Classical Theories

1 Introduction

In this chapter, we study the economy as a whole under classical assumptions. Classical assumptions include the following: (1) people are individually rational, meaning that consumers maximize utilities and firms maximize profits; (2) prices (including wages and interest rates) are flexible, so that markets always clear; (3) markets for final goods/services and factor inputs are competitive; (4) and people have access to perfect information.

Under classical assumptions, the aggregate supply (AS) curve would be vertical since changes in the general price level cannot fool entrepreneurs to increase or cut output. Furthermore, competition between entrepreneurs will always push production to the point where almost all capital and labor are employed. This amounts to the same as the full employment of capital stock and labor supply.

Given a vertical AS, the demand side does not matter in the determination of the aggregate output. As the famous Say's law says, demand always accommodates supply. Indeed, if there are no fluctuations in factor inputs and productivity, then there would be no business cycles. In particular, there would be no unemployment problem, beyond a healthy *natural rate of unemployment*. If one group of the population somehow reduces their consumption, the rest will increase consumption at a lower price, keeping all factories running. It is a perfect world.

The assumption of flexible prices can be justified if we take a long-term view. Wages, for example, can be sticky in the short term but flexible in the long run. It is for this reason that classical economists may oppose government stimulus during recessions since prices and wages will adjust to bring the economy back to its potential level in the long run.

In this chapter, we assume as given both factor inputs (labor and capital) and the technology that transform inputs into outputs. That is, the "output potential" is assumed to be constant. We will leave to later chapters the study of economic growth.

2 The Output

In this section, we first introduce the macroeconomic concept of *technology*, which may be characterized by a production function. Then we present a classical AD-AS model to characterize the total output of the economy under classical assumptions.

2.1 Technology

Throughout the book, we assume that there are two factor inputs to the economy as a whole: capital and labor. We let K denote a measure of the capital stock, and let L denote the labor supply (either in the unit of working hours or the number of workers). And we use a production function F to characterize the "technology" of the economy, which is to transform K and L into an aggregate output (Y):

$$Y = F(K, L).$$

We should understand the "technology" of the whole economy in general terms. It is determined not only by the scientific and engineering know-how but also manufacturing organization, marketing skills, transportation, communication, and so on.

Assumptions on the Technology

The production function used in macroeconomics generally satisfies the following assumptions:

(a) Constant return to scale: for any $z > 0, F(zK, zL) = zY$.

(b) Increasing in both K and L:

$$F_1 \equiv \frac{\partial F}{\partial K} > 0, \quad \text{and} \quad F_2 \equiv \frac{\partial F}{\partial L} > 0.$$

(c) Decreasing marginal product of capital and labor:

$$F_{11} = \frac{\partial^2 F}{\partial K^2} < 0, \quad F_{22} = \frac{\partial^2 F}{\partial L^2} < 0.$$

(e) Capital-labor complementarity:

$$F_{12} = \frac{\partial^2 F}{\partial L \partial K} > 0.$$

Note that F_1 is the *marginal product of capital* (MPK) and F_2 is the *marginal product of labor* (MPL). It is readily accepted that, as in microeconomics, F should be increasing in both K and L, and that F should exhibit decreasing MPK and decreasing MPL. The capital-labor complementarity means that capital and labor are complementary inputs, in the sense that adding one of them would make the other more productive. Note that, for most production functions, $F_{12} = F_{21}$[1], meaning that the effect of an additional unit of labor on the MPK is equal to the effect of an additional unit of capital on MPL.

The assumption of constant return to scale requires more argument. If F does not have constant return to scale, then the performance of an economy would depend on its size. (We may measure the performance of an economy by per capita GDP, average life expectancy, and so on.) If F has increasing return to scale, for example, big countries would have advantages. In our real world, however, there is no evidence that size plays any crucial role in the contest of economic performance in the per capita sense. High-income countries include big ones and small ones. The same is true for low-income countries.

Perhaps the most famous production function is the Cobb-Douglas function, which is given by

$$F(K, L) = AK^{\alpha}L^{\beta},$$

where A is a constant that denotes the level of production efficiency. To satisfy the constant-return-to-scale assumption, we must impose $\alpha+\beta = 1$. As such, we rewrite the production function as

$$F(K, L) = AK^{\alpha}L^{1-\alpha}. \tag{1}$$

While the preceding production functions are static, we can easily make them dynamic, reflecting technological progress. Let A_t be the level of efficiency at time t. There are three ways to incorporate A_t into the production function:

- Labor augmenting: $Y_t = F(K_t, A_tL_t)$,
- Capital augmenting: $Y_t = F(A_tK_t, L_t)$,
- Total-factor augmenting: $Y_t = A_tF(K_t, L_t)$.

Another name for total-factor augmenting is *Hicks-neutral*. If technological progress is Hicks-neutral, then the marginal products of both factors increase at the same proportion. Obviously, the Cobb-Douglas technology in (1) is Hicks-neutral.

For simplicity, we assume in this chapter that both K and L are fixed, that $K = \bar{K}$ and $L = \bar{L}$, and that $F(\cdot, \cdot)$ is a fixed function. We define *output potential* as the level of total output that utilizes the current technology and all capital and labor. Letting $\bar{Y}$ denote output potential, we have

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

2.2 A Classical AD-AS Model

Since macroeconomics studies the economy as a whole, it is useful to introduce the concepts of aggregate demand (AD) and aggregate supply (AS). AD is the "sum" of all demand for goods and services. We can decompose AD into four major components: consumption demand, investment demand, government demand, and net foreign demand. And the AS is the "sum" of all supply of goods and services. Both AD and AS are in the "real" sense: When we say AD or AS changes, it is the quantity of goods and services that changes.

The quotation marks around "sum", however, signify the difficulty of summation of heterogeneous goods and services. If there is only one good that consumers and firms desire, then AD is simply the total quantity of the good people want to buy. In reality, however, there are almost infinite different goods and services. To understand aggregate demand (supply) of heterogeneous goods and services, we may imagine adding up the value of these goods and services in demand (supply) using constant prices just as we do calculating real GDP.

2.2.1 The AS Curve

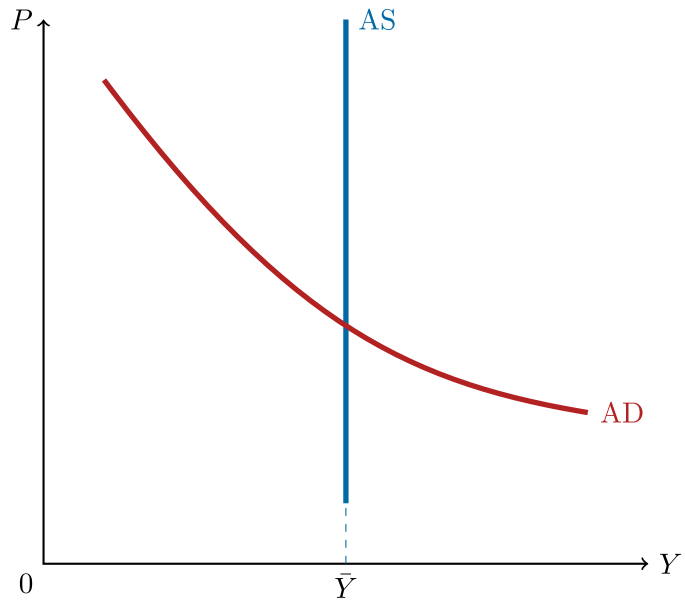
Generally, both AD and AS may be functions of the general price level (P). The AD curve is a relationship between AD and the general price level (P). And the AS curve is a relationship between AS and P. We first discuss the AS curve, which is more important than the AD curve under classical assumptions.

Under classical assumptions (in particular, that people have access to full information and prices are flexible), the AS curve may be vertical. Firms in the classical world know the difference between changes in relative prices, to which they would respond by increasing production, and changes in the general price level, to which they do not respond. Hence the aggregate supply does not change with the general price level.

The vertical AS curve is contrary to the easy conjecture that the AS curve is upward sloping since supply curves for individual products are generally upward sloping. This gives us an example of the so-called "fallacy of composition", which says that what is true for parts does not necessarily hold for the whole.

The next question is where the AS curve is located. We may conjecture that, in the classical world, competition between entrepreneurs will always push production *close to* the output potential ($\bar{Y}$ in Figure 1). Notice here that I use "close" to accommodate the fact that capacity utilization is always below the maximum level (e.g., due to option value of extra capacity) and that there is a natural level of unemployment (e.g., due to the fact that it takes time for people to switch jobs). To put it more precisely, firms will expand production to the point where output ceases to be elastic, which is equivalent to the almost full employment of capital stock and labor supply.

Figure 1: A classical AD-AS model



2.2.2 The AD Curve

The AD curve is widely believed to be downward sloping. However, it does not follow from the microeconomic *law of demand*, which states that the demand curves for individual goods are generally downward sloping. We need a *macroeconomic argument* for this claim. One famous argument made by Arthur Pigou (1877-1959) is that as the general price level declines, the purchasing power of money holding increases. Becoming wealthier, money holders would increase spending, thus boosting aggregate demand. Since a decline in the price level corresponds to an increase in AD, we have a downward sloping AD curve.

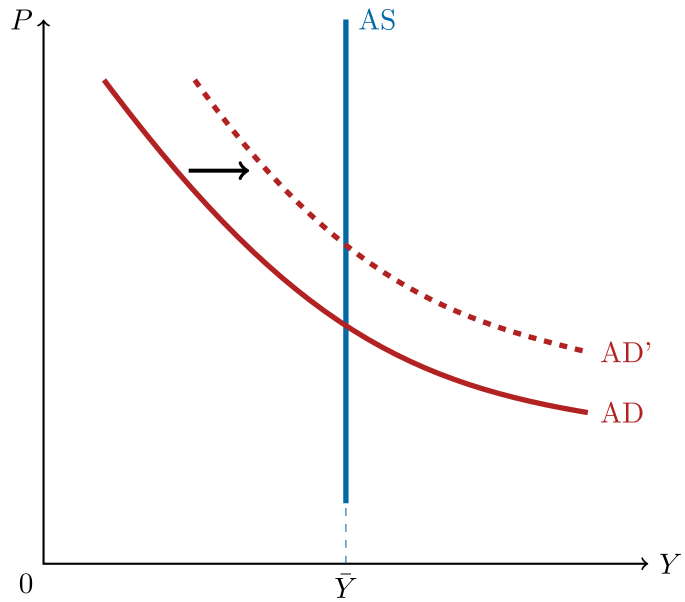
The point where the AD curve crosses the AS curve gives the equilibrium of the economy, as shown in Figure 1. Since the AS curve is vertical, it solely determines the equilibrium output, which is equal to the output potential:

$$Y = \bar{Y} = F(\bar{K}, \bar{L}). \tag{2}$$

2.2.3 The Effect of Demand Shocks

Demand shocks move the AD curve. A positive demand shock (e.g., fiscal stimulus) moves the AD curve to the right. That is, the positive shock increases aggregate demand at every price level. In the classical AD-AS model, the shifting of the AD curve only changes the general price level. For example, if the government expands its welfare program, then the AD curve would shift to the right. As shown in Figure 2, the fiscal expansion would lead to inflation, while failing to raise output.

Figure 2: The effect of a positive shock to the aggregate demand



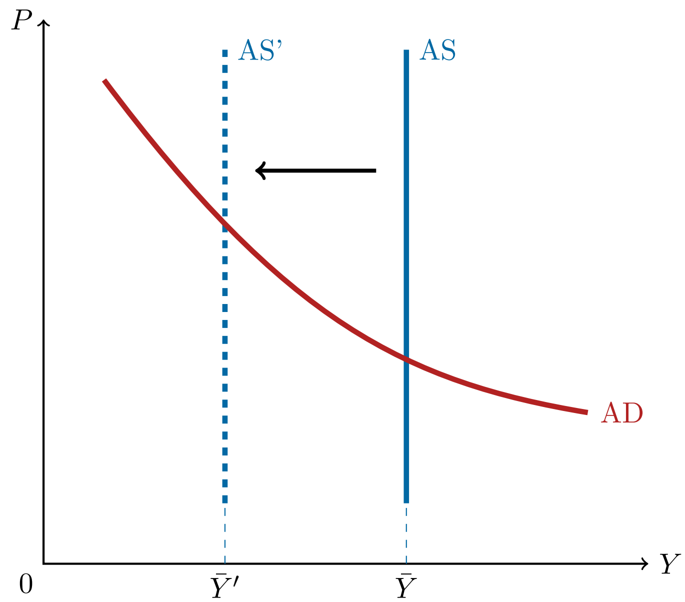
That AD always matches AS at the level of output potential relies on the crucial assumption that prices are flexible. If the aggregate demand falls short (the AD curve shifts to the left), the price level immediately declines to ensure the balance between demand and supply at the output potential $\bar{Y}$. In the real world, it often takes time for firms to adjust prices and wages. That is, many prices are sticky in the short run. But eventually prices and wages will adjust. It is in this sense that we may call classical models in this chapter *long-run* models.

2.2.4 The Effect of Supply Shocks

Supply shocks move the AS curve. Positive supply shocks (e.g., technological progress) move the AS curve to the right, leading to higher output and lower prices. Using the classical AD-AS model, we may argue that technological progress has a deflationary effect on the economy. In the same vein, we may argue that the opening of the Chinese economy, especially after the accession to WTO (World Trade Organization) in 2001, has a deflationary effect on the rest of the world.

Equally interesting, negative supply shocks (e.g., supply chain disruptions) move the AS curve to the left, leading to lower output and higher prices (Figure 3). The combination of stagnation and inflation is known as *stagflation*.

Figure 3: The effect of a negative shock to the aggregate supply



2.2.5 A Special AD Curve

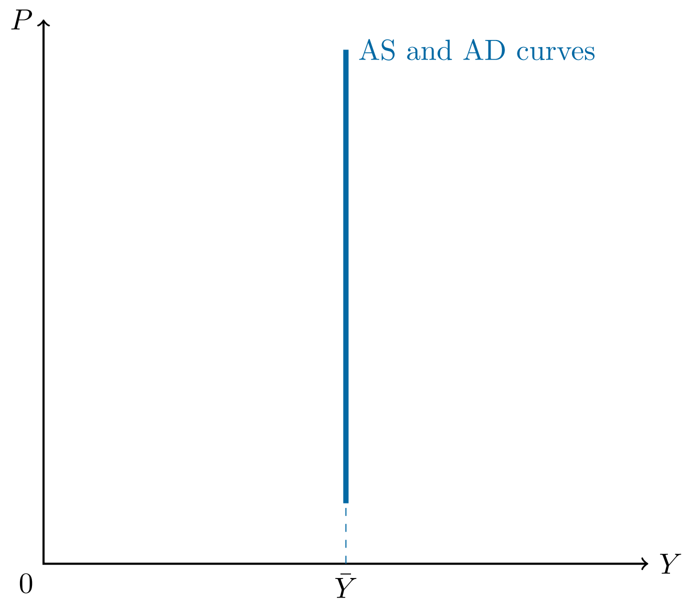
The AD curve may not slope downwards. To counter Pigou's argument, suppose that a substantial number of people are in debt. Then a price decline would make their debt heavier in real terms and they become less willing to spend. The net effect of a price decline on the AD may be zero or even negative.

The case of zero net effect is interesting, since this corresponds to a vertical AD curve. In this case, since both AD and AS curves are vertical, they must overlap to make markets clear, as in Figure 4. Any point on the AS or AD curve is an equilibrium, corresponding to some general price level, which is indeterminate in this model. And to understand how the aggregate demand can accommodate aggregate supply at any price level, imagine that in a barter economy, people sell something to buy something else. As a result, we have "supply creates its own demand", a classical doctrine called Say's law.

3 Unemployment

As discussed previously, in an idealized world where classical assumptions hold, unemployment should be minimal. However, the unemployment rate would still be nonzero simply because it takes time for workers to switch jobs. For example, after a worker quits his job, he/she typically cannot find a new job immediately. It would take some time for him/her to search for vacancies, submit resumes, conduct

Figure 4: When both AD and AS curves are vertical.



interviews, and so on. Between quitting the old job and accepting a new job offer, he/she is unemployed.

We may call the minimal rate of unemployment existing in a healthy economy as the *national rate of unemployment*. In this section, we first present a simple model that relates the natural unemployment rate to the ease (difficulty) of finding and losing jobs. We then discuss the reason why it takes time to find jobs, which results in so-called *frictional unemployment*. Finally we discuss the *structural unemployment* arising from wage rigidity.

3.1 A Model of Natural Unemployment

Let L denote the labor force, E the number of the employed, and U the number of the unemployed. We know that $L = E + U$ and U/L is the unemployment rate.

Let s be the rate of job separation with $0 \leq s < 1$. We assume that in a given period (say, a month), there are sE of those employed losing their jobs. Note that if people who lost jobs can find jobs immediately, then $s = 0$. From a macroeconomic point of view, no one is losing a job when people can switch jobs instantly.

Similarly, let f denote the rate of job finding with $0 \leq f < 1$, and we assume that there are fU of the unemployed finding jobs in the same period.

And we assume that the job market is in a *steady state*, in which the number of job loss (sE) equals the number of job finding (fU),

$$sE = fU.$$

Then, in the steady state, we have

$$s\left(1 - \frac{U}{L}\right) = f\frac{U}{L},$$

which yields

$$\frac{U}{L} = \frac{1}{1 + f/s}.$$

As long as $s > 0$, which means that some of the unemployed cannot immediately find jobs, the unemployment rate will be positive.

Any policy aiming to lower the unemployment rate must make it easier to find jobs. The policies that would make it more difficult to fire workers, however, can easily backfire. Such policies would make employers reluctant to employ workers in the first place.

3.2 Frictional Unemployment

If $s = 0$ and $f > 0$, the unemployment rate in the above model is zero. We may describe such a labor market as *frictionless*. And the unemployment due to the fact that it takes time to find new jobs is called *frictional unemployment*.

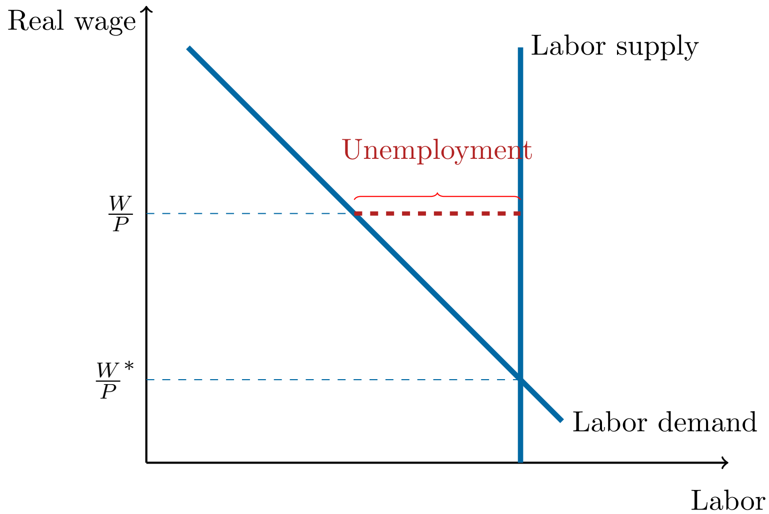
The fundamental reason for the friction is the *heterogeneity* of jobs and workers, meaning that each worker, and also each vacancy, is different. And the problems of asymmetric information, imperfect labor mobility, and so on, would make the job matching even more difficult and time consuming.

Furthermore, there may be industrial or sectoral shifts happening in the economy. When the horse-wagon industry was declining, for example, workers in this industry would find their skills obsolete. To find a new job, say in the automobile industry, it takes time to learn new skills.

To reduce frictional unemployment, the government can help disseminate information about jobs and even provide training programs. The private sector can do at least equally well on information dissemination, especially in the current Internet age. But the government may be especially helpful in supporting training programs since training has a positive *externality*: If a company trains a group of workers, the company incurs the full cost of training, but the company cannot realize all of the benefits since some of the workers may go to other companies after training.

A prevalent policy regarding the job market is *unemployment insurance*, which helps to soften the economic hardship of the unemployed. As a result, especially when the unemployment insurance is overgenerous, workers may have less incentive

Figure 5: Structural unemployment



to look for jobs urgently, and hence it may contribute to higher natural unemployment. However, unemployment insurance may help to achieve better matching between workers and jobs, hence enhancing the efficiency of the labor market.

3.3 Structural Unemployment

In the real world, the estimated natural rate of unemployment also includes the unemployment arising from wage rigidity, which we call *structural unemployment*. As shown in Figure 5, structural unemployment occurs when wages are higher than the market-clearing level and remain rigid.

Wage rigidity may come from the law of minimum wage. Minimum wage may increase structural unemployment among individuals with low or impaired skills (e.g., young or disabled people), whose market-clearing wage may be lower than the law of minimum wage dictates. Empirically, however, economists find that increasing the minimum wage does not necessarily lead to fewer jobs.[2]

Wage rigidity may also be due to strong labor unions. In industries with a strong union presence, union members ("insiders") may, through collective bargaining, manage to keep their wage artificially high. As a result, firms in the industry may tend to reduce employment.

Wage rigidity may also come from the practice of *efficiency wage*. Efficiency wage refers to the practice to pay employees more than equilibrium wage to increase the productivity of workers, or reduce costs associated with turnover. High wage mitigates the problem of adverse selection since higher wage attracts and retains able employees. High wages also mitigate the problem of moral hazards since high wage increases the monetary loss of workers getting fired for shirking. If a large number of firms practice efficiency wage, however, the overall wage level of the economy would be higher than the market-clearing level, causing structural unemployment.

Like frictional unemployment, structural unemployment presumably does not fluctuate with economic cycles. And the actual unemployment rate is composed of the non-fluctuating natural rate of unemployment and a fluctuating component that we call *cyclical unemployment*. The relationship among these definitions of unemployment may be summarized as follows:

$$\begin{aligned}\text{natural unemployment} &= \text{frictional unemployment} + \text{structural unemployment}\\ &= \text{actual unemployment} - \text{cyclical unemployment}\end{aligned}$$

Note that only the actual unemployment rate is observable. We need to estimate the natural rate of unemployment to obtain the cyclical component. Cyclical unemployment will be studied in Chapter **??**.

4 Income Distribution

As previously discussed, the total output and total income must equal the output potential $\bar{Y}$ under classical assumptions. The remaining question is how the income would be distributed among owners of factor inputs, that is, those who provide capital and those who provide labor. As we can imagine, factor prices (real wage and real rental price of capital) would be crucial for the determination of the distribution.

The real wage is the payment to labor measured in units of output, W/P, where W is nominal wage and P is the price of output. (In empirical studies, P would be CPI or GDP deflator).

The real rental price of capital is the rental price paid to the owner of capital in units of output, R/P, where R is the nominal rent. In most cases, firm owners also own the capital stock. But we can imagine that the firm rents capital from the owner of capital and pays rent, just like paying wages to the owner of labor (i.e., workers).

To study how the factor prices are determined, we introduce a representative-firm model.

4.1 A Representative-Firm Model

We assume that the markets for goods and services are competitive and that the markets for factors of production (labor and capital) are also competitive. Note that a market is competitive if no participant is large enough to affect prices. In other words, all market participants are price takers.

To determine the real wage and real rental price of capital, we look at the decision of a *representative firm*. We may imagine that the economy is composed of many small firms with the same technology $F(K_i, L_i)$, where K_i and L_i are capital and labor inputs to the i-th firm, respectively. These firms produce the same product consumed by consumers with the same taste (utility function). As a result, the total production of the economy can be characterized by a representative firm with the production function $F(K, L)$, where K and L are the total capital and labor of the economy, respectively. Here, the constant-return-to-scale assumption on F is crucial, making possible the aggregation of firm-level technology into a macroeconomic production function.

The representative firm takes as given the price of its output (P), wage (W), and real rental price of capital (R), and solves the following problem:

$$\max_{K,L} P \cdot F(K, L) - W \cdot L - R \cdot K.$$

That is, the representative firm tries to maximize *economic profit* by choosing an optimal combination of capital and labor.

Concepts: Economic Profit and Accounting Profit

Consider a firm with two factor inputs: labor and capital. Economic profit is defined as income (revenue) minus costs of labor and capital. In the preceding problem for the representative firm, $P \cdot F(K, L)$ is revenue, $W \cdot L$ is the cost of labor, and $R \cdot K$ is the cost of capital (or return to capital).

Accounting profit is defined by the sum of economic profit and the return to capital. Since most firms own capital rather than rent them, return to capital is part of the accounting profit.

The first-order condition for the maximization problem with respect to K is as follows:

$$F_1(K, L) = \frac{R}{P}, \tag{3}$$

where $F_1 \equiv \partial F / \partial K$ denotes the partial derivative of F with respect to the first argument, K. Equation (3) says that the firm would employ capital up to the point where the MPK equals the real rental price of capital.

And the first-order condition for the maximization problem with respect to L is:

$$F_2(K, L) = \frac{W}{P}, \tag{4}$$

where $F_2 \equiv \partial F / \partial L$ denotes the partial derivative of F with respect to the second argument, L. Equation (4) says that the firm would employ labor up to the point where the marginal product of labor (MPL) equals the real wage.

Note that if we fix $K = \bar{K}$, the first-order condition for L gives us the *demand curve for labor*, i.e., the relationship between real wage (W/P) and the labor demanded (L): $F_2(\bar{K}, L) = W/P$. We can check that, since we assume a *decreasing* marginal product of labor, a lower real wage corresponds to a higher demand for labor.

4.2 Income Distribution

Recall that the classical economy fully employs the total capital $\bar{K}$ and labor supply $\bar{L}$ (omitting the natural rate of unemployment), which implies that $\bar{K}$ and $\bar{L}$ must solve (3) and (4). That is to say, the representative firm maximizes its profit when $K = \bar{K}$ and $L = \bar{L}$. As a result, the owner of labor receives $F_2(\bar{K}, \bar{L}) \cdot \bar{L}$, the owner of capital receives $F_1(\bar{K}, \bar{L}) \cdot \bar{K}$.

Interestingly, there is no economic profit left for the whole economy. To see this, note that under the constant-return-to-scale assumption on the production function, we have $F(zK, zL) = zF(K, L)$ for any $z > 0$. Then it follows from $\frac{dF(zK,zL)}{dz} = \frac{d(zF(K,L))}{dz}$ that

$$F_1(zK, zL)K + F_2(zK, zL)L = F(K, L).$$

Now let $z = 1$ and use the fact that $K = \bar{K}$ and $L = \bar{L}$, we have

$$F_1(\bar{K}, \bar{L}) \cdot \bar{K} + F_2(\bar{K}, \bar{L}) \cdot \bar{L} = \bar{Y}.$$

To understand this intuitively, imagine an economy with many small firms with the same technology. Since the technology has constant return to scale, tiny would-be firms (say, workshops) can enter the market and compete with existing ones. As a result, we may deduce that there would be no economic profit for the existing firms.

Income Distribution in the Cobb-Douglas Economy

Suppose that a classical economy is characterized by the Cobb-Douglas production function, $F(K, L) = AK^\alpha L^{1-\alpha}$, we have

$$\begin{aligned} \text{MPK} &= F_1(K, L) = \alpha \frac{AK^\alpha L^{1-\alpha}}{K} = \alpha \frac{F(K, L)}{K} \\ \text{MPL} &= F_2(K, L) = (1-\alpha) \frac{AK^\alpha L^{1-\alpha}}{L} = (1-\alpha) \frac{F(K, L)}{L} \end{aligned}$$

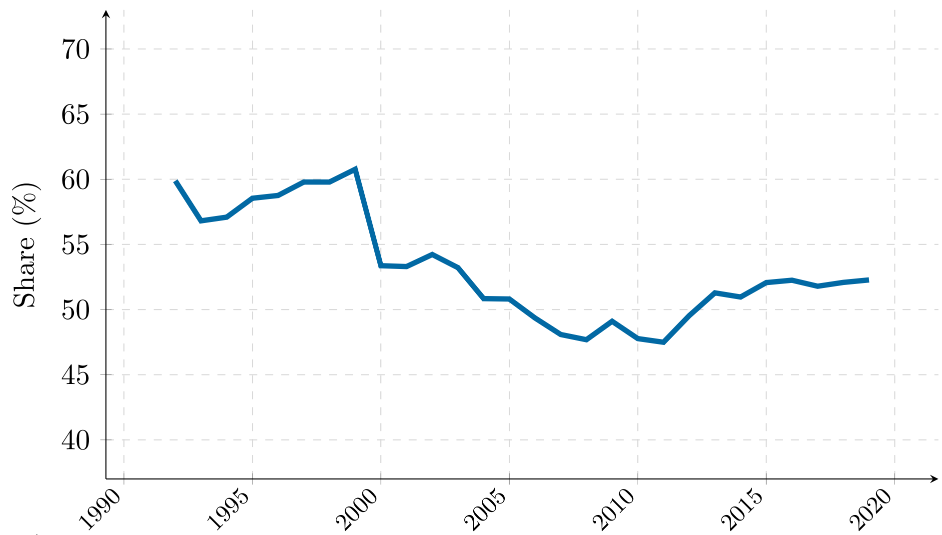
The capital share of income is

$$F_1(\bar{K}, \bar{L}) \cdot \bar{K} = \alpha \bar{Y}.$$

And the labor share of income is

$$F_2(\bar{K}, \bar{L}) \cdot \bar{L} = (1-\alpha) \bar{Y}.$$

Figure 6: Labor share of income in China



†Data source: National Bureau of Statistics, China

It would be an interesting empirical exercise to check whether the shares of capital and labor are constant. Figure 6 shows, however, that the labor share of income in China changes substantially over time.[3] During the 1990s, the labor share fluctuates around 60 percent. It dropped substantially in the first half of 2000s. The labor share reached the lowest point (47.5 percent) in 2011, after which we see a slow rebound. In 2019, the labor share of income in China stood at 52.3 percent.

The United States has a much longer data set on labor's share of income. Figure 7 shows the ratio of employee compensation in the national income. From 1929 to 1970, we can see a trend of increasing labor's share of income. From 1970 to recent years, we can see a downward trend.

In addition to trends, we can also see cycles in labor's share of income, which often peaks in the depth of recessions. To understand this, note that the return to capital often declines faster than the return to labor during recessions.

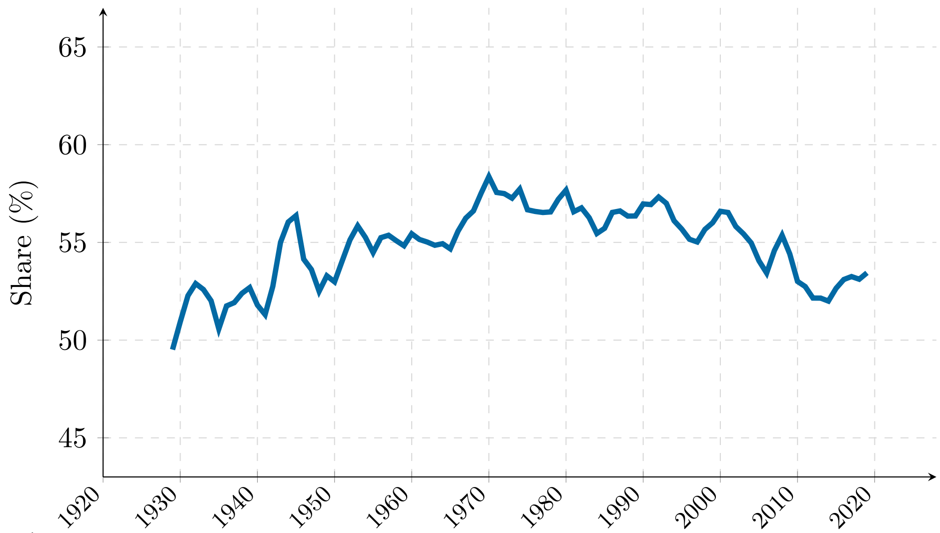
4.3 Labor Productivity and Real Wage

The *average labor productivity* (or simply, labor productivity) of an economy is defined by the average output, Y/L. In the Cobb-Douglas economy, we have

$$\text{MPL} = F_2(K, L) = (1-\alpha)\frac{AK^{\alpha}L^{1-\alpha}}{L} = (1-\alpha)\frac{Y}{L}.$$

Hence the MPL is proportional to the average labor productivity in the Cobb-Douglas economy. Once again, it would be interesting to investigate whether this is the case in the real economy. Table 1 shows that, in the United States where long

Figure 7: Labor's share of income in the United States



†The labor's share of income is measured by the ratio of total employee compensation to GDI. Data source: WIND.

Table 1: Growth in Labor Productivity and Real Wage in the United States

Period	Average growth in labor productivity	Average growth in real nonfarm compensation
1959-2019	2.1	1.3
1959-1972	2.8	2.3
1973-1994	1.6	0.7
1995-2007	2.7	1.6
2008-2019	1.3	0.8

†Data source: FRED.

data are available, the growth rates of labor productivity and real wage are positively correlated. At the same time, however, the growth of real wage lags behind that of labor productivity. This observation is consistent with the fact that the labor's share of income has been declining in the United States during the sample period.

5 Interest Rate

Interest is payment from a borrower to the lender for the price of using of borrowed money. The interest rate (or rate of interest) is interest per amount due *per period*, which is often a year. Even if the borrowing is for a shorter term, say a month, we still quote the interest rate in annual percentages. For example, if the interest payment on a one-month loan of 100 is 1, then the interest rate is 1% per month, or an annualized 12.68%[4]. It is convenient to use annualized rates to compare interest rates on loans with different maturities.

Without explicit qualifications, interest rate in macroeconomics refers to *risk-free* interest rate, the interest rate on loans or bonds without credit risk. For example, the interest rate on China's government bonds is a risk-free rate for savers who do not worry about the exchange rate risk.[5]

Note that interest is a different concept from rental price of capital. Roughly speaking, interest is return to money, while rental price of capital is return to capital.[6] A dramatic macroeconomic phenomenon in the past four decades is that interest rates in the Western world have declined to zero or even negative values, while the return to capital remains stable.[7]

5.1 Real Interest Rate

Interest rate quotes in the practical world are nominal interest rates, which do not account for inflation. The real interest rate is the interest rate a lender receives (or expects to receive) after accounting for the effect of inflation. Given a nominal interest rate, if inflation is high (or is expected to be high), then the real interest rate is low. In economics, we often assume that people care about the real interest rate, which is the "real" opportunity cost of money.

The concept of the real interest rate is best understood in the context of "real" borrowing. For example, if I borrow 100 kg of rice from my neighbor and I have to pay a debt of 110 kg of the same rice, then the real interest rate of my borrowing is 10%.

If I borrow money (and then use money to buy rice), however, then the problem of calculating the real interest rate becomes more difficult. The difficulty lies in how to account for inflation. For example, if I borrow 1,000 CNY from my neighbor and buy 100 kg of rice (the rice price is 10 CNY/kg). If I pay a debt of 1,100 CNY next year, then the nominal interest rate is 10%. If the rice price does not change, then the real interest rate is also 10%. But if the rice price rises to 11 CNY/kg, then the real interest rate is zero. A zero real interest rate means that the borrowed money has the same purchasing power as the money paid back.

Unlike nominal interest rate, which we can directly observe, the real interest rate needs to be estimated. There are two ways to define real interest rates. One is called the *ex post* real interest rate or *realized* real interest rate,

$$r = i - \pi, \tag{5}$$

where i is the nominal interest rate, r is the real interest rate, and π is the inflation rate. For example, if the nominal interest rate on a loan is 5% and the inflation rate turns out to be 3%, then the ex post real interest rate is 2%.

The other definition gives the *ex ante* real interest rate,

$$r = i - E\pi, \tag{6}$$

where $E\pi$ is the expectation of inflation. The ex ante real interest rate is useful when loaners and debtors negotiate a (nominal) interest rate and they need to form an expectation about the future inflation. For example, if a loaner and a debtor agrees on a nominal interest rate of 5% on a one-year loan and they expect that there will be an inflation of 3% over the next year, then the ex ante real interest rate is 2%.

Note that (5) is called the Fisher equation (named after Irving Fisher) and (6) is called the modified Fisher equation.

5.2 A Classical Model of Interest Rate

In the modern world, central banks determine one or more key interest rates such as the federal funds rate of the US Federal Reserve, the main refinancing operations (MRO) rate of the European Central Bank, and so on. Although other interest rates (e.g., long-term government bonds, corporate bonds, bank loans, etc.) are mostly equilibrium outcomes of the market demand and supply, they are immensely influenced by the policy rates that the central banks control.

Classical economists, however, live in the era of small government with very limited central banking. They generally view the interest rate as a price that brings demand and supply of funds into equilibrium, without much influence from any monetary authority. In this section, we present a model that captures such a view. The model specifies a set of behavioral assumptions and imposes an equilibrium condition. We will use the model to examine the effects of external shocks (e.g., change in fiscal policy).

For simplicity, we assume that the net export equals zero, $X = 0$. This implies either a closed economy or an open economy with balanced trade. Then the national income accounts identity becomes,

$$Y = C + I + G, \tag{7}$$

where Y represents GDP, C represents consumption expenditure, I represents investment expenditure, and G represents government expenditure. To define national saving $S = Y - C - G$, we may rewrite (7) as

$$S = I.$$

This equation states that "saving must equal investment." If we regard saving as the supplier of funds and investment as the demander of funds, then the equation may be interpreted as an equilibrium condition in a financial market. We will build a model on this equilibrium condition.

In the following, we make a set of behavioral assumptions on the consumption expenditure (C) and investment expenditure (I). Specifically, we introduce a consumption function and an investment function to characterize consumption and investment in the economy, respectively. And we regard the government expenditure (G) and tax (T) as exogenous variables, whose values are given outside the model. After building the model, we may change exogenous variables and see what happens to endogenous variables (in this case, the real interest rate). We may call such analysis a *virtual experiment*.

5.2.1 Consumption Function

Assume that the tax T is levied on household income. The disposable income is then $Y - T$, the toal income minus tax. The *consumption function* characterizes the total consumption expenditure (C) by a function of the disposable income, $C = C(Y - T)$. We assume that $C(\cdot)$ is an increasing function. That is, more disposable income leads to more consumption.

We define the *marginal propensity to consume* (MPC) as the amount of additional consumption given unit increase in disposable income. Mathematically, MPC is the first derivative of the consumption function with respect to Y:

$$MPC = \frac{dC(Y)}{dY}.$$

For example, if $C(\cdot)$ is a linear function, e.g.,

$$C(Y - T) = 100 + 0.7(Y - T),$$

then MPC is a constant and $MPC = 0.7$.

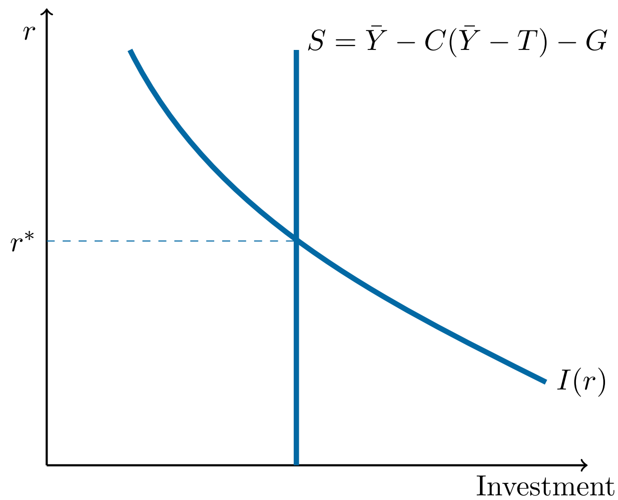
5.2.2 Investment Function

Since higher real interest rates discourage borrowing and hence investment, we assume that the investment expenditure of the economy is a decreasing function of the real interest rate, $I = I(r)$ with $I'(r) < 0$.

5.2.3 Fiscal Policy

The fiscal policy determines how much the government tax and spend. In this model, we capture the fiscal policy by two exogenous variables: the tax revenue of the government (T), and the government expenditure (G). If $G = T$, we have a *balanced budget*; if $G > T$, we have a *budget deficit*; and if $G < T$, we have a *budget surplus*.

Figure 8: Determination of real interest rate



The budget surplus $(T - G)$ is also called the *public saving*. A negative public saving means budget deficit. And we may define the private (non-government) saving as

$$S_{ng} = Y - C - T.$$

By adding the public saving and private saving together, we obtain national saving: $S = Y - C - G$.

5.2.4 Equilibrium in the Financial Market

We assume there exists a simple financial market for loanable funds. Those with savings would lend their savings to borrowers (investors) in the financial market. We assume that the national savings, $Y - C - G$, is the supply of loanable funds in the financial market. On the other hand, the demand for loanable funds comes from the investment need, $I(r)$.

In equilibrium, the real interest rate (r) must adjust so that saving (supply of loanable funds) equals investment (demand for loanable funds):

$$\bar{Y} - C(\bar{Y} - T) - G = I(r). \tag{8}$$

Note that the left-hand side is the saving (S). The unknown real interest rate is the only endogenous variable in the model. All the remaining variables, T, G, and $\bar{Y}$, are exogenous variables. Recall that $\bar{Y}$ is the output potential of the economy and that, under the classical assumptions, the total output of the economy equals the output potential. The solution of the preceding equilibrium equation is illustrated in Figure 8. Note that in the model, saving does not depend on the interest rate, hence a vertical supply (or saving) curve.

5.3 Virtual Experiment

Models allow us to conduct virtual experiments on the economy without actually fiddling with it. We now use the preceding model to study how exogenous shocks would affect the equilibrium real interest rate.

From a mathematical point of view, the equilibrium condition in (8) defines an *implicit function* of r, the only endogenous variable. When exgenous variables change ($\bar{Y}$, T, G), the equilibrium r changes. The implicit function characterizes the dependence of the equilibrium r on exogenous variables, which we may denote by $r(\bar{Y}, T, G)$.

A virtual experiment is thus a study of the implicit function. For example, we may be interested in the question of how r would change if G increases, holding $\bar{Y}$ and T fixed. Assuming that the implicit function $r(\bar{Y}, T, G)$ is differentiable,[8] we just need to study the partial derivative of r with respect to G,

$$\frac{\partial r(\bar{Y}, T, G)}{\partial G} \equiv \lim_{\Delta \to 0} \frac{r(\bar{Y}, T, G + \Delta) - r(\bar{Y}, T, G)}{\Delta}.$$

And the *implicit function theorem* (see Section **??** for a review) can be employed to study the preceding partial derivative. For example, we may apply the implicit function theorem to (8) and obtain:

$$\frac{\partial r(\bar{Y}, T, G)}{\partial G} = -\frac{-1}{-I'(r)} > 0.$$

That is, if the government increases spending, then the real interest rate would rise.

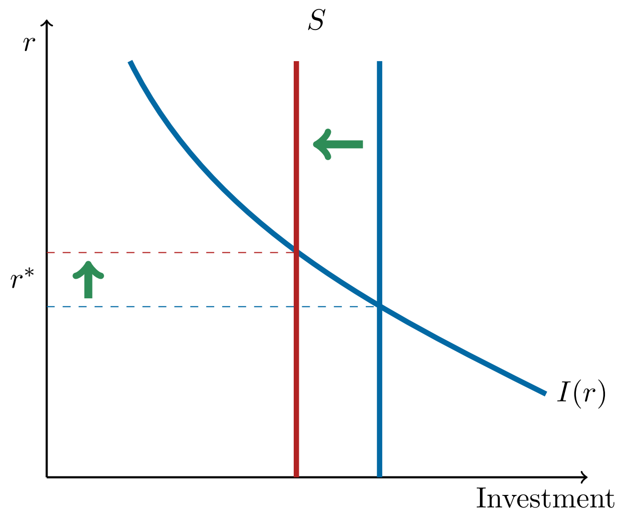
Graphically, the increased government spending shifts the saving curve to the left (Figure 9), resulting in a higher equilibrium real interest rate. The same would be true if there is a tax cut ($T \downarrow$). Both the increase in G and the decrease in T would reduce national savings. The former reduces national savings by reducing public savings ($T - G$), and the latter reduces national savings by increasing private consumption.

A higher real interest rate corresponds to lower investment. Economists would say that fiscal stimulus ($G \uparrow$ or $T \downarrow$) *crowds out* the private investment. And under classical assumptions, the crowding out is complete, meaning that the stimulus would fail to increase total output or employment.

5.4 Explain the Decline of Real Interest Rate

We can use this model for economic explanation. For example, the real interest rate in the Western world has been declining for almost fourty years (see Figure 10 for the US case). We may conjecture that the lack of investment enthusiam, which itself

Figure 9: The effect of fiscal stimulus



may be due to the lack of investment opportunities, may be to blame. We now check whether the model prediction is consistent with the conjecture.

We introduce an exogenous variable, d, that enters the investment function, $I = I(r, d)$. We assume that d measures investment enthusiasm and that $\partial I/\partial d > 0$. The model in (8) becomes

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

We need to check whether a declining d leads to a declining r, holding other exogenous variables fixed. Since the left-hand side is fixed and I is increasing in d and decreasing in r, r must decline as d declines in order for the equation to hold. Applying the implicit function theorem, we obtain

$$\frac{\partial r(\bar{Y}, T, G, d)}{\partial d} = -\frac{-I_2(r, d)}{-I_1(r, d)} > 0,$$

where $I_1 \equiv \partial I/\partial r < 0$ and $I_2 \equiv \partial I/\partial d > 0$. Hence r must decline when d declines.

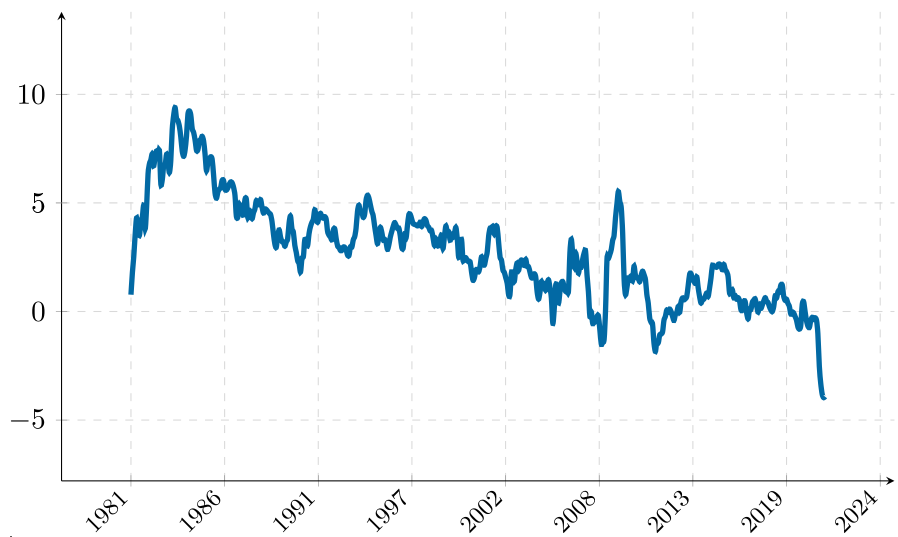
Graphically, the decline of d corresponds to the shift the investment curve to the left (Figure 11), meaning that the investment demand decreases at every r. This results in a lower equilibrium real interest rate.

6 Money and Inflation

6.1 Money

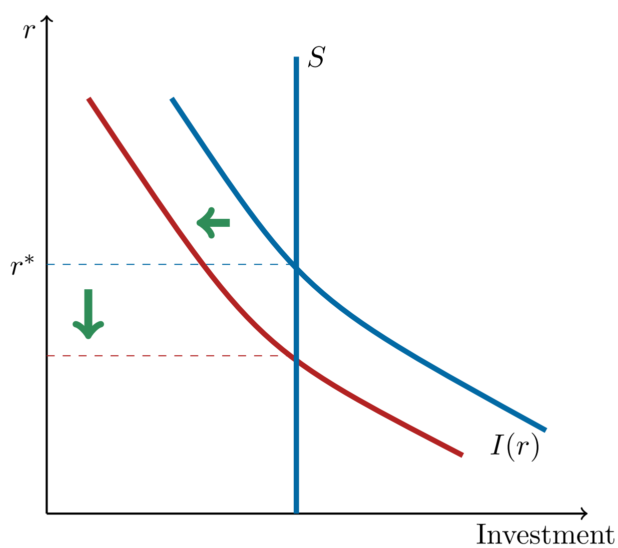
Money is the stock of assets that can be readily used to make transactions. Functions of money include store of value, unit of account, and medium of exchange. The last

Figure 10: The US real interest rate (1981-2021)



†Data source: FRED. The real interest rate is measured by interest rate on ten-year Treasury minus CPI YoY inflation.

Figure 11: The effect of declining investment enthusiasm



function, to intermediate the exchange of goods and services, is especially important for classical thinkers. If there is no money and people barter goods and services with each other, trading opportunities would be much reduced, since it is highly unlikely to find *double coincidence of wants*. The introduction of money, which is acceptable by everybody for everything, solves this problem.

Furthermore, money makes pricing simple. Imagine a market with N different goods without money, we need $N(N-1)/2$ pairs of price quotes. But if money is used to intermediate exchanges, only N price quotes are needed.

Given such convenience afforded by money, it is not surprising that human society has used money, in one form or another, from very early history. At first, people use commodity money such as shells, gold, silver, and so on. But transactions using commodity money (say gold) is costly since the purity and weight of a piece of gold have to be examined in every transaction. To reduce transaction costs, a bank (possibly with authorization from the government) may mint gold coins of known purity and weight. To further reduce cost, the bank may issue gold certificates, which can be redeemed for gold. The gold certificate eventually becomes gold-backed paper money.

In modern times, especially after the industrial revolution, economic growth sped up, outpacing the growth of the gold supply. Hence the limited supply of gold has a deflationary effect on the economy if a country keeps using gold as money. Eventually, it is realized that if people do not care about the option of redeeming gold, the bank can issue certificates that are not backed by gold in the vault. The modern central bank does exactly this, and these certificates become fiat money. Fiat money is valued because people expect it to be valued by everyone else.

In the traditional sense, money includes cash, central bank reserves, and bank deposits. As financial markets evolve, more financial assets take on "moneyness" and become money (e.g., government bonds). The supply of cash and reserves is controlled by the monetary authority (the central bank), which deliberate and implement monetary policies. The supply of bank deposits is determined by bank lending, which is in turn determined by prevailing interest rates, risk appetite, investment opportunities, and so on.

The monetary authority in China is the People's Bank of China (PBC). In the United States, the monetary authority is the Federal Reserve (the Fed). A monetary authority may serve more than one sovereign nation. The European Central Bank (ECB), for example, serves as the monetary authority for the entire eurozone, a group of sovereign European countries.

6.2 Inflation

Inflation is sustained increase in the general price level of goods and services. Temporary fluctuations in price level do not constitute inflation. For example, seasonal increase in price before and during the Spring Festival in China may not be inflation, since the price would often decline after the holiday, as demand wanes and supply recovers. Price increases in some particular goods or services are also not regarded as inflation unless they are accompanied by the rise in the general price level.

If there is a sustained decrease in the general price level, we call it *deflation*. A related concept is *disinflation*, which refers to the case where the inflation rate declines. As discussed in the previous chapter, we use CPI or GDP deflator to measure inflation.

When there is inflation, the purchasing power of the money declines. And there are losers and winners from inflation. Losers include people who save, people who hold bonds, and, generally, people who receive fixed incomes. Retired pensioners are especially vulnerable. Unexpected inflation is equivalent to redistribution of wealth from savers to borrowers, who are winners of inflation. Unexpected inflation also increases a sense of uncertainty in the economy, discouraging investment.

Even expected inflation has costs. First, high inflation leads to a high frequency of price changes, which are costly because sellers and buyers have to renegotiate prices, and new menus have to be printed (metaphorically, menu costs). Second, high inflation leads to high opportunity cost of holding cash, causing inconveniences of insufficient cash holding. It can be metaphorically called "shoe-leather cost", meaning that more frequent visits to banks would cause one's shoes to wear out more quickly. Third, high inflation makes price signal noisy, affecting the ability of the "invisible hand" to allocate resources. Fourth, tax brackets are often in nominal terms (e.g., the minimum taxable monthly salary is 3,500 CNY in China), and high inflation would make tax burden heavier than intended.

When prices lose control and inflation skyrockets, the economy may fall into a full crisis. According to a loose definition, if inflation exceeds 50% per month, we call it *hyperinflation*. All the costs of moderate inflation previously described become prohibitive under hyperinflation. Money may cease to function as a store of value, and may not serve its other functions (unit of account, medium of exchange). People may have to barter or use a stable foreign currency.

What causes hyperinflation? An easy and classical answer is that hyperinflation is caused by excessive money supply growth. When the central bank prints money, the price level rises. If it prints money rapidly enough, the result is hyperinflation. But why would a central bank print money like crazy? In most cases, it would be due to fiscal problems. When a government experiences fiscal crisis due to either extraordinary expenditure (war, indemnity, etc.) or impaired tax power or both, the government may resort to excessive money printing.

However, there is one benefit of moderate inflation that proves important for the health of macroeconomy. It is a well-supported fact that nominal wages are rigid, even during recessions. Inflation allows real wages to reach equilibrium levels without cutting the nominal wage. Therefore, moderate inflation makes the labor market less frictional.

On the other hand, deflation may look good, since it implies increased purchasing power of money. But deflation is almost intolerable in a modern economy, as it makes debts more difficult to service, discourages investment, and thus aggravates unemployment. This is why, after the 2008 global financial crisis, central banks around the world had conducted aggressive monetary policies (quantitative easing, negative interest rates, etc.) to maintain positive inflation.

6.3 Quantity Theory of Money

The quantity theory of money is the most important classical theory about money and inflation.

Suppose there is only one good in the economy with price P. Let T be the total number of transactions during a period and M the money in circulation. We may define the transaction velocity of money in this simple economy by

$$V \equiv PT/M.$$

The quantity theory of money is thus stated as an identity,

$$MV = PT.$$

In reality, there is an almost infinite number of goods and services with different prices. It is obviously not appropriate to count the transactions of different goods equally. For example, the purchase of a car uses a lot more money than that of a cell phone. A natural way to count transactions of heterogenous goods and services is to *weight* each transaction with constant prices, just like the calculation of real GDP. This gives us a more practical quantity theory of money:

$$MV = PY, \tag{9}$$

where Y denotes real GDP. Note that the new version of quantity theory is nothing but an alternative definition of the velocity of money.

We may also interpret the equation in (9) as an equilibrium condition in the money market. Rewrite (9) as

$$M/P = kY, \tag{10}$$

where $k \equiv 1/V$. If V is, as usual, assumed to be a constant, so is k. We may interpret the left-hand side as the real money supply and the right-hand side as money demand, which is assumed to depend on the total income only. We can read the equilibrium condition as follows:

$$\text{real money supply} = \text{money demand}.$$

Note that k characterizes how much money people wish to hold for each unit of income. It is by definition inversely proportional to V: When people hold lots of money relative to their incomes, money changes hands infrequently.

6.3.1 Money and Output

According to the classical AD-AS model, the output of a classical economy always matches the output potential, $Y = \bar{Y}$. Thus, money does not affect output, and monetary policy is ineffective.

For this to be compatible with the quantity theory of money, the general price level should be perfectly flexible. When the money supply rises, price should rise proportionally so that M/P remain constant. If P rises slower than M, the quantity theory of money would predict a rise in Y since k is assumed to be a constant.

Of course, the flexibility of the general price level is implied by the classical assumption of flexible prices.

6.3.2 Money and Inflation

We take total differentiation of (9) and obtain

$$\frac{dM}{M} + \frac{dV}{V} = \frac{dP}{P} + \frac{dY}{Y}.$$

Note that dM/M and dY/Y are growth rates of money supply and real GDP, respectively, and that dP/P is the inflation rate. If we assume that the velocity V is constant, then $dV/V = 0$. The quantity theory of money implies that, given the real GDP growth rate, a higher growth rate of the money supply leads to higher inflation.

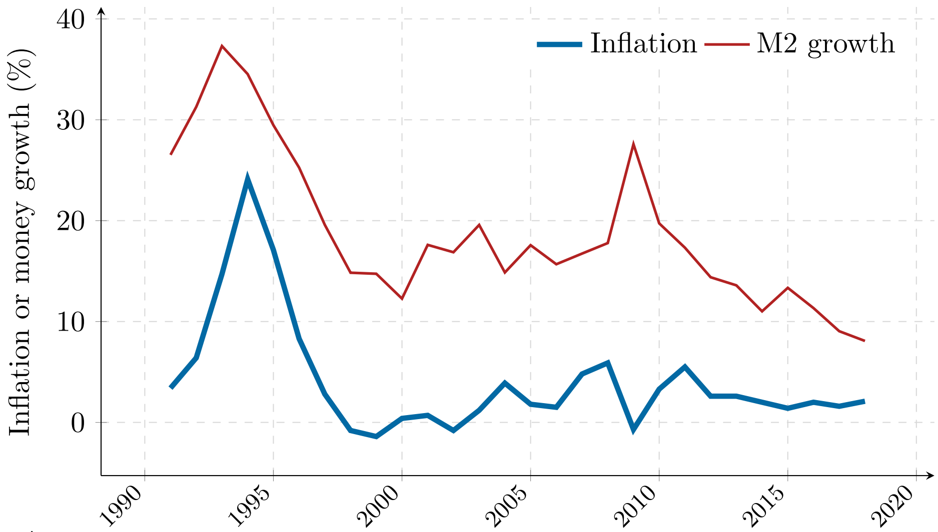
In the real world, inflation does not necessarily co-move with the growth in the money supply. Figure 12 shows the annual inflation and growth of money supply in China. Although there are periods when the two move roughly in tandem, there are also periods when they move in opposite directions. For example, M2 grew rapidly in 2009 thanks to the Four-Trillion-Stimulus program. But inflation declined to negative territory amidst the global recession.

However, when money supply is excessive, as would happen during episodes of hyperinflation (e.g., China in the late 1940s, Germany in 1923), the co-movement between the growth of money supply and inflation is clear.

6.4 China's Hyperinflation in the 1940s

In the following, we document the Chinese hyperinflation in the 1940s. The hyperinflation had a long-lasting impact on the Chinese economy, especially on the weight of price stability among policy priorities.

Figure 12: China's annual inflation and M2 growth



†Data source: National Bureau of Statistics, China

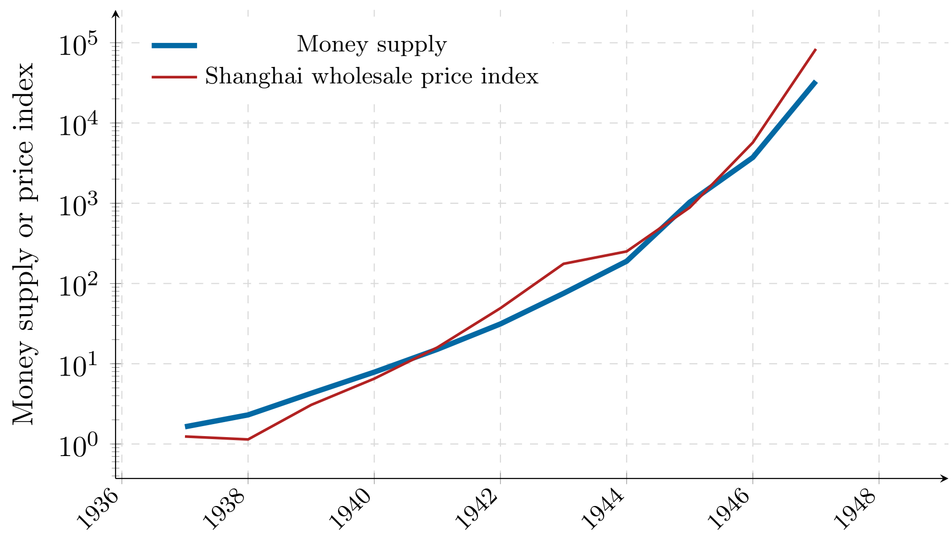
6.4.1 Background

In 1935, China's Nationalist government carried out a currency reform with two major tasks. First, the reform centralized the bank-note issuance to four major national banks and started to issue a unified currency: Fabi ("legal note"). Second, the reform gave up the former silver-based currency and started to peg Fabi to the British pound, which served as the reserve currency. In 1936, the US dollar became another reserve currency for Fabi. However, the Nationalist government soon found that it was unable to maintain the link to pound or dollar, with its fiscal capacity much impaired by the full-scale Japanese invasion starting from 1937.

6.4.2 Hyperinflation: Stage One

The currency reform was to tackle the problem of deflation due to the outflow of silver, which was, in turn, caused by the US effort to increase its silver reserve. For this objective, the reform was a success. Soon enough, inflation picked up. And as the war started and the fiscal condition became desperate, seigniorage became the major fiscal revenue for the Nationalist government. As a result, inflation turned from moderate to extreme. From 1937 to 1945, the year when the war with Japan ended, the money supply grew from 1.6 billion to 1 trillion Yuan (unit of Fabi), and the price level in Shanghai (measured by the wholesale price index) increased from 1.2 to 885 (Figure 13).

Figure 13: China's hyperinflation in the 1940s (1937-1948)



†Data source: Liu (1999), Research on China's Central Banking.

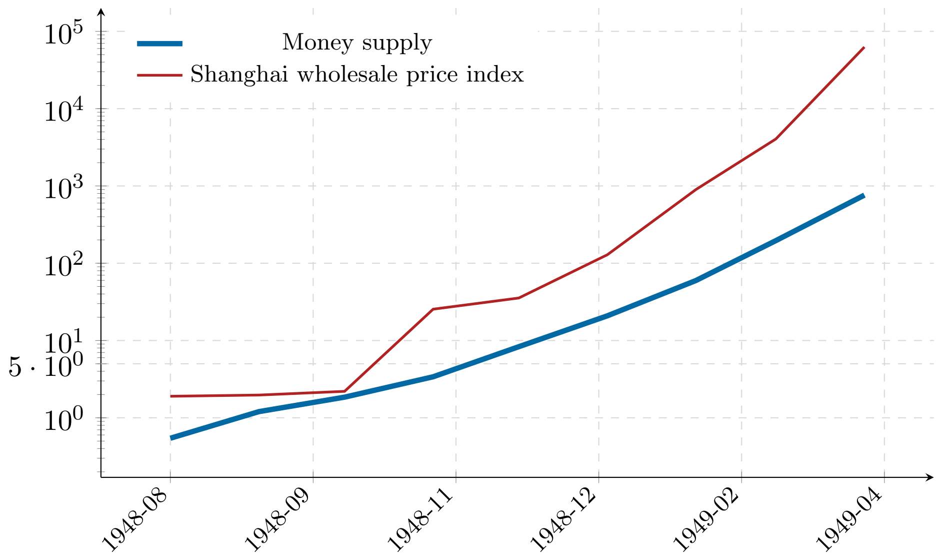
6.4.3 Hyperinflation: Stage Two

And this was not the end of the story. China's civil war soon broke out. Since the Nationalist government was unable to support its war with tax revenue, it continued to rely on seigniorage. By the end of 1947, the money supply already multiplied to 33.2 trillion Yuan. In the meantime, inflation accelerated. By August 1948, the money supply reached 604.5 trillion, and inflation in Shanghai more than doubled in eight months (Figure 13).

At this point, the Nationalist government conducted another currency reform: issuing a new currency to replace Fabi. This new currency, called the Gold-Yuan Coupon (GYC),[9] was designed to be backed by the gold reserve. But the Nationalists did not have enough gold reserve. Nor were they able to stabilize, let alone expand, fiscal revenue. It was a monetary reform without the companion of a fiscal reform. And making matters worse, the Nationalist army was losing on almost all of the battlefields. The reform was doomed.

From August 1948 to the end of the year, the supply of GYC ballooned from 0.5 billion to 8.3 billion, and the price level in Shanghai went up from 1.9 (a different index from that mentioned previously) to 35.5. By April 1949, the GYC supply multiplied to 760 billion, and inflation averaged 130% every month (Figure 14). The economy was in full collapse.

Figure 14: China's hyperinflation in the 1940s (1948-8:1949-4)



†Data source: Liu (1999), Research on China's Central Banking.

6.4.4 Legacies

It is widely believed that the hyperinflation contributed to the end of the Nationalists' regime on the mainland. But hyperinflation itself reflected deep problems in the Nationalists' rule. For one thing, the central government did not have effective control over all of China, even before the Japanese invasion. Given the limited fiscal capacity, the excessive extraction of seigniorage revenue became a necessity for defending the country against the Japanese.

After the Japanese surrendered, the Nationalists had the opportunity to consolidate its fiscal position. But its paramount leader, Chiang Kai Shek, continued to rely on seigniorage to wage an unpopular civil war, believing in an easy victory. The hyperinflation in 1947 and onward reflected the failure of the Nationalists' economic management as well as their military failure.

After the Communists won the civil war and established the People's Republic of China, they introduced a new monetary regime, and the price soon stabilized. Knowing that its dramatic success was partly due to the hyperinflation during the Nationalists' reign, the new rulers of China made maintaining price stability one of their highest policy priorities.

6.5 Classical Dichotomy

We can combine the classical AD-AS model in (2), the classical model of real interest rate in (8), the quantity theory of money in (10), and the Fisher equation in (5):

$$
\begin{aligned}
Y &= \bar{Y}, \\
I(r) &= Y - C(Y-T) - G, \\
\frac{M}{P} &= kY, \\
i &= r + \pi.
\end{aligned}
$$

Note that in this integrated model, real variables (e.g., Y and r) are determined without considering money. Money supply only influences the general price level, which in turn determines the nominal values such as nominal GDP, nominal interest rate (i), and so on. The idea of separating "real" from "nominal" analysis is called the *classical dichotomy*. If the classical dichotomy holds, we also say that money is *neutral*.

Naturally, monetary policy is irrelevant if money is neutral. The expansion of the money supply, according to the classical theory, only drives up the price level and the nominal interest rate. It does not reduce the real interest rate, or influence the output or employment. In the real world, however, evidence abounds that monetary policy has real effects.

7 Exchange Rate

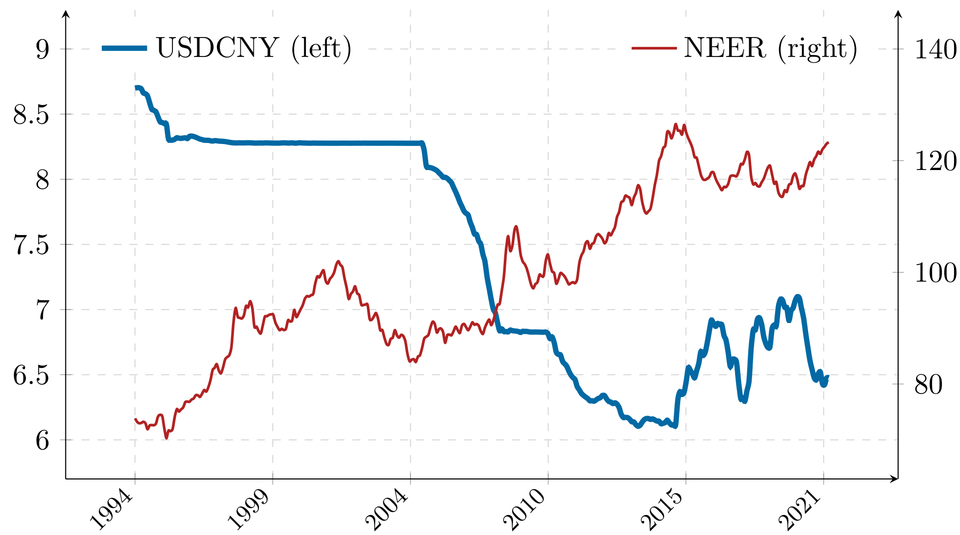
The exchange rate (also known as the foreign-exchange rate, or forex rate) between two currencies is the rate at which one currency exchanges for another.

In this course, we adopt the convention that the exchange rate is in units of foreign currency per domestic currency. Under this convention, a rise in the exchange rate is called *appreciation*; a fall in the exchange rate is called *depreciation*. Appreciation is also called *strengthening*, while depreciation is also called *weakening*.

Because a country trades with many countries, it is often useful to calculate the effective exchange rate, an index measuring the value of the domestic currency against a basket of foreign currencies. Figure 15 shows the exchange rate of the Chinese Yuan (CNY) with respect to the US dollar (USD) and the nominal effective exchange rate (NEER) of CNY. Note that USDCNY represents the amount of CNY a USD can exchange. When USDCNY declines, it means that CNY appreciates against USD.

It is interesting to note that during 2015, CNY depreciated about 8% against USD. But in terms of effective exchange rates, CNY appreciated approximately 10% relative to its trading partners. So looking at only one bilateral exchange rate,

Figure 15: The exchange rates of Chinese Yuan (CNY)



†Data source: WIND. USDCNY represents the amount of CNY a USD can exchange. NEER represents the nominal effective exchange rate of CNY.

however important it is, may make us miss the big picture of a currency's exchange rate movement.

7.1 Real Exchange Rate

The real exchange rate is the purchasing power of a currency relative to another currency at current nominal exchange rates and prices. Let e be the nominal exchange rate, P the domestic price in domestic currency, and P^* the foreign price in foreign currency. Since e is in units of foreign currency per domestic currency, the domestic price in foreign currency is eP. (Here we may imagine that there is only one good in the world. Thus the price level is just the price for the good, making price levels in two countries comparable.) Then the real exchange rate (ε) is defined by the ratio of the domestic price in foreign currency to the foreign price in foreign currency:

$$\varepsilon = \frac{eP}{P^*}. \tag{11}$$

The lower the real exchange rate, the less expensive domestic goods and services are relative to foreign ones.

Example: Real Exchange Rate

Suppose both China and the United States produce and consume one good, the Big Mac. And suppose that the Big Mac costs 20 CNY in China and 4 USD in the United States and that the nominal exchange rate is 6 USDCNY.

Then the real exchange rate between China and the United States is

$$\varepsilon = \frac{\frac{1}{6} \cdot 20}{4} = \frac{5}{6}.$$

Since the real exchange rate is less than 1, we say that PPP does not hold, and CNY is undervalued: It is profitable to buy Big Macs in China, sell them in the United States, and convert the USD proceeds back to CNY.

7.2 Purchasing Power Parity

Examining the definition of the real exchange rate in (11), we can see that if domestic and foreign currencies have identical purchasing power, then the real exchange rate (ε) should be exactly one. Indeed, if $\varepsilon = 1$, we say that the exchange rates are at *purchasing power parity* (PPP). Theoretically, PPP is implied by the *law of one price*, which states that the same good should have the same price. If $\varepsilon > 1$, the domestic currency is overvalued in terms of purchasing power, meaning that domestic prices are higher than foreign prices in general. If $\varepsilon < 1$, the domestic currency is under-valued in terms of purchasing power.

If PPP holds, we have

$$e_t = P_t^*/P_t. \tag{12}$$

In practice, since P_t and P_t^* are measured by price indices, they are not comparable. Thus (12) is not useful for empirically testing whether PPP holds. Instead, we may take log difference of (12). Since $\pi_t = \log(P_t/P_{t-1})$, we have

$$\log\left(\frac{e_t}{e_{t-1}}\right) = \pi_t^* - \pi_t, \tag{13}$$

where π_t^* and π_t are foreign and domestic inflations, respectively. Note that $\log\left(e_t/e_{t-1}\right)$ represents the rate of appreciation of the domestic currency from time $t-1$ to t. Equation (13) says that, under PPP, if foreign inflation is higher than domestic inflation, the domestic currency would appreciate by the inflation gap $(\pi_t^* - \pi_t)$.

If we further assume a common real interest rate in the two economies, then we have

$$\log\left(\frac{e_t}{e_{t-1}}\right) = i_t^* - i_t, \tag{14}$$

where i_t^* and i_t are foreign and domestic nominal interest rates, respectively. Equation (14) says that if the foreign nominal interest rate is higher than the domestic one, the domestic currency tends to appreciate. The equation in (14) is often called *uncovered interest rate parity*, which characterizes an equilibrium where investors of the weak currency have to be compensated with a higher interest rate.

PPP does not generally hold in practice, especially in the short term. First, not all goods are tradable. Second, there are trading barriers and trading costs. These make cross-country arbitrage of price differences incomplete and costly. As a result, researchers find little empirical support for PPP if they use short-term data to test implications of PPP, say Equation (13). However, more support of PPP can be found in long-term data.

7.3 Trade Balance and Capital Flow

In an open economy, domestic spending need not equal its output. The difference is the net export, which is the total value of export minus that of import. According to the national income identity,

$$Y - (C + I + G) = X = EX - IM,$$

where Y is output, $(C + I + G)$ represents domestic spending, X stands for net export, EX stands for export, and IM stands for import. All these variables are in the *real* sense.

If domestic spending is less than the output, then $X > 0$ and the surplus of goods and services is lent to foreigners. If the domestic spending exceeds the output, then $X < 0$ and the country borrows goods and services worth $(-X)$ from abroad. The net export is also called the trade balance.

The flow of goods and services is mirrored by capital flow. Let $S = Y - (C+G)$ be the national saving, we have

$$S - I = X. \tag{15}$$

The left hand side is the difference between the national saving (S) and investment (I), which is the *excess saving* of the economy. Since the excess saving has to flow out of the country, we may also call $(S - I)$ the *net capital outflow*.

Equation (15) says that the net capital outflow always equals the net export. If $S - I = NX > 0$, the country lends its excess saving $(S - I)$ to foreigners. And if $S - I = NX < 0$, then the country borrows $(I - S)$ from abroad.

To understand this identity more intuitively, we examine an imagined example. If BYD (a Chinese auto company) sells an electric car to a US consumer for 10,000

USD, how does the sale change China's trade and capital flow? On trade, the Chinese export rises by 10,000 USD. On capital flow, if BYD invests the 10,000 USD in the US securities (e.g., stocks or bonds), then Chinese capital outflow rises by 10,000 USD. The same is true even if BYD keeps the cash, which is the most liquid US "security" issued by the US government. If BYD converts the 10,000 USD into CNY at a local Chinese bank, then the bank also has to do something about it. If the bank chooses to purchase the US securities or to keep the dollar cash, then Chinese capital outflow again rises by 10,000 USD. If the bank sells the dollar to the central bank, which uses the 10,000 USD to purchase US treasury bills, we would still see a 10,000 USD rise in capital outflow.

7.4 A Model of a Small Open Economy

Now we introduce a model that characterizes the determination of the real exchange rate, which further determines net export or net capital outflow. The model is classical in the sense that output is given (exogenous) under classical assumptions and that the real exchange rate, the only endogenous variable, is assumed to be flexible.

7.4.1 The Model

We consider a small open economy and make the following assumptions:

(i) There is a common real interest rate (r^*) in the world.

(ii) The capital flow of the small economy does not affect the world interest rate.

(iii) The net export is a decreasing function of the real exchange rate, $X(\varepsilon)$, with $X'(\varepsilon) < 0$.

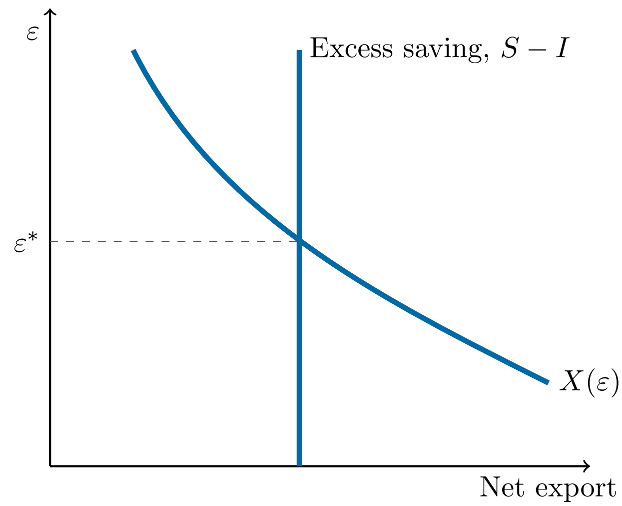
To justify (i), we may assume that capital is perfectly mobile across borders. As a result, global arbitragers would make sure the real interest rate is the same across the world. The assumption (ii) is a key characteristic of a small economy. The excess saving of the small economy does not affect the world interest rate. In other words, the small economy is a "price taker" of the world interest rate r^*. The combination of (i) and (ii) makes the real interest rate *exogenous* in the model. Finally, (iii) is a reasonable assumption since a higher real exchange rate encourages imports and makes the export sector less competitive.

Building on (15), we have

$$S - I(r^*) = X(\varepsilon), \tag{16}$$

where $S = \bar{Y} - C(\bar{Y} - T) - G$. We may interpret the net capital outflow, which is the left-hand-side of (16), as the demand for foreign currency. The net export, on

Figure 16: A model of small open economy



the other hand, represents the supply of foreign currency. Then we may interpret (16) as an equilibrium condition on the foreign exchange market, where exporters would sell their foreign currency to those who want to hold foreign financial assets.

Note that both S and I are exogenous, since they are functions of exogenous variables. Consequently, the demand side of the foreign exchange market (the left-hand side of (16)) is given. The real exchange rate adjusts the right-hand side to make supply equal to demand. We assume that the real exchange rate is flexible so that the foreign exchange market always clears. Figure 16 illustrates the solution of the model graphically.

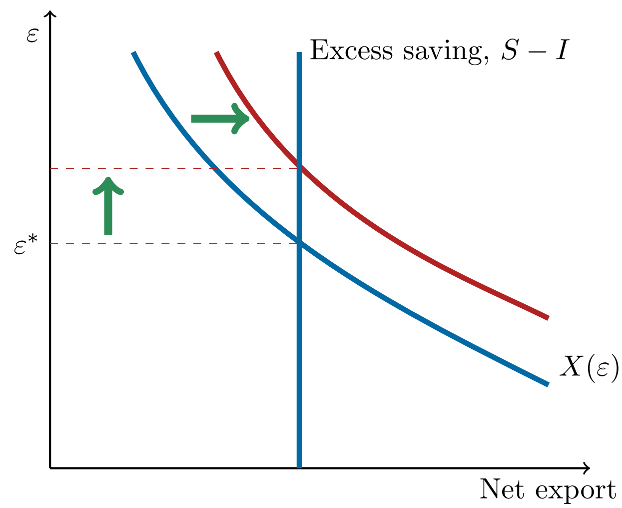
Next we conduct three virtual experiments on the model: a fiscal stimulus, a rise in the world interest rate, and implementation of protectionist trade policy. We use graphical analysis and leave algebraic analysis (using the implicit function theorem) for exercises.

7.4.2 Fiscal Stimulus

Fiscal stimulus may take two forms, cutting taxes ($T \downarrow$) or increasing government spending ($G \uparrow$). Both would reduce national saving (S), thus reducing the excess saving ($S - I$), which constitutes demand for foreign currency. As a result, the real exchange rate must appreciate (the foreign currency becomes cheaper) so that the foreign exchange market can get back to equilibrium.

Graphically, the reduction of national savings would shift the excess-saving curve (Figure 16) to the left, resulting in a higher equilibrium real exchange rate. That is, the domestic currency would appreciate.

Figure 17: The effect of a protectionist shock



The appreciation of the domestic currency would depress exports and stimulate imports. As the total output remains at the level of output potential, the reduction of net exports must be such that the fiscal stimulus would fail to stimulate the total output. This prediction is similar to the complete "crowding-out" of the investment by a fiscal stimulus in the closed economy.

7.4.3 A Rise in the World Interest Rate

If the world interest rate rises ($r^* \uparrow$), the investment expenditure would decline and the excess savings ($S - I$) would increase. This would shift the excess-saving curve (Figure 16) to the right, resulting in a lower equilibrium real exchange rate.

The depreciation of the domestic currency would stimulate the net exports. Since the total output remains at the potential level, when a rising world interest rate depresses the investment demand, a rising foreign demand fully compensates for the loss of aggregate demand.

7.4.4 Protectionist Policy Shock

Suppose that the government implements a protectionist policy that discourages imports and encourages exports. At every real exchange rate (ε), the policy would make the net export $X(\varepsilon)$ bigger. As a result, the $X(\varepsilon)$ curve would shift to the right (Figure 17), and the equilibrium real exchange rate will rise. Thus the classical model predicts that the protectionist policy would fail to lift the net export. The only effect of the policy is the appreciation of the domestic currency.

The reason why we reach such a dramatic conclusion is that we assume the excess savings ($S-I$) does not depend on the exchange rate. And the excess savings alone determine the net export in our model. To increase net export or decrease the trade deficit, classical economists would argue that the government should increase national savings by, for example, cutting government expenditure.

7.5 A Model of a Large Open Economy

In the small open economy model, we assume that the economy is a price taker of the world interest rate. That is, the excess saving of the economy does not affect the world interest rate. This makes the world interest rate an exogenous variable in the small-open-economy model. If this condition does not hold, meaning that the capital outflow of the economy does affect the world interest rate, then we have to develop a model with two endogenous variables: the world (real) interest rate and the real exchange rate. We call it the model of a large open economy. Presumably, the savings and investment behavior of a large economy would have an impact on the world interest rate.

7.5.1 The Model

For the large-open-economy model, we make the following assumptions:

(i) There is a common real interest rate (r) in the world.

(ii) The world interest rate declines when the net capital outflow from the economy increases. Equivalently, the net capital outflow is a decreasing function of the world interest rate, $F(r)$, with $F'(r) < 0$.

(iii) The net export is a decreasing function of the real exchange rate, $X(\varepsilon)$, with $X'(\varepsilon) < 0$.

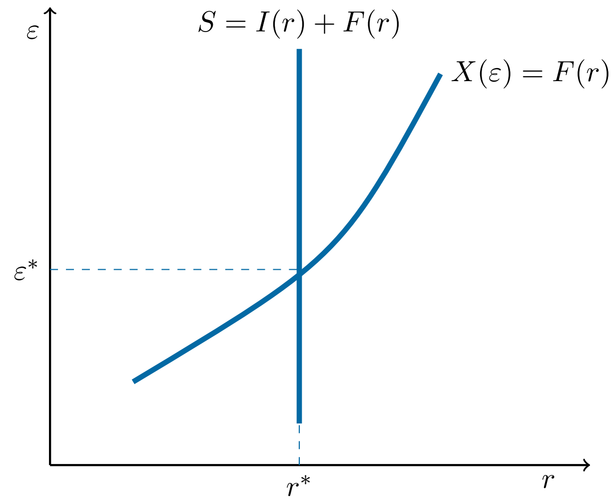
Assumptions (i) and (iii) are the same as in the small open economy model. The second assumption speaks to the "largeness" of the large open economy. To see why it is reasonable, note that an increase in the net capital outflow of a large economy would make capital more abundant in the world capital market, depressing the world (real) interest rate.

By definition, the net capital outflow equals the excess saving, $F = S - I$. Rearranging the terms, we have

$$S = I + F.$$

We may interpret S as the supply side of loanable funds in the (domestic) financial market. The demand side has two components: the investment demand and the capital-outflow demand. We may imagine that savers supply loanable funds to the financial market, entrepreneurs borrow funds to invest in the economy, and the

Figure 18: A model of large open economy



excess saving goes to people who want to hold foreign financial assets. Note that F can be negative. In this case, entrepreneurs borrow funds from abroad to invest in the economy. The identity $S = I + F$ gives us the first equilibrium condition.

Recall that the net export always equals the net capital outflow,

$$X = F.$$

We may interpret X as the supply side in the foreign exchange market and F as the demand side. In the foreign exchange market, exporters would sell their foreign currency (obtained from the sale of goods to foreigners) to those who want to hold foreign financial assets. The identity $X = F$ gives us the second equilibrium condition.

Building on the preceding two equilibrium conditions, we have the following model for a large open economy:

$$S = I(r) + F(r), \tag{17}$$

$$X(\varepsilon) = F(r). \tag{18}$$

We have two endogenous variables in the model of two equations: the world real interest rate (r) and the real exchange rate (ε). But the analysis of the model is straightforward. Note that there is only one endogenous variable (r) in (17), which solely determines the equilibrium real interest rate r^*. Next, we can analyze the equilibrium exchange rate ε, treating r^* as given.

Graphically, (17) corresponds to the vertical line on the two-dimensional diagram in Figure 18. On the other hand, (18) dictates that a bigger r must accompany a bigger ε. Thus the curve corresponding to (18) must be upward-sloping.

Using the model, we can conduct virtual experiments on a large open economy. We first analyze the impact of a fiscal stimulus on the economy. Then we analyze what would happen if the government implements a protectionist policy.

7.5.2 Fiscal Stimulus

The fiscal stimulus, whether in the form of increased government expenditure or tax cuts, is a negative shock to the national savings (S). We first analyze the impact of the shock on the equilibrium interest rate r^* by inspecting (17). Then we analyze the impact on ε^*, treating the change in r^* as given.

Since both $I(r)$ and $F(r)$ are decreasing functions of r, r^* must rise when S declines. Graphically, the vertical line in Figure 18 shifts to the right. As a result, the equilibrium exchange rate also rises.

We may verify the second prediction by inspecting (18). Since $F(r^*)$ has declined after the negative shock to S, $X(\varepsilon^*)$ should also decline. Since $X(\varepsilon)$ is decreasing in ε, the equilibrium exchange rate ε^* must rise (appreciate). In conclusion, a negative shock to national savings would result in a higher real interest rate and an appreciation of the domestic currency.

7.5.3 Protectionist Shock

The protectionist shock would have an impact on the net export function $X(\cdot)$, which appears only in the second equation (18). The vertical curve corresponding to the first equation does not shift. We now analyze how the upward-sloping curve $(X(\varepsilon) = F(r))$ shifts under the shock.

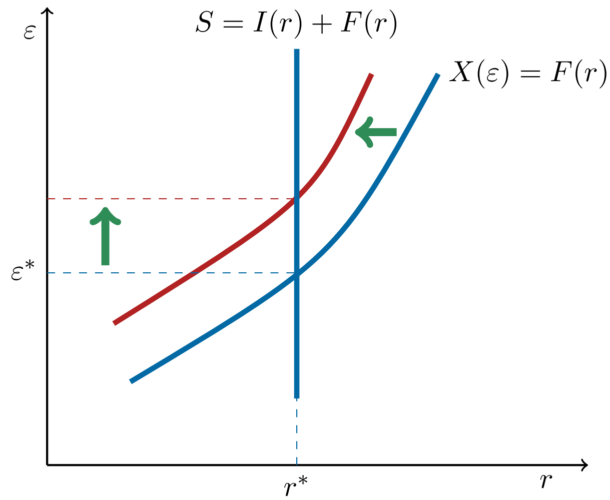
If we fix any ε, $X(\varepsilon)$ would increase as the protectionist shock takes place. To make $F(r)$ increase as well, r must decline. As this is true for every ε, we conclude that the curve $(X(\varepsilon) = F(r))$ must shift to the left (Figure 19). Hence a protectionist shock (e.g., raising import tariffs) would result in the appreciation of the domestic currency.

8 Concluding Remarks

The classical models in this chapter deal with the long-run equilibrium, assuming that the productive capacity of the economy does not grow. Although the theories are intellectually satisfying and the arguments are sometimes convincing, they do not directly deal with two of the most important questions in macroeconomics, economic growth and fluctuations. On the need for thinking about short-term fluctuations, Keynes famously made the following remark:

> In the long run we are all dead. Economists set themselves too easy, too

Figure 19: The effect of a protectionist shock



useless a task if in tempestuous seasons they can only tell us that when the storm is long past the ocean is flat again.

Nonetheless, understanding classical models is still useful. First, they may serve as the benchmark, the starting point from which we may conduct further research. Second, when prices are flexible (i.e., during periods of high inflation), the long-term equilibrium analysis may shed light on the short-term trends. Finally, thanks to its simplicity, the classical theories are often influential. For example, the quantity theory of money almost dominates in the popular media. To have productive dialogues with nonprofessionals, economists should understand classical theories, both their strengths and weaknesses.

Notes

[1]This result is known as Schwarz's theorem, Clairaut's theorem, or Young's theorem.

[2]For a famous empirical study of the minimum wage, read Card and Krueger (1994): Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. American Economic Review, September, 84(4), 772-793.

[3]We calculate the labor share of income using the income to the household sector in the flow-of-funds table (nonfinancial transactions).

[4]This is obtained by $(1 + 0.01)^{12} - 1$.

[5]In China, market participants call risk-free bonds "interest-rate bonds" (利率债). And bonds with credit risk are called "credit bonds" (信用债)

[6]Classical economists may use the word "capital" in place of money. Alfred Marshall, for example, define interest as the price paid for the use of capital in any market. Here the word capital means money.

[7] Paul Gomme, B. Ravikumar, and Peter Rupert, "Secular Stagnation and Returns on Capital", Economic Synopses, No. 19, 2015.

[8] The implicit function is differentiable if both sides of the equation are differentiable. In this case, we need to assume that both $C(\cdot)$ and $I(\cdot)$ are differentiable.

[9] 金圆券

References

[1] Card, David and Krueger, Alan B., 1994, Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. American Economic Review, September, 84(4), 772-793.

[2] Liu, Huiyu, 1999, Research on China's Central Banking (in Chinese, 中国中央银行研究). Beijing: Economic Press China.

[3] Akerlof, George A., Yellen, Janet L., 1986, Efficiency Wage Models of the Labor Market. Cambridge: Cambridge University Press.

[4] Fisher, Irving, 1930, The Theory of Interest (as Determined by Impatience to Spend Income and Opportunity to Invest It). New York: The Macmillan Company. Available online at `https://fraser.stlouisfed.org/title/theory-interest-6255`.

[5] Marshall, A., 1920, Principles of Economics, London: Macmillan.