Chapter 2

A framework for macroeconomics

The purpose of this chapter is threefold. First, the chapter begins by reviewing major macroeconomic time series with a focus on their long-run behavior, primarily using U.S. data (we encourage the reader to examine the corresponding graphs for other countries).

The second, and key, purpose of the chapter is to gradually introduce, rather heuristically and with the formal details postponed until the ensuing chapters, the basic framework—the macroeconomic model—that will constitute the core tool in the textbook. Thus, each graph will be interpreted from the perspective of the proposed framework. An underlying assertion is that it is hard, if not impossible, to account for the data except with the kind of framework we use. This framework goes back to Solow's growth model and then builds in conscious choices, such as firms' choices of inputs, consumers' choices for saving and hours worked, later on the purposeful development of human capital and technology, and so on. The emphasis on explicit choices necessitates a microeconomic approach, not just in terms of theory but also in terms of the data we will look at; nowadays, much macroeconomic research directly studies cross-sectional data (for households, firms, etc.). There is no presumption that markets work perfectly. Rather, we emphasize identifying specific market imperfections, especially as they relate to macroeconomic policy. Finally, a key feature of the core framework is that it is *quantitative*: the aim is to formulate a model that can account for the magnitudes of macroeconomic phenomena and not just their qualitative features.

The third purpose of the chapter is to be a stepping-stone into the rest of the text. To that end, it provides a brief overview of the topics covered in later chapters.

2.1 The facts and interpretations: real aggregates

In this section we document some basic facts relevant for macroeconomic analysis. The facts are presented in a stylized manner; for example, the unemployment series will be described as "stationary" and this term should not be interpreted in a statistical sense but rather as a series that does not have a marked trend (toward, say, zero or one). Of course, the swings in the series will be pointed out, including rather persistent ones. The main facts we go over in this section, moreover, emphasize the longer run; short-run facts are discussed in more detail later. The growth facts we will focus most on are from the United States, but we will show some data from other countries as well. They are, for the most part, referred to as the Kaldor facts, but there is no strict adherence here to the facts originally pointed to in Kaldor (1957).

2.1.1 Output grows steadily

One of the most remarkable facts in economics is the steady growth of output over the last centuries. The path for (the logarithm of) real U.S. output is shown in Figure 2.1.[1] The figure reveals almost constant growth over more than a century and the swings up and down seem minor from a bird's-eye perspective.[2] The notable exception is the Great Depression episode and the rebound after that, but after that hiccup the economy lands on "the same" growth path again. As a result, typical business-cycle movements, including the most notable recent recession (the Great Recession, 2007-2009), are barely visible.

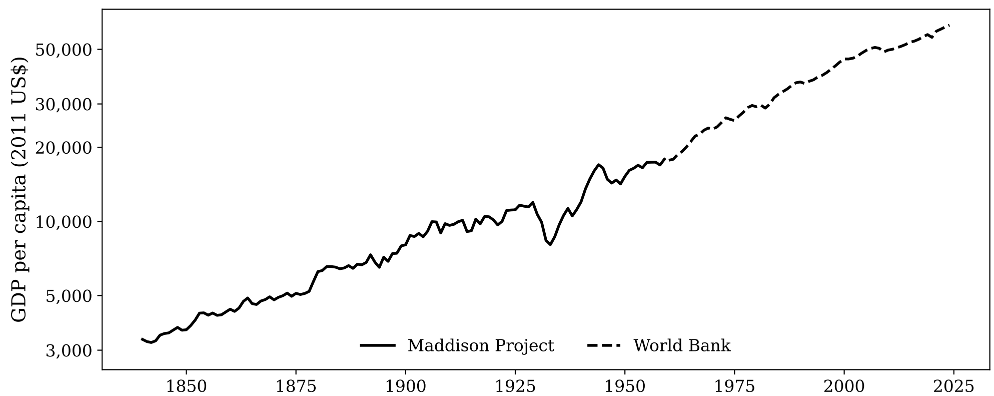

Figure 2.1: GDP per capita in the U.S.

Notes: The figure plots GDP per capita in 2011 prices in the U.S. 1840-2018.
Source: Maddison project.

One of our key goals now is to try to "account" for output growth, i.e., to provide a theory that offers a deeper understanding of the remarkable fact in Figure 2.1. We do so by focusing on the production side, i.e., how the basic inputs into production have evolved over time. We also look at their prices.

2.1.2 The basic resources behind output—and their prices

We begin by looking at capital.

Capital input

As we shall argue, the process of growth is driven, at least in part, by capital accumulation. Figure 2.2 shows the capital-output ratio in the U.S. since the late 1920s. Apart from a

[1]As discussed in the measurement section, systematic measurement did not begin until about a third into the twentieth century. The Maddison database goes much further back and then output estimates are based on available time series for, e.g., production, employment, and prices.

[2]A linear fit suggests that the series is well approximated by a annual growth rate of 1.86 percent. (Such a linear fit obtains an R^2 of more than 0.98.)

marked jaggedness early on—during the Great Depression especially—we also see clear stability at a value of around 3. The measure for capital here is the standard one: "accumulated investments, minus depreciation."

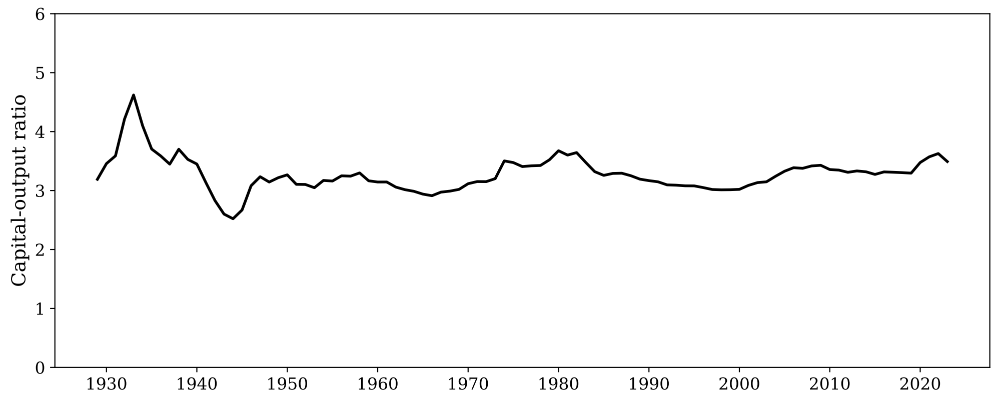

Figure 2.2: Capital-output ratio in the U.S., 1929-2022.

Source: FRED. Numerator: Current-Cost Net Stock of Fixed Assets and Consumer Durable Goods (K1WTOTL1ES000), Annual, Not Seasonally Adjusted, converted to billions of dollars. Denominator: Nominal GDP (GDPA), Annual, Not Seasonally Adjusted, reported in billions of dollars. The figure plots the ratio between fixed capital and consumer durables relative to the GDP.

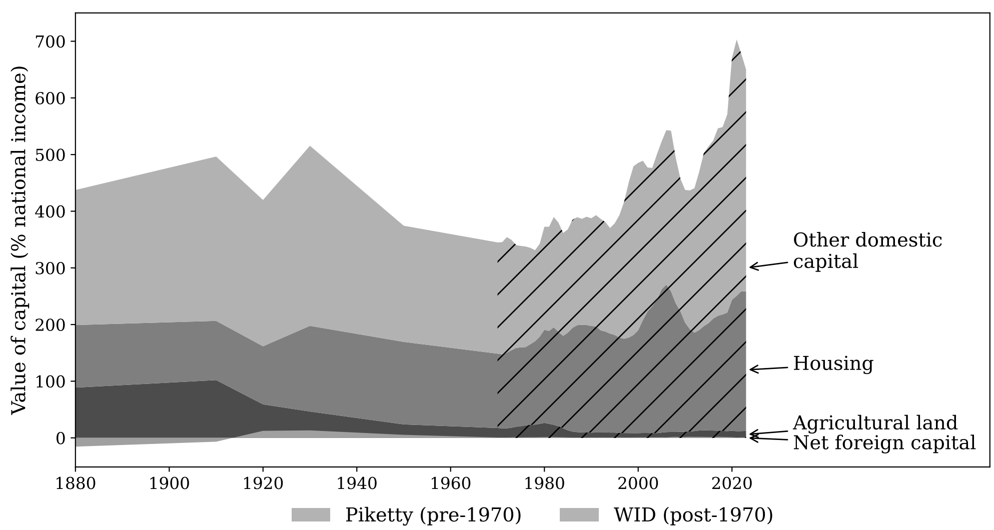

Figure 2.3: Wealth-output ratio in the U.S.

Source: Piketty (2014), Chapter 4, Figure 4.10[3].

The focus here is capital as an input into production. It is nevertheless interesting to note that a broader interpretation of capital is wealth, which would include the value of land, housing, and so on. In Figure 2.3, we show the wealth-output data as computed in

[3]Data can be downloaded from http://piketty.pse.ens.fr/files/capital21c/en/xls/.

Piketty (2014).[4] We see marked stability again, though large changes in the composition of the capital stock, toward manufacturing capital and, especially, housing, and a total that is 4–5 rather than 3.

The price paid for using capital

How expensive has it been for producers to use capital over time? Most commonly, firms buy capital and use it until they scrap it, or sell it in market for used capital goods. It has become increasingly common for firms to instead rent capital (such as machinery or buildings), in which case the price paid for the use of capital is clearly the rent. But due to lacking systematic historical data on rents, measures of the cost of capital are instead constructed based on (a minimum of) theory. One way is to look at a measure of the returns on investments: on the margin, under competitive markets, this return should equal the marginal cost of the investment: the price we are looking for. Thus, we can look at stock-market returns as one measure of capital's price. Using binned data on stock-market returns—a return to investments in the capital of firms whose shares are publicly traded—from three long time-periods, Figure 2.4 shows no strong trends. If one were to look at shorter time periods, the variations in the stock market returns are of course very noticeable and large. These fluctuations are likely due to changes in asset valuations more than to changes in costs, which is why longer-run averages seem more appropriate.

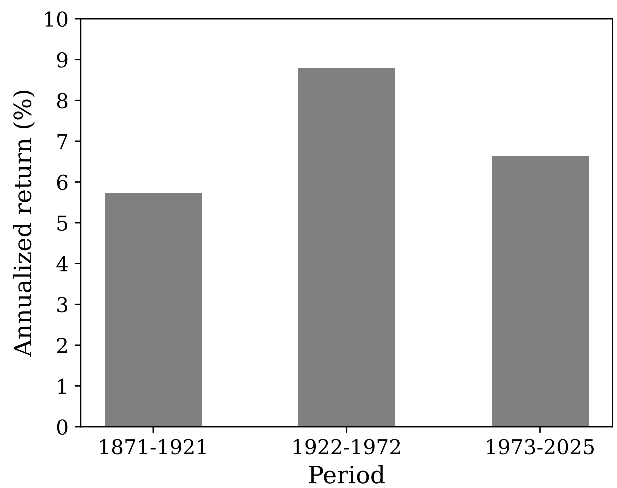

Figure 2.4: Return on capital.

Source: Shiller (2000). Data is annual, geometrically compounded returns to U.S. stock markets[5].

Alternatively, one can attempt to measure the cost side, but also using theory. The "user cost of capital" is based on the foregone return to saving: by buying and owning capital, a firm is losing the return it would have received by merely saving the money. This measure also takes into account depreciation: a part of the capital is lost by using it, or needs

[4]We use his data starting 1880.

[5]Data can be downloaded from http://www.econ.yale.edu/~shiller/data.htm.

maintenance to be kept in good shape. Moreover, it takes into account capital's change in value over time; computers, while not physically depreciating, lose value on the used market since new computers always out-compete old computers. The user cost is, roughly speaking, a market interest rate plus depreciation plus the fall in value.[6] The user cost is also stationary, but of course also includes short-run swings.

Labor input

The second major input into production is labor. Employment is one measure of input, but it is often relevant to take into account how many hours each employee works given that hours worked vary widely and many people have more than one job. Hence, a common measure of labor input is hours worked (in the marketplace) per adult. Various measures are available and we will show two here. First, Figure 2.5 shows that, since the beginning of the last century, hours worked per week have fallen, from around 28 hours to around 23 hours. Looking more closely at the graph, we see that since the end of World War II, hours look rather stable, without a net downward trend. This is a fact that is often referred to—that U.S. hours are stationary—but actually only accurate over the postwar period. Second, we see very large departures from trend in the figure; during the Great Depression, extreme unemployment rates account for the low hours, with a subsequent war-related upswing. In the U.S., unemployment movements, which are large, account for a big share of the movements in hours.

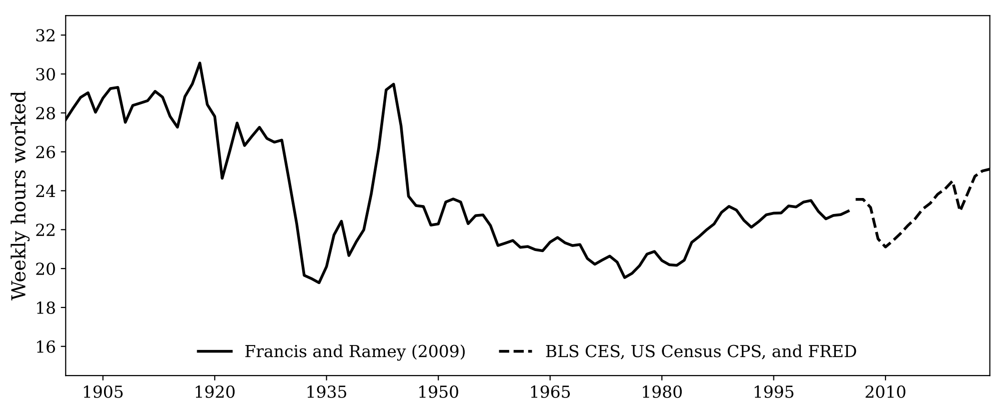

Figure 2.5: Average weekly hours worked in the U.S. (age 14+).

Source: Ramey and Francis (2009)[7].

Figure 2.6 shows hours worked over a longer time period, along with average real wages. Here, the downward trend in hours is even clearer (the graph depicts hours per employed, so the numbers are overall higher and do not account for changes in participation). The cumulative decline is almost 50%.

[6]Other factors can appear in a user-cost formula, such as the role taxation plays for capital income and in deductions for depreciation.

[7]Data can be downloaded from https://econweb.ucsd.edu/˜ vramey/research/Century_Public_Data.xls.

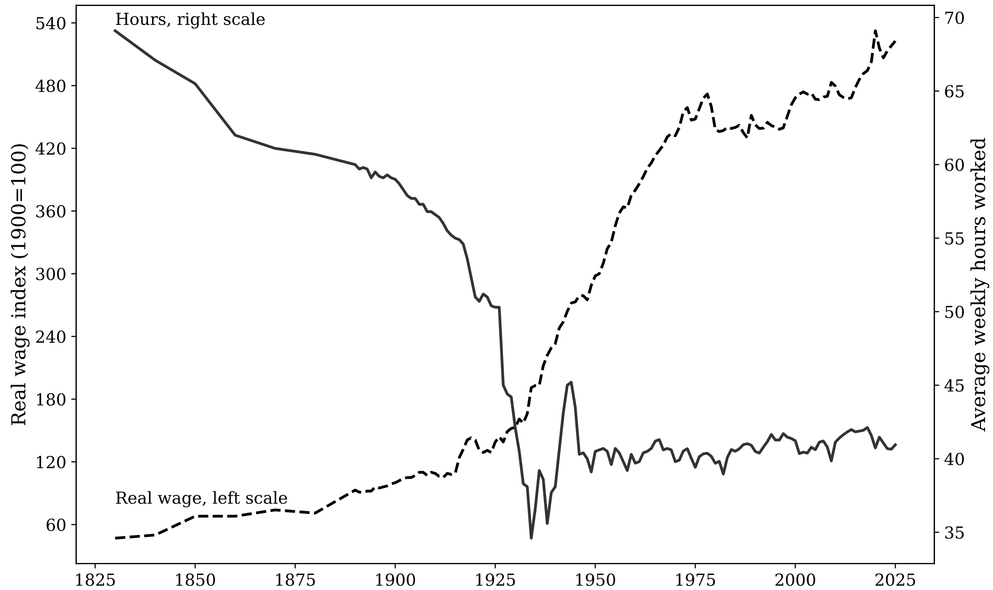

Figure 2.6: Average weekly hours worked in the U.S. (manufacturing).

Sources: Period 1830-1880: Whaples (1990), Table 2.1. Period 1890-1970: Bureau (1975) Series 765 and 803. Period 1970-2023: FRED, monthly series AWHMAN, annualized. Wage data source: Period 1830-1888: Williamson (1995) Table A1.1, (1900=100). Period 1890-2023: FRED, quarterly series LES1252881600Q, annualized (1982-84 CPI Adjusted Dollars) **Note**: Converted wage index in 1982-84 dollars from FRED to 1900=100 index by multiplying the FRED series by the ratio of wage index in 1983 from Williamson (1995) with wage index in 1983 from the FRED.

Wages

Figure 2.6 also shows real wages. Wages have risen at a remarkable average rate of around (or even above) the rate of output, with an exception in the most recent period. The secular increase is not surprising in some general sense: the standards of living have, slowly but surely, risen steadily for most Americans, and the major source of this rise comes from higher and higher earnings. Given that hours are not showing an upward trend—in fact, the opposite—it must be that real wages have risen steadily. There are also some movements in how total output is divided up into capital and labor income; we will discuss these later, but the first-order aspect here is that the shares have been quite stable.

2.1.3 Taking stock: a "neoclassical" picture emerges

The data on capital and output had puzzled economists; in particular, the constancy of the capital-output ratio at a value around 3 suggested something quite stark—it suggested a rigid technology structure where labor played no role and capital and output were always in the same proportions—and this was hard to square with how we knew that production took place in practice.

Solow (1956) however, took a less immediate production perspective on these facts. He noted what you saw in the previous two sections: there has been a steady rise in output, an equally steady rise in capital with stationary, or even declining, hours worked. At the same time, the price of capital has been stationary while wages have had a significant upward trend. These facts did not, by themselves, appear mysterious. Solow in particular noted that it is natural that as the price of an input rises, the use of the input declines. More precisely, from the perspective of production theory, at a higher relative input price, a firm uses less of that input relative to other inputs. Labor has become more and more expensive relative to capital, and its use has fallen, again relative to capital: the capital-labor ratio has risen at the rate of output growth, or even slightly more.

The other side of the coin, which was important to address, is about accounting for these changes in relative input prices: what made labor more and more expensive relative to capital? One natural explanation would have its roots in technological change. In particular, suppose it is directed toward labor: labor becomes more productive over time per unit of hour worked. If capital and labor are complementary, this factor would lead to an increased value of capital on the margin. However, as we saw, the market return to capital has remained stationary—it has not risen. Solow noted, however, that the stationarity could be explained by *neoclassical* forces. A production function that has decreasing marginal returns to each input is labeled neoclassical and with such a production function, as the relative amount of capital—capital used per worker—rises, the marginal value of capital would fall. If firms buy inputs in competitive input markets, moreover, this would be reflected directly in the market return to capital. A story emerges where technological change directed toward labor, along with neoclassical forces, can, at least potentially, account for the historical data on relative input quantities and relative input prices.

The neoclassical features led Solow to describe *aggregate* output as being generated by an *aggregate* production function, with *aggregate* capital and *aggregate* labor as inputs. Solow also posited, along the lines above, that the production function likely changed nature over time—that there was some technological progress—and built a model around these ideas; we will briefly describe this model momentarily but let us first focus on how, given Solow's perspective, one could take the next natural step: accounting for the sources of aggregate growth as coming from growth in inputs and growth in technology. This accounting procedure, along with the development of Solow's neoclassical growth model, would allow us to obtain a much less mysterious, and in fact quite natural and operational, account of how output grows over time. As a matter of microeconomic theory, the existence of an aggregate production function, i.e., a functional mapping from the total (economy-wide) quantities of inputs only into some measure of aggregate output, is not easy to establish in general and has also, as the box below discusses, been subject to some controversy.

The existence of an aggregate production function

There are assumptions under which a functional relationship can be established between input quantities and a price-independent output measure. However, these assumptions are extremely specific, indeed knife-edge, cases. Think of a static economy producing two goods, x_1 and x_2, both from capital and labor but with different production functions; also imagine that there are no restrictions on how inputs can be allocated across the two sectors. Then if a relative price between x_1 and x_2 is assumed—let us call it

p—it is straightforward to see that a competitive equilibrium will generate a mapping between the total input quantities and the value of total output: perfectly competitive markets would allocate, or a planner could equivalently allocate, capital and labor so as to maximize output. As the total amounts of capital and labor would vary, total output would change. Thus, we would obtain a mapping from inputs to output. But this mapping would nontrivially involve p: it would not be a pure production-function relationship. Hence, one would need to add a demand side, endogenizing p, to obtain a pure mapping from input quantities to a measure of output. But then preferences (or whatever gives rise to demand) need to be described, and they will generally, as would p in a direct sense, influence the mapping. Thus, a pure production mapping is hard to imagine.[a]

The existence and usefulness of an aggregate production function was hotly debated in the so-called Cambridge capital controversy during the 1950s. This controversy, which had its two head quarters in the two well-known Cambridges (the University of Cambridge, U.K., chiefly represented by Joan Robinson and Piero Sraffa, vs. Paul Samuelson and Robert Solow at MIT, Cambridge, Mass., U.S.), also involved the notion of "aggregate capital," but in essence it focused on the aggregate production function. The controversy on the existence conditions can be summarized as having been won by Cambridge, England, whereas the usefulness controversy was arguably won by Cambridge, Mass. With the tools of modern macroeconomics, one can solve large models and see to what extent departures from aggregation play an important role quantitatively. Some such endeavors have been undertaken and point to limited departures, but no fully systematic analyses have been conducted yet.

[a] If you assume that the two production functions have identical isoquants—one is a scalar multiplication of the other—then it is possible to construct the mapping, as a relative price p is implied from the production technologies alone. Try to show this as an exercise!

2.1.4 Growth accounting

Solow (1956) introduced "growth accounting" as a way to implement the notion that one could break down aggregate output growth into sub-components. With the use of a rather limited amount of theory, Solow was able to quantify the relative importance of different sources of U.S. growth.

Solow's growth accounting made the following assumptions: (i) aggregate output Y_t is generated from an aggregate production function $F_t(K_t, L_t)$, where K_t is aggregate capital input, L_t aggregate hours worked, and the subscript t denotes that technological change may move the production function upwards over time; (ii) F has constant returns to scale (CRS) and neoclassical properties; (iii) there is perfect competition for inputs and, hence, firms maximize profits. The CRS assumption, familiar from microeconomics, implies zero pure profits in a perfectly competitive equilibrium. Solow viewed this as a reasonable approximation. Moreover, CRS ensures that doubling all inputs leads to a doubling of output, making it more plausible than assuming decreasing returns at the aggregate level.

Throughout the text (when not otherwise noted), we will use the convention that lower-case letters are in per-capita real terms. Thus, in this chapter we use y_t to denote per-capita output, i.e., Y_t divided by the size of the population; denoting population by N_t, we have

$y_t = Y_t/N_t$. Note that, since F_t is CRS, we can write $y_t = F_t(k_t, \ell_t)$, where k and ℓ are capital and hours worked per capita terms. Most of the time in this chapter, beginning here and now, we will also abstract from population growth and simply consider a population of constant size ($N_t = N$). It is sometimes convenient to normalize the population size to 1 so that we have $y_t = Y_t$.

We will now derive Solow's growth-accounting equation and for this we will use a continuous-time formulation where all variables are functions of time: we will write $y(t) = F(k(t), \ell(t), t)$. We will assume differentiability of these functions of time, so that

$$dy = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial k}dk + \frac{\partial F}{\partial \ell}d\ell = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial k}k\frac{dk}{k} + \frac{\partial F}{\partial \ell}\ell\frac{d\ell}{\ell},$$

where all the partials are evaluated at $(t, k(t), \ell(t)$. Dividing by output, using r and w to denote capital's rental rate and the wage (the prices, as well as the quantities, also depend on time but this dependence is omitted for notational convenience), and then using firm profit maximization, we obtain

$$\frac{dy}{y} = \frac{\partial F}{\partial t}\frac{dt}{y} + \frac{rk}{y}\frac{dk}{k} + \frac{w\ell}{y}\frac{d\ell}{\ell}.$$

Here, r and w (both with time dependence suppressed) replaced the two marginal products: this is what follows from taking first-order conditions of the profit maximization problem

$$\max_{k,\ell} F(k, \ell, t) - r(t)k - w(t)\ell.$$

We finally let $1 - \alpha$ denote labor's share of income, which we know from NIPA; α here of course may also depend on time. Then the CRS assumption gives us the capital share as α: total input costs equal output, delivering zero profits. This equation allows us to obtain an account of the sources of growth:

$$\frac{dy}{y} = \frac{\partial F}{\partial t}\frac{dt}{F} + \alpha\frac{dk}{k} + (1-\alpha)\frac{d\ell}{\ell}. \tag{2.1}$$

The term $(\partial F/\partial t)(dt/F)$ is labeled the *Solow residual*, because it can be calculated as a residual of the growth in output that cannot be accounted for by the growth in capital and labor (the second and the third term of the right-hand side). We see that output growth is a weighted sum (over small intervals in time, as we have used derivatives) of capital input growth and labor input growth, where the weights are their respective income, or cost, shares, plus the Solow residual, which is a measure of how the production possibility frontier has moved out over time.

The Solow residual in (2.1) expresses the direct effect of time on production (in percentage terms). In general, this effect depends on the input pair (k, ℓ) at which F is evaluated. If one makes further assumptions on how technology shifts the production function, it is however possible to derive specific series for technological change that are independent of the economy's current capital-labor mix. We will discuss two such assumptions because they are commonly used; they are by no means the only ones imaginable, but especially the second one will play a key role later.

One assumption is to let $F(k, \ell, t)$ take the form $zF(k, \ell)$, where F is a time-independent function and only z depends on time. In this case, equation (2.1) can be expressed as

$$\frac{dy}{y} = \frac{dz}{z} + \alpha\frac{dk}{k} + (1-\alpha)\frac{d\ell}{\ell}. \tag{2.2}$$

Here, z is TFP: total factor productivity. It can equivalently be thought of as a common, "Hicks-neutral" factor multiplying both inputs.[8]

Alternatively, we can make the assumption that technology is *labor-augmenting*: we define $F(k, \ell, t) = F(k, z\ell)$, again with a time-independent function F. Now z has a different meaning. In this case, since $(\partial F/\partial t)dt = (\partial F/\partial \ell)\ell dz$, we can write the original growth-accounting equation (2.1) as

$$\frac{dy}{y} = (1-\alpha)\frac{dz}{z} + \alpha\frac{dk}{k} + (1-\alpha)\frac{d\ell}{\ell}, \tag{2.3}$$

after having replaced $\partial F/\partial \ell$ by w/z and recognized that $w\ell = (1-\alpha)F$.[9]

We now show some results from growth accounting, implemented the way Solow came up with it. In practice, it is important to take into account how the quality of the inputs change over time; in particular, one may want to adjust labor input to account for human capital accumulation, or else part of the Solow residual will reflect the increasing quality of this input. Also, since growth accounting in practice is carried out using data from discrete time periods, like years, one must be clear on which input shares to use, e.g., taking an average over the shares in t and $t+1$ when accounting for growth between these two years.

We first look at the traditional measure of productivity: labor productivity. Figure 2.7 shows an average growth rate of a little below two percent per year, with significant ups and downs; currently, we are in a down period. Figure 2.8 shows a full time series (along with those for some other countries) over a longer time period. These series are smoothed.

We thus see stable, positive labor productivity growth, hovering around two and a half percent per year. How about the growth in total factor productivity? Figure 2.9 gives the answer from two sources, now in un-smoothed form. Here too, we see growth at a little below two percent per year, with significant movements up and down that we will return to later. Figure 2.10 shows smoothed data for TFP growth over a longer time period. The patterns are similar.

2.1.5 The dynamic system

We are now equipped with measures of output and input aggregates, as well as with a measure of aggregate technology, in the form of TFP. Solow's next step was to use his framework to probe further into the mystery of the all but constant capital-output ratio.

Step one in this endeavor was to explicitly link time periods (years) by noting that tomorrow's capital aggregate stock is today's stock, plus new investment minus the part of

[8] This is true since we have assumed that F is homogeneous of degree 1 in (k, ℓ).

[9] Similarly, one could define a capital-augmenting technology series by letting the z multiply capital.

[10] Gordon defines "labor productivity as real GDP divided by an unpublished quarterly BLS series on hours for the total economy, including the private economy, government, and institutions."

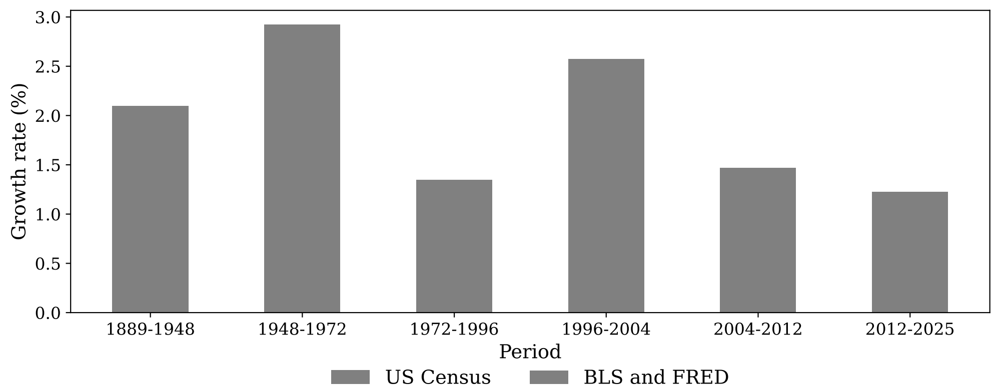

Figure 2.7: Labor productivity in the U.S., sub-periods.

Source: Period 1889-1948: Bureau (1975), Part II, Series W 1, pp. 948.. Period 1948-2023: U.S. Bureau of Labor Statistics and FRED Quarterly total economy hours worked (in billions of hours) series are from BLS "Hours Worked in Total U.S. Economy and Subsectors." Quarterly real GDP (in billions of 2017 dollars) is from FRED Real gross domestic product (GDPC1).
Note 1: Data constructed following Gordon (2012)[10] Percentage logarithmic growth rates are calculated between the first quarter of each of the listed years, e.g., 1948:Q1 to 1972:Q1. To extend the series back from 1948 to 1891, annual NIPA data on real GDP prior to 1929 are ratio-linked to the real GDP data of Balke and Gordon (1989), and the BLS hours data prior to 1948 are ratio-linked to the man-hours data of Kendrick (1961) (see pp. 330-32)." Unlike Gordon (2012), to extend the series back from 1948, we used Census Bureau Hisitorical Statistics. Then the series for both periods are then re-indexed to 1948=100.

capital that has depreciated.[11] With a constant rate of capital depreciation δ, this yields

$$k_{t+1} = (1-\delta)k_t + i_t. \tag{2.4}$$

We see an equation that is consistent with the long-run data, if all the variables appearing in it grow at a common rate. We can rewrite the above equation as

$$\frac{k_{t+1}}{y_{t+1}}\frac{y_{t+1}}{y_t} = (1-\delta)\frac{k_t}{y_t} + \frac{i_t}{y_t}$$

and denoting the (net) growth rate of output as γ_t, this equation can be expressed as

$$\frac{k_{t+1}}{y_{t+1}}(1+\gamma_t) = (1-\delta)\frac{k_t}{y_t} + \frac{i_t}{y_t}.$$

Clearly, if capital, investment, and output all grow at a same constant rate γ and i_t/y_t is equal to a constant value s, then this equation is consistent with capital-output ratio that does not change over time:

$$\frac{k_t}{y_t} = \frac{s}{\gamma+\delta}$$

[11]A similar accumulation equation can be formulated for human capital. We delay this discussion until our Chapter 13 on growth.

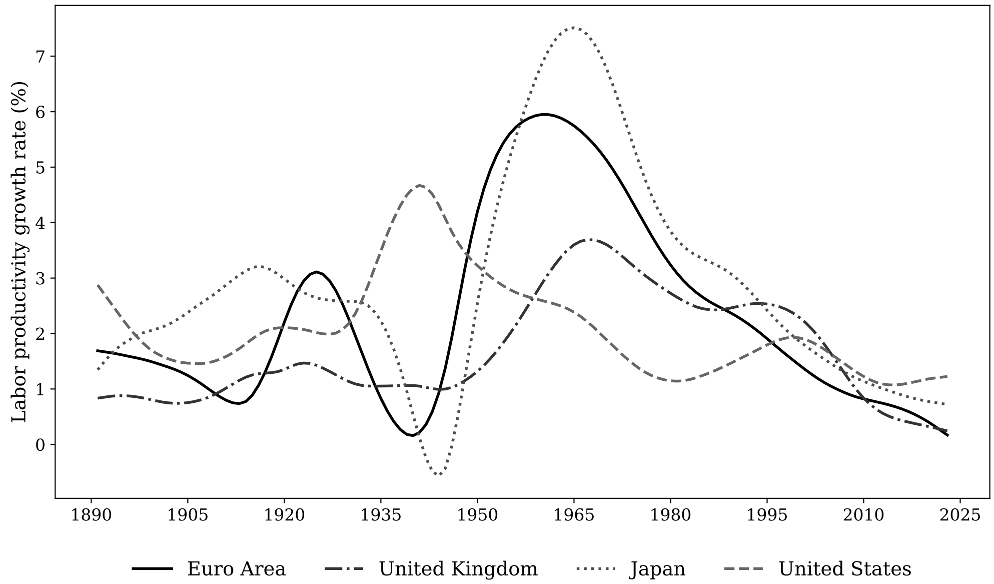

Figure 2.8: Labor productivity for a selection of countries.

Note: Hodrick-Prescott-filtered annual growth of labor productivity per hours worked. Following Bergeaud et al. (2016), we focus on 30-year cycles, which implies an HP-filter value of 500 for lambda.

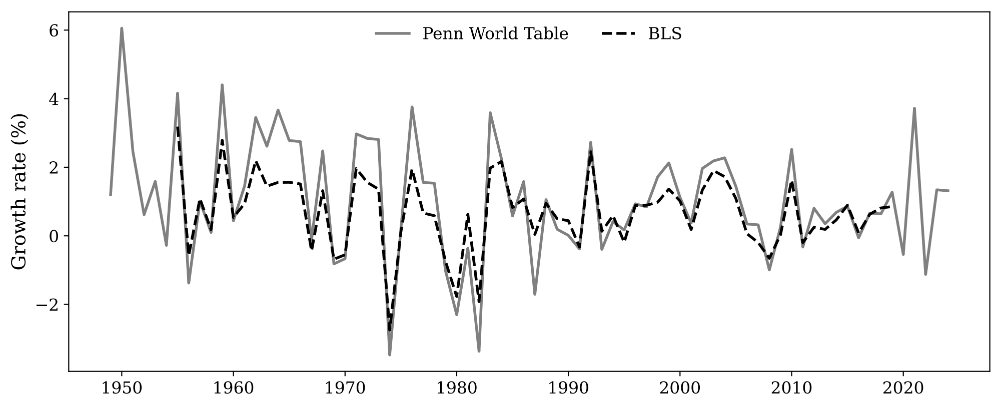

Figure 2.9: TFP in the U.S., two measures.

Sources: Series 1: FRED, "Total Factor Productivity at Constant National Prices for United States (RTFPNAUSA632NRUG)," reported by Penn World Tables. Available for 1955-2019 Series 2: Utilization-adjusted quarterly TFP series for the U.S. Business Sector, 1948-2022 from Fernald (2012).

for all t. The investment-output ratio is not constant over time; in particular, it fluctuates significantly. The consumption-output ratio for the U.S. is plotted in Figure 2.11; here, consumption is defined as private plus government. One minus this measure is close to the ratio of investment (again, private plus government) to output, since net exports are near zero in the U.S. as a fraction of GDP.[12] We see significant movements in s early in the 20th

[12]This comes from the national accounting identity $Y = C + I + NX$, where C and I both contain private

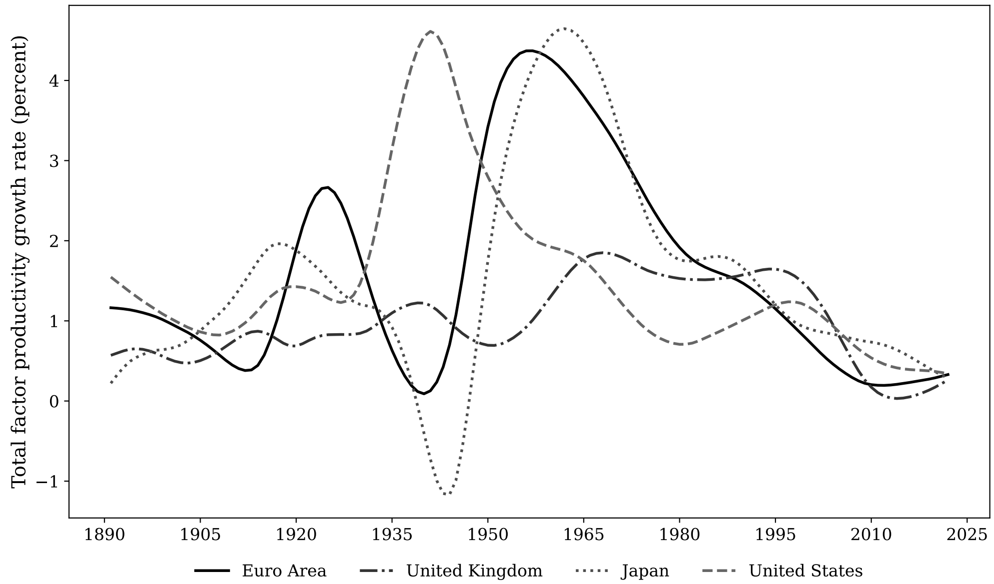

Figure 2.10: Historical TFP for a broader set of countries.

Source: Following Bergeaud et al. (2016), we focus on 30-year cycles, which implies an HP-filter value of 500 for lambda. TFP growth rate for years after 2012 is from the OECD data series Multifactor productivity. Aggregate TFP growth rate for EU19 countries is calculated by taking a weighted average of the growth rates of each country where weights are the share of each country in the total GDP of the EU19 in each year. Data obtained from OECD Gross domestic product (GDP).

century and thereafter small movements, possibly with a slight downward trend. But the overall assumption of a constant investment, or saving, rate appears to be a good one.

The above discussion allows a mechanical account of how one can interpret the capital-labor ratio: it is a simple function of the rate of saving, the economy's growth rate, and the rate of depreciation of capital, along what has been labeled the *balanced growth path*. The numerical values can be squared, too: with a depreciation rate of around 0.08 and a net growth rate of around 0.02, a saving rate of 0.30 delivers a capital-output ratio of 3.[13]

However, this account still does not give an answer to the question why: why is the economy always (almost) at this value? That is, it explains if the capital-output ratio is 3 at some point in time, it will remain 3. But why is it 3 to start with? Solow found an answer.

Solow considered the following dynamic system, which is the logical implication of the above reasoning:

$$k_{t+1} = sF(k_t, (1+\gamma)^t \ell) + (1-\delta)k_t$$

for all t. This system is almost exactly what we have looked at before. First, we have replaced i by sy, since we take saving to be a constant fraction of output, in consistency with the above evidence. Second, there are two additional assumptions: we have set hours worked to a constant, ℓ, and we have made technology growth appear only in a *labor-augmenting*

as well as government expenditures.

[13] The saving rate of 30% is larger than what is implied in Figure 2.11. This discrepancy is largely because Figure 2.11 includes the government spending in consumption.

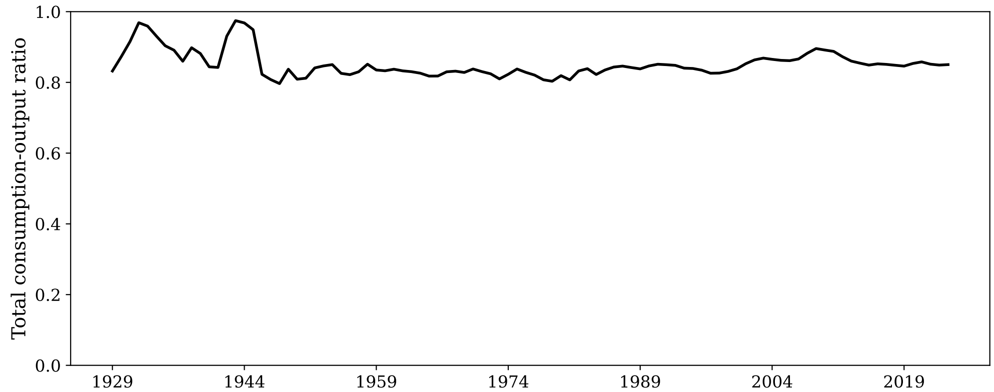

Figure 2.11: Ratio of total consumption to output.

Source: BEA, NIPA Table 1.1.5 Calculated as the ratio of two series: Numerator: Sum of Personal consumption expenditures (DPCERC) and Government consumption expenditures and gross investment (A822RC), Annual, Millions of dollars Denominator: Gross domestic product (A191RC), Annual, Millions of dollars The figure plots the ratio between total consumption expenditures (sum of all goods and services) relative to the GDP.

form, i.e., technical change is equivalent to raising labor input, or the quality of labor input. This was based on the hunch above: if labor units become more and more productive due to technological change, the return to capital does not have to fall due to a higher ratio of capital to hours worked. It turns out that this assumption is actually crucial. Namely, Uzawa (1961) proved that, for a production function of two inputs where one is limited—labor hours, in this case, are constant—to admit exact balanced growth in a model of this kind, technological change has to take this form.[14] So we can obtain balanced growth if and only if the above assumptions are met; whether labor input is constant or declining does not matter.

The dynamic system can be rewritten

$$(1+\gamma)\tilde{k}_{t+1} = sF(\tilde{k}_t, \ell) + (1-\delta)\tilde{k}_t;$$

the equation is obtained by means of a simple variable transformation—$\tilde{k}_t \equiv k_t/(1+\gamma)^t$, a stationary variable if k grows at the net rate γ—and division by $(1+\gamma)^t$.

Now note, first, that there is a constant solution to this system: $\bar{\tilde{k}}$. It is the unique value that (under some minimal conditions) solves

$$(1+\gamma)\bar{\tilde{k}} = sF(\bar{\tilde{k}}, \ell) + (1-\delta)\bar{\tilde{k}}.$$

So in a situation where $\tilde{k}_0 = \bar{\tilde{k}}$, we obtain $\tilde{k}_1 = \tilde{k}_2 = \cdots = \bar{\tilde{k}}$. That is, capital grows at a constant rate γ, since $k_t = \tilde{k}_t(1+\gamma)^t$. The same is then true for investment and output, since $y_t = F(k_t, (1+\gamma)^t\ell) = (1+\gamma)^t F(\bar{\tilde{k}}, \ell)$ follows when F is CRS.

We have established that if the initial capital stock has a particular value, this economy will grow at a constant rate. However, the second, and most remarkable, thing to note about

[14] The proof of Uzawa's important theorem can be found in Appendix 3.A, which discusses the Solow model in detail.

this system is a convergence property. Solow showed, under very weak conditions, that for any given initial condition on capital, the ensuing capital sequence, and hence the economy's aggregate variables will *converge* to the balanced growth path. We will explain this in detail in the next chapter. Intuitively, it is the neoclassical production function—the very feature that was used to make sense of why a higher input price is consistent with lower input use, and at the same time, under competition, a higher marginal productivity—that explains convergence: when capital is comparatively low, growth is comparatively fast, because its marginal productivity is high, delivering higher capital accumulation per unit of capital. Conversely, when capital is high, its growth rate is low, so there is movement back toward the balanced growth rate no matter where the initial capital stock is.

The neoclassical convergence mechanism de-mystifies why the value of the total capital stock on average is worth roughly 3 times annual output: it doesn't have to be exactly 3 at all times, but deviations bring it back toward 3. It also produces a dynamic framework that, as we shall see in the rest of the textbook, has become the workhorse model for macroeconomics, much because it offers a coherent account of how the macroeconomic aggregates have evolved historically. The analysis of business cycles, for example, builds on the neoclassical framework with various stochastic shocks added to the dynamic system. These shocks could be "supply shocks" or "demand shocks," but they have in common that their *propagation* through the economy—how the macroeconomic variables respond in the short run, which can differ greatly across different kinds of shocks—eventually takes us back toward the balanced growth path. This is ensured by the convergence property of the system.

The final, unresolved, issue is that some of our assumptions above are mere mechanical descriptions of the data: investment is a constant rate s of output, and hours worked are constant at ℓ. In the real economy, these two features should be the results of conscious choices made by households. Consumption choices are made continuously, and people can influence how much they work. Moreover, a theory that adds consumer choice will also allow us to make welfare statements, which the analysis so far does not. We will turn to households' choices momentarily: we will "rationalize" the s and the ℓ based on microeconomic theory—utility maximization. However, let us first briefly revisit the movements of input shares.

2.1.6 Input shares

On a balanced growth path, the shares of income paid to capital and to workers, respectively, are constant. This is because in the theory Solow proposed, rk grows at the rate of output (r is constant but k grows at the rate of output) and $w\ell$ does as well (w grows at the rate of output and ℓ is constant, at least over the postwar period). Therefore, as shares of output, they should be constant. Have they been? We see in Figure 2.12 that there is rather remarkable constancy over time.

In the most recent years, a downward trend of the labor share can be noticed, however, as is evident if the time interval is restricted. Figure 2.13 illustrates, for a few developed countries, that the labor share has been declining. Figure 2.14 shows the same fact as a global average. The downward trend has been subject to much scrutiny and research recently, but for now we will maintain the stylized fact as "the labor share is close to constant over time."

Virtually all applied macroeconomic studies employ an aggregate production function that is of the Cobb-Douglas variety, i.e., we have $F(k, \ell) = Ak^{\alpha}\ell^{1-\alpha}$, where α is constant

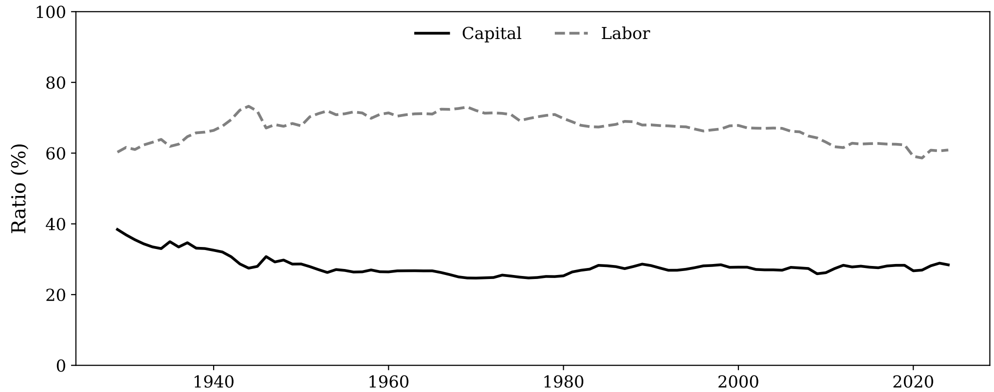

Figure 2.12: U.S. factor shares over time.

Source: NIPA Table 2.1 Factor shares are calculated as compensation of employees and capital income divided by personal income. Capital income is calculated as proprietors income with inventory valuation and capital consumption adjustments plus rental income of persons with capital consumption adjustment plus personal income receipts on assets.

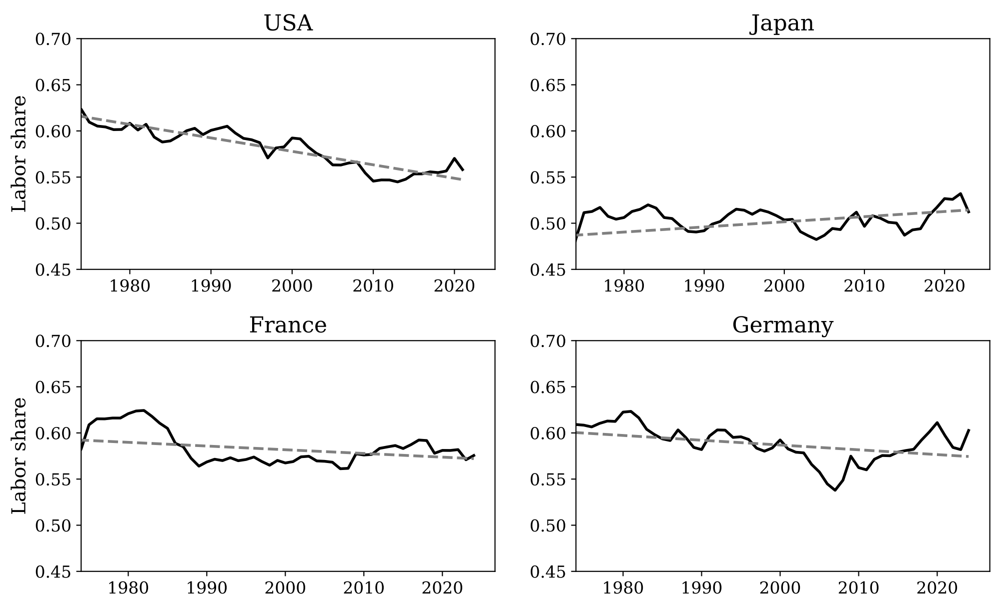

Figure 2.13: Labor share.

Source: Karabarbounis and Neiman (2014). **Note**: No corporate labor share data is available for Japan, thus total labor share is plotted instead.

over time. This function has the property that $F_k(k,\ell)k/F(k,\ell) = \alpha$, i.e., the income shares under perfect competition are independent of the values of k and ℓ. We see in the data that although the shares are not literally constant, their movements are relatively minor, even as economies go through recessions and booms. The Cobb-Douglas function conveniently generates a decent approximation to the data, which is why it is so often used.

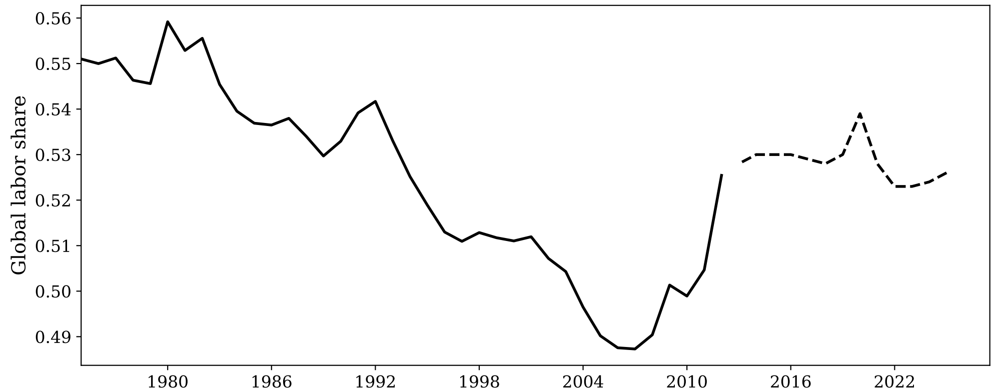

Figure 2.14: Global labor share.

Source: Solid line (1975-2012): Karabarbounis and Neiman (2014). Global labor share is the average of country-level labor shares weighted by GDP measured in US dollars at market exchange rates. Dashed line (2013-2025): International Labour Organization (ILO), Sustainable Development Goals Indicator 10.4.1, world.

2.1.7 Summing up

We summarize the main stylized (notice the repeated appearance of the word "roughly" in the descriptions below) facts:

1. output per capita has grown at a roughly constant rate
2. the capital-output ratio (where capital is measured using the perpetual inventory method based on past consumption foregone) has remained roughly constant
3. consumption as a fraction of output has been roughly constant
4. the wage rate has grown at a roughly constant rate equal to the growth rate of output
5. the real interest rate has been roughly constant, seen over a longer period of time
6. labor income as a share of output has remained roughly constant
7. hours worked per capita have been roughly constant over the recent half a century.

These facts are consistent with aggregates obeying a neoclassical structure whose core is a CRS production function with labor-augmenting technical change and decreasing marginal products in each input, constant labor supply, a constant rate of capital depreciation, and a constant investment-output (saving) ratio.

2.1.8 Rationalizing saving and labor-supply choices

We now discuss how macroeconomists theorize further to make sense of households' observed choices for saving and labor supply. The aim is to add a richer microeconomic structure that is qualitatively, and quantitatively, consistent with the data. The central object here will be utility functions and a key question is whether there are utility functions consistent with the observed choices. A follow-up question is whether, given such utility functions, the main dynamic properties—in particular convergence—will be maintained. We briefly address the first of these questions in the present chapter and postpone the second until a later chapter.

Before we introduce utility functions, we will briefly discuss the population structure, along with our general approach.

Time and people

We begin with time and then describe our people.

Time The dynamic system above was described in discrete time, i.e., time periods are integers. It is also possible to describe the system in continuous time (here, we would use t as an argument of our functions and not as a subscript, e.g., $y(t)$ vs. y_t). The main workhorse model developed in this book is using discrete time, mostly because we consider it somewhat easier to teach.[15] There is, however, no substantive difference between the two approaches and both are common in practice.

Second, we will almost always assume that time is infinite; the main exception is when we illustrate mechanisms in, say, two-period models. The reason for adopting infinite time is mostly practical but to some extent also conceptual. The same economic structures as we study under the infinite-horizon assumption can also be analyzed under the assumption that time is finite. However, then there is a built-in non-stationarity: time itself obtains importance—because it captures how far we are from "the end" and, certainly, if the end would be near, many decisions in the economy would change radically. So long as decisions and outcomes today are not more than marginally affected by the exact end date, it is simply more convenient to consider time to be infinite; then, time is not important in itself. Most macroeconomic models with an end date T have the feature that T is virtually irrelevant for decisions far from T (i.e., decisions at $t \ll T$); when the models we look at do not have this property, we will point it out.

People At any point in time, the economy is inhabited by individuals, or households, and our maintained assumption will be that they are utility maximizers. By assuming utility maximizers we mean we "rationalize" the choices we observe in the data by treating these choices as optimal given well-defined utility maximization problems. We back out how consumers must value consumption (and leisure) based on their own behavior. This approach is powerful, since it allows us to make statements that directly reflect the welfare of the population. In the analysis of economic policy, being able to make welfare statements is crucial, since it allows us to compare policies on normative grounds.[16]

[15] The reason is that it is easier to make concrete; for example, dynamic optimization can be more straightforwardly connected to basic optimization theory in one or more (a finite number of) variables; when there are stochastic shocks, continuous time requires more investment still, and becomes a bit more abstract as well.

[16] There is a strand of macroeconomic literature that builds in behavioral elements; after all, the literal interpretation of the data that consumers make perfect decisions at all time is of course very strong. One interpretation of this research is as a robustness check: if minor departures from rationality create major changes in the results, we should perhaps worry. This research is ongoing and we have no systematic treatment of it in this text.

The most common macroeconomic model has a very stark, and highly stylized, description of the population: there are many identical individuals alive at any one point in time, and these individuals, moreover, live forever. "Identical" here means that they have the same utility functions and constraints and, therefore, face identical maximization problems. "Many" is important in that it allows us to make sense of price-taking behavior, but the number itself does not matter. "Live forever" sounds much sillier than it is: the notion here is that people are representatives in a dynasty, i.e., a consumer at time t values not only her own consumption but also that of her children, grandchildren, and so on. Given how people appear to care about their children and grand-children—given the resources they spend on them while alive and also in the form of bequests—it does seem a very natural starting point. It should be pointed out that the assumption here is not just that people care about their offspring: they are truly altruistic about their offspring. In other words, their preferences are aligned. We will touch on the possibility that they are not later on in the text.

Of course, macroeconomic models with more explicit and realistic structures are also common. First, building in a life-cycle pattern—since, at the very least, as a person ages, needs and abilities change—is common.[17] Then, the addition of children and altruism toward them would still deliver a model that in its macroeconomic features is quite similar to the simpler model. If life-cycle models do not have altruism toward children, such as in the simple "overlapping-generations model," they can (but do not necessarily) behave differently than than dynastic models. We will discuss this in some detail too.

Second, nowadays a "heterogeneous-agent" framework has been developed and become a very commonly used workhorse for macroeconomic analysis. In it, households are different in a variety of ways (income, wealth, preferences, etc.). Such models share many properties with the simplest dynastic setting, but of course also add richness. One aspect of these models is that they allow us to jointly study macroeconomics and inequality. Another one is that they behave quite differently (and, arguably, in ways more closely aligned with the data) in some respects, in particular in terms of how the economy responds to various shocks and to policy changes. From our perspective here, however, the key observation is that heterogeneous-agent models can be viewed as extensions of the dynastic representative-agent setting rather than as fundamentally different. Third, some macroeconomic models also have richer models of the household structure, explicitly incorporating couples and children. This is, so far, a smaller literature, however.

Preferences

Households will be assumed to have utility functions that are time-additive, with consumption in periods t and $t+1$ evaluated as $u(c_t) + \beta u(c_{t+1})$, where u is strictly increasing and strictly concave. The same function u is used for consumption in both periods, but there is a weight β on $t+1$ consumption. The fact that u is the same for both consumption goods implies that consumption in both periods are normal goods: with more income, consumers would like to consume more of both goods, which seems very reasonable in this application.[18] Another aspect of this setting is an element of *consumption smoothing*: there is decreasing

[17] Death can then be modeled as deterministic or stochastic.

[18] Try to verify this by showing that, if the income allocated to the two goods is increased and consumers can choose between the goods freely, given a fixed relative price, both consumption levels will rise.

marginal utility to consumption in both periods so spending all of an income increase in one period will in general never be optimal. Furthermore, $\beta < 1$ captures impatience, or a probability of death—or any other reason for down-weighting future utility—and is a typical assumption.

Choice

An important part of the text will explain intertemporal choice from first principles: different methods for solving intertemporal problems, with and without uncertainty, along with a number of important macroeconomic applications. Here, the purpose is to very briefly explain the key steps, heuristically, in order to account for the growth facts.

Conceptually, the way consumers make decisions—if able to choose when to consume their income—is according to basic microeconomic principles: so as to set their marginal rate of substitution equal to the relative price. We will now go through the two key choices using these principles. Before looking at the specific choice examples, let us note that by rationality, in the present context, we include the notion of *perfect foresight*: given that no shocks are occurring, consumers know what prices prevail not only today but also in the future. If there are shocks—and we will study shocks later in the text—rationality is interpreted as *rational expectations*, i.e., knowing the probability distribution for variables in the future.

Consumption vs. saving The relative price between consumption at t and $t+1$ is the *real interest rate*: it is the amount of goods at $t+1$ that a consumer can buy for one unit of the good at t. We will denote the gross real interest rate between t and $t+1$ R_{t+1} here. The marginal rate of substitution between the goods can be obtained by defining an indifference curve relating to these two goods. Thus, write $u(c_t) + \beta u(c_{t+1}) = \bar{u}$, take total differentials, i.e., $u'(c_t)dc_t + \beta u'(c_{t+1})dc_{t+1} = 0$, and then solve for $-dc_{t+1}/dc_t$. Setting the resulting expression equal to the gross real interest rate, we obtain

$$\frac{u'(c_t)}{\beta u'(c_{t+1})} = R_{t+1}.$$

This equation, which equivalently can be written

$$u'(c_t) = \beta u'(c_{t+1})R_{t+1},$$

is commonly referred to as the *Euler equation* and it is a central element of macroeconomic theory. It says that an optimizing consumer sets the marginal utility loss of saving one consumption unit for tomorrow (the left-hand side) equal to the gain tomorrow in consumption terms (the right-hand side), that is, R_{t+1} (the return on the savings) times the marginal utility of each unit of tomorrow's consumption, $\beta u'(c_{t+1})$.

We argued above that a constant saving rate will imply convergence toward a constant level of capital relative to technology, i.e., $\hat{k}_t$ becomes constant—this is implied by Solow's analysis, which we will elaborate more on later. In particular, a constant saving rate is associated with aggregate consumption growing at a constant rate. We also saw that balanced growth requires a constant real interest rate. So the question now is whether individuals, when faced with a constant interest rate, would choose a consumption path that grows at a constant rate, despite the desires to smooth consumption over time. The question boils down to whether the Euler equation could hold for constantly growing consumption, and we now address this question.

To preview the conclusion: a sharp characterization exists showing that balanced growth arises in equilibrium if and only if the utility function u is a power function. It is easy to verify the "if" part: $u'(c_t)/u'((1+\gamma)c_t)$ becomes constant if $u(c)$ is a power function, and that constant contains the growth rate. Hence, under a power utility function a constant interest rate will lead the consumer to choose a constant consumption growth rate. What that growth rate is precisely depends on the interest rate, on β, and on the curvature of u, but not on how wealthy the consumer is. The precise class of functions is captured by

$$u(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma}, \tag{2.5}$$

with $\sigma > 0$ and $\sigma \neq 1$; the case where σ approaches 1 yields $u(c) = \log c$.[19] The "only if" result is harder to prove, but within reach; the proof is in the appendix to Chapter 4.

The above discussion has been carried out heuristically and in terms of simply selecting two adjacent time periods, without reference to how many periods the consumer lives in total. This discussion also means that the arguments above hold whether in dynastic or overlapping-generation economies, or some combination of these, so that the restrictions placed on preferences in order to be consistent with balanced growth hold rather generally.

Note that the utility-function characterization came from a requirement of exact balanced growth. One could imagine growth paths that are asymptotically balanced and still match our historical data. In the limit, however, the underlying utility functions would then look like our u defined in equation (2.5).[20]

Let us summarize: in macroeconomic modeling, where consumption-saving choices are viewed to come from optimizing consumers, the utility function employed in almost all applications is a power function. This choice is made because we want our frameworks to account for the basic historical facts. For utility functions that are not in this class, we would therefore, for example, see very different saving rates 100 years ago than today; typically, saving rates would go to zero or to one over time as the economy grows, and balanced growth as observed in the data would not be possible.

Labor vs. leisure Turning to labor supply, the idea is to allow households to choose how much to work. But do people choose how much they work? In many countries, work hours are regulated, and your own specific employer may not offer you much choice. From the historical, macroeconomic perspective, however, there is no doubt that labor supply is a choice. First, we have seen that in the longer run, work hours per adult have fallen appreciably. Second, looking across countries, a similar pattern emerges (one we will revisit later):

[19] To understand the $\sigma = 1$ case, take the limit as σ goes to 1 but use l'Hôpital's rule. One obtains

$$\lim_{\sigma \to 1} \frac{d(c^{1-\sigma} - 1)/d\sigma}{d(1-\sigma)/d\sigma} = \lim_{\sigma \to 1} \frac{-c^{1-\sigma} \log c}{-1} = \log c.$$

[20] An example is $u(c + 0.01)$, where u satisfies (2.5).

households in rich countries work significantly less than do households in poor countries. These patterns reflect differences (over time and across space) that involve both the intensive margin, i.e., how many hours each individual works when she works, and the extensive margin, i.e., whether a given individual works at all and the fraction of her lifetime the individual works positive hours. We see differences across time and space as reflecting choice and, for example, labor-market regulations stipulating a 40-hour workweek should be seen as an outcome reflecting people's choice at the time these regulations were decided upon.[21] In sum, and especially with our long-run perspective, we consider the choice of how much labor to supply as a very natural one.

In more concrete terms, and focusing on the intensive margin only, the period utility function has to allow the agent to explicitly value leisure. Assume that the time endowment is 1 and the consumer chooses between working hours ℓ and leisure l: $\ell + l = 1$. We can choose to have leisure as an argument: u would depend on (c, l), or equivalently, $(c, 1-\ell)$. Below, we write the utility function as $u(c, 1-\ell)$ to represent this dependence.

Now we need to insist on balanced growth in consumption jointly with a constant labor supply in the long run. We therefore need the condition $u_2(c_t, 1-\ell_t)/u_1(c_t, 1-\ell_t) = w_t$, where $u_i(\cdot,\cdot)$ represents the partial derivative with respect to ith argument, to be met at all points in time on a balanced growth path where, as shown above, c and w grow at the same rate.

Let us begin with the requirement that work hours are constant along a balanced path, as motivated by the postwar U.S. data. Then, just like in the above case, there is a sharp characterization of what preferences are consistent with an exact balanced-growth path. In particular, they are consistent with an exact balanced-growth path if and only if the utility function is of the form $u(c, 1-\ell) = ((cv(1-\ell))^{1-\sigma} - 1)/(1-\sigma)$, where $v(l)$ is strictly increasing (in leisure $l = 1-\ell$) and such that $cv(l)$ is strictly quasiconcave. It is again straightforward to show the "if" part—it is a matter of looking at the Euler equation and the first-order condition for the hours choice jointly and verifying that balanced growth is consistent with the derived equations—but, as before, more demanding to show the "only if" part.

A second possibility is that we would like our theory to be consistent also with the longer-run data and across countries with very different standards of living. This is actually possible, since it also seems—though it was perhaps not clear from the earlier graphs—that as output has grown at a roughly constant rate, hours worked have declined at a roughly constant rate. The rate of hours decline is only, say, a third of a percent per year, but over time these small changes accumulate and become visible. It turns out, moreover, that an outcome with exact balanced growth where hours fall at a constant rate is possible as an outcome within our framework if and only if the utility function satisfies $u(c, 1-\ell) = (c^{1-\sigma} g(c^{\frac{\nu}{1-\nu}} \ell) - 1)/(1-\sigma)$, $\nu > 0$, and $g(\cdot)$ is a decreasing function. With technology and output growth proceeding at a net rate γ, we then have hours grow at $(1+\gamma)^{-\nu}$ and consumption at $(1+\gamma)^{1-\nu}$. Wages now grow slightly faster than output, and the labor share remains a constant share of output.[22]

The case just mentioned collapses to the one described earlier with $\nu = 0$. With $\nu = 0$, as

[21] Of course, in countries that are not democratic, one could imagine that work hours are dictated. But even then people's preferences would likely play a role.

[22] Constantly declining, or growing, hours worked do not pose a problem from the perspective of the production side.

wages grow along a balanced path, the substitution effect (making it more beneficial to work at higher wages) exactly cancels the income effect (making it more attractive to choose more leisure as income rises); if $\nu > 0$, the income effect is slightly stronger. The interpretation is that at lower levels of income, people work more because consumption is more important to them.

In conclusion, we have now arrived at a utility-function specification that is (the only one) consistent with choosing a constant saving rate and constant (or constantly declining, at a low rate) labor supply in the long run. The precise population structure and the length of households' lives can, however, satisfy a variety of assumptions and our main two applications below will be the representative-agent dynasty and the simplest overlapping-generations model.

2.2 The rest of the text

In the next five chapters, we will go over the main macroeconomic tools: (i) the Solow growth model (Chapter 3); (ii) dynamic optimization (Chapter 4); (iii) dynamic equilibrium theory (Chapter 5); (iv) welfare (Chapter 6); (v) uncertainty (Chapter 7); and (vi) empirical methods (Chapter 8). The methods presented in these chapters are core material that will then be used and applied over and over. The methods will also make the material discussed in the present chapter more precise; for example, convergence in the Solow model will be discussed in detail, the maximization problems of consumers will be fully described and a convergence theorem under optimal saving will also be provided, a dynamic competitive equilibrium will be defined and characterized in its application to the growth model, the welfare properties of such equilibria will be discussed in some detail, and it will be made clear under what assumptions the methods and results extend straightforwardly to the case of uncertainty.

The two methodological chapters, Chapters 9 and 10, describe some basic mathematical and computational tools not described in other chapters.

We now very briefly discuss some of the key issues and contents in the applied chapters that then follow in the second part of our text.

The applied issues: Chapters 11–25

Chapter 11: consumption This chapter looks more in detail at how key consumer choices are made on the individual level. It makes clear that the simplest consumption-saving model delivers predictions for marginal propensities to consume—these are a key component of macroeconomic propagation mechanisms—that are hard to square with micro data: they are too low. Imperfect insurance against individual shocks to labor-market outcomes, an a priori plausible addition to the basic model, will change this and bring the model closer in line with the data. At this point, heterogeneity will become an important element of the analysis and the connection between macroeconomics and inequality will appear for the first time.

Chapter 12: labor supply As just discussed in the present chapter, the choice of labor supply can be seen as an extension to basic consumption theory. Allowing for both an extensive and an intensive margin of choice is of particular relevance, the former referring to whether or not to work at all and the latter to how many hours to work, conditional on working. Given this foundation, it is of significant interest to understand how labor choice varies with wages: the wage elasticity of labor supply. There are different notions of elasticities and they are useful in different contexts, with a particular distinction between how hours vary over time in response to wage changes and how hours vary across countries. Another key distinction that will be discussed is individual vs. aggregate labor supply.

Chapter 13: growth Now the wide disparity of incomes across countries will be addressed. What explains why many countries remain poor, and more generally how do the distributions of GDP, consumption, and other key variables across all countries in the world evolve over time? What is the role of technological progress, and what determines it? Human capital accumulation and its role in the growth process are also discussed. Relative to the material in the earlier chapters, this chapter introduces some additional features and, at the same time, flesh out quantitative predictions and compare them with data. At the end of the chapter, there will be an attempt at providing quantitative answers to the core growth questions.

Chapter 14: real business cycles Already in Chapters 3 and 7 there are some glimpses into how the neoclassical model fares when the economy is hit by random shocks. The so-called *real business cycle* (RBC) model, as laid out by Kydland and Prescott in their 1982 paper, proposed that macroeconomic fluctuations are indeed to an important extent a result of unforeseen changes (mostly increases in) technology. In addition, and even more importantly, their paper introduced quantitative, microeconomics-based macroeconomic theory as a tool. Their idea was to formulate a stochastic, dynamic general-equilibrium (DSGE) model, solve it numerically for parameter values that were plausible given micro data and long-run facts, and then use it to address macroeconomic phenomena such as fluctuations and the effects of policy changes. Virtually all analyses of macroeconomic fluctuations since have adopted their approach, though many of the frameworks that came later departed in various ways from Kydland and Prescott's basic RBC model. In particular, many other shocks were proposed, money was introduced, and numerous frictions to how markets operate were included. This chapter begins by discussing how to filter the raw data in order to focus on the business-cycle frequencies and describes the key facts: how macroeconomic aggregate fluctuate and correlate. It then describes the core model, if nothing else because many of the later models treat it as a benchmark. What is the jury's verdict on the role of technology shocks in accounting for business-cycle fluctuations? This question will be addressed in the later chapters.

Chapter 15: government So far, very little has been said about the government. The U.S. government is sizable, and many other governments are much larger still as a share of total expenditures or employment. Governments spend resources and make transfers aimed at redistribution, e.g., from rich to poor and across age groups. The chapter first describes and discusses the key facts pertaining to government variables. Second, the chapter uses our basic theory to examine, given some specified objectives, how different government financing schemes compare, both positively and normatively. Taxation, for example, tends to involve distortions, so some effort will be devoted to understanding its effects. Is it important that the government runs a balanced budget, and does debt management even matter? As a part of this effort, the chapter will give us an introduction to how one can formulate optimal policy problems aiming to maximize consumer welfare while being restricted to the use of distortionary taxes.

Chapter 16: asset prices The previous chapters will have studied many intertemporal issues, including borrowing and lending, but the analysis of asset markets, and the determination of asset prices in particular, is important in its own right. The chapter will describe the key facts—for example, asset prices fluctuate "wildly" and risky assets pay a much higher return on average than do riskless assets such as U.S. Treasury bonds—and then proceed to analyze these facts through the lens of our basic theories. A core framework is the so-called consumption Capital Asset Pricing Model (CAPM), which derives asset prices and their stochastic features in relation to explicit household choices over its stochastic consumption path.

Chapter 17: money A very particular asset is fiat money. "Fiat" refers to the fact that money is intrinsically without value and, nowadays, the value of money is in no sense backed by any real objects (as it was historically in many economies: its value was backed by gold). This raises the question of why it has value at all. From this perspective, inflation means that money loses value. The chapter will go over basic data on inflation and basic theories of how the value of money can be determined, all in the absence of price-setting frictions. It will briefly touch on the determination of exchange rates too; although the value of money in terms of real goods in any given economy does not necessarily fluctuate much, exchange rates do and the chapter will briefly address this volatility. The chapter, finally, will provide a bridge into the next chapter by explaining how money is introduced in the so-called New-Keynesian model of business cycles, where price-setting frictions are central, as well as a discussion of the interdependence of fiscal and monetary policy.

Chapter 18: nominal frictions and business cycles Prices and wages appear to move sluggishly on in the micro data. The chapter will begin by documenting some key facts on this and also review some evidence suggesting that monetary policy can have real effects because prices and wages are "sticky." The New-Keynesian model will then be introduced here. This model has become a workhorse for central banks around the world. It builds on the RBC model but adds nominal frictions: costs associated with changing prices and wages. The extension to the RBC model involves introducing long-lived firms with market power: these firms set prices knowing that prices will be costly to change in the future. The framework also has a description of how the central bank behaves; in particular it introduces a notion of monetary "policy shocks," as an additional source of macroeconomic fluctuations. The chapter will discuss the evidence on the role of monetary policy in accounting for aggregate fluctuations.

Chapter 19: frictional credit markets By many economists, the Great Depression is viewed to have in part been caused by frictions in the credit market, i.e., impediments to borrowing for firms. Similarly, the 2007-2009 Great Recession is also considered to have had its roots in financial-market malfunctioning. The chapter will begin by documenting some correlations that suggest that financial frictions might be important. It will then show how such frictions can be introduced into the core framework and how macroeconomic propagation sometimes, but not always, changes nature in the presence of such frictions.

Chapter 20: frictional labor markets Often, the rate of unemployment is even used to define the business cycle: it is highly countercyclical—rises in recessions and falls in booms. The chapter begins by reviewing not only the key facts on aggregate unemployment but also on individuals' movements in and out of jobs over time. It then introduces the most common framework for analyzing unemployment: the search and matching model. It begins by looking at worker search and then introduces a full general-equilibrium model with matching frictions as in Pissarides (1985). The resulting model is then confronted with data and the so-called Shimer puzzle is introduced and discussed. Finally, the chapter shows how the Pissarides model can be extended so as to incorporate capital accumulation and, as such, can be seen as an important extension of our basic macroeconomic framework.

Chapter 21: inequality In this chapter, we discuss inequality between households. The chapter views inequality as interesting in its own right and, hence, reviews both data and theory. The focus is broad, covering labor-market inequalities (wages, earnings, and hours) as well as inequality in consumption and in wealth, and for each variable of interest it surveys the main theories. The discussion, again, aims to be quantitative, i.e., the theories are evaluated based on how much of the observed inequality they can plausibly account for. The chapter also asks how inequality might matter for macroeconomic aggregates. We have already touched on one way in which it could: to the extent a model with significant inequality generates a marginal propensity to consume that is higher on average, it will alter many of the model's predictions. This is an active research area in macroeconomics; the HANK model—a Heterogeneous-Agent New-Keynesian setting—in particular has already had significant impact in applied monetary policy contexts.

Chapter 22: heterogeneous firms The introduction of the workhorse model is based on an aggregate production function. Clearly, this is an abstraction, at the same time as it hopefully offers a good approximation to the properties of a more realistic framework with a multitude of firms. This chapter examines this issue, both by looking at data on firms and by constructing models with firm heterogeneity. Like the chapter on household heterogeneity, it discusses how firm heterogeneity suggests new mechanisms and add insight into how the macroeconomy works. Two channels are studied in particular. One involving misallocation of input factors across firms when there are frictions. The other makes specific assumptions about firm size and discusses granularity: a notion of extreme firm inequality where some very large firms can be relevant to the whole economy. The chapter also briefly touches on markups and the degree of competition.

Chapter 23: international macroeconomics Many readers of this textbook will perhaps not primarily feel at home in the "large, closed economy" version of our macroeconomic theories. There are, in fact, even strong arguments to suggest that the U.S. economy of today is much more dependent on the rest of the global economy than it used to be and, therefore, issues of trade, exchange rates, and international borrowing and lending ought to take a more central place than it does in many textbooks. The present chapter thus tries to make amends. In particular, it builds toward an up-to-date international business cycle model with monetary and other frictions.

Chapter 24: sovereign debt and default risk This chapter focuses on studying sovereign default, the possibility that governments renege on their debt obligations. The chapter begins by reviewing the key facts on sovereign debt crises and their relationship to macroeconomic fluctuations. It then introduces a canonical sovereign default model in which a government borrows abroad but retains the option to renege. The framework highlights the central trade-off: borrowing helps smooth consumption, but default remains a tempting option when debt burdens grow large. The model provides a foundation for understanding how sovereign risk shapes borrowing limits, interest rates, and macroeconomic volatility.

Chapter 25: sustainability The final chapter should concern all readers, independent of country of origin, as it deals with global topics that have risen to the top of the political agenda virtually everywhere: areas where human economic activity causes environmental problems, such as climate change. The specific question of climate change concerns macroeconomics, as macroeconomic activity, at least historically, is closely tied to carbon dioxide emissions—as a byproduct of using fossil fuels for energy generation. The chapter mostly focuses on climate change and goes through the necessary basic natural-science background and the way in which climate and economics interact. A simple "integrated assessment model" is developed and used to study how policy can be used to address the issue. A brief discussion of natural resource use is also included.