labour market. It characterises the natural rate of unemployment, which is the unemployment rate to which the economy tends to return in the medium run. Think about what happens when firms respond to an increase in demand by increasing production. Higher production leads to higher employment. Higher employment leads to lower unemployment. Lower unemployment leads to higher wages. Higher wages increase production costs, leading firms to increase prices. Higher prices lead workers to ask for higher wages. Higher wages lead to further increases in prices, and so on.

So far, we have simply ignored this sequence of events. By assuming a constant price level in the IS – LM model, we in effect assumed that firms were able and willing to supply any amount of output at a given price level. So long as our focus was on the *short run*, this assumption, explore how prices and wages adjust over time, and how this, in turn, affects output.

7.1 A Tour of the Labour Market

The **total EU population** in 2014 was 498.7 million. The **population in working age** was 328.1 million. The **labour force**, which is the sum of those either working or looking for work, was only 245.5 million. The other 85.6 million people were **out of the labour force**, neither working in the marketplace nor looking for work. The **participation rate**, which is defined as the ratio of the labour force to the population in working age, therefore was **73.9%.** Of those in the labour force, 217.7 million were employed, and 24.8 million unemployed, looking for work. The **unemployment rate**, which is defined as the ratio of the unemployed to the labour force, therefore was **11.4%.**

A given unemployment rate may reflect two different realities. It may reflect an active labour market, with many **separations** and many **hires**, and so with many workers entering and exiting unemployment, or it may reflect a sclerotic labour market, with few separations, few hires and a stagnant unemployment pool.



The figure reported on the side has three striking features:

- **The flows of workers in and out of employment are large**. At any given time, some firms are suffering decreases in demand and decreasing their employment, while other firms are enjoying increases in demand and increasing employment.
- **The flows in and out of unemployment are large relative to the number of unemployed**. The **average duration of unemployment**, which is the average length of time people spend unemployed, is between two and three months. This fact has an important implication. You should not think of unemployment as a stagnant pool of workers waiting indefinitely for jobs. For most (but obviously not all) of the unemployed, being unemployed is more a quick transition tan a long wait between jobs.
- **The flows in and out of the labour force are also surprisingly large**. You might have expected these two flows to be composed, on one side, of those finishing school and entering the labour force for the first time, and, on the other side, of workers entering retirement. Bu each of these two groups actually represents a small fraction of the total flows.

What this fact implies is that many of those classified as "out of the labour force" are in fact willing to work and move back and forth between participation and non-participation. Indeed, among those classified as out of the labour force, a large proportion report that, although they are not looking, they "want a job". What they really mean by this statement is unclear, but the evidence is that many do take jobs when offered them.

This fact has another important implication. The sharp focus on the unemployment rate by economists, policy makers and news media is partly misdirected. Some of the people classified as out of the labour force are much like unemployed. They are in effect **discouraged workers**. And although they are not actively looking for a job, they will take it if they find one. This is why economists sometimes focus on the **employment rate**, which is the ratio of employment to the population available for work, rather than on the unemployment rate. We shall follow tradition in this text and focus on the unemployment rate as an indicator of the state of the labour market, but you should keep in mind that the unemployment rate is not the best estimate of the number of people available for work.

7.2 Movements in Unemployment

Years-to-years movements in the unemployment rate are closely associated with recessions and expansions.
How do these fluctuations in the *aggregate unemployment rate* affect *individual workers*?
Let's start by asking ho firms can decrease their employment in response to a decrease in demand. They can hire fewer new workers, or they can lay off the workers they currently employ. Typically, firms prefer to slow or stop the hiring of new workers first, relying on quits and retirements to achieve a decrease in employment. But doing only this may not be enough if the decrease in demand is large, so firms may then have to lay off workers.

- If the adjustment takes place through fewer hires, the chance that an unemployed workers will find a job diminishes. Fewer hires means fewer jobs openings; higher unemployment means more job applicants. Fewer openings and more applicants combine to make it harder for the unemployed to find jobs.
- If the adjustment takes place instead through higher layoffs, ten employed workers are at a higher risk of losing their job.

In general, as firms do both, higher unemployment is associated with both a lower chance of finding a job if one is unemployed and a higher chance of losing it if one is employed.

The relation between movements in the proportion of unemployed workers finding jobs and the unemployment rate is striking:

- Periods of higher unemployment are associated with much lower proportions of unemployed workers finding jobs.
- Higher unemployment implies a higher separation rate, that is, a higher chance of employed workers losing their jobs.

To summarise, when unemployment is high, workers are worse off in two ways:

- Employed workers face a higher probability of losing their job.
- Unemployed workers face a lower probability of finding a job, equivalently, they can expect to remain unemployed for a longer time.

7.3 Wage Determination

Having looked at unemployment, let's turn to wage determination, and to the relation between wages and unemployment. Wages are set in many ways. Sometimes they are set by **collective bargaining**, which is bargaining between firms and unions. The higher the skills needed for the job, the more likely there is to be bargaining. Although institutional differences influence wage determination, there are common forces at work in all countries. Two sets of facts stand out:

- Workers are typically paid a wage that exceeds their **reservation wage**, which is the wage that would make the indifferent between working or being unemployed.
- Wages typically depend on **labour market conditions**. The lower the unemployment rate, the higher the wages.

How much **bargaining power** workers have depends on two factors. The first is how costly it would be for the firm to find other workers, were they to leave the firm. The second is how hard it would be for them to find another job, were they to leave the firm. The costlier it is for the firm to replace them, and the easier it is for them to find another job, the more bargaining power they will have. This has two implications:

- How much bargaining power a worker has depends first on the **nature of the job**.
- How much bargaining power a worker has also depends on labour market conditions. When the unemployment rate is low, it is more difficult for firms to find acceptable replacement workers. At the same time, it is easier for workers to find other jobs. Under these conditions, workers are in a stronger bargaining position and may be able to obtain a higher wage. Conversely, when the unemployment rate is high, finding good replacement workers is easier for firms, whereas finding another job is harder for workers. Being in a weak bargaining position, workers may have no choice but to accept a lower wage.

Regardless of workers' bargaining power, firms may want to pay more than the reservation wage. They may want their workers to be productive, and a higher wage can help them achieve that goal. Economists call the theories that link the *productivity* or the *efficiency* of workers to the wage they are paid **efficiency wage theories**.
Like theories based on bargaining, efficiency wage theories suggest that wages depend on both the nature of the job and labour market conditions:
- Firms, such as high-tech firms, that see employee morale and commitment as essential to the quality of their work will pay more than firms in sectors where workers' activities are more routine.
- Labour market conditions will affect wage. A low unemployment rate makes it more attractive for employed workers to quit. When unemployment is low, it is easy to find another job. That means, when unemployment decreases, a firm that wants to avoid an increase in quits will have to increase wages to induce workers to stay with the firm. When this happens, lower unemployment will again lead to higher wages. Conversely, higher unemployment will lead to lower wages.

We can capture our discussion of wage determination by using the following equation:

$$\boldsymbol{W = P^e F(u, z)}$$
$$(-,+)$$

The aggregate nominal wage W depends on three factors:
- The expected price level, $\boldsymbol{P^e}$
- The unemployment rate, $\boldsymbol{u}$
- A catch-all variable, $\boldsymbol{z}$, that stands for all other variables that may affect the outcome of wage setting.

Why does the **price level** affect nominal wages? Because both workers and firms care about *real wages*, not nominal wages:
- Workers do not care about how many dollars they receive but about how many goods they can buy with those dollars. In other words, they do not care about the nominal wages they receive, but about the nominal wages (W) they receive relative to the price of the goods thy buy (P). They care about W/P.
- In the same way, firms do not care about the nominal wages they pay but about the nominal wages (W) they pay relative to the price of the goods they sell (P). So they also are about W/P.

Why do wages depend on the *expected price level*, rather than the *actual price level*? Because wages are set in nominal terms, and when they are set, the relevant price level is not yet known.

Also affecting the aggregate wage is the **unemployment rate**, u. The minus sign under u indicates that an increase in the unemployment rate decreases wages.

The third variable, **z**, is a catch-all variable that stands for all the factors that affect wages given the expected price level and the unemployment rate. By convention, we will define z so that an increase in z implies an increase in the wage. Take, for example, **unemployment insurance**, which is the payment of unemployment benefits to workers who lose their job; by making the prospects of unemployment less distressing, more generous unemployment benefits do increase wages at a given unemployment rate. In this case, we can think of z as representing the level of unemployment benefits. At a given unemployment rate, higher unemployment benefits increase the wage. Another example could be an increase in **employment protection**, which makes it more expensive for firms to lay off workers. Such a change is likely to increase the bargaining power of workers covered by this protection, increasing the wage for a given unemployment rate.

7.4 Price Determination

Having looked at wage determination, let's now turn to price determination. The prices set by firms depend on the costs they face. These costs depend, in turn, on the nature of the **production function**, which is the relation between the inputs used in production and the quantity of output produced, and on the prices of these inputs. For the moment, we will assume firms produce goods using labour as the only factor of production. We will write the production function as follows:

$$Y = AN$$

Where Y is output, N is employment and A is labour productivity. This way of writing the production function implies that **labour productivity**, which is output per worker, is constant and equal to A. Given these assumptions, we can make one further simplification. We can choose the units of output so that one worker produces one unit of output, in other words, so that A = 1. With this assumption, the production function becomes:

$$Y = N$$

The production function implies that the cost of producing one more unit of output is the cost of employing one more worker, at wage W. If there were perfect competition in the goods market, the price of a unit of output would be equal to marginal cost: P would be equal to W. But many goods market are not competitive, and firms charge a price higher than their marginal cost. A simple way of capturing this fact is to assume that firms set their price according to:

$$P = (1 + m)W$$

Where m is the **mark-up** of the price over the cost. If goods market were perfectly competitive, m would be equal to zero, and the price, P, would simply equal the cost, W. TO the extent that they are not competitive and firms have market power, m is positive and the price, P, will exceed the cost, W, by a factor equal to (1+m).

7.5 The Natural Rate of Unemployment

Let's now look at the implications of wage and price determination for unemployment. We shall do so under the assumption that nominal wages depend on the actual price level P, rather than on the expected price level P^e. Under this additional assumption, wage setting and price setting determine the equilibrium rate of unemployment.

Give the assumption that nominal wages depend on the actual price level (P), the equation that characterises wage determination, becomes:

$$W = PF(u, z)$$

Dividing both sides by the price level:

$$\frac{\boldsymbol{W}}{\boldsymbol{P}} = \boldsymbol{F}(\boldsymbol{u}, \boldsymbol{z})$$
$$(-, +)$$

Wage determination implies a negative relation between the real wage, W/P, and the unemployment rate, u: the higher the unemployment rate, the lower the real wage chose by wage setters. The intuition is straightforward. The higher the unemployment rate, the weaker the workers' bargaining position, and the lower the real wage will be. This relation between the real wage and the rate of unemployment – let's call it the **wage – setting relation** – is draw in the following figure as the downward-sloping curve **WS**.

Let's now look at the implication of price determination. If we divide both sides of the price determination equation by nominal wage, we get:

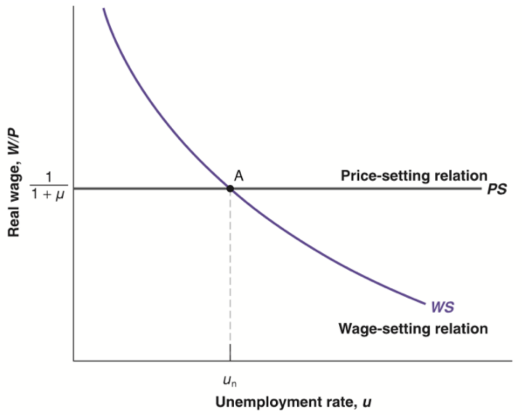
$$\frac{P}{W} = (1 + m)$$

The ratio of the price level to the wage implied by the price-setting behaviour of firms equals one plus the mark-up. Now invert both sides of this equation to get the implied real wage:

$$\frac{W}{P} = \frac{1}{(1 + m)}$$

Note what this equation says: price-setting decisions determine the real wage paid by firms. An increase in the mark-up leads firms to increase their prices given the wage they have to pay; equivalently, it leads to a decrease in the real wage. How price setting actually determines the real wage paid by firms may not be intuitively obvious. Think of it this way: suppose the firm you work for increases its mark-up and therefore increases the price of its product. Your rea wage does not change much; you are still paid the same nominal wage, and the product produced by the firm is at most a small part of your consumption basket. Now, suppose that the firm you work for and all the firms in the economy increase their mark-up. All the prices go up. Even if you are paid the same nominal wage, your real wage goes down. So, the higher the mark-up set by firms, the lower your (and everyone else's) real wage will be.

The **price-setting relation** is drawn as the horizontal line PS. The real wage implied by price setting is $1/(1 + m)$; it does not depend on unemployment rate.



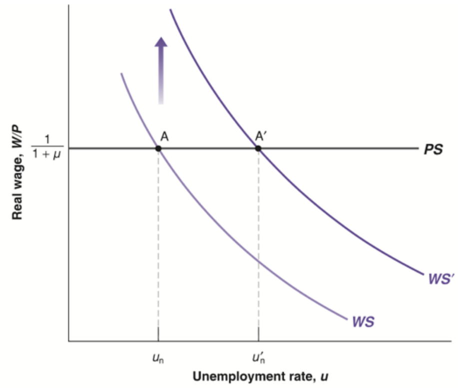
Equilibrium in the labour market requires that the real wage chosen in wage setting be equal to the real wage implied by price setting. We can also characterise the equilibrium unemployment rate algebraically. Eliminating W/P between equations, gives:

$$\frac{1}{(1 + m)} = F(u_n, z)$$

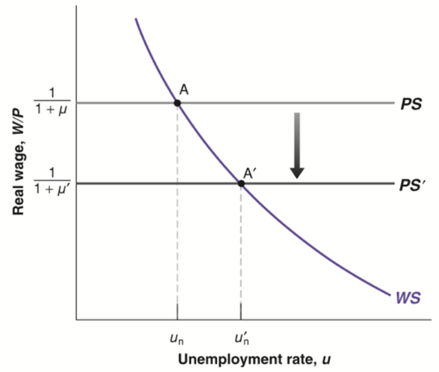
The equilibrium unemployment rate, u_n is called the **natural rate of unemployment**. The word suggests a constant of nature, one that is unaffected by institutions and policy. As its derivation makes clear, however, the natural rate of unemployment is anything but natural. The positions of the wage-setting and price-setting curves, and thus the equilibrium unemployment rate, depend on both z and m.

Consider two examples:

- **An increase in unemployment benefits**. An increase in unemployment benefits can be represented by an increase in z. Because an increase in benefits makes the prospect of unemployment less painful, it increases the wage set by wage setters at a given unemployment rate. So it shifts the wage-setting relation up, from WS to WS'. The economy moves along the PS line, from A to A'. The natural rate of unemployment increases from u_n to $u_n{}'$.



- **A less stringent enforcement of existing antitrust legislation**. To the extent that this allows firms to collude more easily and increase their market power, it will lead to an increase in their mark-up, that is, an increase in m. The increase in m implies a decrease in the real wage paid by firms, and so it shifts the price-setting relation down, from PS to PS'. The economy moves along WS. The equilibrium moves from A to A' and the natural rate of unemployment increases from u_n to $u_n{}'$.



Factors like the generosity of unemployment benefits or antitrust legislation can hardly be thought of as the result of nature. Rather, they reflect various characteristics of the structure of the economy For that reason, a better name for the equilibrium rate of unemployment would be the **structural rate of unemployment**, but so far the name has not caught on.

7.6 Where We Go From Here

In the short run, the factors that determine movements in output are the factors we focused on in the preceding four chapters: monetary policy, fiscal policy, and so on.
In the medium run, output tends to return to the natural level. The factors that determine unemployment and, by implication, output are the factors we have focused on in chapter 7.

Chapter 8 – The Phillips Curve, the Natural Rate of Unemployment and Inflation

In 1958, A.W. Phillips drew a diagram plotting the rate of inflation against the rate of unemployment in the United Kingdom for each year from 1861 to 1957. He found clear evidence of a negative relation between inflation and unemployment. When unemployment was low, inflation was high, and when unemployment was high, inflation was low, often even negative. This relation, labelled the **Phillips Curve**, rapidly became central to macroeconomic thinking an policy. It appeared to imply that countries could choose between different combinations of unemployment and inflation. A country could achieve lower unemployment if it were willing to tolerate higher inflation, or it could achieve price-level stability – zero inflation – if it were willing to tolerate higher unemployment. In the 1970s, however, this relation broke down. In the United States and most OECD countries, there were both high inflation and high unemployment, clearly contradicting the original Phillips curve. A relation reappeared, but it reappeared as a relation between the unemployment rate and the change in the inflation rate.

8.1 Inflation, Expected Inflation and Unemployment

This section shows how the model of the labour market implies a relation between inflation, expected inflation and unemployment.

Wage Determination: $\boldsymbol{W = P^e F(u, z)}$
Price Determination: $\boldsymbol{P = (1 + m)W}$

We want to explore what happens if we avoid the previously used assumption that the price level was equal to the expected price level. Replacing the nominal wage in the second equation by its expression from the first gives:

$$\boldsymbol{P = P^e(1 + m)F(u, z)}$$

An increase in the expected price level leads to an increase in nominal wages, which in turn leads firms to increase their prices, and thus leads to an increase in the price level. An increase in the unemployment rate leads to a decrease in nominal wages, which in turn leads to lower prices, and a decrease in the price level.
It will be convenient to assume a specific form for the function, F:

$$\boldsymbol{F(\mathrm{u}, \mathrm{z}) = 1 - \alpha \mathrm{u} + \mathrm{z}}$$

This captures the notion that the higher the unemployment rate, the lower the wage; and the higher z, the higher the wage. The parameter α captures the strength of the effect of unemployment on the wage. Replacing the function, F, by this specific form in the equation above gives:

$$\boldsymbol{P = P^e(1 + m)(1 - \alpha \mathrm{u} + \mathrm{z})}$$

This gives us a relation between the price level, the expected price level and the unemployment rate. Our next step is to derive a relation between inflation, expected inflation and the unemployment rate. Let π denote the inflation rate and π^e denote the expected inflation rate.
Then the equation can be rewritten as:

$$\boldsymbol{\pi = \pi^e + (\mathrm{m} + \mathrm{z}) - \alpha \mathrm{u}}$$

- **An increase in expected inflation leads to an increase in actual inflation**. An increase in the expected price level leads, one for one, to an increase in the actual price level. If wage setters expect a higher price level, they set a higher nominal wage, which leads in turn to an increase in the price level. Now note that, given last period's price level, a higher price level this period implies a higher rate of increase in the price level from last period to this period – that is, higher inflation. Similarly, given last period's price level, a higher expected price level this period implies a higher expected rate of increase in the price level from last period to this period – that is, higher expected inflation.
- **Given expected inflation, an increase in the mark-up, or an increase in the factors that affect wage determination (z) leads to an increase in actual inflation**. Given the expected price level, an increase in either m or z increases the price level. And as before…
- **Given expected inflation, a decrease in the unemployment rate leads to an increase in the actual inflation**. Given the expected price level, a decrease in the unemployment rate leads to a higher nominal wage, which leads to a higher price level. And as before…

8.2 The Phillips Curve and Its Mutations

Assume that inflation varies from year to year around some value $\bar{\pi}$. Assume also that inflation is not persistent, so that inflation this year is not a good predictor of inflation next year. In such an environment, it makes sense for wage setters to assume that, whatever inflation was last year, inflation this year will simply be equal to $\bar{\pi}$. In this case $\pi_t^e = \bar{\pi}$ and equation becomes:

$$\boldsymbol{\pi_t = \bar{\pi} + (m + z) - \alpha u_t}$$

In this case, we observe a negative relation between unemployment and inflation. This is precisely the negative relation between unemployment and inflation that Phillips found for the United Kingdom and Solow and Samuelson found for the United States. When unemployment was high, inflation was low, even sometimes negative. When unemployment was low, inflation was positive.

When these findings were published, they suggested that policy makers faced a trade-off between inflation and unemployment. If they were willing to accept more inflation, they could achieve lower unemployment. This looked like an attractive trade-off and, starting in the early 1960s, US macroeconomic policy aimed at steadily decreasing unemployment.
From 1961 to 1969, the unemployment rate declined steadily from 6.8% to 3.4%, and the inflation rate steadily increase, from 1.0% to 5.5%. Around 1970, however, the relation between the inflation rate and the unemployment rate broke down.
Why did the original Phillips curve vanish? Because wage setters changed the way they formed their expectations about inflation. This change came, in turn, from a change in the behaviour of inflation. The rate of inflation became more persistent. High inflation in one year become more likely to be followed by high inflation in the next year. As a result, people, when forming expectations, started to take into account the persistence of inflation. In turn, this change in expectation formation changed the nature of the relation between unemployment and inflation.

Let's look at the argument in the previous paragraph more closely. Suppose expectations of inflation are formed according to:

$$\boldsymbol{\pi_t^e = (1 - \theta)\bar{\pi} + \theta\pi_{t-1}}$$

In words, expected inflation this year depends partly on a constant value, $\bar{\pi}$, with weight, $(1 - \theta)$, and partly on inflation last year, which we denote by π_{t-1}, with weight θ. The higher the value of θ, the more last year's inflation leads workers and firms to revise expectations of what inflation will be this year, and so the higher is the expected inflation rate.

We can then think of what happened in the 1970s as an increase in the value of θ over time.

- When θ equals zero, we get the original Phillips curve, a relation between the inflation rate and the unemployment rate: $\boldsymbol{\pi_t = \bar{\pi} + (m + z) - \alpha u_t}$

- When θ equals one, the relation becomes: $\boldsymbol{\pi_t - \pi_{t-1} = (m + z) - \alpha u_t}$

  So, when $\theta = 1$, the unemployment rate affects not the inflation rate, but rather the change in the inflation rate. High unemployment leads to decreasing inflation; low unemployment leads to increasing inflation.

Although there is a clear negative relation between unemployment and the change in the inflation rate, you can see that the relation is far from tight. Some points are far from the regression line. The Phillips curve is both a crucial and a complex economic relation.

The natural unemployment rate is the unemployment rate at which the inflation rate remains constant. When the actual unemployment rate exceeds the natural rate of unemployment, the inflation rate typically decreases; when the actual unemployment rate is less than the natural unemployment rate, the inflation rate typically increases.

8.3 …

Chapter 9 – Putting all markets together: from the short to the medium run

Previously, we looked at equilibrium in the goods and financial markets and saw how, in the short run, output is determined by demand. Then we looked at equilibrium in the labour market and derived how unemployment affects inflation. We now put the two parts together and use them to characterise the behaviour of output, unemployment and inflation, in both the short and the medium runs. When confronted with a macroeconomic question about a particular shock or a particular policy, this model, which we shall call the **IS-LM-PC** model, is typically the model we use or start from.

9.1 The IS – LM – PC Model

$$Y = C(Y - T) + I(Y, r + x) + G$$

In the short run, as we already saw, output is determined by demand as described in the above equation. While the relation between inflation and unemployment is called Phillips Curve.

$$\pi - \pi^e = -a(u - u_n)$$

When the unemployment rate is lower than the natural rate, inflation turns out to be higher than expected. If the unemployment is higher than the natural rate, inflation turns out to be lower than expected. Given that the first relation is in terms of output, our first step must be to rewrite the Phillips Curve in terms of output rather than unemployment rate and employment.

$$u = U/L = (L - N)/L = 1 - N/L$$

where N denotes employment and L denotes the labour force.
The unemployment rate is equal to one minus the ratio of unemployment to the labour force. Rearranging to express N as a function of u gives:

$$N = L(1 - u)$$

Employment is equal to the labour force times one minus the unemployment rate. Turning to output:

$$Y = N = L(1 - N)$$

where the second equality follows from the previous equation.
Thus, when the unemployment rate is equal to the natural rate, u_n, employment is given by $N_n = L(1 - u_n)$ and output is equal to $Y_n = L(1 - u_n)$. Call N_n the natural level of employment (natural employment for short) and Y_n the natural level of output (natural output for short), Y_n is also called **potential output**, and we shall often use that expression in what follows. It follows that we can express the deviation of unemployment from its natural level as:

$$Y - Y_n = L\big((1 - u) - (1 - u_n)\big) = -L(u - u_n)$$

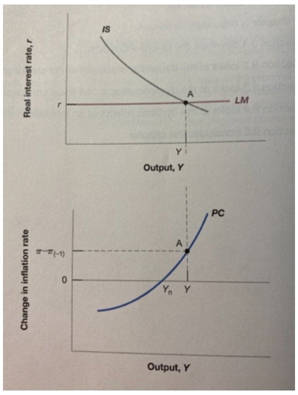
This gives us a simple relation between the deviation of output from potential and the deviation of unemployment from its natural rate. The difference between output and potential output is called the **output gap**. If unemployment is equal to the natural rate, output is equal to potential and the output gap is zero; if unemployment is above the natural rate, output is below potential and the output gap is negative; and if unemployment is below the natural rate, output is above potential and the output gap is positive.

Replacing $(u - u_n)$ in the previous equation gives:

$$\pi - \pi^e = -(a/L)(Y - Y_n)$$

We need to take one last step. We saw earlier how the way wage setters from expectations has changed through time. We shall work in this chapter under the assumption that they assume that inflation this year is the same as last year. This assumption implies that the Phillips Curve relation is given by:

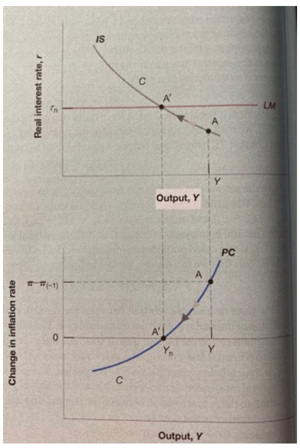
$$\pi - \pi_{(-1)} = (a/L)(Y - Y_n)$$



In words, when output is above potential and therefore the output gap is positive, inflation increases. When the output is below potential and therefore the output gap is negative, inflation decreases. The positive relation between output and the change in inflation is drawn at an upward-sloping curve. When output is equal to potential, equivalently when the output gap is equal to zero, the change in inflation is equal to zero. Thus, the Phillips Curve crosses the horizontal axis at the point where output is equal to potential.

9.2 Dynamics and the Medium-Run Equilibrium

Suppose that the policy rate chosen by the central bank is equal to r. The top part of the figure tells us that, associated with this interest rate, the level of output is given by Y. The bottom part of the figure tells us that this level of output Y implies a change in inflation equal to $\pi - \pi_{(-1)}$. Given the way we have drawn the figure, Y is larger than Y_n, so output is above potential. This implies that inflation is increasing. Put less formally, the economy is overheating, putting pressure on inflation. This is the short-run equilibrium. What happens over time if there is no change in the policy rate, nor in any of the variables which affect the position of the IS curve? Then output remains above potential and inflation keeps increasing. At some point, however, policy is likely to react to this increase in inflation. If we focus on the central bank, sooner or later it will increase the policy rate so as to decrease output back to potential and there is no longer pressure on inflation. The adjustment process and the medium-run equilibrium are represented in the picture aside. Let the initial equilibrium be denoted by point A, in both the top and bottom graphs. You can think of the central bank as increasing the policy rate over time, so the economy moves along the IS curve up from A to A'. Output decreases.



Now turn to the bottom graph. As output decreases, the economy moves down the PC curve down from A to A'. At point A', the policy rate is equal to r_n, output is equal to Y_n and, by implication, inflation is constant. This is the medium-run equilibrium. Output is equal to potential and, as a result, there is no longer any pressure on inflation.

You may have the following reaction to the description of the dynamics. If the central bank wants to achieve stable inflation and keep output at Y_n, why does it not increase the policy rate to r_n right away, so that the medium-run equilibrium is reached without delay? The answer is that the central bank would indeed like to keep economy at Y_n. But, although it looks easy to do on a graph, reality is more complex. The reasons parallel the discussion we had about the adjustment of the economy output is exactly, and thus how far output is from potential. The change in inflation provides signal of the output gap, the distance between actual and potential output, but in contrast to the simple equation, the signal is noisy. The central bank may thus want to adjust the policy rate slowly and see what happens. Second, it takes time for the economy to respond.

The fact that it takes time for output to go back to its natural level raises an issue about inflation. During the process of adjustment, output is consistently above potential, and so inflation is consistently increasing. Thus, when the economy reaches point A', inflation is higher than it was at point A. If the central bank cares not only about stable inflation, but about the level of inflation, it may well decide that it has both to stabilise and to reduce inflation. To do so, it needs to increase the policy rate beyond r_n to generate a decrease in inflation, until inflation is back at a level acceptable to the central bank. In other words, if the central bank wants to achieve a constant level of inflation over the medium run, then the initial boom must be followed by a recession.

The role of expectations revisited

The previous discussion depends on the way people form expectations and on the specific form of the Phillips Curve. To see this, return to our earlier discussion of expectation formation and instead of assuming that expected inflation is equal to last year's inflation, $\pi_{(-1)}$, assume instead that people think that inflation will be equal to some constant $\bar{\pi}$, irrespective of what inflation was last year. In this case, the equation becomes:

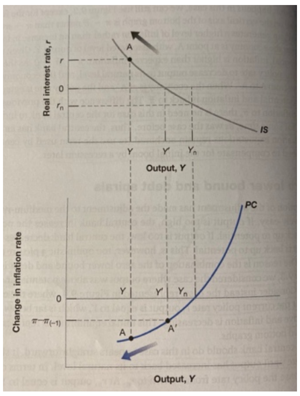
$$\pi - \bar{\pi} = (a/L)(Y - Y_n)$$

A positive output gap generates a higher level of inflation rather than an increase in inflation. Now suppose that the economy is at point A, with associated level of output Y. Given that output is above potential, inflation is higher than expected inflation: $\pi - \bar{\pi} > 0$. As the central bank increases the policy rate to decrease output to its natural level, and the economy moves along the IS curve from A to A'. When the economy is at A' and the policy rate is equal to r_n, output is back to potential and inflation to $\bar{\pi}$, there is no need in this case for the central bank to increase the rate beyond r_n for some time, as was the case before. Thus, the central bank has an easier job. So long as inflation expectations remain anchored (to use the term used by central banks), it does not need to compensate for the initial boom by a recession later.

The zero lower bound and debt spirals

Our description of the adjustment has made the adjustment to the medium-run equilibrium look relatively easy. If output is too high, the central bank increases the policy rate until output is back down to potential. If output is too low, the central bank decreases the policy rate until output is back up to potential. This is, however, too optimistic a picture and things can go wrong. The reason is the combination of the zero lower bound and deflation.

In the previous figure, we considered the case where output was above potential, and inflation was increasing. Consider instead the case, represented in the following picture, where the economy is in a recession. At the current policy rate r, output is equal to Y, which is far below Y_n. The output gap is negative, and inflation is increasing. What the central bank should do in this case appears straightforward. It should decrease the policy rate until output has increased back to its natural level.



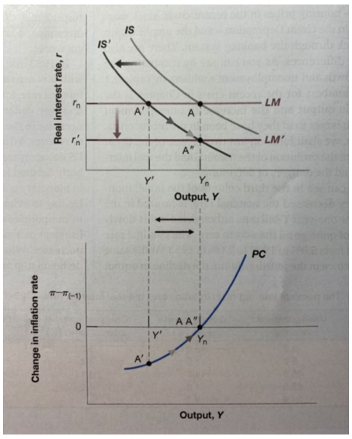
Note that, if the economy is sufficiently depressed, the real policy rate, r_n needed to return output to its natural level may be negative. If the zero-lower bound prevents monetary policy from increasing output back to potential the result may be a deflation

spiral. More deflation leads to a higher real policy rate in turn leads to lower output and more deflation.
The zero lower bound constraint may, however, make it impossible to achieve the negative real policy rate. Suppose, for example, that initial inflation is zero. Because of the zero-lower bound, the lowest the central bank can decrease the nominal policy rate is 0%, which combined with zero inflation, implies a real policy rate of 0%. In terms of the figure, the central bank can decrease the real policy rate only down to 0%, with associated level of output Y'. At Y', output is still below potential, and thus inflation is still decreasing. This starts what economists call a **deflation spiral**, or **deflation trap**.
Let's continue to assume that inflation expectations are such that wage setters expect inflation to be the same as last year, so a negative output gap implies decreasing inflation. If inflation was equal to zero to start with, it becomes negative. Zero inflation turns into deflation. This implies in turn that even if the nominal rate remains equal to zero, the real policy rate increases, leading to even lower demand and lower output. Deflation and low output feed on each other. Lower output leads to more deflation, and more deflation leads to a higher real interest rate and lower output.
There is little the central bank can do, and the economy goes from bad to worse.
This scenario is not just a theoretical concern. This is very much the scenario which played out during the Great Depression.

The recent crisis gave rise to similar worries. With the policy rate down to zero in the major advanced countries, the worry was that inflation would turn negative and start a similar spiral. This did not happen. Inflation decreased and, in some countries turned to deflation. As we saw previously, this limited the ability of the central banks to decrease the real policy rate and increase output. But deflation remained limited and the deflation spiral did not happen. One reason, which connects to our previous discussion of expectations formation, is that inflation expectations remained largely anchored.

9.3 Fiscal Consolidation Revisited

In this section, we go back to the fiscal consolidation we discussed previously. We can now look not only at the short-run effects, but also at its medium-run effects as well.
Suppose that output is at potential, so the economy is at point A in both the top and bottom graphs of the figure. Output Y is equal to Y_n, the policy rate is equal to r_n, and inflation is stable. Now, assume that the government, which was running a deficit, decides to reduce it by, say, increasing taxes. This increase in taxes shifts the IS curve to the left, from IS to IS'. The new short-run equilibrium is given by point A' in both the top and bottom graphs. At the given policy rate r_n, output decreases from Y_n to Y' and inflation starts decreasing. In other words, if output was at potential to start with, the fiscal consolidation, as desirable as it may be on other grounds, leads to a recession. This is the short-run equilibrium. Note that, as income comes down and taxes increase, consumption decreases on both counts. Note also that, as output decreases, so does investment. In the short run, on macroeconomic grounds, fiscal consolidation looks rather unappealing: both consumption and investment go down.



Let's, however, turn to the dynamics and to the medium run. As output is too low, and inflation is decreasing, the centra bank is likely to react and decrease the policy rate until output is back to potential. In terms of the precedent figure, the economy moves down the IS curve in the top graph and output increases. As output increases, the economy moves up the PC curve in the bottom graph, until output is back to potential. Thus, the medium-run equilibrium is given by point A'' in both the top and the bottom graphs. Output is back at Y_n and inflation is again stable. The policy rate needed to maintain output at potential is now lower than before, equal to $r_n{}'$ rather than r_n. Now look at the composition of output in this new equilibrium. As income is the same as it was before fiscal consolidation but taxes are higher, consumption is lower, although not as low as it was in the short run. As output is the same as before but the interest rate is lower, investment is higher than before. In other words, the decrease in consumption is offset by an increase in investment, so demand, by implication, is unchanged. This is in sharp contrast to what happened in the short run and makes fiscal consolidation look more attractive. Although consolidation may decrease investment in the short run, it increases it in the medium run.

This discussion raises some of the same issues we discussed in the previous section. First, it looks as if fiscal consolidation could take place without a decrease in output in the short run. All that is needed is for the central bank and the governments to coordinate carefully. As fiscal consolidation takes place, the central bank should decrease the policy rate so as to maintain output at the natural level. In other words, the proper combination of fiscal and monetary policy can achieve the medium-run equilibrium outcome in the short-run. Such coordination indeed happens sometimes. But it does not happen always. One reason is that the central bank may be unable to decrease the policy rate sufficiently. The central bank may have limited room to decrease the policy rate. This indeed has been the case in the euro area in the recent crisis. With the nominal policy rate at zero in the euro area, monetary policy was unable to offset the adverse effects of fiscal consolidation on output. The result was a stronger and longer lasting adverse effect of fiscal consolidation on output than would have been the case, had the European Central Bank been able to decrease the policy rate further.

9.4 The Effects of an Increase in the Price of Oil

So far we have looked at shocks to demand, shocks that shifted the IS curve but left potential output and this the position of the PC curve unaffected. There are other shocks, however, that affect both demand and potential output and play an important role in fluctuations. An obvious candidate is movements in the price of oil. What matters for economic decisions is not the dollar price, but the real price of oil; that is, the dollar price of oil divided by the price level.

Twice over the last 40 years, the US economy was hit with a five-fold increase in the real price of oil, the first time in the 1907s and the second time in the 2000s. The crisis then led to a dramatic drop in late 2008, followed by a partially recovery. And since 2014, the price has again dropped to pre-200 levels. What was behind the two large increases? In the 1970s, the main factors were the formation of **OPEC** (the Organization of Petroleum Exporting Countries), a cartel of oil producers that was able to act as a monopoly and increase prices, and disruptions due to wars and revolutions in the Middle East. In the 2000s, the main factor was quite different, namely the rapid growth of emerging economies, in particular China, which led to a rapid increase in the world demand for oil and, by implication, a steady increase in real oil prices.

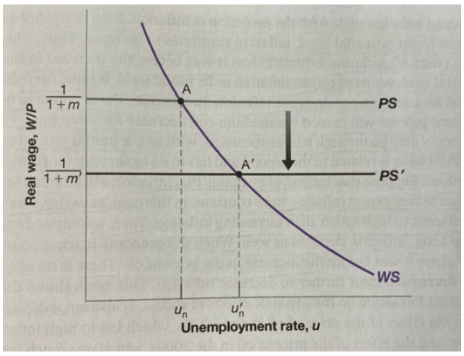
What was behind the two large decreases? The sudden drop in the price at the end of the 2008 was a result of the crisis, which led to a large recession and in turn to a large and sudden decrease in the demand for oil. The causes of the more recent drop since 2014 are still being debated. Most observers believe that it is a combination of increased supply because of the increase in shale oil production in the United States and the partial breakdown of the OPEC cartel.

Let’s focus on the two large increases. Although the causes were different, the implications for US firms and consumers was the same: more expensive oil. The question is: What would we expect the short-and-medium run effects of such increase to be? It is clear, however, that, in answering the question, we face a problem. The price of oil appears nowhere in the model we have developed so far! The reason is that, until now, we have assumed that output was produced using only labour. One way to extend our model is simply to capture the increase in the price of oil by an increase in m, the mark-up of the price over the nominal wage. The justification is straightforward. Given wages, an increase in the price of oil increases the cost of production, forcing firms to increase prices to maintain the same profit rate. Having made this assumption, we can then track the dynamic effects of an increase in the *mark-up* on output and inflation.

Effects on the Natural Rate of Unemployment
Let's start by asking what happens to the natural rate of unemployment when the real price of oil increases. The wage setting relation is downward sloping; a higher unemployment rate leads to lower real wages. The price-setting relation is represented by the horizontal line at:

$$W/P = 1/(1+m)$$

The initial equilibrium is at point A and the initial natural unemployment rate is u_n. An increase in the mark-up leads to a downward shift of the price-setting line, from PS to PS'. The higher the mark-up, the lower the real wage implied by price setting.



The equilibrium moves from A to A'. The real wage is lower and the natural unemployment rate is higher. Think of it this way: because firms have to pay more for the oil, the wage they can pay is lower. Getting workers to accept the lower real wage requires an increase in unemployment. The increase in the natural rate of unemployment leads in turn to a decrease in the natural level of employment. If we assume that the relation between employment and output is unchanged – that is, each unit of output still requires one worker in addition to the energy input – then the decrease in the natural level of employment leads to an identical decrease in potential output. Putting things together, an increase in the price of oil leads to a decrease in potential output.

We can now go back to the IS – LM – PC model. Assume the initial equilibrium is at point A in both the top and bottom panels, with output at potential, so Y is equal to Y_n, inflation is stable and the policy rate is equal to r_n. As the price of oil increases, the natural level of output decreases, say from Y_n to Y_n'. The PC curve shifts up, from PC to PC'. If the ISS curve does not shift and the central bank does not change the policy rate, output does not change, but the same level of output is now associated with higher inflation. For given wages, the price of oil leads firms to increase their prices, so inflation is higher. The short-run equilibrium is given by point A' in the top and bottom panels. In the short-run, output does not change, but inflation is higher.
Turning to the dynamics, if the central bank were to leave the policy rate unchanged, output would continue to exceed the now lower level of potential output and inflation would keep increasing. Thus, at some point, the central bank will increase the policy rate to stabilise inflation. As it does so, the economy moves up from A' to A'' along the IS curve in the top panel, and down from A' to A'' along the PC curve in the bottom panel. As output decreases to its lower level, inflation continue to increase, although more and more slowly until eventually it becomes stable again. Once the economy is at point A'', the economy is in its medium-run equilibrium. Because potential output is lower, the increase in the price of oil is reflected in a permanently lower level of output. Note that along the way, lower output is associate with higher inflation, a combination that economists call **stagflation** (i.e. a contraction of stagnation and inflation).

As in the previous sections, this description raises a number of issues. The first is our assumption that the IS curve does not shift. In fact, there are many channels through which the increase in the price of oil may affect demand and shift the IS curve. The higher the price of oil may lead firms to change their investment plans, cancelling some investment projects, shifting to less energy-intensive equipment. The increase in the price of oil also redistributes income from oil buyers to oil producers. Oil producers may spend less than oil buyers, leading to a decrease in demand. So it may well be that the IS curve shifts to the left, leading to a decrease in output not only in the medium run, but in the short run as well.

A second issue has to do with the evolution of inflation. Note that, until output decreases to its new lower potential level, inflation continues to increase. Thus, when the economy reaches point A'', inflation is higher than it was before the increase in the price of oil. If the central bank wants to return inflation to its initial level, it must decrease output below potential for some time to decrease inflation. In this case, the decrease in output along the adjustment process will exceed the medium-run decrease for some time. Put more simply, the economy may go through a large recession, with only a partial recovery.

The third issue is related to the second and has to do again with the formation of inflation expectations. Suppose that instead of assuming that inflation will be equal to last year's inflation, wage setters expect inflation to be constant. In this case, as we have seen, output above potential leads to high rather than increasing inflation. Then, as output declines to its lower potential level, inflation declines as well. When the economy reaches point A'', inflation is back to where it was before the increase in the price of oil. There is no need for the central bank to decrease output further to decrease inflation.

The third hypothesis regards the evolution of energy use in the last three decades follows the trend in energy intensity of GDP. All countries have seen a decrease in this parameter, which means that for every unit of output, the power requirements have decreased in almost every country in the world, hence the impact of an increase in the price of oil is lower today than it was in the 1970s. In many cases, these reductions can be attributed to strong improvements in energy efficiency following the introduction of technology and modern production processes.

9.5 Conclusions

This chapter has covered a lot of ground. Let us repeat some key ideas and develop some of the conclusions.

The Short-Run versus the Medium-Run

One key message of this chapter is that shocks or changes in policy typically have different effects in the short-run and in the medium-run. If you are worried about output and investment in the short-run, you might be reluctant to proceed with fiscal consolidation. But if your focus is on the medium and long-run, you will see the consolidation as helping investment and eventually, through higher investment and thus capital accumulation, increasing output.

Shocks and Propagation Mechanisms

This chapter gives you a general way of thinking about **output fluctuations** (sometimes called **business cycles**) – movements in output around a trend.

You can think of the economy as being constantly hit by **shocks**. These shocks may be shifts in consumption coming from changes in consumer confidence, shifts in investment, and so on. Or they may come from changes in policy – from the introduction of a new tax law, to a new scheme of infrastructure investment, to a decision by the central bank to fight inflation.

Each shock has dynamic effects on output and its components. These dynamic effects are called the **propagation mechanism** of the shock. Propagation mechanisms are different for different shocks. The effects of a shock on activity may build up over time, affecting output in the medium-run. Or the effects may build up for a while and then decrease and disappear. At times, some shocks are sufficiently large or come in sufficiently bad combinations that they create a recession. The two recession of the 1970s were due largely to increases in the price of oil; the recession of the early 1980s was due to a sharp contraction in money; the recession of the early 1990s was due primarily to a sudden decline in consumer confidence; the recession of 2001 was due to a sharp drop in investment spending. The recent crisis and the sharp decrease in output in 2009 had its origins in the problems of the housing market, which then led to a major financial shock and in turn to a sharp reduction in output. What we call economic fluctuations are the result of these shocks and their dynamic effects on output. Typically, the economy returns over time to its medium-run equilibrium. But, as we have seen when discussing for example the interaction between the zero lower bound and deflation, things can get quite bad form some time.

Chapter 14 – Financial Markets and Expectations

This chapter focuses on the role of expectations in financial markets. It introduces the concept of expected present discounted value, which plays a central role in the determination of asset prices and in consumption and investment decisions. Using this concept, it studies the determination of bond prices and bond yields. It shows how we can learn about the course of expected future interest rates by looking at the yield curve. It then turns to stock prices and shows how they depend on expected future dividends and interest rates. Finally, it discusses whether stock prices always reflect fundamentals or may instead reflect bubbles or fads.

We discussed the role of expectations informally at various point in the core. It is now time to do it more formally. As you will see, not only are these asset prices affected by current and expected future activity, but they in turn affect decisions that influence current economic activity. Understanding their determination is thus central to understanding fluctuations.

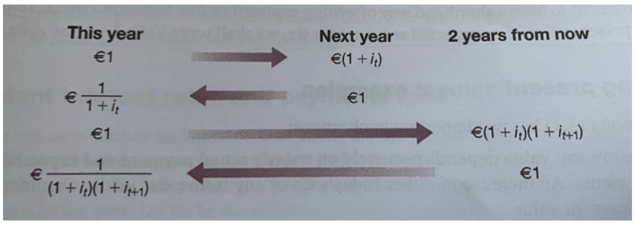
14.1 Expected Present Discounted Values

To understand why present discounted values are important, consider the problem facing a manager who is deciding whether or not to buy a new machine. On the one hand, buying and installing the machine involves a cost today. On the other, the machine allows for higher production, higher sales and higher profits in the future. The question facing the manager is whether the value of these expected profits is higher than the cost of buying and installing the machine. This is where the concept of expected present discounted value comes in handy. The **expected present discounted value** of a sequence of future payment is the value today of this expected sequence of payments. Once the manager has computed the expected present discounted value of the sequence, the problem becomes simpler. The manager compares the two numbers, the expected present discounted value and the initial cost. If the value exceeds the cost, the manager should go ahead and buy the machine. If it does not, the manager should not. The practical problem is that the expected present discounted values are not directly observable. They must be constructed from information on the sequence of expected payments and expected interest rates. Let's look first at the mechanics of construction.

Computing Expected Present Discounted Values

Denote the one-year nominal interest rate by i_t, so lending one euro this year implies getting back $(1+ i_t)$ euros next year. In this sense, one euro this year is worth $(1+ i_t)$ euros next year.
Turn the argument around and ask: How much is one euro next year worth this year? The answer is $1/(1+ i_t)$ euros. Think of it this way: if you lend $1/(1+ i_t)$ euros this year, you will receive $1/(1+ i_t)$ times $(1+ i_t)$ = 1 euros next year. Equivalently, if you borrow $1/(1+ i_t)$ euros this year, you will have to repay exactly one euro next year. So, one euro next year is worth $1/(1+ i_t)$ euros this year.
More formally, we say that $1/(1+ i_t)$ is the **present discounted value** of one euro next year. The word present comes from the fact that we are looking at the value of a payment next year in terms of euro today. The word discounted comes from the fact that the value next year is discounted, with $1/(1+ i_t)$ being the **discount factor** (the rate at which you discount, in this case the nominal interest rate, i_t, is sometimes called the **discount rate**).
The higher the nominal interest rate, the lower the value today of a euro received next year.
$1/(1+ i_t)$: **discount factor**. If the discount rate goes up, the discount factor goes down. Now apply the same logic to the value today of a euro received *two years from now*. For the moment, assume

that current and future one-year nominal interest rates are known with certainty. Let i_t be the nominal interest rate for this year, and i_{t+1} be the one-year nominal interest rate next year.
If today, you lend one euro for two years, you will get $(1+ i_t)(1+ i_{t+1})$ euros two years from now. Put another way, one euro today is worth $(1+ i_t)(1+ i_{t+1})$ euros two years from now.
What is one euro two years from now worth today? By the same logic as before, the answer is $1/(1+ i_t)(1+ i_{t+1})$ euros. If you lend $1/(1+ i_t)(1+ i_{t+1})$ euros this year, you will get exactly one euro in two years. So, the **present discounted value of a euro two years from now** is equal to $1/(1+ i_t)(1+ i_{t+1})$ euros.



A General Formula
Having gone through these steps, it is easy to derive the present discounted value for the case where both payments and interest rates can change over time.
Consider a sequence of payments in euros, starting today and continuing into the future. Assume for the moment that both future payments and future interest rates are known with certainty. Denote today's payment by €z_t, the payment next year by €z_{t+1}, the payment two years from today by €z_{t+2}, and so on.
The present discounted value of this sequence of payments – that is, the value in today's euros of the sequence of payments – which we shall call €V_t is given by:

$$€V_t = €z_t + \frac{1}{(1+ i_t)}€z_{t+1} + \frac{1}{(1+ i_t)(1+i_{t+1})}€z_{t+2} + \cdots$$

Each payment in the future is multiplied by its respective discount factor. The more distant the payment, the smaller the discount factor, and thus the smaller today's value of that distant payment. In other words, future payments are discounted more heavily, so their present discounted value is lower.
We have assumed that future payments and future interest rates were known with certainty. Actual decisions, however, have to be based on expectations of future payments rather than on actual values for these payments. In our previous example, the manager cannot be sure of how much profit the new machine will actually bring, nor does the manager know what interest rates will be in the future. The best thing to do is get the most accurate forecasts available and then compute the expected present discounted value of profits based on these forecasts.
How do we compute the **expected present discounted value** when future payments and interest rates are uncertain? Basically, in the same way as before, but by replacing the **known** future payments and **known** interest rates with **expected** future payments and **expected** interest rates.

$$€V_t = €z_t + \frac{1}{(1+ i_t)}€z^e_{t+1} + \frac{1}{(1+ i_t)(1+i_{t+1})}€z^e_{t+2} + \cdots$$

Expected present discounted value is a heavy expression to carry; instead, for short, we will often just use **present discounted value**, or even just **present value**. To denote the present value of an expected sequence for €z_t, we shall $V(€z_t)$, or just $V(€z\ )$.

Using Present Values: Examples

The present value depends positively on today's actual payment and expected future payments. An increase in either today's €z or any future €z^e leads to an increase in the present value.
The present value depends negatively on current and expected future interest rates. An increase in either current i or in any future i^e leads to a decrease in the present value.

14.2 Bond Prices and Bond Yields

Bonds differ in two basic dimensions:

- **Maturity**, it is the length of time over which the bond promises to make payments to the holder of the bond.
- **Risk**, it may be default risk, the risk that the issuer of the bond (it could be a government or a company) will not pay back the full amount promised by the bond. Or it may be price risk, the uncertainty about the price you can sell the bond for if you want to sell it in the future before maturity.

Both risk and maturity matter in the determination of interest rates. As we want to focus here on the role of maturity and, by implication, the role of expectations, we shall ignore risk, to start with and reintroduce it later.
Bonds of different maturities each have a price and an associated interest rate called **yield to maturity**, or simply the **yield**. Yields on bonds with a short maturity, typically a year or less, are called **short-term interest rates**. Yields on bonds with a longer maturity are called **long-term interest rates**. On any given day, we observe the yields on bonds of different maturities, so we can trace graphically how the yield depends on the maturity of a bond. This relation between maturity and yield is called the **yield curve**, or the **term structure of interest rates**.
To think about the determination of the yield curve and the relation between short-term interest rates and long-term interest rates, we proceed in two steps:

- We derive **bond prices** for bonds of different maturities.
- We go from bond prices to **bond yields** and examine the determinants of the yield curve and the relation between short- and long-term interest rates.

Bond Prices as Present Values

In much of this section, we shall look at just two types of bonds:

- A bond that promises one payment of €100 in one year – a one-year bond.
- A bond that promises one payment of €100 in two years – a two-year bond.

Let's start by deriving the prices of the two bonds:

- Given that the one-year bond promises to pay €100 next year, it follows, from the previous section, that its price, call it €P_{1t}, must be equal to the present value of a payment of €100 next year. Let the current one-year nominal interest rate be i_{1t}. Note that we now denote the one-year interest rate in year t by i_{1t} rather than simply i_t, as we did in previous chapters. This is to make it easier for you to remember that it is the one-year interest rate. So:
-

$$€P_{1t} = \frac{100}{1 + i_{1t}}$$

The price of the one-year bond varies inversely with the current one-year nominal interest rate.

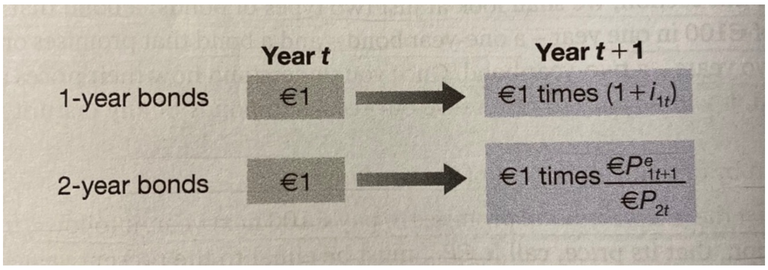
- Given that the two-year bond promises to pay €100 in two years, its price, call it €P_{2t}, must be equal to the present value of €100 two years from now:

$$€P_{2t} = \frac{100}{(1 + i_{1t})(1 + i^{e}_{1t+1})}$$

Where i_{1t} denotes the one-year interest rate this year and i^{e}_{1t+1} denotes the one-year expected by financial markets for next year. The price of the two-year bond depends inversely on both the current one-year rate and the one-year rate expected for next year.

Arbitrage and Bond Prices

Before further exploring the implications of these equations, let us look at an alternative derivation of $€P_{2t}$ equation. This alternative derivation will introduce you to the important concept of **arbitrage**. Suppose you have the choice between holding one-year bonds or two-year bonds and what you care about is how much you will have one year from today. Which bonds should you hold? Suppose you, and other financial investors, care only about the expected return and do not care about risk. This assumption is known as the **expectation hypotheses**. It is a simplification. You and other investors are likely to care not only about the expected return, but also about the risk associated with holding each bond. If you hold a one-year bond, you know with certainty what you will get next year. If you hold a two-year bond, the price at which you will sell it next year is uncertain; holding the two-year bond for one year is risky. As we indicated previously, we disregard this for now but we shall come back to it later.



Under the assumption that you and other financial investors care only about expected return, it follows that the two bonds must offer the same expected one-year return. Suppose this condition was not satisfied. Suppose that, for example, the one-year return on one-year bonds was lower than the expected one-year return on two-year bonds. In this case, no one would want to hold the existing supply of one-year bonds, and the market for one-year bonds could not be in equilibrium. Only if the expected one-year return is the same on both bonds will you and other financial investors be willing to hold both one-year bonds and two-year bonds.
If the two bonds offer the same expected one-year return, it follows from the previous figure that:

$$1 + i_{1t} = €P^{e}_{1t+1} / €P_{2t}$$

The left side of the equation gives the return per euro from holding a one-year bond for one year; the right side gives the expected return per euro from holding a two-year bond for one year. From which:
1

$$€P_{2t} = \frac{€P^{e}_{1t+1}}{1 + i_{1t}}$$

Arbitrage implies that the price of a two-year bond today is the present value of the expected price of the bond next year. This naturally raises the next question: What does the expected price of one-year bonds next year ($€P^{e}_{1t+1}$) depends on?
The answer is straightforward. Just as the price of a one-year bond this year depends on this year's one-year interest rate, the price of a one-year bond next year will depend on the one-year interest rate

next year. Writing $€P^e_{1t+1}$ equation for next year (year t + 1) and denoting expectations in the usual way, we get:

$$€P^e_{1t+1} = \frac{100}{(1+\ i^e_{1t+1})}$$

The price of the bond next year is expected to equal the final payment, €100, discounted by the one-year interest rate expected for next year.
Substituting $€P^e_{1t+1}$ in $€P_{2t}$ we get:

$$€P_{2t} = \frac{100}{(1+i_{1t})(1+\ i^e_{1t+1})}$$

This expression is the same as the equation we derived earlier. What we have shown is that **arbitrage** between one- and two-year bonds implies that the price of two-year bonds is the **present value** of the payment in two years, namely €100, discounted using the current and next year's expected one-year interest rates.

From Bond Prices to Bond Yields
Having looked at bond prices, we now go on to bond yields. The basic point is that bond yields contain the same information about future expected interest rates as bond prices. They just do so in a much cleaner way.
To begin, we need a definition of the yield to maturity. The **yield to maturity** on an **n-year** bond or, equivalently, the **n-year interest rate** is defined as that constant annual interest rate that makes the bond price today equal to the present value of future payments on the bond. This definition is simpler than it sounds. Take, for example, the two-year bond we introduced previously. Denote its yield by i_{2t}, where the subscript 2 is there to remind us that this is the **yield to maturity** on a two-year bond or, equivalently, the two-year interest rate. Following the definition of the yield to maturity, this yield is the constant annual interest rate that would make the present value of €100 in two years equal to the price of the bond today. So, it satisfies the following relation:

$$€P_{2t} = \frac{100}{(1+i_{2t})^{\,2}}$$

What is the relation of the two-year interest rate to the current one-year interest rate and the expected one-year interest rate? To answer this question, look at the previous two equations. Eliminating $€P_{2t}$ between the two gives:

$$\frac{100}{(1+i_{1t})(1+\ i^e_{1t+1})} = \frac{100}{(1+i_{2t})^{\,2}}$$

Rearranging:

$$(1+i_{2t})^{\,2} =\ (1+i_{1t})(1+\ i^e_{1t+1})$$

A useful approximation to this relation is given by:

$$i_{2t} \approx \frac{1}{2}(i_{1t} +\ i^e_{1t+1})$$

This equation simply says that the two-year interest rate is (approximately) the average of the current one-year interest rate and next year's expected one-year interest rate.
For example, the yield on a 10-year bond is (approximately) equal to the average of the current one-year interest rate and the one-year interest rates expected for the next nine years.
The general principle is clear: long-term interest rates reflect current and future expected short-term interest rates.

Reintroducing Risk
We have assumed so far that investors did not care about risk. But they do care. Go back to the choice between holding a one-year bond for one year or holding a two-year bond for one year. The first option is riskless. The second is risky because you do not know the price at which you will sell the bond in a year. You are thus likely to ask for a risk premium to hold the two-year bond, and the arbitrage equation takes the form:

$$1 + i_{1t} + x = \frac{€P^e_{1t+1}}{€P_{2t}}$$

The expected return on the two-year bond (the right-hand side) must exceed the return on the one-year bond by some risk premium x. Rearranging gives:

$$€P_{2t} = \frac{€P^e_{1t+1}}{1 + i_{1t} + x}$$

The price of the two-year bond is the discounted value of the expected price of a one-year bond next year, with the discount rate now reflecting the risk premium. As one-year bonds have a known return and are therefore not risky, the expected price of a one-year bond next year is still given by:

$$€P^e_{1t+1} = \frac{100}{(1 + i^e_{1t+1})}$$

So replacing in the previous equation gives:

$$€P_{2t} = \frac{100}{(1 + i_{1t})(1 + i^e_{1t+1} + x)}$$

Now, to go from prices to yields, let's go through the same steps as before. Using two expressions for the price of the two-year bond gives:

$$\frac{100}{(1 + i_{2t})^2} = \frac{100}{(1 + i_{1t})(1 + i^e_{1t+1} + x)}$$

Manipulating the equation gives:

$$(1 + i_{2t})^2 = (1 + i_{1t})(1 + i^e_{1t+1} + x)$$

Finally, using the same approximation as before gives:

$$i_{2t} \approx \frac{1}{2}(i_{1t} + i^e_{1t+1} + x)$$

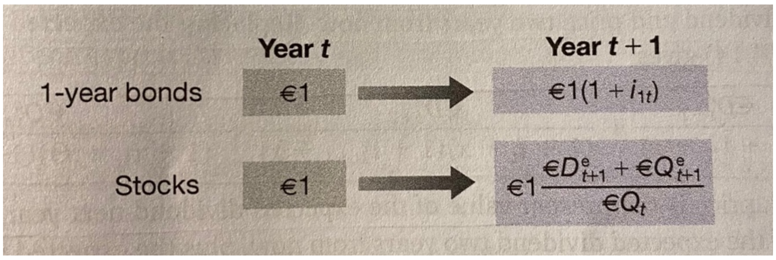
The two-year rate is the average of the current and expected one-year rate plus a risk premium. As the price risk increases with the maturity of the bonds, the risk premium typically increases with maturity, typically reaching 1 to 2% for long-term bonds. This implies that, on average, the yield curve is slightly upward sloping, reflecting the higher risk involved in holding longer maturity bonds.

Interpreting the Yield Curve
When investors expect interest rates to be constant over time, the yield curve should be slightly upward sloping, reflecting the fact that the risk premium increases with maturity. Thus, the fact that the yield curve was downward sloping, something relatively rare, tells us that investors expected interest rates to go down slightly over time, with the expected decrease in rates more than compensating for a rising term premium.

Let's summarize what we have learned in this section. We have seen how arbitrage determines the price of bonds. We have seen how bond prices and bond yields depend on current and future expected interest rates and risk premiums and what can be learned by looking at the yield curve.

14.3 The Stock Market and Movements in Stock Prices
So far, we have focused on bonds. But although governments finance themselves by issuing bonds, the same is not true of firms. Firms finance themselves in four ways. First, they rely on **internal finance**, that is they use some of the earnings; second, and this is the main channel of **external finance** for small firms, through bank loans. Third, through **debt finance** – bonds and loans; and fourth, through **equity finance**, issuing **stocks** – or **shares**, as stocks are also called. Instead of paying predetermined amounts as bonds do, stocks pay **dividends** in an amount decided by the firm. Dividend are paid from the firm's profits. Typically, dividends are less than profits because firms retain some of their profits to finance their investment. But dividends move with profits. When profit increase, so do dividends.

Stock Prices as Present Values
What determines the price of a stock that promises a sequence of dividends in the future? The stock price must be equal to the present value of future expected dividends.



Returns from holding one-year bonds or stocks for one year
Just as we did for bonds, let's derive the result from looking at the implications of arbitrage between one-year bonds and stocks. Suppose you face the choice of investing either in one-year bonds or in stocks for a year. What should you choose?

- Suppose you decide to hold one-year bonds. Then for every euro you put in one-year bonds, you will get $(1 + i_{1t})$ euros next year.
- Suppose you decide instead to hold stocks for a year. Let €Q_t be the price of the stock. Let €D_t denote the dividend this year, €D^e_{t+1} the expected dividend next year. Suppose we look at the price of the stock after the dividend has been paid this year; this price is known as the **ex-dividend price**, so that the first dividend to be paid after the purchase of the stock is next year's dividend.

Holding the stock for a year implies buying a stock today, receiving a dividend next year and then selling the stock. As the price of a stock is €Q_t, every euro you put in stocks buys you €$1/$€Q_t stocks.

And for each stock you buy, you expected to receive, $€D^e_{t+1} + €Q^e_{t+1}$, the sum of the expected dividend and the stock price next year. Therefore, for every euro you put in stocks, you expect to receive $(€D^e_{t+1} + €Q^e_{t+1} + 1)/€Q_t$.

Let's use the same arbitrage argument we used for bonds. It is clear that holding a stock for one year is risky, much riskier than holding a one-year bond for a year (which is riskless). Rather than proceeding in two steps as we did for bonds, let's take risk into account from the start and assume that financial investors require a risk premium to hold stocks.
In the case of stocks, the risk premium is called the **equity premium**. Equilibrium then requires that the expected rate of return from holding stocks for one year be the same as the rate of return on one-year bonds plus the equity premium:

$$\frac{€D^e_{t+1} + €Q^e_{t+1}}{€Q_t} = 1 + i_{1t} + x$$

Where x denotes the equity premium. Rewrite this equation as:

$$€Q_t = \frac{€D^e_{t+1} + €Q^e_{t+1}}{1 + i_{1t} + x}$$

Arbitrage implies that the price of the stock today must be equal to the present value of the expected dividend plus the present value of the expected stock price next year.
The next step is to think about what determines $€Q^e_{t+1}$, the expected stock price next year.
Next year, financial investors will again face the choice between stocks and one-year bonds. Thus, the same arbitrage relation will hold. Writing the previous equation, but now for time t + 1 and taking expectations into account gives:

$$€Q^e_{t+1} = \frac{€D^e_{t+2} + €Q^e_{t+2}}{1 + i^e_{1t+1} + x}$$

The expected price next year is simply the present value next year of the sum of the expected dividend and price two years from now. Replacing the expected price $€Q^e_{t+1}$ in the previous equation gives:

$$€Q_t = \frac{€D^e_{t+1}}{1 + i_{1t} + x} + \frac{€D^e_{t+2}}{(1 + i_t + x)(1 + i^e_{1t+1} + x)} + \frac{€Q^e_{t+2}}{(1 + i_{1t} + x)(1 + i^e_{1t+1} + x)}$$

The stock price is the present value of the expected dividend next year, plus the present value of the expected dividend two years from now, plus the expected price two years from now.
If we replace the expected price in two years as the present value of the expected price and dividends in three years, and so on for n years, we get:

$$€Q_t = \frac{€D^e_{t+1}}{1 + i_{1t} + x} + \frac{€D^e_{t+2}}{(1 + i_t + x)(1 + i^e_{1t+1} + x)} + \dots + \frac{€D^e_{t+n}}{(1 + i_t + x)(1 + i^e_{1t+n-1} + x)}$$

$$+ \frac{€Q^e_{t+n}}{(1 + i_{1t} + x)(1 + i^e_{1t+n-1} + x)}$$

Look at the last term in the last equation, the present value of the expected price is n years. As long as people do not expect the stock price to explode in the future, then, as we keep replacing $€Q^e_{t+n}$ and n increases, this term will go to zero. To see why, suppose the interest rate is constant and equal to i. The last term becomes:

$$\frac{€Q_{t+n}^{e}}{(1+\ i_{1t}+x)(1+i_{1t\,+\,n\,-\,1}^{e}+x)}=\frac{€Q_{t+n}^{e}}{(1+\ i\ +x)^{\ n}}$$

Suppose further that people expect the price of the stock to converge to some value, call it $€Q$, in the far future. Then, the last term becomes:

$$\frac{€Q_{t+n}^{e}}{(1+\ i\ +x)^{\ n}}=\frac{€\overline{Q}}{(1+\ i\ +x)^{\ n}}$$

If the interest rate is positive, this expression goes to zero as n becomes large. Then the $€Q_t$ equation reduces to:

$$€Q_t=\frac{€D_{t+1}^{e}}{(1+i_{1t}+x)}+\frac{€D_{t+2}^{e}}{(1+i_{1t}+x)(1+i_{1t+1}^{e}+x)}+\cdots+\frac{€D_{t+n}^{e}}{(1+i_{1t}+x)(1+i_{1t+1}^{e}+x)}$$

The price of the stock is equal to the present value of the dividend next year, discounted using the current one-year interest rate plus the equity premium, plus the present value of the dividend two years from now, discounted using both this year's one-year interest rate and next year's expected one-year interest rate, plus the equity premium, and so on.
The previous equation gives the stock price as the present value of nominal dividends, discounted by nominal interest rates. Finally, we can rewrite the real stock price as:

$$€Q_t=\frac{€D_{t+1}^{e}}{(1+r_{1t}+x)}+\frac{€D_{t+2}^{e}}{(1+r_{1t}+x)(1+r_{1t+1}^{e}+x)}+\cdots$$

Q_t and D_t, without a euro sign, denote the real price and real dividends at time t. The real stock price is the present value of future real dividends, discounted by the sequence of one-year real interest rates plus the equity premium.
This relation has three important implications:

- Higher expected future real dividends lead to a higher real stock price.
- Higher current and expected future one-year real interest rates lead to a lower real stock price.
- A higher equity premium leads to a lower stock price.

The Stock Market and Economic Activity
There is a saying in economics that it is a sign of a well-functioning stock market that movements in stock prices are unpredictable. The saying is too strong. At any moment, a few financial investors will have better information or simply better at reading the future. If they are only a few, they may not buy enough of the stock to bid its price all the way up today. Thus, they may get large expected returns. But the basic idea is nevertheless correct. The financial market gurus who regularly predict large imminent movements in the stock market are quacks. Major movements in stock prices cannot be predicted. If movements in the stock market cannot be predicted, if they are the result of news, where does this leave us? We can still do two things:

- We can criticize the actions or decisions or others after the fact, looking back and identify the news to which the market reacted.
- We can ask "what if" questions. For example, what would happen to the stock market if the FED were to embark on a more expansionary policy, or if consumers were to become more optimistic and increase spending?

14.4 Risk, Bubbles, Fads and Asset Prices

Do all movements in stock and other asset prices come from news about future dividends or interest rates? The answer is no, for two reasons. The first is that there is variation overtime in perceptions of risk. The second is deviations of prices from their fundamental value, namely bubbles or fads. Let's look at each one in turn.

Stock Prices and Risk

In the previous section, we assumed that the equity premium x was constant. It is not. A lot of the movement in stock prices comes not just from expectations of future dividends and interest rates, but also from shifts in the equity premium.

Asset Prices, Fundamentals and Bubbles

In the previous section, we assumed that stock prices were always equal to their fundamental value, defined as the present value of expected dividends. Do stock prices always correspond to their fundamental value? Most economists doubt it.

Instead, they argue that stock prices are not always equal to their **fundamental value**, defined as the present value of expected dividends and that stocks are sometimes under-priced or over-priced. Overpricing eventually comes to an end, sometimes with a crash, as in October 1929, or with a long slide, as in the case of the Nikkei Index.

Under what conditions can such mispricing occur? The surprising answer is that it can occur even when investors are rational and when arbitrage holds. To see why, consider the case of a truly worthless stock. Putting $€D_{t+1}^e, €D_{t+2}^e, \ldots$ equal to zero, yields a simple and unsurprising answer: the fundamental value of such a stock is equal to zero.

Might you nevertheless be willing to pay a positive price for this stock? Maybe. You might if you expect the price at which you can sell the stock next year to be higher than this year's price. And the same applies to a buyer next year. The buyer may well be willing to buy at a high price if he or she expects to sell at an even higher price in the following year. This process suggests that the stock prices may increase just because investors expect them to. Such movements in stock prices are called **rational speculative bubbles**. Financial investors might well be behaving rationally as the bubble inflates. Even those investors who hold the stock at the time of the crash, and therefore sustain a large loss, may have been rational. They may have realized there was a chance of a crash but also a chance that the bubble would continue, and they could sell at an even higher price.

To make things simple, our example assumed the stock to be fundamentally worthless. But the argument is general and applies to stocks with a positive fundamental value as well. People might be willing to pay more than the fundamental value of a stock if they expect its price to increase further in the future. And the same argument applies to other assets, such as housing, gold and paintings.

Are all deviations from fundamental values in financial markets rational bubbles? Probably not. The fact is that many investors are not rational. An increase in stock prices in the past, say due to a succession of good news, often creates excessive optimism. If investors simply extrapolate from past returns to predict future returns, a stock may become "hot" (high priced) for no reason other than its price has increased in the past. This is true not only of stocks, but also of houses. Such deviations of stock prices from other fundamental value are sometimes called **fads**. We are all aware of fads outside of the stock market; there are good reasons to believe they exist in the stock market as well.

We have focused in this chapter on the determination of asset prices. The reason why this belongs to a macroeconomic text is that asset prices are more than just a sideshow. They affect economic activity, by influencing consumption and investment spending.

Lecture Notes on IS-LM Model with Expectations

A two-period model (Consumption)

The representative agent in the economy lives for two periods. She receives net income Y – T and Y' – T' in period 1 and period 2, respectively, and chooses consumption C and C' in period 1 and period 2, respectively, so as to maximize:

$$U = \max_{C,C'} \; log\, C + \; \beta \; logC'$$

subject to the intertemporal budget constraint (BC):

$$pC + \frac{p'C'}{1+i} = p(Y-T) + \frac{p'(Y'-T')}{1+i}$$

Here p and p' denote, respectively, the price of the period-1 good and the price of the period-2 good and i is the nominal interest rate.
From the definition of inflation rate:

$$\pi = \frac{p'}{p} - 1$$

Thus, we can rewrite the BC as follows:

$$C + \frac{C'}{\frac{1+i}{1+\pi}} = Y - T + \frac{Y'-T'}{\frac{1+i}{1+\pi}}$$

Note that the ratio $\frac{1+i}{1+\pi}$ is actually the real interest rate. We can therefore define:

$$1 + r = \frac{1+i}{1+\pi}$$

To solve the agent's maximization problem, we solve for C and then replace it in the utility function. The problem becomes:

$$U = \max_{C'} \; log\,(Y - T + \frac{Y'-T'}{1+r} - \frac{C'}{1+r}) + \; \beta \; logC'$$

This is a concave problem in one variable. Therefore, the following first-order condition (FOC) is necessary and sufficient for the solution:

$$\frac{1}{Y - T + \frac{Y'-T'}{1+r} - \frac{C'}{1+r}}\left(-\frac{1}{1+r}\right) + \beta\frac{1}{C'} = 0$$

Which can be rewritten as:

$$\frac{1}{C} = (1+r) + \beta\frac{1}{C'}$$

This equation is known as the **Euler equation**. The interpretation is simple: If the agent decides to give up one unit of current consumption and save it, she will lose $\frac{1}{C}$ in terms of utility. By saving this unit, however, the agent obtains $(1+r)$ units of consumption in the future, which results in a gain of $\beta\frac{1}{C\prime}$ in terms of future utility. The Euler equation says that, at the optimum, the agent must be indifferent between consuming the unit in the present or saving it for the future. If this was not the case, in fact, the agent could increase her utility by either consuming more in the present or by consuming more in the future, which would contradict optimality.

Finally, we can combine the Euler equation and the BC to get a nice expression for current consumption C:

$$C = \frac{1}{1+\beta}(Y-T) + \frac{1}{1+\beta}\frac{1}{1+r}(Y'-T')$$

This expression deserves a few comments. First, if we collect $\frac{1}{1+\beta}$, we can rewrite C as:

$$C = \frac{1}{1+\beta}[(Y-T) + \frac{1}{1+r}(Y'-T')]$$

The term in square brackets is the present discounted value (PDV) of future (net) income. This is known as **human capital**. Therefore, the agent consumes a constant fraction of her human capital. Second, note that if we define $c_1 = \frac{1}{1+\beta}$ and $c_0 = \frac{1}{1+\beta}\frac{1}{1+r}(Y'-T')$, then we have the standard consumption function assumed in the IS – LM model:

$$C = c_0 + c_1(Y-T).$$

Thus, utility maximization can be used to justify one of the main assumptions of the IS – LM model. What is more, by deriving C from first principles (utility maximization), we have learnt that the constant c_0 depends on future variables and on the interest rate. Therefore, we can augment the standard IS – LM model by incorporating expectations of future variables. In particular, we can write:

$$C = (Y-T, Y'-T', r)$$

Where Y – T and Y' – T' have positive sign, while r has negative.

A two-period model (Investment)

We now consider the problem of the typical firm in the economy. As for the consumer, we assume that also the firm lives for two periods. The firm maximizes the PDV of profits:

$$\prod + \frac{\prod'}{1+r}$$

Where $\prod$ and denote $\prod'$, respectively, current and future profits. In turn, profits equal revenues minus costs.

To produce output Y a firm must employ labour and capital. Formally, $Y = AK^{\alpha}N^{1-\alpha}$ and, in the future, $Y' = A'(K')^{\alpha}(N')^{1-\alpha}$. Here, A and A' capture the productivity of the firm. Finally, we set the price of the good produced by the firm equal to 1.

We assume that the firm begins its life with some capital K, which it cannot change. However, the firm can change **future** capital K' by investing an amount I. Current profits are given by:

$$\prod = Y - (\omega N + I) = AK^{\alpha}N^{1-\alpha} - (\omega N + I)$$

Furthermore, by investing I today, the firm can create future capital K' according to the following equation:

$$K' = (1 - \delta)K + I$$

Here, δ is the depreciation rate: After using capital K, a fraction δ of it is lost. Thus, only the remaining fraction $1 - \delta$ can be used for production in the next period.

Remember that the firm lasts only for two periods. Thus, after producing in the last period, it will sell all the remaining capital and shut down. Formally, future profits are given by:

$$\prod' = Y' + (1 - \delta)K' - \omega' N' = A'(K')^{\alpha}(N')^{1-\alpha} + (1 - \delta)K' - \omega' N'$$

Putting things together, the firm chooses current labour, future labour, and investment as follows:

$$\max_{N,N',I} \; [AK^{\alpha}N^{1-\alpha} - (\omega N + I)] + \frac{A'(K')^{\alpha}(N')^{1-\alpha} + (1 - \delta)K' - \omega' N'}{1 + r}$$

It turns out that we don't need to solve this problem fully. Instead, it is enough only to solve for the optimal choice of I. Formally, taking the derivative of $\prod$ with respect to I gives the FOC:

$$-1 + \frac{1}{1 + r} [\alpha A'(K')^{\alpha-1}(N')^{1-\alpha} + (1 - \delta)] \frac{\partial K'}{\partial I} = 0$$

Where $\frac{\partial K'}{\partial I}$ is the (partial) derivative of K' with respect to I. And we have that $\frac{\partial K'}{\partial I} = 1$.
Therefore, the FOC for the firm is:

$$-1 + \frac{1}{1 + r} [\alpha A' \frac{(N')^{1-\alpha}}{(K')^{1-\alpha}} + (1 - \delta)] = 0$$

If we multiply and divide the first term in the square bracket by $(K')^{\alpha}$, we have:

$$-1 + \frac{1}{1 + r} [\alpha A' \frac{(N')^{1-\alpha}(K')^{\alpha}}{(K')^{1-\alpha}(K')^{\alpha}} + (1 - \delta)] = 0$$

or, using the fact that $Y' = A'(K')^{\alpha}(N')^{1-\alpha}$,

$$-1 + \frac{1}{1 + r} [\alpha A' \frac{Y'}{K'} + (1 - \delta)] = 0$$

Finally, we can rewrite the latter as:

$$\alpha \frac{Y'}{K'} = r + \delta$$

This condition is easy to interpret. It says that, at the optimum, the firm will invest up to the point where the marginal product of one extra unit of capital ($\alpha \frac{Y'}{K'}$) is equal to the marginal cost of financing the capital plus the loss due to depreciation ($r + \delta$).

Using $K' = (1-\delta)K + I$ we have:

$$\alpha \frac{Y'}{K(1-\delta)K + I} = r + \delta$$

or

$$I = \frac{\alpha Y'}{r+\delta} - K(1-\delta)$$

Note that investment increases with future output Y' and decreases with the interest rate (r). It also decreases with current capital K (if the firm already has capital, it needs to invest less for the future). We can thus write:

$$I(Y', r)$$

This is almost identical to the investment function in the standard IS – LM model. However, investment now depends on future output, not the current one. This is very intuitive: the firm invests to produce more in the future, thus, it cares about future production/income, not current one.

IS – LM with Expectations

We can now use the new consumption function $C = (Y - T, Y' - T', r)$ and the new investment function $I(Y', r)$ to incorporate expectations into our old IS – LM model. Before doing that, we can also add the risk-premium into our formulas. The quickest and easiest way to do this is to assume that, since the environment is risky and investors are risk-averse, they require an extra compensation x to invest in risky activities. Basically, this means that we simply need to add x to the interest rate r in the previous equations.

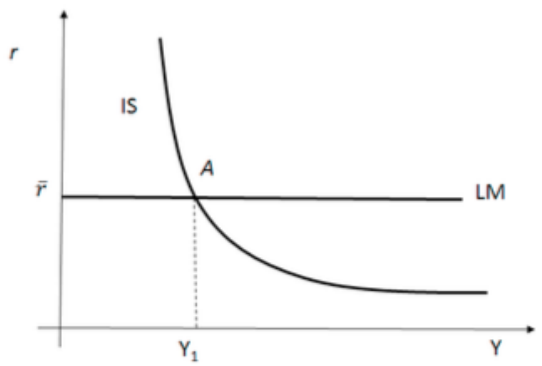
We are now ready for the new IS – LM model. We assume that the central bank has still control over the real interest rate, hence, the LM curve is still:

$$r = \bar{r}$$

Instead, the IS curve is now:

$$Y = C(Y - T, Y' - T', r + x) + I(Y', r + x) + G$$

Since, Y and r **have similar effects on consumption and investment as in the old IS – LM model**, the graphical representation of the model is basically the same.



Chapter 17 – Openness in Goods and Financial Markets

The chapter discusses the implications of openness in goods markets and financial markets. Openness in goods markets allows people to choose between domestic goods and foreign goods. An important determinant of their decisions is the real exchange rate – the relative price of domestic goods in terms of foreign goods. Openness in financial markets allows people to choose between domestic assets and foreign assets. This imposes a tight relation between the exchange rate, both current and expected, and domestic and foreign interest rates – a relation known as the **interest parity condition**.

We have assumed until now that the economy we looked at was closed, that is it did not interact with the rest of the world. We had to start this way to keep things simple and to build up intuition for the basic macroeconomic mechanisms.
Openness has three distinct dimensions:

- **Openness in Goods Markets**, which is the ability of consumers and firms to choose between domestic goods and foreign goods.
- **Openness in Financial Markets** is the ability of financial investors to choose between domestic assets and foreign assets.
- **Openness in Factor Markets** is the ability of firms to choose where to locate production, and of workers to choose where to work.

In the short run and in the medium run, openness in factor markets plays much less of a role than openness in either goods markets or financial markets. Thus, we shall ignore openness in factor markets and focus on the implications of the first two dimensions of openness here.

17.1 Openness in Goods Markets

Let's start by looking at how much Europe sells to and buys from the rest of the world. Then we shall be better able to think about the choice between domestic goods and foreign goods and the role of the relative price of domestic goods in terms of foreign goods – the real exchange rate.

Export and Imports
The EU economy is becoming more and more open over time. Exports and Imports, which were equal to 19% of GDP in the early 1960s, are now equal to about 40% of GDP. As such, it is much more open than the United States – the largest world economy – which exports around 13.5% of GDP (compared with only 5% in the early 1960s), and as such the United States has one of the smallest ratios of exports to GDP among the rich countries of the world.
Imports and Exports have followed the same upward trend, and only since 2009 have exports started to exceed imports by more than 1% (and by 3% in 2014, which means the EU runs a trade surplus).

This strongly differs from what happened in the United States, where since the early 1980s imports have consistently exceeded exports. Put another way, for the last 30 years, the United States has consistently run a trade deficit.

The Choice Between Domestic Goods and Foreign Goods

How does openness in goods markets force us to rethink the way we look at equilibrium in the goods market? Until now, when we were thinking about consumers' decisions in the goods market, we focused on their decision to save or to consume. When goods markets are open, domestic consumers face a second decision: whether to buy domestic or foreign goods.

In an open economy, they face two spending decisions: save or buy; and buy domestic or buy foreign. Central to this second decision is the price of domestic goods relative to foreign goods. We call this relative price the **real exchange rate**. The real exchange rate is not directly observable and you will not find it in the newspapers. What you will find in newspapers are **nominal exchange rates**, the relative prices of currencies. So, we start by looking at nominal exchange rates and then see how we can use them to construct real exchange rates.

Nominal Exchange Rates

Nominal exchange rates between two currencies can be quoted in one of two ways:

- As the price of the domestic currency in terms of the foreign currency.
- As the price of the foreign currency in terms of the domestic currency.

From now on, we shall adopt the first definition. We shall define the **nominal exchange rate** as the price of the domestic currency in terms of foreign currency, and denote it by **E**. When looking, for example, at the exchange rate between the euro area and the United Kingdom (from the viewpoint of the euro area, so the euro is the domestic currency), E will denote the price of a euro in terms of pounds. Exchange rates between the euro and most foreign currencies are determined in foreign exchange markets and change every day – indeed every minute of the day. These changes are called **nominal appreciations** or **nominal depreciations**.

- An **appreciation** of the domestic currency is an increase in the price of the domestic currency in terms of a foreign currency. Given our definition of the exchange rate, an appreciation corresponds to an **increase** in the exchange rate.
- A **depreciation** of the domestic currency is a decrease in the price of the domestic currency in terms of a foreign currency. So given our definition of the exchange rate, a depreciation of the domestic currency corresponds to a **decrease** in the exchange rate, E.

You may have encountered two other words to denote movements in exchange rates: "revaluations" and "devaluations". These two terms are used when countries operate under **fixed exchange rates** – a system in which two or more countries maintain a constant exchange rate between their currencies. If we are interested, however, in the choice between domestic goods and foreign goods, the nominal exchange rate gives us only part of the information we need. To tourists from the euro area thinking of visiting the United Kingdom, the question is not only how many pounds they will get in exchange for their euros but how much goods will cost in the United Kingdom relative to how much they cost in the euro area. This takes us to our next step – the construction of real exchange rates.

From Nominal to Real Exchange Rates

We want to construct a real exchange rate that reflects the relative price of all the goods produced in the euro area in terms of all the goods produced in the United Kingdom. This is exactly what the GDP deflators we introduced earlier do. They are, by definition, price indexes for the set of final goods and services produced in the economy. Let P be the GDP deflator for the euro area, **P*** be the GDP deflator for the United Kingdom and E be the euro-pound nominal exchange rate.

- The price of euro area goods in euros is P. Multiplying it by the exchange rate, **E**, the price of euros in terms of pounds, gives us the price of euro area goods in pounds, **EP**.
- The price of British goods in pounds is **P***. The real exchange rate, the price of euro area goods in terms of British goods, which we shall call ε, is thus given by:

$$\varepsilon = \frac{EP}{P^*}$$

The real exchange rate is constructed by multiplying the domestic price level by the nominal exchange rate and then dividing by the foreign price level. Note, however that the real exchange rate is an index number, that is, its level is arbitrary and therefore uninformative. It is uninformative because the GDP deflators used to construct the real exchange rate are themselves index numbers.
Although the level of the real exchange rate is uninformative, the rate of change of the real exchange rate is informative. If, for example, the real exchange rate between the euro area and the United Kingdom increases by 10%, this tells us that euro area goods are now 10% more expensive relative to British goods than they were before.
Like nominal exchange rates, real exchange rates move overtime. These changes are called real appreciations or real depreciations:
- An increase in the real exchange rate – that is, an increase in the relative price of domestic goods in terms of foreign goods – is called a **real appreciation**.
- A decrease in the real exchange rate – that is, a decrease in the relative price of domestic goods in terms of foreign goods – is called a **real depreciation**.

Large fluctuations in the nominal exchange rate also show up in the real exchange rate. This is not surprising. Price levels move slowly. So year-to-year movements in the price ratio P/P* are typically small compared with the often sharp movements in the nominal exchange rate E. Thus, from year to year, or even over the course of a few years, movements in the real exchange rate ε tend to be driven largely by movements in the nominal exchange rate E.

From Bilateral to Multilateral Exchange Rates
The euro area does most of its trade with three sets of countries:
- Its neighbours within the EU but outside the euro area.
- Asia
- United States

How do we go from **bilateral exchange rates**, like the real exchange rate between the euro area and the United Kingdom we focused on previously, to **multilateral exchange rates** that reflect this composition of trade? We want the weight of a given country to incorporate not only how much the country trades with the euro area, but also how much it competes with the euro area in other countries. The variable constructed in this way is called the **real effective exchange rate (REER)**, or the euro real exchange rate for short.

17.2 Openness in Financial Markets
Openness in financial markets allows financial investors to hold both domestic assets and foreign assets, to diversify their portfolios, to speculate on movements in foreign interest rates versus domestic interest rates, on movements in exchange rates, and so on.
Diversify and speculate they do. Given that buying or selling foreign assets implies buying or selling foreign currency – sometimes called **foreign exchange** – the volume of transactions in foreign exchange markets gives us a sense of the importance of international financial transactions.

The volume of transactions in foreign exchange markets is not only high, but also rapidly increasing. The volume of foreign exchange transactions has more than quintupled since 2001.
For a country, openness in financial markets has an important implication. It allows the country to run trade surpluses and trade deficits. Recall that a country running a trade deficit is buying more from the rest of the world than it is selling to the rest of the world. To pay for the difference between what it buys and what it sells, the country must borrow from the rest of the world. It borrows by making it attractive for foreign financial investors to increase their holdings of domestic assets – in effect, to lend to the country.

The Balance of Payments
A country's transactions with the rest of the world, including both trade flows and financial flows, are summarized by a set of accounts called the **Balance of Payments**.
The balance of payments is composed of three parts:
- **Current Account**, records the value of exports and imports of both goods and services.
- **Capital Account**, records the net flow of investment transaction into an economy.
- **Financial Account**

The current account and the capital account are mirror images of each other. Leaving aside **statistical discrepacy** the current account plus the capital account must sum to zero. A current account deficit is financed by net capital flows from the rest of the world, thus by a capital account surplus. Similarly, a current account surplus corresponds to a capital account deficit.

Now that we have looked at the current account, we can return to an issue we touched on earlier: the difference between GDP, the measure of output we have used so far, and gross national product (GNP), another measure of aggregate output.
GDP measures **value added domestically**. GNP measures the **value added by domestic factors of production**. When the economy is closed, the two measures are the same. When the economy is open, however, they can differ. Some of the income from domestic production goes to foreigners, and domestic residents receive some foreign income. Thus, to go from GDP to GNP, one must start from GDP, add income received from the rest of the world, and subtract income paid to the rest of the world. Put another way, GNP is equal to GDP plus net payments from the rest of the world. More formally, denoting these net income payments by NI: $GNP = GDP + NI$.
In most countries, the difference between GNP and GDP is small (relative to GDP).

The Choice between Domestic and Foreign Assets
Openness in financial markets implies that people face a new financial decision: whether to hold domestic assets or foreign assets. It would appear that we actually have to think about at least two new decisions: the choice of holding domestic money versus foreign money; and the choice of holding domestic interest-paying assets versus foreign interest-paying assets. But remember why people hold money: to engage in transactions. For someone who lives in the euro area and whose transactions are mostly or fully in euros, there is little point in holding foreign currency. Foreign currency cannot be used for transactions in the euro area, and if the goal is to hold foreign assets, holding foreign currency is clearly less desirable than holding foreign bonds, which pay interest. This leaves us with only one new choice to think about: the choice between domestic interest-paying assets and foreign interest-paying assets.

$$(1 + i_t) = (1 + i_t^*)(\frac{E_t}{E_{t+1}^e})$$

This equation takes the name of **Uncovered Interest Parity (UIP)** relation, or simply the **Interest Parity Condition**.

The assumption that financial investors will hold only the bonds with the highest expected rate of return is obviously too strong, for two reasons:

- It ignores transaction costs.
- It ignores risk.

But as a characterization of capital movements among the major world financial markets (New York, Frankfurt, London, Tokyo), the assumption is not far off. Small changes in interest rates and rumors of impending appreciation or depreciation can lead to movements of billions of dollars within minutes. For the rich countries of the world, the arbitrage assumption in the previous equation is a good approximation of reality.

Interest Rates and Exchange Rates

Let's get better sense of what the interest parity condition implies. First rewrite $\frac{E_t}{E_{t+1}^e}$ as $1/\,[1 + (E_{t+1}^e - E_t)/E_t]$. Replacing in the previous equation gives:

$$(1 + i_t) = \frac{(1 + i_t^*)}{[1 + (E_{t+1}^e - E_t)/E_t]}$$

This gives us a relation between the domestic nominal interest rate, i_t, the foreign nominal interest rate, i_t^*, and the expected rate of appreciation of the domestic currency $(E_{t+1}^e - E_t)/E_t$. As long as interest rates or the expected rate of depreciation are not too large – sat below 20% a year – a good approximation to this equation is given by:

$$i_t \approx i_t^* - \frac{E_{t+1}^e - E_t}{E_t}$$

This is the form of the **UIP** relation you must remember. Arbitrage by investors implies that **the domestic interest rate rule must be equal to the foreign interest rate minus the expected appreciation rate of the domestic currency**. Note that the expected appreciation rate of the domestic currency is also the expected depreciation rate of the foreign currency. So, in the previous equation can be equivalently states as saying that **the domestic interest rate must be equal to the foreign interest rate minus the expected depreciation of the foreign currency**.

Let's apply this equation to German Bonds versus UK Bonds. Suppose the one-year nominal interest rate is 2% in Germany and 5% in the United Kingdom. Should you hold UK bonds or German ones?

- It depends whether you expect the pound to depreciate relative to the euro over the coming year by more or less than the difference between the German interest rate and the UK interest rate, or 3% in this case (5% - 2%).
- If you expect the pound to depreciate by more than 3%, then, despite the fact that the interest rate is higher in the UK than in Germany. Investing in UK bonds is less attractive than investing in German bonds. By holding UK bonds, you will get higher interest payments next year, but the pound will be worth less in terms of euros next year, making investing in UK bonds less attractive than investing in German bonds.
- If you expect the pound to depreciate by less than 3% or even to appreciate, then the reverse holds, and UK bonds are more attractive than German bonds.

Looking at it another way, if the UIP condition holds, and the German one-year interest rate is 3% lower than the UK interest rate, it must be that financial investors are expecting, on average, an appreciation of the euro relative to the pound over the coming year of about 3% and this is why they are willing to hold German bonds despite the lower interest rate.

The arbitrage relation between interest rates and exchange rates suggests that, unless countries are willing to tolerate large movements in their exchange rate, domestic and foreign interest rates are likely to move together. Take an extreme case of two countries that commit to maintaining their bilateral exchange rates at a fixed value. If markets have faith in this commitment, they will expect the exchange rate to remain constant, and the expected depreciation will be equal to zero. In this case, the arbitrage condition implies that interest rates in the two countries will have to move exactly together.

17.3 Conclusions and Look Ahead

We have now set the stage for the study of the open economy.

- Openness in goods markets allows people and firms to choose between domestic goods and foreign goods. This choice depends primarily on the **real exchange rate** – the relative price of domestic goods in terms of foreign goods.
- Openness in financial markets allows investors to choose between domestic assets and foreign assets. This choice depends primarily on their relative rates of return, which depend on domestic interest rates and foreign interest rates, and on the expected rate of appreciation of the domestic currency.

First, we look at the implications of openness in goods markets. We then further explore openness in financial markets and, following that, we discuss the pros and cons of different exchange rate regimes.

Chapter 18 – The Goods Market in an Open Economy

In 2009, countries around the world worried about the risk of a recession in the United States. But their worries were not so much for the United States as they were for themselves. To them, a US recession meant lower export to the United States, a deterioration of their trade position and weaker growth at home. The US recession clearly led to a world recession. To understand what happened, we must expand the treatment of the goods market in the core and account for openness in the analysis of goods markets. This is what we do in this chapter.

18.1 The IS Relation in the Open Economy

When we were assuming the economy was closed to trade, there was no need to distinguish between the **domestic demand for goods** and the **demand for domestic goods**; they were clearly the same thing. Now, we must distinguish between the two. Some domestic demand fails on foreign goods and some of the demand for domestic goods comes from foreigners.

The Demand for Domestic Goods

In an open economy, the **demand for domestic goods**, Z, is given by:

$$Z = C + I + G - \frac{IM}{\varepsilon} + X$$

The first three terms – Consumption, C, Investment, I, and Government Spending, G, constitute the total **domestic demand for goods**, domestic or foreign. We must be careful here. Foreign goods are different from domestic goods, so we cannot just subtract the quantity of imports, IM. We must first express the value of imports in terms of domestic goods. This is what $\frac{IM}{\varepsilon}$ stands for, $\frac{1}{\varepsilon}$ is the price of foreign goods in terms of domestic goods. So $\frac{IM}{\varepsilon}$ is the value of import in terms of domestic goods.

The Determinants of C, I and G

Having listed the five components of demand, the next task is to specify their determinants. How much consumers decide to spend still depends on their income and their wealth. Although the real exchange rate surely affects the composition of consumption spending between domestic goods and foreign goods, there is no obvious reason why it should affect the overall level of consumption. The same is true of investment; the real exchange rate may affect whether firms buy domestic machines or foreign machines, but it should not affect total investment.

$$\text{Domestic Demand: } C + I + G = C(Y - T) + I(Y, r) + G$$

Consumption depends positively on disposable income, Y – T, and investment depends positively on production, Y, and negatively on the real policy rate, r.

The Determinants of Imports

Imports are the part of domestic demand that falls on foreign goods. What do they depend on? They clearly depend on domestic income. Higher domestic income leads to a higher demand for all goods, both domestic and foreign. So a higher domestic income leads to higher imports. They also clearly depend on the real exchange rate – the price of domestic goods in terms of foreign goods. The more expensive domestic goods are relative to foreign goods, the higher is the domestic demand for foreign goods. So a higher real exchange rate leads to higher imports.

$$IM = IM(Y, \varepsilon)$$

An increase in domestic income, Y, leads to an increase in imports.
An increase in the real exchange rate, ε (a real appreciation), leads to an increase in imports, IM.

The Determinants of Exports

Exports are the part of foreign demand that falls on domestic goods. What do they depend on? They depend on foreign income. Higher foreign income means higher foreign demand for all goods, both foreign and domestic. So, higher foreign income leads to higher exports. They depend also on the real exchange rate. The higher the price of domestic goods in terms of foreign goods, the lower the foreign demand for domestic goods. In other words, the higher the real exchange rate, the lower the exports. Let Y* denote foreign income. We therefore write exports as:

$$X = X(Y^*, \varepsilon)$$

An increase in foreign income, Y*, leads to an increase in exports.
An increase in the real exchange rate, ε, leads to a decrease in exports.

Putting the Components Together

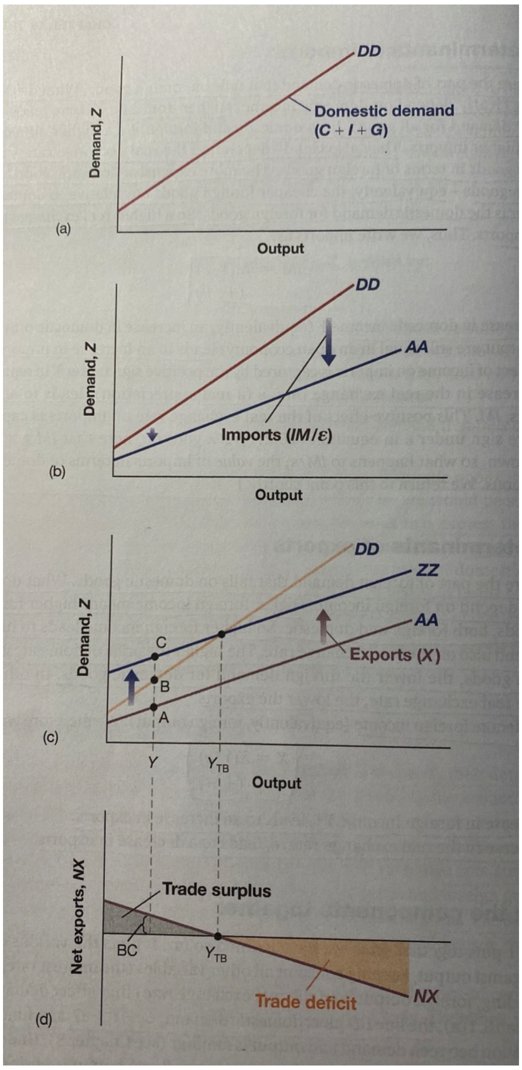
- The line DD plots domestic demand, C + I + G, as a function of output Y.

To arrive at the demand for domestic goods, we must first subtract imports.

- The line AA represents the domestic demand for domestic goods.
- The distance between DD and AA equals the value of imports, $\frac{IM}{\varepsilon}$. Because the quantity of imports increases with income, the distance between the two lines increases with income.
- AA is flatter than DD. As income increases, some of the additional domestic demand falls on foreign goods rather than on domestic goods. In other words, as income increases domestic demand for domestic goods increases less than total domestic demand.
- As long as some of the additional demand falls on domestic goods, AA has a positive slope. An increase in income leads to some increase in the demand for domestic goods.

Finally, we must add exports.

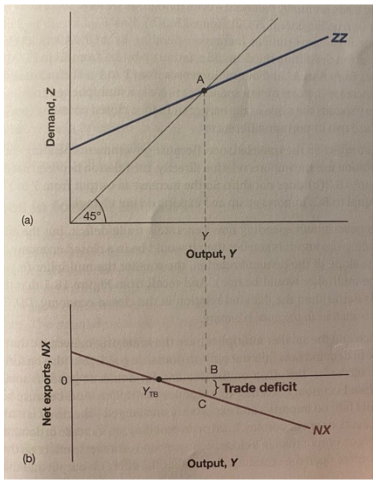
- The line ZZ represents the demand for domestic goods.
- The distance between ZZ and AA equals exports. Because exports do not depend on domestic income, the distance between ZZ and AA is constant.
- Because AA is flatter than DD, ZZ is also flatter than DD.



We are now able to characterize the behaviour of **Net Exports** – the difference between exports and imports – as a function of output. At output level Y, for example, exports are given by the distance AC and imports by the distance AB, so net exports are given by the distance BC.

Net exports are a decreasing function of output. As output increases, imports increase and exports are unaffected, so net export decrease. Call Y_{TB} (TB for Trade Balance) the level of output at which the value of imports equals the value of exports, so that net exports are equal to zero. Levels of output above Y_{TB} lead to higher imports and to a trade deficit. Levels of output below Y_{TB} lead to lower imports and to a trade surplus.

18.2 Equilibrium Output and the Trade Balance



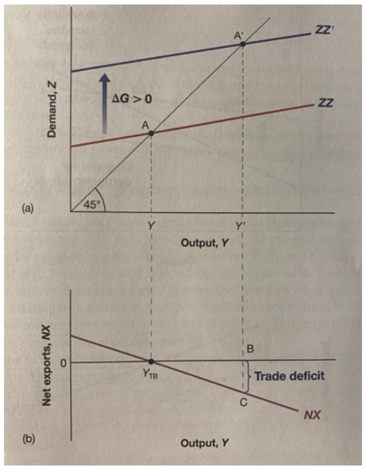
The goods market is in equilibrium when domestic output equals the demand – both domestic and foreign – for domestic goods: Y = Z.
Collecting the relations we derived for the components of the demand for domestic goods, Z, we get:

$$Y = C(Y - T) + I(Y, r) + G - \frac{IM(Y, \varepsilon)}{\varepsilon} + X(Y^*, \varepsilon)$$

This equilibrium condition determines output as a function of all the variables we take as given, from taxes to the real exchange rate to foreign output.
Equilibrium output is at the point where demand equals output, at the intersection of the line ZZ and the 45-degree line. There is in general no reason why the equilibrium level of output, Y, should be the same as the level of output at which trade is balanced, Y_{TB}. As we have drawn the figure, equilibrium output is associated with a trade deficit, equal to the distance BC. Note that we could have drawn it differently, so equilibrium output was associated instead with a trade surplus.

18.3 Increase in Demand – Domestic or Foreign

How do changes in demand affect output in an open economy? Let's start with an old favorite – an increase in government spending – and then turn to a new exercise, the effects of an increase in foreign demand.



Increases in Domestic Demand

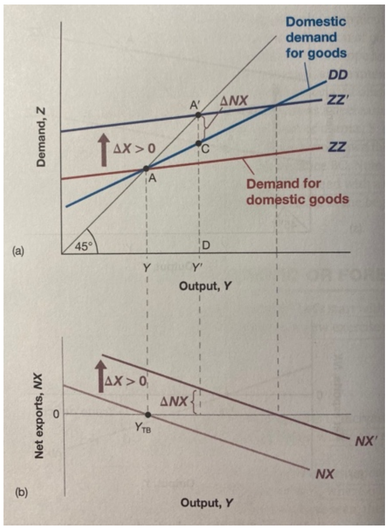
Suppose the economy is in a recession and the government decides to increase government spending to increase domestic demand and, in turn, output. What will be the effects on output and on the trade balance? Before the increase in government spending, demand is given by ZZ, and the equilibrium is at point A, where output equals Y.
What happens if the government increases spending by ΔG? At any level of output, demand is higher by ΔG, shifting the demand relation up by ΔG from ZZ to ZZ'. The equilibrium point moves from A to A' and output increases from Y to Y'. The increase in output is larger than the increase in government spending: there is a multiplier effect. So far, the story sounds the same as the earlier story for closed economy. However, there are two important differences:

- There is now an effect on the trade balance. Because government spending enters neither the exports relation nor the imports relation directly, the relation between net exports and output does not shift. So the increase in output from Y to Y' leads to a **trade deficit** equal to BC: imports go up and exports do not change.
- Not only does government spending now generate a trade deficit, but the effects of government spending on output is smaller than it would be in a closed economy. Recall that the smaller the slope of the demand relation, the smaller the multiplier. And recall that the demand relation, ZZ, is flatter than the demand relation in the closed economy, DD. This means the multiplier is smaller in the open economy.

The trade deficit and the smaller multiplier have the same origin. Because the economy is open, an increase in demand now falls not only on domestic goods, but also on foreign goods. So, when income increases, the effect on the demand for domestic goods is smaller than it would be in a closed economy, leading to a smaller multiplier. And because some of the increase in demand falls on imports – and exports are unchanged – the result is a trade deficit.
These two implications are important. In an open economy, an increase in domestic demand has a smaller effect on output than in a closed economy and an adverse effect on the trade balance. Indeed,

the more open the economy, the smaller the effect on output and the larger the adverse effect on the trade balance.

Increases in Foreign Demand



Consider now an increase in foreign output, which is an increase in Y^*. This could be aresult of an increase in foreign government spending G^* - the policy change we just analysed, but now taking place abroad. The initial demand for domestic goods is given by ZZ. The equilibrium is at point A, with output level Y. It will be useful below to refer to the line that shows the domestic demand for goods C + I + G as a function of income. This line is drawn as DD. Recall that DD is steeper than ZZ. The difference between ZZ and DD equals net exports, so that if trade is balanced at point A, then ZZ and DD intersect at point A.

Now consider the effects of an increase in foreign output, ΔY^*. Higher foreign output means higher foreign demand, including higher foreign demand for European goods. So the direct effect of the increase in foreign output is an increase in European exports by some amount, which we denote by ΔX .

- For a given level of output, this increase in exports leads to an increase in the demand for European goods by ΔX, so the line showing the demand for domestic goods as a function of output shifts up by ΔX, from ZZ to ZZ'.
- For a given level of output, net exports go up by ΔX. So the line showing net exports as a function of output also shifts up by ΔX, from NX to NX'.

The new equilibrium is at point A', with output level Y'. The increase in foreign output leads to an increase in domestic output. The channel is clear. Higher foreign output leads to higher exports of domestic goods, which increases domestic output and the domestic demand for goods through the multiplier.

What happens to the trade balance? We know that exports go up. But could it be that the increase in domestic output leads to such a large increase in imports that the trade balance actually deteriorates? No. The trade balance must improve. To see why, note that, when foreign demand increases, the demand for domestic goods shifts up from ZZ to, but the line DD, which gives the domestic demand for goods as a function of output, does not shift. At the new equilibrium level of output Y, domestic demand is given by the distance DC, and the demand for domestic goods is given by DA'. Net exports are therefore given by the distance CA' – which, because DD is necessarily below ZZ', is necessarily positive. Thus, while imports increase, the increase does not offset the increase in exports, and the trade balance improves.

Fiscal Policy Revisited

We have derived two results so far.

- An increase in domestic demand leads to an increase in domestic output but leads also to a deterioration of the trade balance.
- An increase in foreign demand leads to an increase in domestic output and an improvement in the trade balance.

These results, in turn, have two important implications. Both have been in evidence in the recent crisis.

- First, and most obviously, they imply that shocks to demand in one country affect all other countries. The stronger the trade links between countries, the stronger the interaction, and the more countries will move together.
- Second, these interactions complicate the task of policy makers, especially in the case of fiscal policy.

Start with the following observation. Governments do not like to run trade deficits, and for goods reasons. The main reason is that a country which consistently runs a trade deficit accumulates debt vis-á-vis the rest of the world and therefore has to pay steadily higher interest payments to the rest of the world. Thus, it is no wonder that countries prefer increases in foreign demand (which improve the trade balance) to increases in domestic demand (which worsen the trade balance).
But these preferences can have disastrous implications. Consider a group of countries, all doing a large amount of trade with each other, so that an increase in demand in any one country falls largely on the goods produced in the other countries. Suppose all these countries are in recession and each has roughly balanced trade to start with. In this case, each country might be reluctant to take measures to increase domestic demand. Were it to do so, the result might be a small increase in output but also a large trade deficit. Instead, each country might just wait for the other countries to increase their demand. This way, it gets the best of both worlds: higher output and an improvement in its trade balance. But if all the countries wait, nothing will happen and the recession may last a long time.

Is there a way out? There is – at least in theory. If all countries coordinate their macroeconomic policies so as to increase domestic demand simultaneously, each can increase demand and output without increasing its trade deficit. The reason is clear. The coordinated increase in demand leads to increases both in exports and imports in each country.
In practice, however, **policy coordination** is not easy to achieve. Some countries might have to do more than other and may not want to do so. Suppose that only some countries are in recession. Countries that are not in a recession will be reluctant to increase their own demand, but if they do not, the countries that expand will run a trade deficit vis-á-vis countries that do not. Or suppose some countries are already running a large budget deficit. These countries might not want to cut taxes or further increase spending as this would further increase their deficits. They will ask other countries to take more of the adjustment. Those other countries may be reluctant to do so. Countries also have a strong incentive to promise to coordinate and then not deliver on their promise. Once all countries have agreed, say, to an increase in spending, each country has an incentive not to deliver, so as to benefit from the increase in demand elsewhere and thereby improve its trade position. But if each country cheats, or does not do everything it promised, there will be insufficient demand expansion to get out of the recession. The result is that, despite declarations by governments at international meetings, coordination often fizzles. Only when things are really bad does coordination appear to take hold.

18.4 Depreciation, the Trade Balance and Output

Suppose the UK government takes policy measures that lead to a depreciation of the pound – a decrease in the nominal exchange rate. Recall that the real exchange rate is given by:

$$\varepsilon = \frac{EP}{P^*}$$

The real exchange rate ε (the price of domestic goods in terms of foreign goods) is equal to the nominal exchange rate E (the price of domestic currency in terms of foreign currency) times the domestic price level, P, divided by the foreign price level, P^* as given. This implies that the nominal depreciation is reflected one for one in a real depreciation. More concretely, if the pound depreciates vis-á-vis the yen by 10% (a 10% nominal depreciation), and if the price levels in Japan and the United Kingdom do not change, UK goods will be 10% cheaper compared with Japanese goods (a 10% real depreciation).
Let's now ask how this real depreciation will affect the UK trade balance and output.

Depreciation and the Trade Balance: the Marshall – Lerner Condition
Return to the definition of net exports:

$$NX = X - IM/\varepsilon$$

Replace X and IM by their expressions from their respective equations:

$$NX = X(Y^*, \varepsilon) - IM(Y, \varepsilon)/\varepsilon$$

As the real exchange rate ε enters the right side of the equation in three places, this makes it clear that the real depreciation affects the trade balance through three separate channels.

- **Exports, X, increase**. The real depreciation makes UK goods relatively less expensive abroad. This leads to an increase in foreign demand for UK goods – an increase in UK exports.
- **Imports, IM, decrease**. The real depreciation makes foreign goods relatively more expensive in the United Kingdom. This leads to a shift in domestic demand towards domestic goods and to a decrease in the quantity of imports.
- **The relative price of foreign goods in terms of domestic goods, $\frac{1}{\varepsilon}$, increases**. This increases the import bill, IM/ε. The same quantity of imports now costs more to buy (in terms of domestic goods).

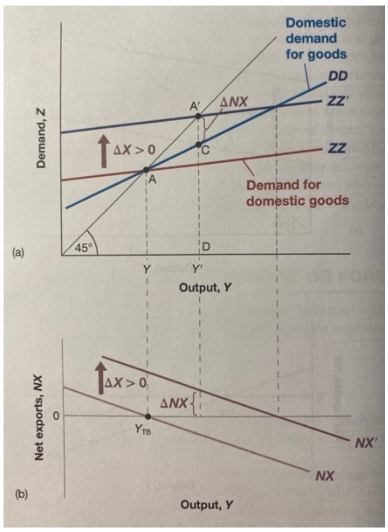
For the trade balance to improve following a depreciation, exports must increase enough and imports must decrease enough to compensate for the increase in the price of imports. The condition under which a real depreciation leads to an increase in net exports is known as the **Marshall-Lerner Condition**. It turns out – with a complication we will state when we introduce dynamics later in this chapter – that this condition is satisfied in reality. So, for the rest of this text, we shall assume that a real depreciation – a decrease in ε – leads to an increase in net exports – an increase in NX.

The Effects of a Real Depreciation

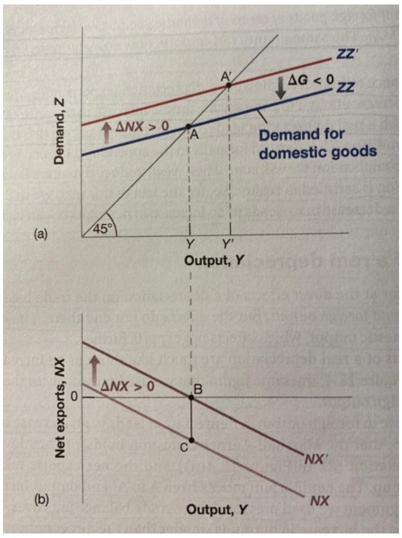
We have looked so far at the direct effects of a depreciation on the trade balance – that is, the effects given UK and foreign output. But the effects do not end there. The change in net exports changes domestic output, which affects next exports further.

Just like an increase in foreign output, a depreciation leads to an increase in net exports (assuming, as we do, that the Marshall-Lerner condition holds), at any level of output. Both the demand relation and the net exports relation shift up. The equilibrium moves from A to A' and output increases from Y to Y'. By the same argument we used previously, the trade balance improves. The increase in imports induced by the increase in output is smaller than the direct improvement in the trade balance induced by the depreciation.

Let's summarize. The depreciation leads to a shift in demand, both foreign and domestic, towards domestic goods. This shift in demand leads, in turn, to both an increase in domestic output and an improvement in the trade balance.



Although a depreciation and an increase in foreign output have the same effects on domestic output and the trade balance, there is a subtle but important difference between the two. A depreciation works by making foreign goods relatively more expensive. But this means that, for a given income, people – who now have to pay more to buy foreign goods because of the depreciation – are worse off. This mechanism is strongly felt in countries that go through a large depreciation. Governments trying to achieve a large depreciation often find themselves with strikes and riots in the streets, as people react to the much higher prices of imported goods.



Combining Exchange Rate and Fiscal Policies

Suppose output is at its natural level, but the economy is running a large trade deficit. The government would like to reduce the trade deficit while leaving output unchanged, so as to avoid overheating. What should it do?

A depreciation alone will not do. It will reduce the trade deficit, but it will also increase output. Nor will a fiscal contraction do. It will reduce the trade deficit, but it will decrease output. What should the government do? The answer: use the right combination of depreciation and fiscal contraction.

Suppose the initial equilibrium is at A, associated with output Y. At this level of output, there is a trade deficit, given by the distance BC. If the government wants to eliminate the trade deficit without changing output, it must do two things.

- It must achieve a depreciation sufficient to eliminate the trade deficit at the initial level of output. So, the depreciation must be such as to shift the net exports relation from NX to NX'. The problem is that this depreciation, and the associated increase in net exports, also shifts the demand relation from ZZ to ZZ'. In the absence of other measures, the equilibrium would move from A to A', and output would increase from Y to Y'.
- To avoid the increase in output, the government must reduce government spending so as to shift ZZ' back to ZZ. This combination of a depreciation and a fiscal contraction leads to the same level of output and an improved trade balance.

To the extent that governments care about both the level of output and the trade balance, they have to use both fiscal policy and exchange rate policies. We just saw one such combination.

Table 18.1 Exchange rate and fiscal policy combinations

Initial conditions	Trade surplus	Trade deficit
Low output	ε ? G ↑	ε ↓ G ?
High output	ε ↑ G ?	ε ? G ↓

Take, for example, the cell in the top right corner of the table: initial output is too low (put another way, unemployment is too high) and the economy has a trade deficit. A depreciation will help on both the trade and the output fronts. It reduces the trade deficit and increases output. But there is no reason for the depreciation to achieve both the correct increase in output and the elimination of the trade deficit. Depending on the initial situation and the relative effects of the depreciation on output and the trade balance, the government may need to complement the depreciation with either an increase or a decrease in government spending. This ambiguity is captured by the question mark in the cell.

18.6 Saving, Investment and the Current Account Balance

You saw earlier how we could rewrite the condition for equilibrium in the goods market as the condition that investment was equal to saving. We can now derive the corresponding condition for the open economy. Start from our equilibrium condition:

$$Y = C + I + G - \frac{IM}{\varepsilon} + X$$

Move consumption, C, from the right side to the left side of the equation, subtract taxes, T, from both sides, denote net exports ($\frac{IM}{\varepsilon} + X$) by NX to get:

$$Y - T - C = I + (G - T) + NX$$

Recall that, in an open economy, the income of domestic residents is equal to output, Y, plus net income from abroad, NI, plus net transfers received. Denote these transfers by NT and add NI and NT to both sides of the equation:

$$(Y + NI + NT - T) - C = I + (G - T) + (NX + NI + NT)$$

Note that the term in brackets on the left is equal to disposable income, so the left side is equal to disposable income minus consumption (i.e. saving, S). Note also that the sum of net exports, net income from abroad and net transfers on the right side is equal to the current account. Denote the current account by CA and rewrite the previous equation as:

$$S = I + (G - T) + CA$$

Rearrange the equation to read:

$$CA = S + (T - G) - I$$

The current account balance is equal to saving – the sum of private saving and public saving – minus investment. A current account surplus implies that the country is saving less than it invests. One way of getting more intuition for this relation is to go back to the discussion of the current account and the capital account. There we saw that a current account surplus implies net lending from the country to the rest of the world. So, consider a country that invests more than it saves, so that S + (T – G) – I is negative. That country must be borrowing the difference from the rest of the world; it must therefore be running a current account deficit. Symmetrically, a country that lends to the rest of the world is a country that saves more than it invests.

- An increase in investment must be reflected either in an increase in private saving or public saving, or in a deterioration of the current account balance – a smaller current account surplus, or a larger current account deficit, depending on whether the current account is initially in surplus or in deficit.
- A deterioration in the government budget balance – either a smaller budget surplus or a larger budget deficit – must be reflected in an increase in private saving, or in a decrease in investment, or else in a deterioration of the current account balance.
- A country with a high saving rate (private plus government) must have either a high investment rate or a large current account surplus.

Note also, however, what the equation does not say. It does not day, for example whether a government budget deficit will lead to a current account deficit, or instead to an increase in private

saving, or to a decrease in investment. To find out what happens in response to a budget deficit, we must explicitly solve for what happens to output and its components using the assumption that we have made about consumption, investment, exports and imports. That is, we need to do the complete analysis laid out in this chapter. Using only the previous equation can, if you are not careful, be very misleading. To see how misleading, consider, for example, the following argument:
"It is clear the United Kingdom cannot reduce its large current account deficit through a depreciation". Why should a depreciation affect either saving or investment? So, how can a depreciation affect the current account deficit? The argument might sound convincing, but we know it is wrong. We showed that a depreciation leads to an improvement in a country's trade position and by implication, given net income and transfers, an improvement in the current account. So, what is wrong with the argument? A depreciation actually does affect saving and investment. It does so by affecting the demand for domestic goods, thereby increasing output. Higher output leads to an increase in saving over investment or, equivalently, to a decrease in the current account deficit.

Chapter 19 – Output, the Interest Rate and the Exchange Rate

Previously, we treated the exchange rate as one of the policy instruments available to the government. But the exchange rate is not a policy instrument. Rather, it is determined in the foreign exchange market – a market where, as we saw, there is an enormous amount of trading. This fact raises two obvious questions: What determines the exchange rate? And how can policy makers affect it?

These questions motivate this chapter. To answer them, we reintroduce financial markets, which we had left aside earlier in the text. We examine the implications of equilibrium in both the goods market and financial markets, including the foreign exchange market. This allows us to characterize the joint movements of output, the interest rate and the exchange rate in an open economy. The model we develop is an extension to the open economy of our IS – LM model and known as the **Mundell – Fleming Model** (the model presented here retains the spirit of the original but differs in its details).

19.1 Equilibrium in the Goods Market

Equilibrium in the goods market was the focus in the previous chapter, where we derived the equilibrium condition equation:

$$Y = C(Y - T) + I(Y, r) + G - \frac{IM(Y, \varepsilon)}{\varepsilon} + X(Y^*, \varepsilon)$$

It will be convenient in what follows to regroup the last two terms under "net exports", defined as exports minus the value of imports:

$$Y = C(Y - T) + I(Y, r) + G - NX(Y, Y^*, \varepsilon)$$

For our purposes, the main implication of this equation is that both the real interest rate and the real exchange rate affect demand, and in turn equilibrium output:

- An increase in the real interest rate leads to a decrease in investment spending and, as a result, to a decrease in the demand for domestic goods. This leads, through the multiplier, to a decrease in output.
- An increase in the real exchange rate leads to a shift in demand towards foreign goods and, as a result, to a decrease in net exports. The decrease in net exports decreases the demand for domestic goods. This leads, through the multiplier, to a decrease in output.

- Given our focus on the short run, we assumed that the domestic price level was given. We shall extend this assumption to the foreign price level, so the real exchange rate $\varepsilon = \frac{EP}{P^*}$ and the nominal exchange rate E move together. A decrease in the nominal exchange rate – a nominal depreciation – leads, one for one, to a decrease in the real exchange rate – a real depreciation. Conversely, an increase in the nominal exchange rate – a nominal appreciation – leads, one for one, to an increase in the real exchange rate – a real appreciation. If, for notational convenience, we choose P and P^* so that $\frac{P}{P^*} = 1$, then $\varepsilon = E$ and we can replace ε by E in the previous equation.
- Because we take the domestic price level as given, there is no inflation, neither actual nor expected. Therefore, the nominal interest rate and the real interest rate are the same, and we can replace the real interest rate, r, in the previous equation by the nominal interest rate, i.

With these two simplifications, the equation becomes:

$$Y = C(Y - T) + I(Y, i) + G - NX(Y, Y^*, E)$$

In words, goods-market equilibrium implies that output depends negatively on both the nominal interest rate and the nominal exchange rate.

19.2 Equilibrium in Financial Markets

As we look at the choice between domestic bonds and foreign bonds, we shall rely on the assumption we introduced earlier: that financial investors, domestic or foreign, go for the highest expected rate of return, ignoring risk. This implies that, in equilibrium, both domestic bonds and foreign bonds must have the same expected rate of return; otherwise, investors would be willing to hold only one or the other, but not both, and this could not be an equilibrium. This assumption implies that the following arbitrage relation – the **interest parity condition** – must hold:

$$(1 + i_t) = (1 + i_t^*)(\frac{E_t}{E_{t+1}^e})$$

The left side of the equation gives the return, in terms of domestic currency, from holding domestic bonds. The right side of the equation gives the expected return, also in terms of domestic currency, from holding foreign bonds. In equilibrium, the two expected returns must be equal.
Multiply both sides by E_{t+1}^e and rearrange to get:

$$E_t = \frac{(1 + i_t)}{(1 + i_t^*)} E_{t+1}^e$$

For now, we shall take the expected future exchange rate as given and denote it as $\bar{E}^e$. Under this assumption, and dropping time indexes, the interest parity condition becomes:

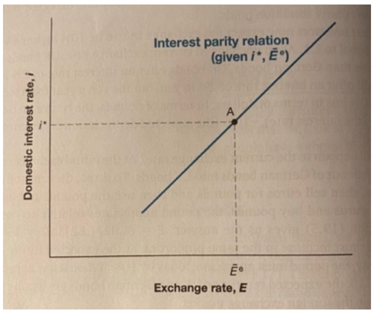
$$E = \frac{(1 + i)}{(1 + i^*)} \bar{E}^e$$

This relation tells us that the current exchange rate depends on the domestic interest rate, on the foreign interest rate and on the expected future exchange rate:

- An increase in the domestic interest rate leads to an increase in the exchange rate.

- An increase in the foreign interest rate leads to a decrease in the exchange rate.
- An increase in the expected future exchange rate leads to an increase in the current exchange rate.

The graph plots the relation between the domestic interest rate and the exchange rate implied by the interest parity relation. The relation is drawn for a given expected future exchange rate, and a given foreign interest rate, and is represented by an upward – sloping line.
The higher the domestic interest rate, the higher the exchange rate. It also implies that when the domestic interest rate is equal to the foreign interest rate, the exchange rate is equal to the expected future exchange rate.



19.3 Equilibrium in Financial Markets

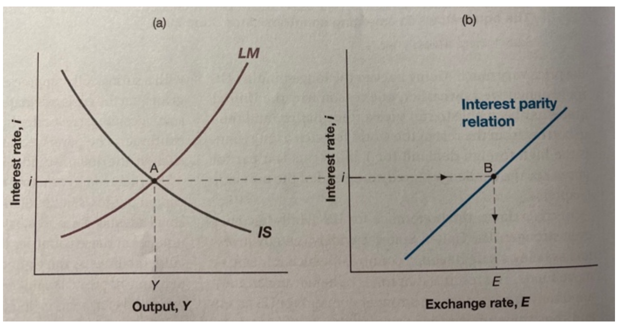
We now have the elements we need to understand the movements of output, the interest rate and the exchange rate. Goods-market equilibrium implies that output depends, among other factors, on the interest rate and exchange rate. Let's think of the interest rate as the policy rate, set by the central bank. The interest parity condition implies a positive relation between the domestic interest rate and the exchange rate. Together, these three relations determine output, the interest rate and the exchange rate. Working with three equations and three variables is not easy. But we can easily reduce them to two by using the interest parity condition to eliminate the exchange rate in the goods-market equilibrium relation. Doing this gives us the following two equations, the open economy versions of our familiar IS and LM relations:

$$IS:\ Y = C(Y-T) + I(Y,i) + G - NX(Y, Y^*, \frac{(1+\ i)}{(1+\ i^*)}\ \bar{E}^e)$$
$$LM: i = \ \bar{\imath}$$

Take the IS relation first and consider the effects of an increase in the interest rate on output. An increase in the interest rate now has two effects:

- The first effect, which was already present in a closed economy, is the direct effect on investment. A higher interest rate leads to a decrease in investment, a decrease in the demand for domestic goods and a decrease in output.
- The second effect, which is only present in the open economy, is the effect through the exchange rate. A higher interest rate leads to an increase in the exchange rate – an appreciation. The appreciation, which makes domestic goods more expensive relative to foreign goods, leads to a decrease in net exports, and therefore to a decrease in the demand for domestic goods and a decrease in output.

Both effect work in the same direction. An increase in the interest rate decreases demand directly and indirectly – through the adverse effect of the appreciation on demand. The IS curve is downward sloping. An increase in the interest rate leads to lower output. The curve looks much the same as in



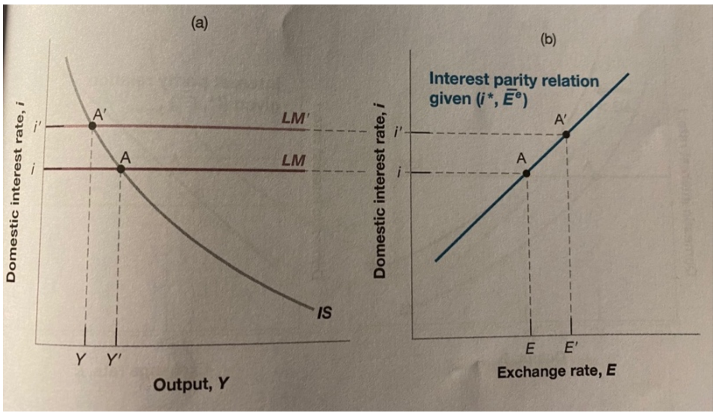
the closed economy, but it hides a more complex relation than before. The interest rate affects output not only directly, but also indirectly through the exchange rate. The LM relation is the same as in the closed economy; it is a horizontal line, at the level of the interest rate i set by the central bank. Equilibrium in the goods and financial markets is attained at point A, with output level Y and interest rate i. The equilibrium value of the exchange rate cannot be read directly from the graph. But it is easily obtained from figure b, which gives the exchange rate associated with a given interest rate found at point B, given also the foreign interest rate and the expected exchange rate. The exchange rate associated with the equilibrium interest rate i is equal to E.

Let's summarize. We have derived the IS and the LM relations for an open economy. The IS curve is downward sloping. An increase in the interest rate leads both directly and indirectly (through the exchange rate) to a decrease in demand and a decrease in output. The LM curve is horizontal at the interest rate set by the central bank. Equilibrium output and the equilibrium interest rate are given by the intersection of the IS and the LM curves. Given the foreign interest rate and the expected future exchange rate, the equilibrium interest rate determines the equilibrium exchange rate.

19.4 Equilibrium in Financial Markets

The Effects of Monetary Policy in an Open Economy

Let's start from the effects of the domestic central bank's decision to increase the domestic interest rate. At a given level of output, with a higher interest rate, the LM curve shifts up, from LM to LM'. The IS curve does not shift. The equilibrium moves from point A to point A'. In figure b, the increase in the interest rate leads to an appreciation. A monetary contraction shifts the LM curve up. It shifts neither the IS curve nor the interest parity curve. So, in the open economy, monetary policy works through two channels: first, as in the closed economy, it works through the effect of the interest rate on spending; second, it works through the effect of the interest rate on the exchange rate and the effect of the exchange rate on exports and imports. Both effects work in the same direction. In the case of monetary contraction, the higher interest rate and the appreciation both decrease demand and output.

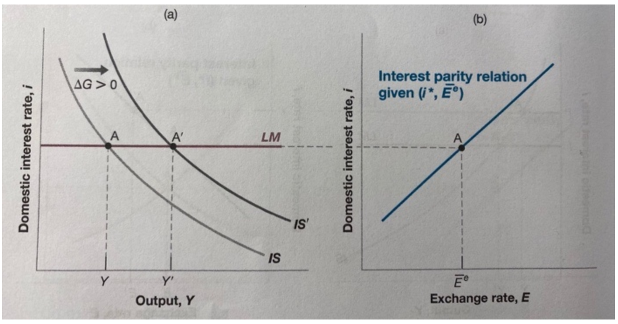


The Effects of Fiscal Policy in an Open Economy

Let's look now at a change in government spending. Suppose that the governments decides to increase defense spending without raising taxes, and so runs a budget deficit.

Let's first assume that, before the increase in government spending, the level of output, Y, was below potential. If the increase in government spending moves output towards potential, but not above potential, the central bank will not be worried that inflation might increase and will keep the interest rate unchanged.

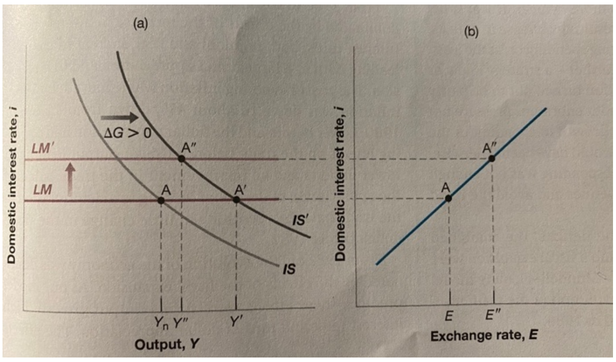
The economy is initially at point A. The increase in government spending increases output at a given interest rate, shifting the IS curve to the right, from IS to IS'. Because the central bank does not change the policy rate, the LM curve does not shift. The new equilibrium is at point A', with a higher level of output, Y'. In panel b, because the interest rate



has not changed, neither has the exchange rate. So, an increase in government spending, when the central bank keeps the interest rate unchanged, leads to an increase in output with no change in the exchange rate. Can we tell what happens to the various components of demand?

- Consumption and government spending both increase.
- Investment rises.
- What about net exports? Foreign output is unchanged, because we are assuming that the rest of the world does not respond to the increase in domestic government spending. The exchange rate is also unchanged, because the interest rate does not change. We are just left with the effect of higher domestic output; as the increase in output increases imports at an unchanged exchange rate, net exports decrease. As a result, the budget deficit leads to a deterioration of the trade balance.

Now, assume instead that the increase in government spending happens in an economy where output is close to potential output, Y_n. In this case, the central bank will worry that the increase in G, by moving the economy above potential output, might push inflation up. It is likely to respond by raising the interest rate. At an unchanged interest rate, output would increase from Y_n to Y' and the exchange rate would not change. But if the central bank accompanies the increase in government spending with an increase in the interest rate, output will increase by less, from Y_n to Y'', and the exchange rate will appreciate, from E to E''.



Again, can we tell what happens to the various components of demand?

- Consumption and government spending both increase.
- What happens to investment is ambiguous. Investment depends on both output and the interest rate. Here output rises but so does the interest rate.
- Net exports decrease, increasing imports and decreasing exports. The budget deficit leads to a trade deficit. (Whether, however, the trade deficit is larger than if the policy rate remained constant is ambiguous).

19.5 Fixed Exchange Rates

We have assumed so far that the central bank chose the interest rate and let the exchange rate adjust freely in whatever manner was implied by equilibrium in the foreign exchange market. In many countries, this assumption does not reflect reality. Central banks act under implicit or explicit exchange rate targets and use monetary policy to achieve those targets. The targets are sometimes implicit, sometimes explicit; they are sometimes specific values, sometimes bands or ranges. These exchange rate arrangements (or **regimes**, as they are called) come under many names.

Pegs, Crawling Pegs, Bands, the EMS and the Euro.

At one end of the spectrum are countries with flexible exchange rates. These countries have no explicit exchange rate targets. Although their central banks do not ignore movements in the exchange rate, they have shown themselves quite willing to let their exchange rates fluctuate considerably.

At the other end are countries that operate under **fixed exchange rates**. Those countries maintain a fixed exchange rate in terms of some foreign currency. Some **peg** their currency to the dollar. Still other countries peg their currency to a basket of foreign currencies, with the weights reflecting the composition of their trade. The label fixed is a bit misleading. It is not the case that the exchange rate in countries with fixed exchange rates never actually changes. But changes are rare. Because of that, economists use specific words to distinguish them from the daily changes that occur under flexible exchange rates. A decrease in the exchange rate under a regime of fixed exchange rates is called a **devaluation** rather than a depreciation, and an increase in the exchange rate under a regime of fixed exchange rates is called a **revaluation** rather than an appreciation.

Between these extremes are countries with various degrees of commitment to an exchange rate target. For example, some countries operate under a **crawling peg**. The name describes it well. These countries typically have inflation rates that exceed the US inflation rate. If they were to peg their nominal exchange rate against the dollar, the more rapid increase in their domestic price level above the US price level would lead to a steady real appreciation and rapidly make their goods uncompetitive. To avoid this effect, these countries choose a predetermined rate of depreciation against the dollar. They choose to "crawl" (move slowly) vis-á-vis the dollar.
If domestic inflation is higher than foreign inflation: P increases faster than P*.
If E is fixed, EP/P* steadily increases.
Equivalently, there is a steady real appreciation. Domestic goods become steadily more expensive relative to foreign goods.

Yet, another arrangement is for a group of countries to maintain their bilateral exchange rates (the exchange rate between each pair of countries) within some bands. Perhaps the most prominent example was the **European Monetary System (EMS)**, which determined the movements of exchange rates within the EU from 1978 to 1998. After a major crisis in 1992, which led a number of countries to drop out of the EMS altogether, exchange rate adjustment became more and more infrequent, leading a number of countries to move one step further and adopt a common currency, the **Euro**.

Monetary Policy when the Exchange Rate is Fixed
Suppose a country decides to peg its exchange rate at some chosen value, call it $\bar{E}$. How does it actually achieve this? The government cannot just announce the value of the exchange rate and remain idle. Rather, it must take measures so that its chosen exchange rate will prevail in the foreign exchange market. Let's look at the implications and mechanics of pegging.
Pegging or no pegging, the exchange rate and the nominal interest rate must satisfy the interest parity condition:

$$(1+i_t) = (1+i_t^*)\left(\frac{E_t}{E_{t+1}^e}\right)$$

Now, suppose the country pegs the exchange rate at $\bar{E}$, so the current exchange rate is $E_t = \bar{E}$. If financial and foreign exchange markets believe that the exchange rate will remain pegged at this value, then their expectation of the future exchange rate, E_{t+1}^e, is also equal to $\bar{E}$, and the interest parity relation becomes:

$$(1+i_t) = (1+i_t^*) \Rightarrow i_t = i_t^*$$

In words, if financial investors expect the exchange rate to remain unchanged, they will require the same nominal interest rate in both countries. Under a fixed exchange rate and perfect capital mobility, the domestic interest rate must be equal to the foreign interest rate.
Let's summarize. Under fixed exchange rates, the central bank gives up monetary policy as a policy instrument. With a fixed exchange rate, the domestic interest rate must be equal to the foreign interest rate.

Fiscal Policy when the Exchange Rate is Fixed

If monetary policy can no longer be used under fixed exchange rates, what about fiscal policy? The effects of an increase in government spending when the central bank pegs the exchange rate are identical to those we saw for the case of flexible exchange rates. This is because, if the increase in spending is not accompanied by a change in the interest rates, the exchange rate does not move. The difference between fixed and flexible exchange rates is the ability of the central bank to respond. If the increase in government spending pushed the economy above potential output, thus raising the possibility that inflation might increase, the central bank could respond by raising the interest rate. This option is no longer available under fixed exchange rates because the interest rate must be equal to the foreign rate.

As this chapter comes to an end, a question should have start to form in your mind. Why would a country choose to fix its exchange rate? You have seen a number of reasons why this appears to be a bad idea:

- By fixing the exchange rate, a country gives up a powerful tool for correcting trade imbalances or changing the level of economic activity.
- By committing to a particular exchange rate, a country also gives up control of its policy rate.
- Although the country retains control of fiscal policy, one policy instrument may not be enough.

Chapter 22 – Fiscal Policy: A Summing Up

In most advanced economies, the crisis has led to large budget deficits and a large increase in debt-to-GDP ratios. In a number of countries, investors are worried about whether debt can indeed be repaid and are asking for higher interest rates to compensate for the risk of default. This calls for governments to reduce deficits, stabilize the debt and reassure investors. At the same time, however, the recovery is weak and a fiscal contraction is likely to slow it down further, at least in the short run. Thus, governments face a difficult choice. Reduce deficits rapidly and reassure markets that they will pay their debt at the risk of lower growth or even recession, or reduce deficits more slowly to avoid further slowing the recovery at the risk of not convincing investors that debt will be stabilized.

22.2 The Government Budget Constraint: Deficits, Debt, Spending and Taxes

Suppose that, starting from a balanced budget, the government decreases taxes, creating a budget deficit. What will happen to the debt over time? Will the government need to increase taxes later? If so, by how much?

The Arithmetic of Deficits and Debt

To answer these questions, we must begin with a definition of the budget deficit. We can write the budget deficit in year t as:

$$deficit_t = rB_{t-1} + G_t - T_t$$

All variables in real terms:

- B_{t-1} is government debt at the end of year t – 1, or, equivalently, at the beginning of year t; r is the real interest rate, which we shall assume to be constant here. Thus, rB_{t-1} equals the real interest payments on the government debt in year t.
- G_t is government spending on goods and services during year t.
- T_t is taxes minus transfers during year t.

In words, the budget deficit equals spending, including interest payments on the debt, minus taxes net of transfers.
Note two characteristics of the previous equation:

- We measure interest payments as real interest payments – that is, the product of the real interest rate and existing debt – rather than as actual interest payments – that is, the product of the nominal interest rate and the existing debt. The correct measure of the deficit is sometimes called the **inflation-adjusted deficit**.
- For consistency with our definition of G as spending on goods and services, G does not include transfer payments. Transfers are instead subtracted from T, so that T stands for taxes minus transfers.

These are only accounting conventions. Whether transfers are added to spending or subtracted from taxes makes a difference to the measurement of G and Tr, but clearly does not affect G – Tr and therefore does not affect the measure of the deficit.
The **government budget constraint** then simply states that the change in government debt during year t is equal to the deficit during year t:

$$B_t - B_{t-1} = Deficit_t$$

If the government runs a deficit, government debt increases as the government borrows to fund the part of spending in excess of revenues. If the government runs a surplus, instead, government debt decreases as the government uses the budget surplus to repay part of its outstanding debt.
Using the definition of the deficit, we can rewrite the government budget constraint as:

$$B_t - B_{t-1} = rB_{t-1} + G_t - T_t$$

The government budget constraint links the change in government debt to the initial level of debt and to current government spending and taxes. It is often convenient to decompose the deficit into the sum of two terms:
- Interest payments on the debt, rB_{t-1}.
- The difference between spending and taxes, $G_t - T_t$. This term is called the **primary deficit** (equivalently, $T_t - G_t$ is called the **primary surplus**).

Using this decomposition, we can rewrite the previous equation as:

$$B_t = (1 + r)B_{t-1} + (G_t - T_t)$$

This relation states that the debt at the end of year t equals (1 + r) times the debt at the end of year t – 1 plus the primary deficit during year t, $(G_t - T_t)$.

Current versus Future Taxes
Consider first a one-year decrease in taxes for the path of debt and future taxes. Start from a situation where, until year 1, the government has balanced its budget, so that initial debt is equal to zero. During year 1, the government decreases taxes by one (think 1 billion euros, for example) for one year. Thus, debt at the end of the year 1, B_1, is equal to one. We take up the question: What happens thereafter?

- **Full Repayment in Year 2**
  Suppose the government decides to repay the debt fully during year 2. From the last equation, the budget constraint for year 2 is given by:

$$B_2 = (1+r)B_1 + (G_2 - T_2)$$

If the debt is fully repaid during year 2, then the debt at the end of year 2 is equal to zero, $B_2 = 0$. Replacing B_1 by 1 and B_2 by 0 and transposing terms gives:

$$T_2 - G_2 = (1+r)1 = (1+r)$$

To repay the debt fully during year 2, the government must run a primary surplus equal to (1 + r). it can do so in one of two ways: a decrease in spending or an increase in taxes.

- **Full Repayment in Year t**
  Now, suppose the government decides to wait until year t to repay the debt. From year 2 to year t – 1 the primary deficit is equal to 0; taxes are equal to spending, not including interest payments on the debt. With the primary deficit still equal to zero during year 3, debt at the end of year 3 is:

$$B_3 = (1+r)B_2 + 0 = (1+r)(1+r)1 = (1+r)^2$$

Thus, debt at the end of year t – 1 is given by:

$$B_{t-1} = (1+r)^{t-2}$$

Despite the fact that taxes are cut only in year 1, debt keeps increasing over time, at a rate equal to the interest rate. The reason is simple: although the primary deficit is equal to zero, debt is now positive and so are interest payments on it. In year t, the year in which the government decides to repay the debt, the budget constraint is:

$$B_t = (1+r)B_{t-1} + (G_t - T_t)$$

If debt is fully repaid during year t, then B_t (debt at the end of year t) is zero. Replacing B_t by zero and B_{t-1} by its expression from the previous equation gives:

$$0 = (1+r)(1+r)^{t-2} + (G_t - T_t)$$

Rearranging the terms:

$$T_t - G_t = (1+r)^{t-1}$$

To repay the debt, the government must run a primary surplus equal to $(1+r)^{t-1}$ during year t. If the adjustment is done through taxes, the initial decrease in taxes of one during year 1 leads to an increase in taxes of $(1+r)^{t-1}$ during year t.
This example yields our first set of conclusions:
  - If government spending is unchanged, a decrease in taxes must eventually be offset by an increase in taxes in the future.
  - The longer the government waits to increase taxes, or the higher the real interest rate is, the higher the eventual increase in taxes must be.

- **Debt Stabilization in Year t**
  We have assumed so far that the government fully repays the debt. Let's now look at what happens to taxes if the government only stabilizes the debt. (Stabilizing the debt means changing taxes or spending so that debt remains constant from then on).

Suppose the government decides to stabilize the debt from year 2 on.
From the previous equations, the budget constraint for year 2 is:

$$B_2 = (1 + r)B_1 + (G_2 - T_2)$$

Under our assumption that debt is stabilized in year 2, $B_2 = B_1 = 1$, the preceding equation becomes:

$$1 = (1 + r) + (G_2 - T_2)$$

Rearranging:

$$T_2 - G_2 = (1 + r) - 1 = r$$

To avoid a further increase in debt during year 1, the government must run a primary surplus equal to real interest payments on the existing debt. It must do so in each of the following years as well. Each year, the primary surplus must be sufficient to cover interest payments, leaving the debt level unchanged. Debt remains equal to one from year 1 on. Taxes are permanently higher from year 1 on, by an amount equal to r; equivalently, from year 1 on, the government runs a primary surplus equal to r.
The logic of this argument extends directly to the case where the government waits until year t to stabilize the debt.
This example yields our second set of conclusions:

- The legacy of past deficits is higher government debt today.
- To stabilize the debt, the government must eliminate the deficit.
- To eliminate the deficit, the government must run a primary surplus equal to the interest payments on the existing debt. This requires higher taxes for ever.

The Evolution of the Debt-To-GDP Ratio

We have focused so far on the evolution of the level of debt. But in an economy in which output grows over time, it makes more sense to focus instead on the ratio of debt to output. To see how this change in focus modifies our conclusions, we need to go from the previous equations to an equation that gives the evolution of the **debt-to-GDP ratio** – the **debt ratio**.

First divide both sides of the previous equation by the real output, Y_t, and multiply both numerator and denominator by Y_{t-1}, to get:

$$\frac{B_t}{Y_t} = (1 + r)\left(\frac{Y_{t-1}}{Y_t}\right)\frac{B_{t-1}}{Y_{t-1}} + \frac{G_t - T_t}{Y_t}$$

Note that all the terms in the equation are now in terms of ratios to output, Y. to simplify this equation, assume that output growth is constant and denote the growth rate of output by g, so $\frac{Y_{t-1}}{Y_t}$ can be written as $\frac{1}{(1+g)}$. And use the approximation $\frac{(1+r)}{(1+g)} = 1 + r + g$.
Using these two assumptions, rewrite the preceding equation as:

$$\frac{B_t}{Y_t} = (1 + r - g)\frac{B_{t-1}}{Y_{t-1}} + \frac{G_t - T_t}{Y_t}$$

$$\frac{B_t}{Y_t} - \frac{B_{t-1}}{Y_{t-1}} = (r - g)\frac{B_{t-1}}{Y_{t-1}} + \frac{G_t - T_t}{Y_t}$$

The change in the debt ratio over time (the left side of the equation) is equal to the sum of two terms:

- The first term is the difference between the real interest rate and the growth rate times the initial debt ratio.
- The second term is the ratio of the primary deficit to GDP.

Comparing the last equation with: $B_t - B_{t-1} = rB_{t-1} + G_t - T_t$

The difference is the presence of r – g in the ultimate equation compared with the r in the above equation. The reason for the difference is simple. Suppose the primary deficit is zero. Debt will then increase at a rate equal to the real interest rate, r. But if GDP is growing as well, the ratio of debt-to-GDP will grow more slowly; it will grow at a rate equal to the real interest rate minus the growth of output, r – g. The ultimate equation implies that the increase in the ratio of debt-to-GDP will be larger:

- The higher the real interest rate;
- The lower the growth of output;
- The higher the initial debt ratio;
- The higher the ratio of the primary deficit to GDP.

Building on this relation, we now turn in the next section to describe how governments that inherited high debt ratios at the end of the war steadily decreased them through a combination of low real interest rates, high growth rates and primary surpluses.

History tells us that there are two, not mutually exclusive ways in which a country can reduce its debt ratio. The first is through high primary surpluses. Suppose, for example, that (r – g) was equal to zero. Then the decrease in the debt ratio over some period would just be the sum of the ratios of primary surpluses to GDP over the period. The second is through a low (r – g), so either through low real interest rates or through high growth, or both. This leads to a final question: Why were real interest rates so low? Average inflation was relatively high. This inflation, combined with consistently low nominal interest rates, are what account for the negative real interest rates. Put another way, a large part of the decrease in debt ratios was achieved by paying bond holders a negative real return on their bonds for many years.

So, to summarize, the evolution of the ratio of debt to GDP depends on four factors: the interest rate, the growth rate, the initial debt ratio and the primary surplus.

22.3 Ricardian Equivalence, Cyclical Adjusted Deficits and War Finance

Ricardian Equivalence

How does takin into account the government budget constraint affect the way we should think about the effects of deficits on output? One extreme view is that once the government budget constraint is taken into account, neither deficits nor debt have an effect on economic activity! This argument is known as the **Ricardian Equivalence** proposition. The best way to understand the logic of the proposition is to use the example of tax changes:

- Suppose that the government decreases taxes by one this year. And as it does so, it announces that, to repay the debt, it will increase taxes by (1 + r) next year. What will be the effect of the initial tax cut on consumption?
- One possible answer is: No effect at all. Why? Because consumers realize that the tax cut is not much of a gift. Lower taxes this year are exactly offset, in present value, by higher taxes next year. Put another way, their human wealth – the present value of after-tax labor income – is unaffected. Current taxes go down by one, but the present value of next year's taxes goes up by (1 + r)/(1 + r) = 1, and the net effect of the two changes is exactly equal to zero.
- Another way of coming to the same answer – this time looking at saving rather than consumption – is as follows. To say that consumers do not change their consumption in response to the tax cut is the same as saying that **private saving increases one for one with**

the deficit. So, the Ricardian equivalence proposition says that if a government finances a given path of spending through deficits, private saving will increase one for one with the decrease in public saving, leaving total saving unchanged. The total amount left for investment will not be affected. Over time, the mechanics of the government budget constraint implies that government debt will increase. But this increase will not come at the expense of capital accumulation.

Under the Ricardian equivalence proposition, a long sequence of deficits and the associated increase in government debt are no cause for worry. As the government is dissaving, the argument goes, people are saving more in anticipation of the higher taxes to come. The decrease in public saving is offset by an equal increase in private saving. Total saving is therefore unaffected and so is investment. How seriously should we take the Ricardian Equivalence proposition? Most economists would answer: "Seriously, but surely not seriously enough to think that deficits and debt are irrelevant". So, it is safe to conclude that budget deficits have an important effect on activity, although perhaps a smaller effect than you thought before going through the Ricardian Equivalence argument. In the short run, larger deficits are likely to lead to higher demand and to higher output. In the long run, higher government debt lowers capital accumulation and, as a result, lower output.

Deficits, Output Stabilization and the Cyclically Adjusted Deficit
To stabilize the economy, the government should run deficits during recessions and surpluses during booms. The cyclically adjusted deficit tells us what the deficit would be, under existing tax and spending rules, if output were at its potential level.

Wars and Deficits
Deficits are justified in times of high spending, such as wars. Relative to an increase in taxes, deficits lead to higher consumption and lower investment during wars. They therefore shift some of the burden of war from people living during the war to those living after it. Deficits also help smooth taxes and reduce tax distortions.

22.4 The Dangers of High Debt
We have seen how high debt requires higher taxes in the future. A lesson from history is that high debt can also lead to vicious cycles, making the conduct of fiscal policy extremely difficult.

High Debt, Default risk and Vicious Cycles
High debt ratios increase the risk of vicious cycles. A higher perceived risk of default can lead to a higher interest rate and an increase in debt. The increase in debt can in turn lead to a higher perceived risk of default and a higher interest rate. Together, both can combine to lead to a debt explosion. Governments may have no choice other than to default or to rely on money finance (printing money). Money finance may in turn lead to hyperinflation. In either case, the economic costs are likely to be high.

Chapter 10 – The Facts of Growth
Our perceptions of how the economy is doing are often dominated by year-to-year fluctuations in economic activity. A recession leads to gloom, and an expansion to optimism. But if we step back to get a look at activity over longer periods – say over many decades – the picture changes. Fluctuations fade. **Growth**, which is the steady increase in aggregate output over time, dominates the picture. With this in mind, we now shift our focus from fluctuations to growth. Put another way, we turn from the study of the determination of output in the short and medium run – where fluctuations dominate – to the determination of output in the long run – where growth dominates. Our goal is to understand what

determines growth, why some countries are growing while others are not, and why some countries are rich while many others are still poor.

10.1 Measuring the Standard of Living

The reason we care about growth is that we care about the **standard of living**. Looking across time, we want to know by how much the standard of living has increased. Looking across countries, we want to know how much higher the standard of living is in one country relative to another. Thus, the variable we want to focus on, and compare either over time or across countries, is **output per person**, rather than output itself.

A practical problem then arises: How do we compare output per person across countries? Countries use different currencies; thus output in each country is expressed in terms of its own currency. A natural solution is to use exchange rates. This simple approach will not do, however, for two reasons:

- First, exchange rates can vary a lot.
- The second reason goes beyond fluctuations in exchange rates. In general, the lower a country's output per person, the lower the prices of food and basic services in that country.

So, when we focus on comparing standards of living, we get more meaningful comparisons by correcting for the two effects we just discussed: variations in exchange rates and systematic differences in prices across countries. The numbers for GDP – and hence for GDP per person – are constructed using a common set of prices for all countries. Such adjusted real GDP numbers, which you can think of as measures of **purchasing power** across time or across countries, are called **purchasing power parity (PPP)** numbers.

When comparing rich versus poor countries, the differences between PPP numbers and the numbers based on current exchange rates can be large. Let's end this section with three remarks before we move on and look at growth:

- What matters for people's welfare is their consumption rather than their income. One might therefore want to use **consumption per person** rather than output per person as a measure of the standard of living.
- Thinking about the production side, we may be interested in differences in productivity rather than in differences in the standard of living across countries. In this case, the right measure is **output per worker** – or, even better, **output per hour worked**, if the information about total hours worked is available – rather than output per person.
- The reason we ultimately care about the standard of living is presumably because we care about happiness. We may therefore ask the obvious question: Does a higher standard of living lead to greater happiness? The answer is a qualified yes.

10.2 Growth in Rich Countries since 1950

Let's start by looking, in this section, at growth in rich countries since 1950. The following table shows the evolution of **output per person** for selected European countries including Denmark,

	Annual growth rate output per person (%)	Real output per person (2000 dollars)		
	1950–2004	1950	2004	2004/1950
France	3.3	5920	26 168	4.4
Ireland	3.6	4422	28 956	6.5
Japan	4.6	2187	24 661	11.3
Sweden	2.2	8507	27 072	3.2
UK	2.7	8091	26 762	3.3
USA	2.6	11 233	36 098	3.2
Average	3.2	6727	28 286	4.2

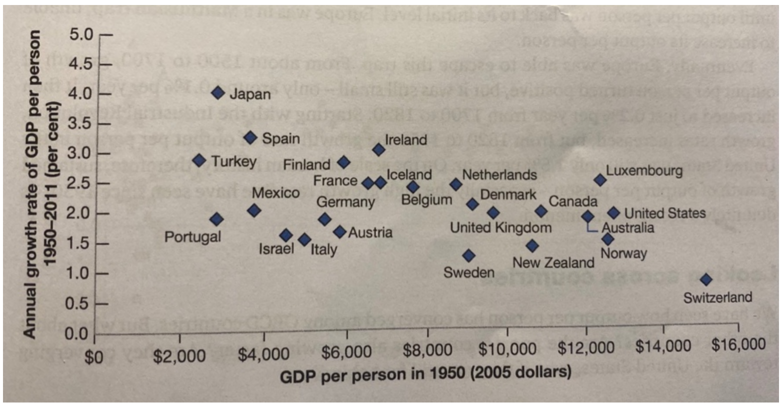
France, Germany, Italy, Spain, Sweden and the United Kingdom, together with Japan and the United States, since 1950. We have chosen these countries because some of them are the world's major economic powers, and also because what has happened to them is broadly representative of what was happened in other advanced countries over the last half-century or so.

The table yields two main conclusion:

- There has been a large increase in output per person.
- There has been a convergence of output per person across countries.

The Large Increase in the Standard of Living since 1950
Output per person has increase by a factor of 3.3 since 1950 in the United States, by a factor of 5.1 in France and by a factor of 11.3 in Japan. These numbers show what is sometimes called the **force of compounding**. Clearly, a better understanding of growth, if it leads to the design of policies that stimulate growth, can have a large effect on the standard of living.

The Convergence of Output per Person.
The second and third column of the table show that the levels of output per person have converged (become closer) over time. The numbers for output per person are much more similar in 2011 than they were in 1950. Put another way, those countries that were behind have grown faster, reducing the gap between them and the United States. In 1950, from the perspective of Europe or Japan, the United States was seen as the land of plenty, where everything was bigger and better. Today these perceptions have faded, and the numbers explain why. Using PPP numbers, US output per person is still the highest, but in 2011 it was only 25% above average output per person in the other countries considered, a much smaller difference than in the 1950, when it was double the average.
This **convergence** of levels of output per person across countries is not specific to the countries we are looking at. It extends to the set of OECD countries. This is shown in the following figure, which plots the average annual growth rate of output per person since 1950 against the initial level of output per person in 1950 for the set of countries that are members of the OECD today. There is a clear negative relation between the initial level of output per person and the growth rate since 1950.



10.3 A Broader Look Across Time and Space

In the previous section, we focused on growth over the last 50 years in rich countries. Let's now put this in context by looking at the evidence both over a much longer time span and a wider set of countries.

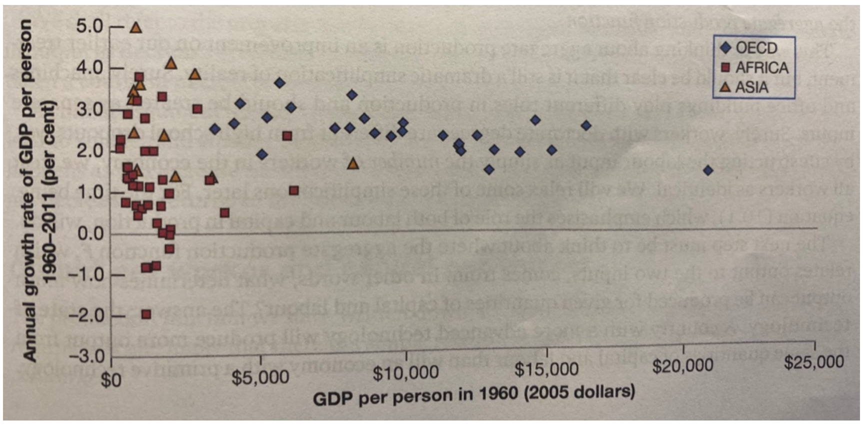
Looking Across Two Millennia
Has output per person in the currently rich economies always grown at rates similar to the growth rates in the previous table? The answer is no. Estimates of growth rate are clearly harder to construct

as we look further back in time. But there is an agreement among economic historians about the main evolutions overs the last 2,000 years.
From the end of the Roman Empire to roughly year 1500, there was essentially no growth of output per person in Europe. This period of stagnation of output per person is often called the **Malthusian era**. Any increase in output would lead to a decrease in mortality, leading to an increase in population until output per person was back to its initial level. Europe was in a **Malthusian trap**, unable to increase its output per person. Eventually, Europe was able to escape this trap. From about 1500 to 1700, growth of output per person turned positive, but it was still small – only around 0.1% per year. It then increase to just 0.2% per year from 1700 to 1820. Starting with the Industrial Revolution, growth rates increased, but from 1820 to 1950 the growth rate of output per person in the United States was still only 1.5% per year.

Looking Across Countries
We have seen how output per person has converged among OECD countries. But what about the other countries? The answer is given in the following figure, which plots the average annual growth rate of output per person since 1960 against output per person for the year 1960, for the 85 countries for which we have the data. The striking feature of this figure is that there is no clear pattern. The cloud of points hides, however, a number of interesting patterns that appear when we put countries into different groups.

- The picture for the OECD countries is much the same as in previous figure, which looked at a slightly longer period of time. Nearly all start at high levels of output per person and there is clear evidence of convergence.
- Convergence is also visible for many Asian countries: most of the countries with high growth rates over the period are in Asia. Japan was the first country to take off. Starting a decade later, in the 1960s, four countries – Singapore, Taiwan, Hong Kong and South Korea, a group of countries sometimes called the **four tigers** – started catching up as well.
- The picture is different, however for the African countries. Most African countries were very poor in 1960, and most have not done well over the period. Many have suffered from either internal or external conflicts. Eight of them have had negative growth of output per person – an absolute decline in their standard of living between 1960 and 2011.

Looking further back in time, the following picture emerges. For much of the first millennium, and until the fifteenth century, China probably had the world's highest level of output per person. For a could of centuries, leadership moved to the cities of northern Italy. But until the nineteenth century, differences across countries were typically much smaller than they are today. Starting in the nineteenth century, a number of countries, first in Western Europe, then in North and South America, started growing faster than others. Since then, a number of countries, most notably in Asia, have started growing fast and are converging. Many others, mainly in Africa, are not.



10.4 Thinking About Growth: A Primer

To think about growth, economists use a framework developed originally by Robert Solow from MIT, in the late 1950s.

The Aggregate Production Function

The starting point for any theory of growth must be an **aggregate production function** which is a specification of the relation between aggregate output and the inputs in production. From now on, we will assume that there are two inputs – capital and labour – and that the relation between aggregate output and the two inputs is given by:

$$Y = F(K, N)$$

As before, Y is aggregate output, K is capital – the sum of all the machines, plant and office buildings in the economy. N is labour – the number of workers in the economy. The function F, which tells us how much output is produced for given quantities of capital and labour, is the **aggregate production function**.

The next step must be to think about where the aggregate production function F, which relates output to the two inputs, comes from. In other words, what determines how much output can be produced for given quantities of capital and labour? The answer: the **state of technology**. A country with a more advanced technology will produce more output from the same quantities of capital and labour than will an economy with primitive technology.

Returns to Scale and Returns to Factors

What restrictions can we reasonably impose on this function? Consider first a thought experiment in which we double both the number of workers and the amount of capital in the economy. This property is called **constant returns on scale**. If the scale of operation is doubled – that is, if the quantities of capital and labour are doubled – then output will also double:

$$2Y = F(2K, 2N)$$

Or, more generally, for any number x:

$$xY = F(xK, xN)$$

We have just looked at what happens to production when both capital and labour are increased. Let's now ask a different question. What should we expect to happen if only one of the two inputs in the economy – say capital – is increased?

Surely output will increase. That part is clear. But it is also reasonable to assume that the same increase in capital will lead to smaller and smaller increases in output as the level of capital increases. In other words, if there is little capital to start with, a little more capital will help a lot. If there is a lot of capital to start with, a little more capital may make little difference.

We shall refer to the property that increases in capital lead to smaller and smaller increases in output as **decreasing returns to capital**. A similar argument applies to the other input, labour. Increases in labour, given capital, lead to smaller and smaller increases in output. There are **decreasing returns to labour** as well.

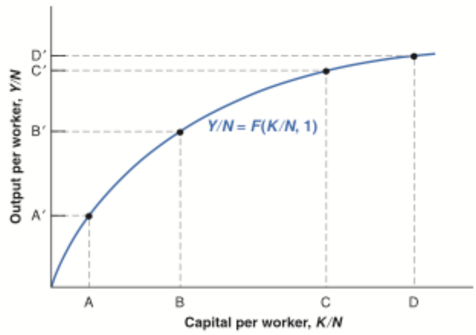
Output per Worker and Capital per Worker

The production function we have written down, together with the assumption of constant returns to scale, imply that there is a simple relation between output per worker and capital per worker.
To see this, set x = 1/N in the previous equation, so that:

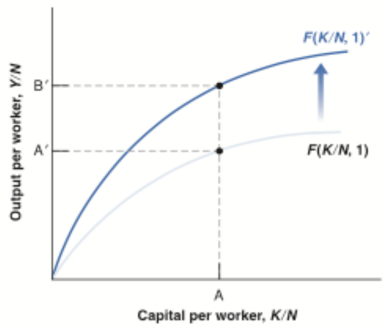
$$\frac{Y}{N} = F(\frac{K}{N}, 1)$$

Note that Y/N is output per worker, K/N is capital per worker. So the previous equation tells us that the amount of output per worker depends on the amount of capital per worker. This relation between output per worker and capital per worker will play a central role in what follows, so let’s look at it more closely.
This relation is drawn in the following figure. Output per worker (Y/N) is measured on the vertical axis and capital per worker (K/N) is measured on the horizontal axis. The relation between the two is given by the upward-sloping curve. As capital per worker increases, so does output per worker. Note that the curve is drawn so that increases in capital lead to smaller and smaller increases in output. This follows from the property that there are **decreasing returns to capital**: at point A, where capital per worker is low, an increase in capital per worker, represent by the horizontal distance AB, leads to an increase in output per worker equal to the vertical distance A’B’. At point C, where capital per worker is larger, the same increase in capital per worker, represented by the horizontal distance CD, leads to a much smaller increase in output per worker, only C’D’.

The Source of Growth
We are now ready to return to our basic question. Where does growth come from? Why does output per worker – or output per person, if we assume the ratio of workers to the population as a whole remains constant over time – go up over time? The previous equation gives a first answer:



- Increases in output per worker (Y/N) can come from increases in capital per worker (K/N). As (K/N) increases – that is, as we move to the right on the horizontal axis – (Y/N) increases.
- Or they can come from improvements in the state of technology that shift the production function, F, and lead to more output per worker given capital per worker. An improvement in the state of technology shifts the production function up, from F(K/N, 1) to F(K/N, 1)’. For a given level of capital per worker, the improvement in technology leads to an increase in output per worker. For example, for the level of capital per worker corresponding to point A, output per worker increases from A’ to B’.



Hence, we can think of growth as coming from **capital accumulation** and from **technological progress** – the improvement in the state of technology. We will see, however, that these two factors play different roles in the growth process.

- Capital accumulation by itself cannot sustain growth.
Because of decreasing returns to capital, sustaining a steady increase in output per worker will require larger and larger increases in the level of capital per worker. At some stage, the economy will be unwilling or unable to save and invest enough to increase capital further. At that stage, output per worker will stop growing. Does this mean that an economy’s **saving rate**, which is the proportion of income that is saved, is irrelevant? No. It is true that a higher

saving rate cannot permanently increase the growth rate of output. But a higher saving rate can sustain a higher level of output. Let's state this in a slightly different way. Take two economies that differ only in their saving rates. The two economies will grow at the same rate, but at any point in time, the economy with the higher saving rate will have a higher level of output per person than the other.
- Sustained growth requires sustained technological progress. Given that the two factors that can lead to an increase in output are capital accumulation and technological progress, if capital accumulation cannot sustain growth forever, then technological progress must be the key to growth. And it is. We will see later that the economy's rate of growth of output per person is eventually determined by its rate of technological progress.

This is important. It means that, in the long run, an economy that sustains a higher rate of technological progress will eventually overtake all other economies. This, of course, raises the next question: What determines the rate of technological progress? Recall the two definitions of the state of technology we discussed previously: a narrow definition, namely the set of blueprints available to the economy; and a broader definition, which captures how the economy is organized, from the nature of institutions to the role of the government. What we know about the determinants of technological progress narrowly defined – the role of fundamental and applied research, the role of patent laws, the role of education and training.

Chapter 11 – Saving, Capital Accumulation and Output

11.1 Interactions Between Output and Capital

At the centre of the determination of output in the long run are two relations between output and capital:
- The amount of capital determines the amount of output being produced.
- The amount of output being produced determines the amount of saving and, in turn, the amount of capital being accumulated over time.

The Effects of Capital on Output

To simplify notations respect to the previous equations, we will rewrite the relation between output and capital per worker as:

$$\frac{Y}{N} = f(\frac{K}{N})$$

Where the function f represents the same relation between output and capital per worker as the function F:

$$f\left(\frac{K}{N}\right) = F(\frac{K}{N}, 1)$$

In this chapter, we shall make two further assumptions:
- The first is that the size of the population, the participation rate and the unemployment are all constant. This implies that employment, N, is also constant.
    - The labour force is equal to population multiplied by the participation rate. So, if population is constant and the participation rate is constant, the labour force is also constant.
    - Employment, in turn, is equal to the labour force multiplied by one minus the unemployment rate.

Under these assumptions, output per worker, output per person and output itself all move proportionately. The reason for assuming that N is constant is to make it easier to focus on how capital accumulation affects growth. If N is constant, the only factor of production that changes over time is capital.

- The second assumption is that there is no technological progress, so the production function f (or, equivalently F) does not change over time.

Again, the reason for making this assumption – which is obviously contrary to reality – is to focus just on the role of capital accumulation. With these two assumptions, our first relation between output and capital per worker, from the production side, can be written as:

$$\frac{Y_t}{N} = f(\frac{K_t}{N})$$

Where we have introduced time indexes for output and capital, but not for labour, N, which we assume to be constant and so does not need the time index.
In words, higher capital per worker leads to higher output per worker.

The Effects of Output on Capital Accumulation
To derive the second relation between output and capital accumulation, we proceed in two steps:

- First, we derive the relation between **output and investment**.
  To derive the relation, we make three assumptions:
  - We continue to assume that the economy is closed, **I = S + (T – G)**.
  - To focus on the behaviour of private saving, we assume that public saving, T – G is equal to zero. **I = S**, that is, Investment is equal to Private Savings.
  - We assume that private saving is proportional to income, so: **S = sY**.
    The parameter, **s**, is the **saving rate**. It has a value between 0 and 1. Combining these two relations and introducing time indexes gives a simple relation between investment and output: $\boldsymbol{I_t = sY_t}$.
    Investment is proportional to output: the higher output is, the higher is saving and so the higher is investment.

- Second, we derive the relation between investment and capital accumulation.
  The second step relates investment, which is a flow, to capital, which is a stock.
  K_t refers to the capital stock at the beginning of year t, K_{t+1} to the capital stock at the beginning of year t + 1, and so on.
  Assume that capital depreciates at rate δ per year. Then: $K_{t+1} = (1 - \delta)K_t + I_t$
  We can now combine the relation between output and investment and the relation between investment and capital accumulation to obtain the second relation we need to think about growth: the relation from output to capital accumulation.
  Replacing investment by its expression from above and dividing both sides by N:

$$\frac{K_{t+1}}{N} = (1 - \delta)\frac{K_t}{N} + s\frac{Y_t}{N}$$

  In words, capital per worker at the beginning of year t + 1 is equal to capital per worker at the beginning of year t, adjusted for depreciation, plus investment per worker during year t, which is equal to the saving rate times output per worker during year t.
  Expanding:

$$\frac{K_{t+1}}{N} - \frac{K_t}{N} = s\frac{Y_t}{N} - \delta\frac{K_t}{N}$$

11.2 The Implications of Alternative Saving Rates

We have derived two relations:

- From the production side, we have seen how capital determines output.
- From the saving side, we have seen how output in turn determines capital accumulation.

We can now put the two relations together and see how they determine the behaviour of output and capital overtime.

Dynamics of Capital and Output

Replacing output per worker ($\frac{Y_t}{N}$) by its expression in terms of capital per worker, gives:

$$\frac{K_{t+1}}{N} - \frac{K_t}{N} = sf(\frac{K_t}{N}) - \delta\frac{K_t}{N}$$

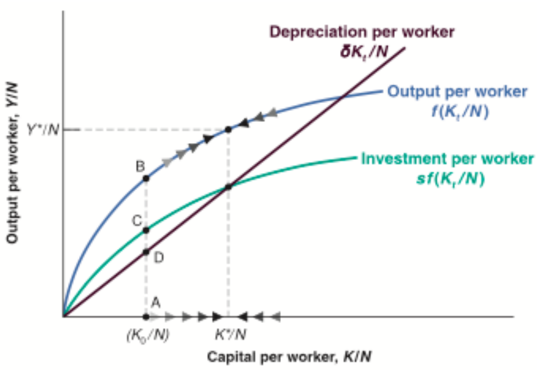
This relation describes what happens to capital per worker. The change in capital per worker from this year to next year depends on the difference between two terms:

- Investment per worker.
- Depreciation per worker.

If investment per worker exceeds depreciation per worker, the change in capital per worker is positive. Capital per worker increases. If investment per worker is less than depreciation per worker, the change in capital per worker is negative. Capital per worker decreases.

The easiest way to interpret them is to use a graph. We do this in the following figure, where output per worker is measured on the vertical axis and capital per worker is measured on the horizontal axis.

- In the figure, look first at the curve representing output per worker, $f(\frac{K_t}{N})$, as a function of capital per worker. When output per worker increases with capital per worker, but, because of decreasing returns of capital, the effect is the smaller, the higher the level of capital per worker.
- The relation representing investment per worker, $sf(\frac{K_t}{N})$, has the same shape as the production function except that it is lower by a factor s (the saving rate).
- The relation representing depreciation per worker, $\delta\frac{K_t}{N}$, is represented by a straight line. Depreciation per worker increases in proportion to capital per worker so the relation is represented by a straight line with slope equal to δ.

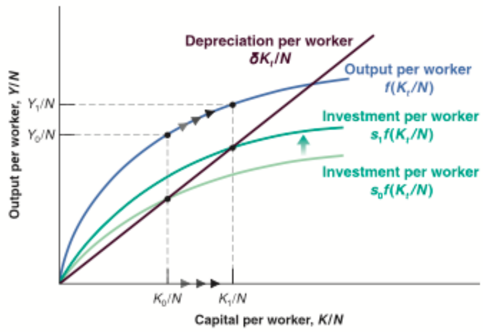


The change in capital per worker is given by the difference between investment per worker and depreciation per worker. At $\frac{K_0}{N}$ the difference is positive; investment per worker exceeds depreciation per worker by an amount represented by the vertical distance CD = AC – AD, so capital per worker increases. As we move to the right along the horizontal axis and look at higher and higher levels of capital per worker, investment increases by less and less, while depreciation keeps increasing in proportion to capital. For some level of capital per worker, $\frac{K^*}{N}$, investment is just enough to cover depreciation, and capital per worker remains constant. To the left of $\frac{K^*}{N}$, investment exceeds depreciation and capital per worker increases. This is indicated by the arrows pointing to the right along the curve representing the production function. To the right of $\frac{K^*}{N}$, depreciation exceeds investment, and capital per worker decreases. This is indicated by the arrows pointing to the left along the curve representing the production function.

Let's look more closely at the levels of output per worker and capital per worker to which the economy converges in the long run. The state in which output per worker and capital per worker are no longer changing is called the **Steady State** of the economy. The steady state value of capital per worker, $\frac{K^*}{N}$, is given by: $sf\left(\frac{K^*}{N}\right) = \delta \frac{K^*}{N}$.

The steady-state value of capital per worker is such that the amount of saving per worker (the left side) is just sufficient to cover depreciation of the capital stock per worker (the right side of the equation). Given steady-state capital per worker ($\frac{K^*}{N}$), the steady state value of output per worker ($\frac{Y^*}{N}$) is given by the production function: $\frac{Y^*}{N} = f\left(\frac{K^*}{N}\right)$.

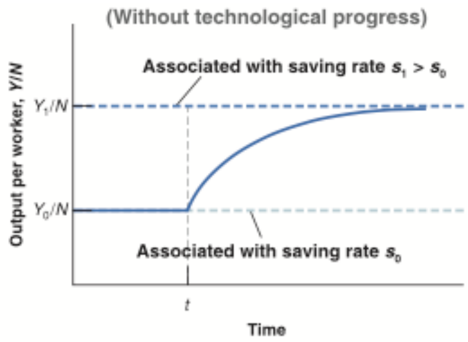
We now have all the elements we need to discuss the effects of the saving rate on output per worker, both over time and in steady state.

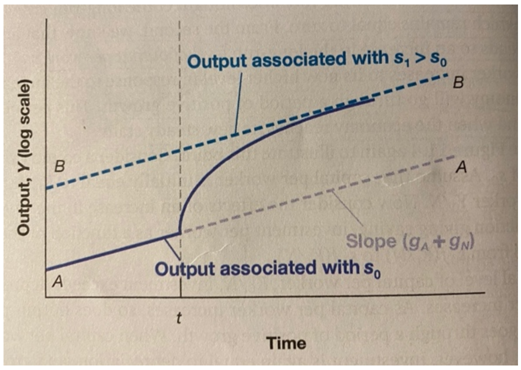


The Saving Rate and Output

How does the saving rate affect the growth rate of output per worker? Our analysis leads to a three part answer.

- The saving rate has no effect on the long-run growth rate of output per worker, which is equal to zero. In other words, in the long run, the growth rate of output is equal to zero, no matter what the saving rate is.
- Nonetheless, the saving rate determines the level of output per worker in the long run. Other things being equal, countries with a higher saving rate will achieve higher output per worker in the long-run.
- An increase in the saving rate will lead to higher growth of output per worker for some time, but not for ever. This conclusion follows from the two proposition we just discussed. From the first, we know that an increase in the saving rate does not affect the long-run growth rate of output per worker, which remains equal to zero. From the second, we know that an increase in the saving rate leads to an increase in the long run level of output per worker. It follows that, as output per worker increases to its new higher level in response to the increase in the saving rate, the economy will go through a period of positive growth. This period of growth will come to an end when the economy reaches its new steady state.





We have derived these three results under the assumption that there was no technological progress, and therefore no growth of output per worker in the long-run. An economy in which there is technological progress has a positive growth rate of output per worker, even in the long-run. This long-run growth rate is independent of the saving rate – the extension of the first result just discussed. The saving rate affects the level of output per worker, however – the extension of the second result. An increase in the saving rate leads to growth greater than steady-state growth rate for some time until the economy reaches its new higher path – the extension of our third result. These three results are illustrated in the figure on the right which extends the figure on the left by plotting the effect than an increase in the saving rate has on an economy with positive technological progress. At the initial saving rate, s_0, the economy moves along AA. If, at time t, the saving rate increases to s_1, the economy experiences higher growth for some time until it reaches its new, higher path, BB. On path BB, the growth rate is again the same as before the increase in the saving rate (i.e. the slope of BB is the same as the slope of AA).

The Saving Rate and Consumption

Governments can affect the saving rate in various ways. First, they can vary public saving. Given private saving, positive public saving – a budget surplus, in other words – leads to higher overall saving. Conversely, negative public saving – a budget deficit – leads to lower overall saving. Second, governments can use taxes to affect private saving. For example, they can give tax breaks to people who save, making it more attractive to save and thus increasing private saving.

What saving rate should governments aim for? To think about the answer, we must shift out focus from the behaviour of output to the behaviour of consumption. It is clear that an increase in saving must come initially at the expense of lower consumption. A change in the saving rate this year has no effect on capital this year, and consequently no effect on output and income this year. So, an increase in saving comes initially with an equal decrease in consumption.

Does an increase in saving lead to an increase in consumption in the long run? Not necessarily. Consumption may decrease, not only initially, but also in the long run. You may find this surprising. After all, we know that an increase in the saving rate always lead to an increase in the level of output per worker. But output is not the same as consumption. To see why not, consider what happens for two extreme values of the saving rate:

- An economy in which the saving rate is (and has always been) zero is an economy in which capital is equal to zero. In this case, output is also equal to zero and so is consumption. A saving rate equal to zero implies zero consumption in the long run.
- Now consider an economy in which the saving rate is equal to one. People save all their income. The level of capital, and thus output, in this economy will be high. But because people save all of their income, consumption is equal to zero. What happens is that the economy is carrying an excessive amount of capital. A saving rate equal to one also implies zero consumption in the long run.

These two extreme cases mean that there must be some value of the saving rate between zero and one that maximizes the steady-state level of consumption. Increases in the saving rate below this value lead to a decrease in consumption initially, but lead to an increase in consumption in the long run. Increases in the saving rate beyond this value decreases consumption not only initially, but also in the long run. This happens because the increase in capital associated with the increase in the saving rate leads to only a small increase in output – an increase that is too small to cover the increased depreciation. In other words, the economy carries too much capital. The level of capital associated with the value of the saving rate that yields the highest level of consumption in steady-state is known as **golden-rule level of capital**. Increases in capital beyond the golden-rule level reduce steady-state consumption. This argument is illustrated in the following figure, which plots consumption per worker in steady-state (on the vertical axis) against the saving rate (on the horizontal axis). A saving rate equal to zero implies a capital stock per worker equal to zero, a level of output per worker equal to zero and, by implication, a level of consumption per worker equal to zero. For s between zero and s_G (G for golden rule), a higher saving rate leads to higher capital per worker, higher output per worker and higher consumption per worker. For s larger than s_G, increases in the saving rate still lead to higher values of capital per worker and output per worker, but they now lead to lower values of consumption per worker. This is because the increase in output is more than offset by the increase in depreciation as a result of the larger capital stock. For s = 1, consumption per worker is equal to zero. Capital per worker and output per worker are high, but all of the output is used just to replace depreciation, leaving nothing for consumption. If an economy already has so much capital that it is operating beyond the golden rule, then increasing saving further will decrease consumption not only now, but also later.

Chapter 12 – Technological Progress and Growth

The conclusion in the previous chapter that capital accumulation cannot by itself sustain growth has a straightforward implication: sustained growth requires technological progress. This chapter looks at the role of technological progress in growth.

12.2 The Determinants of Technological Progress

Technological progress depends on:

- The fertility of R&D, how spending on R&D translates into new ideas and new products;
- The appropriability of the results of R&D, which is the extent to which firms benefit from the results of their R&D.

When designing patent laws, governments must balance their desire to protect future discoveries and provide incentives for firms to do R&D with their desire to make existing discoveries available to potential users without restrictions.

12.3 Institutions, Technological Progress and Growth

Sustained technological progress requires that the right institutions are in place. In particular, it requires well-established and well-protected property rights. Without good property rights, a country is likely to remain poor. But in turn, a poor country may find it difficult to put in place good property rights.

12.4 The Facts of Growth Revisited

France, Japan, the United Kingdom and the United States have experienced roughly balanced growth since 1950. Growth of output per worker has been roughly equal to the rate of technological progress. Growth in China is a combination of a high rate of technological progress and unusually high investment, leading to an increase in the ratio of capital to output.