3

Economic Growth

3.1. Introduction

Of all the issues facing development economists, economic growth has to be one of the most compelling. In Chapter 2, we noted the variety of growth experiences across countries. We've seen that percentage growth rates can look deceptively innocuous, but we also learnt to appreciate its power. A percentage point added or subtracted can make the difference between stagnation and prosperity over the period of a generation. Small wonder, then, that the search for key variables in the growth process can be tempting. For precisely this reason, the theory and empirics of economic growth (along with the distribution of that growth) has fired the ambitions and hopes of academic scholars and policy makers alike. I was certainly inspired by Robert Lucas's Marshall Lectures at the University of Cambridge (Lucas 1988):

> Rates of growth of real per-capita income are diverse, even over sustained periods. Indian incomes will double every 50 years; Korean every 10. An Indian will, on average, be twice as well off as his grandfather; a Korean 32 times.[1]
>
> I do not see how one can look at figures like these without seeing them as representing *possibilities*... The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else.

Never mind the fact that India has grown at far faster rates since these words were penned. The sentiment still makes sense: Lucas captures, more keenly than any other writer, the passion that drives the study of economic growth. We can sense the big payoff, the possibility of change with extraordinarily beneficial consequences, if one only knew the exact combination of circumstances that drives economic growth.

If only one knew..., but to expect a single theory about an incredibly complicated economic universe to deliver that knowledge would be unwise. Yet theories of

[1] As we have seen, this is no longer true of India and Korea post 1990, but the general point, made at a time when India was growing slowly, is still valid.

economic growth do take us some way towards understanding the development process, especially so if we supplement the theories with what we know empirically.

3.2. Modern Economic Growth: Basic Features

Economic growth, as the title of Simon Kuznets' pioneering book on the subject (Kuznets 1966) suggests, is a relatively "modern" phenomenon. Today, we greet 5% annual rates of per-capita growth with approval but no great surprise. But throughout most of human history, such growth — or indeed any growth at all — was the exception rather than the rule. In fact, it isn't an exaggeration to say that modern economic growth was born after the Industrial Revolution in Britain.

Consider the growth rates of the world's leaders over the past four centuries. During the period 1580–1820 the Netherlands was the leading industrial nation; it experienced an average annual growth in real GDP per worker hour of roughly 0.2%. The United Kingdom, leader during the approximate period 1820–90, experienced an annual growth of 1.2%. That's (much) faster than the Netherlands, true, but still tortoise-like compared to today's hares. And since then — with some small-country exceptions — it's been the United States, but from 1890 to the present it has averaged around 2% a year, dropping to more sedate 1.7% over 1990–2011. That is certainly impressive, but it still isn't what we've seen lately: first from Japan and then from East Asia more generally, and more recently South Asia.

A little calculation suggests that you don't even have to look at history to establish the modernity of economic growth. Simply run our trusty formula (from Chapter 2, footnote 12) on doubling times backwards. Let's use what by today's high standards is a pretty moderate number: 3% per year. A country growing at that rate will halve its income every 24 years or so, which means that running back 200 years, that country would have to have an income around 1/350th of what it has today! For the United States, that would mean a princely annual income of around $100 per year in 1800. That was most assuredly not the case. And furthermore, poorer countries extrapolated backwards at this rate would simply vanish.

But of course, this sort of calculation isn't merely theoretical. You can see the acceleration, *even* among now-developed countries. Table 3.1 provides a historical glimpse of the period 1850–2010, and shows how growth has transformed the world. This table displays per-capita real GDP (valued in 1990 international dollars) for selected OECD countries in the equally spaced years 1850, 1930, and 2010. The penultimate column gives us the ratio of per-capita GDP in 1930, at the peak just before the Depression, to its counterpart in 1950. The last column does the same for the years 2010 and 1930. The numbers are pretty stunning. On average, GDP per capita in 1930 was 3 times the figure for 1850, but the corresponding ratio for the equally long period period between 1930 and 2010 is by 1978 is 5.2! A nearly sixteen-fold increase in real per-capita GDP in the space of 150 years cannot but transform societies completely. The developing world, after its own transformation, will be no exception.

Indeed, in the broader sweep of historical time, the development story has only just begun. In the nineteenth and twentieth centuries, only a handful of countries, mostly in Western Europe and North America, and represented by and large by the entries in Table 3.1, could manage the "takeoff into sustained growth," to use a well-known term coined by the economic historian W. W. Rostow. Throughout most of

Country	1850	1930	2010	1930/1850	2010/1930
Austria	1,650	3,586	24,096	2.2	6.7
Belgium	1,847	4,979	23,557	2.7	4.7
Canada	1,330	4,811	24,941	3.6	5.2
Denmark	1,767	5,341	23,513	3.0	4.4
Finland	911	2,666	23,290	2.9	8.7
France	1,597	4,532	21,477	2.8	4.7
Germany	1,428	3,973	20,661	2.8	5.2
Japan	681	1,850	21,935	2.7	11.9
Netherlands	2,355	5,603	24,303	2.4	4.3
Norway	956	3,627	27,987	3.8	7.7
Sweden	1,076	4,238	25,306	3.9	6.0
Switzerland	2,293	8,492	25,033	3.7	2.9
United Kingdom	2,330	5,441	23,777	2.3	4.4
United States	1,849	6,213	30,491	3.4	4.9
Simple Average	1,576	4,668	24,312	3.0	5.2

Table 3.1. Per capita GDP (1990 international dollars) in selected OECD countries, 1850–2010. Source: Maddison [2008] and Bolt and Van Zanden [2013].

what is commonly known as the Third World, the growth experience only began well into this century; for many of them, probably not until the post-World War II era, when colonialism ended. Although detailed and reliable national income statistics for most of these countries were not available until some decades ago, the economically underdeveloped nature of these countries is amply revealed in historical accounts. Table 3.2 records the per-capita incomes of several developing countries (and some now-developed countries as well) *relative* to that of the United States, for the last two decades of the twentieth century. I don't plan to be around to update this table 50 years from now, but I would be very curious to know what it will look like.

The Table makes it obvious that despite the very high growth rates experienced by several developing countries, there is plenty of catching-up to do. Moreover, there is a twist in the story that wasn't present a century ago. Then, the now-developed countries grew (not in perfect unison, of course) unshadowed by nations of far greater economic strength. Today, the story is completely different. There is not just a drive to grow, but to grow at rates that far exceed historical experience. The developed world already exists. Their access to resources is not only far higher than that of the developing countries, but *the power afforded by this access is on display*. The urgency of the situation is heightened by the extraordinary flow of information in the world today. People are ever more quickly aware of new products elsewhere and of changes and disparities in standards of living the world over. Exponential growth at rates of 2% may well have significant long-run effects, but they cannot match the parallel growth of human aspirations, and the increased perception of global inequalities. Perhaps no one country, or group of countries, can be blamed for the emergence of these inequalities, but they do exist, and the need for sustained growth is all the more urgent as a result.

Country	1982	1996	2009	2021	Country	1982	1996	2009	2021
Argentina	20.2	25.7	17.3	15.1	India	1.9	1.3	2.3	3.2
Bangladesh	1.4	1.3	1.5	3.5	Indonesia	4.0	3.8	4.7	6.2
Botswana	6.9	10.2	10.5	9.7	Malaysia	13.3	16.3	15.2	15.8
Brazil	14.7	17.1	18.2	10.7	Mexico	18.1	15.0	17.2	14.3
Chile	14.9	17.9	21.6	23.2	Nigeria	12.8	1.5	4.0	2.9
China	1.4	2.4	8.1	17.9	Pakistan	2.4	1.5	1.9	2.1
Cote d'Ivoire	5.9	4.1	3.5	3.6	Rwanda	1.7	0.7	1.2	1.2
Egypt, Arab Rep.	4.2	3.4	4.7	5.3	South Africa	19.2	12.2	13.7	10.0
Ethiopia	1.4	0.5	0.8	1.3	Sri Lanka	2.1	2.5	4.4	5.7
Ghana	2.2	1.3	2.2	3.4	Thailand	5.3	10.1	8.8	10.1

Table 3.2. Per capita GDP in Selected Developing Countries Relative to that of the United States (%), 1982-1996-2009-2021. Source: The World Bank: National Accounts Data.

3.3. The Beginnings of a Theory

3.3.1. Savings and Investment. In its simplest terms, economic growth is the result of abstention from current consumption. An economy produces goods and services. The act of production generates income, which in turn is used to buy these goods and services. Exactly *which* goods are produced depends on individual preferences and the distribution of income.

Classify all goods and services into two groups. We may think of the first group as *consumption items*, which are produced for the direct purpose of satisfying human preferences. The mangos you eat, the tshirts you wear, or a movie you might watch all come under this category. The second group of commodities consists of *capital items*, which we may think of as commodities that are produced for the purpose of producing *other* commodities. An industrial robot, a conveyor belt, or a screwdriver might come under the second category.[2] Typically, households buy consumer goods, whereas firms buy capital goods to expand their production or to replace worn-out machinery.

That immediately raises a question: if all income is paid out to households, and if households spend their income on consumption goods, where does the money for capital goods come from? The answer to this question is simple, although in many ways deceptively so: households save. By abstaining from consumption, households make available a pool of funds (via deposits, stock purchases, or reinvested dividends) that firms use to buy capital goods. This latter purchase is the act of *investment*. Buying power is channeled from savers to investors through banks, individual loans, governments, and stock markets. How these transfers are actually carried out is a story in itself. Later chapters will tell some of this story.

Figure 3.1 is a circuit diagram with income flowing "out" of firms as they produce and income flowing back "into" firms as they sell. Savings is shown as a leakage from the system: the demand for consumption goods alone falls short of the income that created this demand. Investors fill this gap by stepping in with their

[2]There is an intrinsic ambiguity regarding this classification. Although a mango or an industrial robot may be easily classified into its respective categories, the same is not true of, say, a computer. The correct distinction is between goods that have current consumption value and those that produce future consumption, and many goods embody a little of each.

demand for capital goods. By entering or expanding a business, or by replacing worn-out capital, investment adds to the stock of capital, and so an economy grows. In macroeconomic equilibrium, investment demand just counterbalances the savings leakage. Thus, savings equals investment and forms a tentative starting point of the theory of economic growth, in which capital is central (I say "tentative" because I've ignored international capital flows, but that's all right for starters). Figure 3.2 justifies the centrality of capital. It shows an impressively tight connection between the value of physical capital per worker in a country, and its per-capita GDP. This isn't at all to say that other factors — e.g., natural resources, education, R&D or government policies — are unimportant. They matter both directly for output and less directly by permitting capital accumulation. But a model that puts physical capital on center-stage cannot be far off the track.

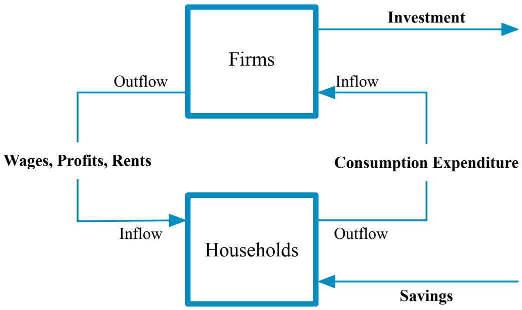

Figure 3.1. Production, consumption, savings, and investment.

3.3.2. Inputs, Outputs and the Production Function. Begin with production, that central activity converting inputs to outputs. This is helpfully summarized by a *production function*. It is a simple mathematical description of how various inputs (such as capital, land, labor, and raw materials) are combined to produce outputs. An easy example is one in which just two inputs — capital and labor — combine to produce a single output. Symbolically, we write

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

where K stands for capital, L for labor, Y for output, and $F(K, L)$ is mathematical notation for some function that converts input pairs (K, L) to output Y. A classic instance is the *Cobb-Douglas* production function, in which we write

$$Y = AK^a L^b, \qquad (3.1)$$

for some positive constants A, a and b. The parameter A measures the extent of technological proficiency. The larger it is, the higher is output for any fixed combination of K and L, so we will call it *total factor productivity*, or TFP for short. The parameters a and b capture the relative importance of each input, as well as whether (and how much) the marginal returns to each input diminish. If a lies between 0 and 1, then there are *diminishing returns to capital*: each additional input of capital increases output, but by a progressively smaller amount. (The same is true of labor, if b lies between 0 and 1.) "Diminishing returns to each input" is a compelling presumption: if more and more of a particular input is added, *without changing the amounts of any other input*, its marginal contribution to output would likely come down.

In contrast, if *all* inputs — or "scale" for short — are increased in the same proportion, it seems reasonable that output should also climb in that proportion. The argument that's usually trotted out in favor of this is "replicability": if you've doubled *every* input,

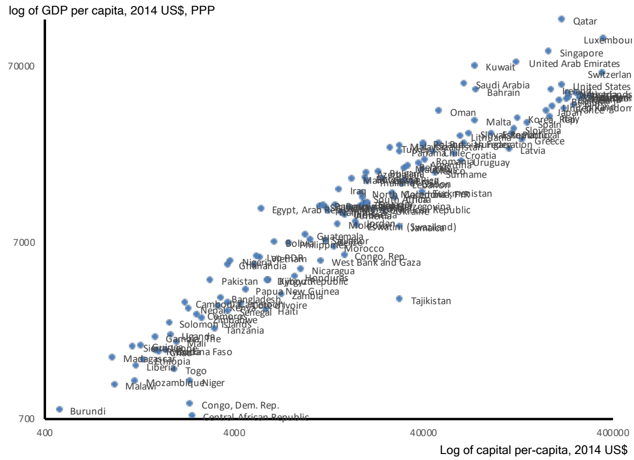

Figure 3.2. Log capital per-capita and log income per-capita, 2014. Source: World Bank *Wealth Accounts* database and *World Development Indicators*

then you've created a perfect twin: how can output not double? That typically kicks off a discussion with some interesting philosophical twists, but we will bypass such matters for now.[3] This phenomenon in which output changes in the same proportion as all inputs is called *constant returns to scale*. It is very easy to verify that in the Cobb-Douglas case described by equation (3.1), "constant returns to scale" is captured by the additional restriction that $a + b = 1$. But there are other possibilities. If output increases less-than-proportionately when all inputs are scaled up, we have *diminishing returns to scale* (with $a + b < 1$ in the Cobb-Douglas special case) or if the increase is more-than-proportionate, we have *increasing returns to scale* ($a + b > 1$). Neither possibility should be summarily dismissed. But the assumption of constant returns to scale, coupled with diminishing returns to each input, is a good starting point.

When there are constant returns to scale, we can typically express all productive activity in per-capita terms; that is, by dividing by the amount of labor being used in production. The Appendix to this chapter shows you how to do this quite generally for production functions, but the Cobb-Douglas case is particularly easy. With $a + b = 1$ under constant returns to scale, change b to $1 - a$ in (3.1) to get

$$Y = AK^a L^{1-a}, \tag{3.2}$$

and now divide by L to see that

$$y = Ak^a \tag{3.3}$$

where the lower case letters y and k stand for the per-capita magnitudes Y/L and K/L respectively. As before, a, typically a number between 0 and 1, is an inverse measure of diminishing returns to capital. The lower the value of a, the greater the "curvature" of the production function and the greater the extent of diminishing returns to capital.

[3] For instance, can we really double *all* inputs? What about, say, that elusive input called entrepreneurship: are we doubling that too? If not, output might less than double. On the other hand, if bricks-and-mortar build a production facility and we double the amount of bricks, the volume of the facility will typically more than double, the former being proportional to surface area and the latter rising at (surface area)$^{3/2}$.

In passing, notice that if we take logarithms on both sides of (3.3), we get

$$\ln y = \ln A + a \ln k,$$

which shows that (under Cobb-Douglas) the relationship between $\ln k$ and $\ln y$ is affine. Figure 3.2 provides significant empirical support for this relationship.

Figure 3.3 depicts a production function in per-capita form $f(k)$, with diminishing returns to per-capita capital. As k increases, so does $y = f(k)$, but in a progressively muted way. Thus the marginal product of capital, and the output-capital *ratio* $\theta \equiv y/k$, will fall as k climbs, driven by a relative shortage of labor. Just how quickly it falls will depend on the extent of diminishing returns to capital.

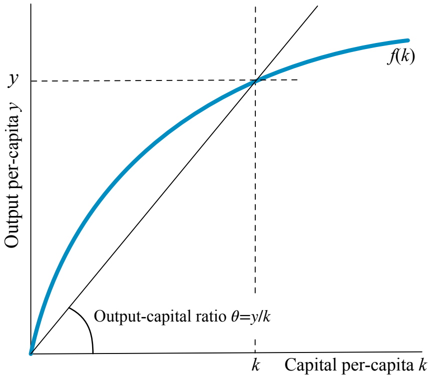

Figure 3.3. The production function in per-capita form.

The concept of a production function already allows to sift and sort some preliminary thoughts about underdevelopment. For instance, staring at equation (3.1), one might wonder where developing countries might come up relatively short. In TFP, maybe? Perhaps it is fundamentally in technical knowledge that development lies. Perhaps patent restrictions and other limitations on the flow of knowledge across countries are fundamentally responsible for underdevelopment. (Though such arguments must contend with the willingness and ability of human beings to take apart new products and simply learn how to rebuild them.)

There are other interpretations of TFP. For instance, it might embody the extent of skills and education possessed by the working population, so that a given amount of capital and raw labor goes a longer way in developed countries. But properly viewed, these differences should not be thrown into the A-term. Rather, we should recognize that the production functions described in equations such as (3.1) are two-input caricatures of reality. More realistically, we could write

$$Y = F(K, \text{labor of different qualities})$$

and recognize that developing countries have lower endowments of some types of labor. That's conceptually different from saying that they are incapable of producing the same output *even with* all relevant labor and nonlabor inputs at hand. One quick example using the Cobb-Douglas formulation is to write

$$Y = AK^a(eL)^b \tag{3.4}$$

where e is schooling per person. That mathematically links up with "lower A" (think of combining the compound term Ae^b into a new A term) but it's conceptually very different, pointing the finger to education and not some productivity difference that applies even after schooling differences are accounted for.

Yet another interpretation of "different A" is that resources or inputs are somehow misallocated to a greater degree in developing countries.[4] For instance, entrepreneurs

[4] See, for instance, Banerjee and Duflo (2005) and Hsieh and Klenow (2009).

in a poor country might not have enough access to capital or credit markets to raise the input outlays for a hew technology. Or the older technology may be in the hands of older, elite groups with political power, who block access to new technologies that could spell their own ruin. Or local communities might oppose new technologies for fear that these will let "foreign interests" into the country.

We will have much to say about these and related issues later in the book. But our quest for explanations starts with the simplest of them: maybe it's just low K (relative to L)? A developing country surely has less per-capita capital — physical and human compared to its developed-country counterpart. Might this, and this *alone*, not explain persistent, ongoing differences in per-capita income across countries? Methodologically, it's a good idea to start small, ask the simplest questions, and expand our inquiry to newer pastures only when that is called for.

So we begin with a theory that emphasizes the systematic accumulation of capital.

3.3.3. The Growth Equation. A little algebra at this stage will make our lives simpler. Divide time into periods $t = 0, 1, 2, 3, \ldots$, tagging our economic variables with the date: $Y(t)$ for output, $I(t)$ for investment, and so on. Investment augments the capital stock after accounting for depreciation, so in symbols:

$$K(t+1) = (1-\delta)K(t) + I(t),$$

where δ is the rate of depreciation. Now recall the famous macroeconomic balance condition, that investment equals savings. It follows that $I(t) = sY(t)$, where s is the *rate* of savings and $Y(t)$ is aggregate output, and using this above, we see that

$$K(t+1) = (1-\delta)K(t) + sY(t), \tag{3.5}$$

which tells us how the capital stock must change over time. We're going to convert all this into per-capita terms by dividing by the total population, which we assume (only for expositional simplicity) to be equal to the active labor force $L(t)$. If we assume that population grows at a constant rate n, so that $L(t+1) = (1+n)L(t)$, (3.5) changes to

$$(1+n)k(t+1) = (1-\delta)k(t) + sy(t), \tag{3.6}$$

where k and y represent per-capita magnitudes K/L and Y/L respectively. Finally, divide through by $(1+n)k(t)$ to get

$$\frac{k(t+1)}{k(t)} = \frac{(1-\delta) + s\theta(k(t))}{1+n}, \tag{3.7}$$

where we've defined $\theta(k) = y/k = f(k)/k$ to be the output-capital ratio in production. This is our basic growth equation. The numerator on its right-hand side measures the rate at which one unit of capital (per-capita) at date t is transformed into capital at date $t+1$. That one unit decays a bit, and we are left with $1-\delta$. In the process, it also produces output $\theta(k) = f(k)/k$, and s of that is saved, so that adds $s\theta(k)$. The total is $(1-\delta) + s\theta(k)$. Meanwhile, the denominator records the drag on per-capita capital caused by population growth: "one unit" of population today becomes $1+n$ units tomorrow. The ratio of these two forces determines whether per-capita capital grows or shrinks; i.e., whether $k(t+1)/k(t)$ on the left-hand side of (3.7) exceeds 1 or falls below it.

Take the growth equation for a spin: start with the per-capita capital stock at any date t, $k(t)$. That produces per-capita output $y(t)$ via the production function $f(k(t))$

from Section 3.3.2, and implies an output-capital *ratio* $\theta(k(t)) = f(k(t))/k(t)$. Now (3.7) tells you what $k(t+1)$ must be, and the story repeats, ad infinitum. It's simple.

3.4. The Solow Model

A much-venerated theory of growth, due to Robert Solow (1957), uses equation (3.7) for its dynamics. In Figure 3.4, we place k on the horizontal axis, and the numerator and denominator of the right-hand side of (3.7) on the vertical axis. The numerator falls in k, because the output-capital ratio $\theta(k)$ *declines* as k *increases*; see the discussion around Figure 3.3. The denominator, $1 + n$ is independent of k and is recorded as a flat line. Thus the line representing the numerator initially lies above this flat line (for low k), and then falls below (for high k).[5]

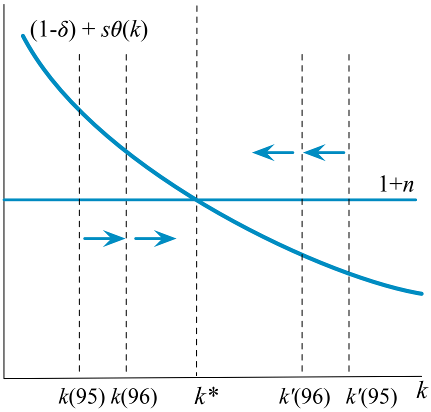

Figure 3.4. Growth dynamics in the Solow model.

Armed with this diagram, we can make some strong predictions about growth. At the left of Figure 3.4 is a low initial level of the per-capita capital stock, labeled $k(95)$, in deference to the year I began writing this book. Because $k(95)$ is low, the corresponding output-capital ratio $\theta(k(95))$ is high and so — using equation (3.7) — we have $k(96) > k(95)$ (capital accumulation outstrips population growth). This process continues through further iterations, and it can be seen that $k(t)$ heads towards k^*, which is a distinguished capital stock where the curved and straight lines meet.

Likewise, you may trace the argument for a high initial capital stock, depicted by the initial value $k'(95)$ above k^*. Now $\theta(k)$ is low enough so that the curved line falls below the flat line, which implies that $k'(96) < k'(95)$. So there is *erosion* of the per-capita stock as time passes, with convergence occurring over time to the same per-capita stock, k^*. In this zone, capital accumulation is eroded by population growth.

3.4.1. The Steady State.

We can therefore think of k^* as a *steady-state* level of the per capita capital stock, to which k starting from *any* initial level converges. Starting at $k(95)$, capital grows faster than labor (so $k(t)$), but diminishing returns to capital lowers the ratio $\theta(k(t))$. That slows the growth of capital into line with that of labor, attenuating the climb in k. In the long run, the capital-labor ratio settles at the steady state k^*. The same process happens in reverse if the economy starts above k^*.

Now, if the per-capita capital stock settles, then so must per-capita income! Indeed, we could have conducted this analysis starting from per-capita income instead of capital, because y and k are related one-for-one via the production function. That is, starting from $y(95) = f(k(95))$, income growth is positive but decelerates as $y \to y^* = f(k^*)$,

[5] If you've been following the argument closely, you will see that this last statement is exactly true if we make the additional assumption that the marginal product of capital is very high when there is very little capital and diminishes to zero as the per-capita capital stock becomes very high.

while the reverse process occurs from $y'(95) = f(k'(95))$. In particular, there is no long-run growth of per-capita output, and *total* output grows precisely at the rate of growth of the population. In algebraic terms, we can put $k(t) = k(t+1) = k^*$ in equation (3.7). Moving terms around a bit, we obtain

$$\frac{y^*}{k^*} = \theta(k^*) = \frac{n+\delta}{s}. \tag{3.8}$$

A unique solution k^* to (3.8) exists, because $\theta(k)$ decreases from very high to very low as we vary k.[6] For instance, when the production function is Cobb-Douglas, so that $y = Ak^a$, the steady state condition (3.15) becomes

$$\frac{Ak^{*a}}{k^*} = \frac{n+\delta}{s},$$

and after a little boring algebra, we see that

$$k^* = \left(\frac{sA}{n+\delta}\right)^{1/(1-a)} \text{ and } y^* = A^{1/(1-a)}\left(\frac{s}{n+\delta}\right)^{a/(1-a)} \tag{3.9}$$

This is a simple and rewarding equation, because you can "see" the steady state explicitly. But at this point you must be scratching your head in some confusion. Or if you're not, you should be. What manner of growth model is a growth model that predicts no per-capita growth in the long run? That's a fair question, and we must address it now.

3.4.2. Technical Progress. The conclusion of the basic Solow model — zero growth per-capita in the long run — is counterfactual, but not counterintuitive. With unchanging technology, the influence of fixed factors must ultimately make itself felt, and growth would vanish in the long term. In the model, labor is that fixed factor, and in the absence of technical progress, growth *per-capita* must slow to a crawl. That's uncomfortable, but not unreasonable. This no-growth scenario disappears if there is continuing technical progress; that is, if TFP rises over time as new knowledge is gained. As long as the optimism of this shift outweighs the impending doom of diminishing returns, per-capita growth can be sustained indefinitely.

One simple way to incorporate technical progress into the theory is to think about new knowledge as contributing to the efficiency, or economic productivity, of labor. Let's make a distinction between the working population $L(t)$ and the amount of labor in "efficiency units"; call it $E(t)$. This distinction is necessary now because in the extension we're about to consider, the productivity of the working population will be constantly increasing. We will write $E(t) = e(t)L(t)$, where $e(t)$ is the productivity of an individual at time t.[7] Not only does population grow over time (at the rate of n, just as before), but we now suppose that efficiency *per person* grows too, say at the rate of π. Thus $e(t+1) = (1+\pi)e(t)$. We will refer to π as the rate of technical progress.

Notice that $E(t+1) = e(t+1)L(t+1) = (1+n)(1+\pi)e(t)L(t) = (1+n)(1+\pi)E(t)$. But because n and π are small numbers such as 0.02 or 0.03, it is a simple and pretty

[6] The careful reader will observe that we're assuming diminishing returns plus suitable end-point conditions on the marginal product of capital.

[7] With a Cobb-Douglas production function, the same specification of technical progress could have several interpretations. See the "second point of interpretation" following equation (3.13) below.

accurate approximation to write $(1+n)(1+\pi) \simeq 1+n+\pi$. We can therefore say that effective labor grows at the rate $n+\pi$, which is close enough for all our calculations.[8]

One more step completes our extended model. Recall how we passed from (3.5) to (3.7) after dividing through by the working population to obtain per-capita magnitudes. We divide here instead by the *effective* population $e(t)L(t)$ to arrive at what looks like per-capita capital and income, but these are expressed *per efficiency unit of labor*. Let's call them $\hat{k}$ and $\hat{y}$ to distinguish them from the earlier per-capita values k and y. Now observe that *all* our calculations between equations (3.7) to (3.9) hold for this case, except that we replace k and y by $\hat{k}$ and $\hat{y}$ wherever we see them, and population growth n by the growth of effective labor, which is $n+\pi$. For instance, Figure 3.5 repeats, in spirit, the analysis done for equation (3.7) in Figure 3.4. *Exactly* the same logic applies, and once again, we have convergence to a steady-state, relative to effective labor, denoted by $\hat{k}^*$.

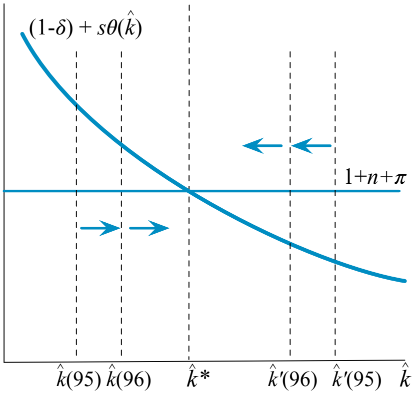

Figure 3.5. The Solow model with technical change. Spot the differences from Figure 3.4.

The formulae describing the steady state are also exact parallels of (3.8) and (3.9), and for completeness we write them here again. First we have:

$$\frac{\hat{y}^*}{\hat{k}^*} = \theta(\hat{k}^*) \simeq \frac{n+\pi+\delta}{s}, \tag{3.10}$$

where the notation "$\simeq$" reminds you of the slight approximation to arrive at $n+\pi$. With Cobb-Douglas production, we follow exactly the same steps to obtain equation (3.9) from (3.8). The production function is

$$Y = AK^a(eL)^b, \text{ or } \hat{y} = A\hat{k}^a, \tag{3.11}$$

once we've divided through by effective labor. Combining (3.11) with (4.6), we see that

$$\hat{k}^* \simeq \left[\frac{sA}{n+\pi+\delta}\right]^{1/(1-a)} \text{ and } \hat{y}^* \simeq A^{1/(1-a)}\left[\frac{s}{n+\pi+\delta}\right]^{a/(1-a)}. \tag{3.12}$$

There *really* is no mathematical difference. The novelty lies in the interpretation. We make two points. First, even though capital and output per efficiency unit converge to a steady state, these continue to increase *per person* at the rate of technical progress.

This is the way it looks in the Cobb-Douglas case. Open up the formula for effective income in (3.12) to see that in steady state, per-capita income has the trajectory:

$$y^*(t) \simeq \hat{y}^*(1+\pi)^t \simeq A^{1/(1-a)}\left(\frac{s}{n+\delta+\pi}\right)^{a/(1-a)}(1+\pi)^t. \tag{3.13}$$

The second point of interpretation has to do with π, which is the growth rate of *labor* productivity. That might look restrictive: are there not other kinds of technical progress that improve the productivity of capital or of all inputs together? The answer is that in

[8]That is, multiply out $(1+n)(1+\pi)$ to get $1+n+\pi+n\pi$ and safely ignore the tiny term $n\pi$.

many situations these are all *mathematically* the same. The Cobb-Douglas function, with $Y(t) = AK(t)^a[e(t)L(t)]^{1-a}$, can be equivalently written as

$$Y(t) = A(t)K(t)^a L(t)^{1-a},$$

where $A(t) = Ae(t)^{1-a}$ can now be viewed as TFP. If e is growing at rate π, $A(t)$ is growing at rate z, where $1 + z = (1 + \pi)^{1-a}$. So, chameleon-like, technical progress in effective labor can also be viewed as technical progress that raises the effectivity of all inputs, or indeed also the effectivity of physical capital. It's just a matter of redefining variables, though we emphasize that the interpretation can be very different.

Figure 3.6 depicts the trajectory (3.13) in logarithmic terms, so that an exponential growth of income appears linear. (Mentally take logarithms in (3.13) to verify that $\log y^*(t)$ has a constant slope in time of $1 + \pi$.) All of the analysis that describes convergence to the steady state in effective labor units can now be re-described as the *convergence* of per-capita income trajectories to the linear steady state trajectory. We will be returning to this important prediction of convergence in Chapter 4.

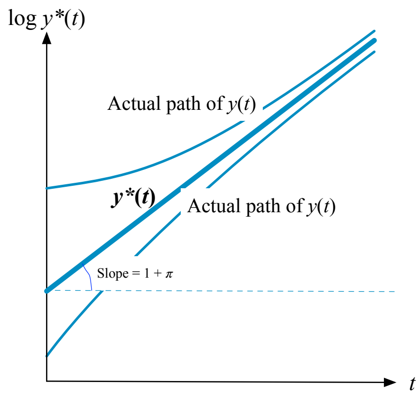

Figure 3.6. Convergence to steady state.

In summary, think of two broad sources of growth: one is via the continued buildup of plant, machinery, and other inputs that bring increased productive power,[9] and the other is through more advanced methods of production (technical progress). The Solow model claims that *in the absence of the second source, the first is not enough for sustained per-capita growth*. Viewed this way, the Solow model points to studying the economics of technological progress, arguing that it is there that one must look for the ultimate sources of growth. This is not to say that such a claim is necessarily true, but it is certainly provocative and not obviously wrong.

3.4.3. Steady State Parameters. The different parameters of the Solow model — the savings rate, population growth rate, or the rate of depreciation — do not affect the long-run growth rate of per-capita income, which continues unperturbed at the exogenous rate of technical progress. But they do affect the long-run *level* of per-capita income. In our logarithmic trajectories depicted in Figure 3.6, these "level effects" appear as parallel shifts of the steady state trajectory. See the dotted parallel lines in Figure 3.7. You can algebraically see the shifts by taking logarithms in equation (3.13). Parameters such as s and n will only create additive shifts in the logarithmic version of the formula, which are the counterparts of these graphical parallel shifts.

This analysis is quite general and does not rely on the specific form of the technology. To underline that, return to the more general formula that describes the effective steady state in equation (3.10). An increase in the savings rate raises the right-hand side of the equation, so the left-hand side must also increase to restore equality. This means

[9]This is not to deny that these two sources are often intimately linked: technical progress may be *embodied* in the new accumulation of capital inputs.

that the new steady-state output-capital ratio — the value $\hat{y}^*/\hat{k}^*$ — must be *lower*. With diminishing returns, that can only happen if the new steady-state level of $\hat{y}^*$ is *higher*. So an increase in the savings rate raises the long-run level of per-capita income in effective labor units. This is tantamount to an upward parallel shift in the steady state trajectory. By the same logic, we can check that an increase in the population growth rate or the rate of depreciation will raise y^*/k^* and so lower long-run per-capita income.[10]

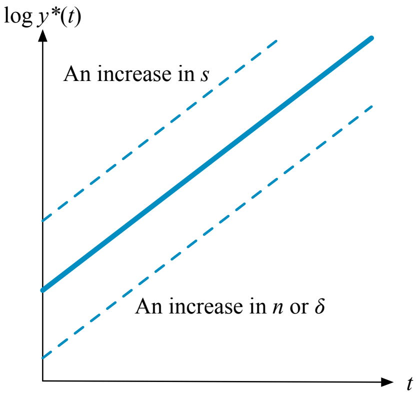

Figure 3.7. s, n and δ have level effects.

All these exercises are completely intuitive, except for one, which has to do with a change in the rate of technical progress π. This is the only parameter which has both a level effect and a "growth rate" effect. After all, we have already seen that the long run growth rate of the economy occurs at the rate π, so its change must of necessity *twist* the steady state trajectory and not just shift it. Figure 3.8 illustrates this change.

Notice how the twist actually *reduces* the level of the steady state path at some points: the dotted line intersects the old steady state path from below. And indeed, the algebra corroborates this. The easiest way to see this is to note that n, π and δ all enter as the sum $n + \pi + \delta$ into the steady state in effective units of labor; inspect equation (3.10). That means that an increase in the rate of technical progress π has the same implication as, say, an increase in the population growth rate n, which is that it reduces the steady state measured in effective units of labor. But of course, the resulting *trajectory* is steeper, and the two combine in an uncomfortably ambiguous way, as shown by the pair of straight lines in Figure 3.8. That discomfort vanishes when you look at the resulting trajectories from the moment the change takes place. *At that moment*, the country has just one per-capita income value, shown by the value $y(0)$ marked at time 0 in the Figure. The two transitional paths to the two steady states, shown by the thin curves, will not cross. The one for the higher value of π will always lie above its lower-π counterpart. In the end, there is no ambiguity at all.

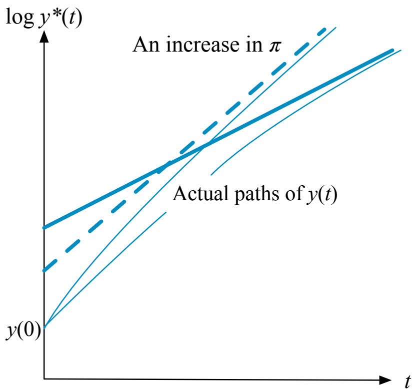

Figure 3.8. A change in π has growth effects.

[10] Make sure you are comfortable with the simple economics behind all the algebra. For instance, a higher rate of depreciation means that more of national savings must go into the replacement of worn-out capital. This means that, all other things being equal, the economy accumulates a smaller net amount of per-capita capital, and this lowers the steady-state. You should similarly "verbally" run through the effects of changes in the savings rate and the population growth rate.

3.5. Exogenous and Endogenous Growth

The Solow model is one of *exogenous growth*. Its steady state growth rate is insensitive to behavioral parameters, and is fully pinned down by the (assumed exogenous) rate of technical progress. In this section, we explore two variants which can be thought of as displaying *endogenous growth*, with equilibrium growth rates driven by behavioral parameters, such as the savings rate, the rate of population change, or deliberately chosen rates of educational investment. To my mind, the essential distinction lies in the word "behavioral". *Human* behavior could affect growth in the variants that we now consider, as opposed to growth in the Solow model, which is fully determined by some exogenously assumed rate of technical progress. This change of focus is important, both conceptually and from the viewpoint of theoretical and empirical analysis. Economic growth does not drop on us as manna from the gods.

3.5.1. The Harrod-Domar Model. This variant on the fundamental growth equation highlights the the accumulation of physical capital. Recall the growth equation:

$$\frac{k(t+1)}{k(t)} = \frac{(1-\delta) + s\theta(k(t))}{1+n}. \tag{3.14}$$

This turns into the Solow model under the *additional* assumption of diminishing returns to capital in production, so that $\theta(k)$ decreases in k over a full range of values. That assumption forces the economy to settle at the distinguished level k^* given by

$$\theta(k^*) = \frac{n+\delta}{s}, \text{ or } k^* = \left(\frac{sA}{n+\delta}\right)^{1/(1-a)}. \tag{3.15}$$

in the Cobb-Douglas special case. There is something about the latter equation to I'd like to draw your attention. Look at the "diminishing returns parameter" a and remember that the smaller it is, the more it is that returns to capital diminish, while at the other end, as a becomes close to 1, the production function becomes almost linear and exhibits constant returns *to the capital input alone*. As we bring a up close to 1, the steady state level of capital becomes ever larger (if $sA > n + \delta$) or ever smaller (if $sA < n + \delta$) and *at a equal to 1*, when the production function is *exactly* linear in capital, there is no steady state: the economy either grows to infinity or shrinks to zero!

This is not some algebraic sleight of hand; on the contrary, it makes intuitive sense. When $a = 1$, the output-capital ratio is fully insensitive to the value of k, and the current scale of the economy becomes irrelevant: whatever rate it can grow at k, it can replicate that at $2k$, $3k$, or a million times k. With constant returns to the capital input, the economy can grow or decline at exactly the same rate, irrespective of capital scale.

The growth equation (3.14) can handle this without a problem, provided we don't go down the garden path looking for steady states where there are none to be found. With $a = 1$, the output-capital ratio is a constant; in fact, it is exactly A in the Cobb-Douglas production function: $y = Ak$. Using this in (3.14) and subtracting 1 from both sides of it, we have

$$\text{Rate of growth} = \frac{k(t+1)-k(t)}{k(t)} = \frac{sA-(n+\delta)}{1+n}. \tag{3.16}$$

This is an influential relationship, known as the *Harrod-Domar equation*. named after Roy Harrod and Evsey Domar, who wrote some of the earliest papers on the subject in 1939 and 1946, respectively. The left-hand side is just the rate of growth of the per-capita capital stock, and by the linearity of the production function it is also the rate of growth of per-capita *income*. Time appears here on the left-hand side of (3.16), but it doesn't in the rest of the equation, which shows that the rate of growth is constant and unchanging. For an easy-to-remember version of (3.16), let g stand for the rate of growth, multiply through by $1 + n$, and note that both g and n are small numbers (analogous to our approach to deriving (3.10)), and their product gn is therefore *extra* small relative to these numbers. That gives us the approximation

$$g \simeq sA - n - \delta, \tag{3.17}$$

which can be used in place of (3.16) without serious loss of accuracy. Observe how (3.17) *endogenizes* the growth rate, making it an function of behavioral parameters such as the savings rate or the population growth rate. In contrast, these parameters have no effect at all on long-run growth in the Solow model. There is no contradiction here, though the difference is striking. In the Solow model with constant returns to scale, the parameter a always lies strictly between 0 and 1, so labor plays an indispensable role in production, therefore constraining per-capita output growth. In the Harrod-Domar variant, $a = 1$, and growth is thereby liberated from the "shackles" of a labor input.

Growth Engineering

It isn't hard to see why the Harrod-Domar growth equation (3.17) was so influential. It has the air of a recipe. Thomas Piketty, in his book *Capital* (2014), calls it the "second fundamental law of capitalism." The equation firmly links the growth rate of the economy to certain parameters, such as the savings rate, the output-capital ratio and the growth rate of population. And capitalism apart, central planning in countries such as India and the erstwhile Soviet Union was deeply influenced by the Harrod–Domar equation.

The Harrod-Domar model served as the conceptual underpinning for large, regulated investments in heavy industry. As we have seen, capital-accumulation fully drives economic growth in the Harrod-Domar setting, with no constraint imposed by labor. The reliance on machinery, and the implied liberation from the everyday drudgery of human work, lies just below the surface. It is a socialist dream — or a market nightmare, now summoned up by the looming dystopia of automation and labor displacement.

The first controlled experiment in "growth engineering" undertaken in the world was in the former Soviet Union, after the Bolshevik Revolution in 1917. The years immediately following the Revolution were spent in a bitter struggle—between the Bolsheviks and their various enemies, particularly the White Army of the previous Czarist regime—over the control of territory and productive assets such as land, factories, and machinery. Through the decade of the 1920s, the Bolsheviks gradually extended their control over most of the Soviet Union (consisting of Russia, Ukraine, and other smaller states). The time had come to use this newly acquired control to achieve the economic goals of the revolutionary Bolsheviks, the foremost among these goals being a fast pace of industrialization.[a]

Under the State Economic Planning Commission (*Gosplan*), a series of draft plans was drawn up. These culminated in the first Soviet Five Year Plan (a predecessor to many more), which covered the period from 1929 to 1933. At the level of objectives, the plan placed a strong emphasis on industrial growth. The resulting need to step up the rate of investment was reflected in the plan target of increasing it from the existing level of 19.9% of national income in 1927–28 to 33.6% by 1932–33. (Dobb [1966, p. 236]).

How did it go? The Table below shows some of the plan targets and actual achievements, and what emerges is quite impressive. Within a space of five years, real national income nearly doubled, although it stayed slightly below the plan target. Progress on the industrial front was spectacular: gross industrial production increased almost 2.5 times. This was mainly due to rapid expansion in the machine producing sector (where growth was far in excess of even planned targets). This can be explained in part by the enormous emphasis on heavy industry in order to expand from a meager industrial base.

	1927–28 (actual)	*1932–33 (plan)*	*1932 (actual)*
National income	24.4	49.7	45.5
Gross industrial production	18.3	43.2	43.3
(a) Producers' goods	6.0	18.1	23.1
(b) Consumers' goods	12.3	25.1	20.2
Gross agricultural production	13.1	25.8	16.6

Table 3.3. Targets and achievements of the first Soviet Five-Year Plan (1928–29 to 1932–33). All entries in 100 million 1926–27 rubles. Source: Dobb (1966).

Note that the production of consumer goods fell way below plan targets. An equally spectacular failure shows up in the agricultural sector, in which actual production in 1932 was barely two-thirds of the plan target and only slightly more than the 1927–28 level. The Bolsheviks' control over agriculture was never as complete as that over industry, and continuing strife with farmers and large landowners from the Czarist era took its toll on crop production. But the implicit reliance on industrialization as *the* key to growth was also a basic factor in this failure, and explains the need for severe agrarian collectivization; Chapter 17 will have more.

Post-Independence India took its economic cue from the Soviets. Like the Soviets, Indian policy makers believed that heavy industry was *the* leading sector to encourage: its growth would pull up the remainder of the economy. Indeed, until the early 1990s, the Indian government deployed powerful instruments (such as the Industrial Licensing Policy) that controlled the allocation of investment to even those sectors largely in private hands. The acquisition of such powers by the Indian government from the birth of the Republic (in 1947) showed an unwillingness to rely purely on the market mechanism.

The Planning Commission of India was established on 15 March 1950, and the First Indian Five Year Plan covered the period 1951–52 to 1955–56. As in the Harrod–Domar theory and in Soviet planning, there was a large emphasis on raising overall rates of investment. The second Five Year Plan (1955–56 to 1960–61) went a step further. One of the main architects of the plan was Professor P. C. Mahalanobis, an eminent statistician and advisor to then Prime Minister Nehru. The so-called Mahalanobis model, which served as the foundation of the Second Plan, bore a close resemblance to a framework enunciated by the Soviet economist Feldman in 1928, on which Soviet planning in the 1930s was largely based. Both models argued that to achieve rapid growth, careful attention was to be given not only to the *size* of investment, but also to its *composition*. In particular, these models stressed the need to make substantial investments in the capital goods sector so as to expand the industrial base. This emphasis on heavy industry in India's second Five Year Plan is illustrated by the fact that 34.4% of planned investment was in the investment goods sector, compared to only 18.2% in consumer goods and 17.2% in agriculture (Hanson 1966, p.126).

During the plan period, we note that national income grew by 4% per annum on average. Given the almost stagnant nature of the Indian economy in the preceding half century or even more, this was pretty dramatic. K.N. Raj (1965) observed that "the percentage increase in national income in the last thirteen years has been higher than the percentage increase realized in India over the entire preceding half a century."[b]

Industry, given the greater share of investment it received, did much better than the nationwide average. Overall industrial production grew at an average rate of 7% per annum over the period of the first two plans. For the second plan period alone, the general index of industrial production grew by roughly 35% between 1955–56 and 1960–61, and that of machine production soared to 250% of its starting level in the meager space of five years (Hanson 1966, p.169). However, there were some serious shortfalls in the infrastructural sector: power production missed its target of 6.9 million kilowatts by 1.2 million, and underinvestment in railways gave rise to bottlenecks and strain toward the end of the period (Hanson 1966).

In conclusion, although the first two Five Year Plans set India on a path of aggregative growth unprecedented in her history, the abysmally poor living conditions for the majority to begin with, coupled with increasing population pressure, hardly left any room for complacence at the beginning of the Third Plan period.

[a] On the eve of the Revolution, Russia lagged behind the industrialized European nations, despite a rich endowment of natural resources. Based on the per-capita use of essential industrial inputs, namely, raw cotton, pig iron, railway services, coal, and steam power, Russia ranked last among nine major European nations in 1910, behind even Spain and Italy; see Nove (1969).

[b] However, population growth during the period exceeded expectations and, more alarmingly, showed a rising trend, mainly due to a fall in the death rate caused by improvements in medical care. Consequently, *per-capita* national income grew by only 1.8% per year, which, though still creditable, is considerably less cheerful.

But what is a to begin with, and how do we get a handle on its quantitative magnitude? There's an interesting way to approach this question. In competitive markets, we know that factors of production are paid their marginal products. A single line of calculus will tell you that the marginal product of capital is given by

$$\frac{\partial Y}{\partial K} = aAK^{a-1}(eL)^{1-a} = a\frac{Y}{K},$$

so that under constant returns to scale, the *share of capital income in total income* is

$$\frac{\partial Y}{\partial K} \cdot \frac{K}{Y} = a!$$

The point is that we do have estimates of the share of capital in national income, and we can apply those to get a sense of a. They vary, of course, as all estimates do. An approximate range for the United States is between a quarter (Parente and Prescott, 2000) and two-fifths (Lucas, 1990). But that places us very far indeed from the Harrod-Domar extreme with $a = 1$. Or does it? That is where the simple Solow model, with its two factors of production, could be misleading. Our next variant illustrates this point.

3.5.2. Human Capital. So far we have considered "labor" to be a single input of production, augmented, perhaps, by exogenous technical progress. But labor can also be *deliberately* accumulated to some degree, along with physical capital. Education, training, experience, learning by doing: these are all ways to augment one's natural endowment, transforming it into new labor with new skills, New labor has the ability to operate ever-more sophisticated machinery. It can generate new ideas and methods for economic activity. It can take on demanding managerial roles. A good name for these "animate" forms of capital is *human capital.* But our extension will also serve as a rough model of endogenous technical progress, where the productivity term in the Solow model is influenced by deliberate human action. All told, this variant will do a better job of reconciling observed income differences across countries.

The basic idea is simple.[11] Add to the Solow model by permitting individuals to "save" in two distinct forms: a fraction s_k of income acquires rights to physical capital K, and a fraction s_h is invested in education, or *human* capital H.[12] We therefore have:

$$K(t+1) - (1-\delta_k)K(t) = s_k Y(t) \text{ and } H(t+1) - (1-\delta_h)H(t) = s_h Y(t), \quad (3.18)$$

where δ_k and δ_h are depreciation rates for K and H respectively. The production function is given by

$$Y = AK^a(eL)^b H^c, \quad (3.19)$$

where $a + b + c = 1$, L is the stock of unskilled labor, growing exogenously at the rate of n, and e is exogenous productivity, growing at π. The interpretation we adopt here, due to Mankiw, Romer and Weil (1992), is that L and H are unskilled and skilled labor inputs, performing different tasks, and that L is not reduced by higher H. One interpretation is that everyone always supplies an unchanging endowment of unskilled labor, and those who are educated separately supply those skills *over and above* their unskilled component. The Appendix to this Chapter considers other interpretations.

Dividing through by eL in these equations and proceeding exactly as we did in our first rendition of the Solow model to obtain (3.6), we can transform (3.18) into:

$$(1+n+\pi)\hat{k}(t+1) \simeq (1-\delta_k)\hat{k}(t) + s_k\hat{y}(t) \quad (3.20)$$

$$(1+n+\pi)\hat{h}(t+1) \simeq (1-\delta_h)\hat{h}(t) + s_h\hat{y}(t), \quad (3.21)$$

where $\hat{k} = K/eL$, $\hat{h} = H/eL$, and $\hat{y} = Ye/L$. Likewise, equation (3.19) becomes

$$\hat{y} = A\hat{k}^a\hat{h}^c. \quad (3.22)$$

This should all be familiar. It's only that we're dealing with *two* accumulable inputs ($\hat{k}$ and $\hat{h}$) instead of just one ($\hat{k}$) as in the basic Solow model. One consequence is that it isn't possible to graphically track — in a simple two dimensional diagram such as Figure 3.4 — the *joint* evolution of $\hat{k}$ and $\hat{h}$ towards their respective steady states. But that aside, the parallel with the basic Solow model is as expected and not surprising: there are steady state values $\hat{k}^*$ and $\hat{h}^*$ (and therefore $\hat{y}^*$) to which the entire system gravitates in the long run. To solve for these, all we need to do is substitute $\hat{k}(t) = \hat{k}(t+1) = \hat{k}^*$ and $\hat{h}(t) = \hat{h}(t+1) = \hat{h}^*$ in equations (3.20) and (3.21) to see that

$$\hat{k}^* \simeq \frac{s_k\hat{y}^*}{n+\pi+\delta_k} \text{ and } \hat{h}^* \simeq \frac{s_h\hat{y}^*}{n+\pi+\delta_h}, \text{ where } \hat{y}^* = A\hat{k}^{*a}\hat{h}^{*b}. \quad (3.23)$$

Now we combine the information in (3.23). Substitute the values of $\hat{k}^*$ and $\hat{h}^*$ from the first two equations in (3.23) into the third. Gather the common $\hat{y}^*$-terms to get

$$\hat{y}^* \simeq A^{1/(1-a-c)}\left(\frac{s_k}{n+\pi+\delta_k}\right)^{a/(1-a-c)}\left(\frac{s_h}{n+\pi+\delta_h}\right)^{c/(1-a-c)}. \quad (3.24)$$

[11]For contributions to this literature, see, for example, Uzawa (1965), Lucas (1988), Barro (1991), Mankiw, Romer and Weil (1992) and Bernanke and Gürkaynak (2001), among many others.

[12]Think of $s_h Y(t)$ as the quantity of physical resources spent on education and training.

Don't be intimidated by this equation just because it is long. You've already derived its sibling in the basic Solow model with one accumulable input. That's equation (3.12), which describes the steady state, and we reproduce part of it here for comparison:

$$\hat{y}^* \simeq A^{1/(1-a)} \left(\frac{s}{n + \pi + \delta} \right)^{a/(1-a)} . \tag{3.25}$$

To make the comparison stark, set $s_k = s_h = s$ and $\delta_k = \delta_h = \delta$ in (3.24) to obtain

$$\hat{y}^* \simeq A^{1/(1-a-c)} \left(\frac{s}{n + \pi + \delta} \right)^{a+c/(1-a-c)} , \tag{3.26}$$

and now it is easy to compare the two scenarios: equation (3.26) is just equation (3.25) with a replaced by $a + c$. The presence of an additional accumulable input is tantamount to making steady state output even more sensitive to the savings and population growth rates, a theme to which we return to later. Intuitively, an increase in the savings rate, say s_k, not only increases the accumulation of physical capital, it has a knock-on effect on the savings of *human* capital as well, creating an echo that amplifies the increase in output. Of course, the argument need not stop at just two accumulable inputs. The greater the *joint* share of accumulable relative to non-accumulable inputs, the sharper the response of output to an higher savings in any one of them.

In the extreme case, if all inputs are accumulable, the system effectively behaves like Harrod-Domar, generating endogenous growth. That is, if $a + c = 1$ in our model, then starting from any initial configuration, all outcome variables will ultimately grow at a common steady state *growth rate*, irrespective of any exogenous technical progress. To figure out what that growth rate is, let r denote the ratio of human to physical capital in the long run. So as not drive each other crazy with lots of notation, we will presume that depreciation rates of physical and human capital, and the rates of exogenous technical progress and population growth, are all zero (but the same methods apply anyway). Divide both sides of equation (3.20) by $\hat{k}(t)$ to note that

$$\frac{\hat{k}(t+1) - \hat{k}(t)}{\hat{k}(t)} = s_k r^{1-a},$$

which gives us the growth rate of physical capital. Likewise, divide both sides of equation (3.21) by $\hat{h}(t)$ to see that

$$\frac{\hat{h}(t+1) - \hat{h}(t)}{\hat{h}(t)} = s_h r^{-a},$$

which gives us the growth rate of human capital. Because these two growth rates are the same in the long run (so that the ratio of human to physical capital also stays constant), we must have $s_k r^{1-a} = s_h r^{-a}$, or simply

$$r = s_h / s_k . \tag{3.27}$$

This equation makes perfect sense. The larger is the ratio of saving in human capital relative to that of physical capital, the larger is the long-run ratio of the former to the latter. We can now use this value of r to compute the long-run growth rate. Use any of the preceding growth-rate equations to do this, because all variables must grow at the same rate in the long run. For instance, the growth-rate equation for k tells us that

$$\frac{\hat{k}(t+1) - \hat{k}(t)}{\hat{k}(t)} = s_k r^{1-a} = s_k^a s_h^{1-a},$$

so that the long-run growth rate of all the variables, including per capita income, is given by the expression $s_k^a s_h^{1-a}$.

These arguments parallel our derivation of the Harrod-Domar equation (3.17), and in case you're interested, I've kept all the details — including the proof that r converges — in the Appendix. So we've arrived at a new model of endogenous growth, one that does not rely on physical capital being the only input in production. In this model, both physical and human capital are accumulated together, eroding the diminishing returns to any one of the two inputs as long as there is constant returns to scale in both inputs *combined*. Of course, one might question the absence of any other fixed input, such as raw unskilled labor. In that case, we are back to something like the three-input model in equation (3.19), where one input (L) cannot be deliberately accumulated. We would be back to the Solow model at a qualitative level; see equation (3.24).

That said, this theory shows that it is perfectly possible for there to be diminishing returns to physical capital and yet for there to be sustained, endogenous growth. The lesson we learn is that physical capital is not a privileged input simply by accident. It is special *because it can be accumulated*. To the extent that there are other inputs that can be accumulated as well, such as human capital in this extension, it is the presence of diminishing returns to that entire complex of accumulable inputs that will determine whether growth is endogenous or exogenous. If, as in this model, there is constant returns to that complex of inputs (two of them here), endogenous growth is possible, In the Harrod-Domar model, that "complex" is just a single input — physical capital.

In the next chapter, we will see how this extended model can throw light on the question of convergence or divergence of cross-country incomes. But for now, we make just two closing observations. The first is a reminder that other extensions are possible that will also lead to endogenous growth. A leading example of such an extension is directed technical progress, progress that occurs because of the deliberate diversion of resources to scientific activity. At a very high level, we can think of three inputs in production — capital, labor and "technical know-how" — or four if you like, if you want to additionally retain the division between physical and human capital. In such a model, the behavioral sources of growth would be threefold: the accumulation of physical capital, the accumulation of human capital, and the deliberate accumulation of technical knowledge. If there is constant or even increasing returns in this complex of accumulable inputs, economic growth will be endogenous, just as in the Harrod-Domar model or in the extension with human capital that we've just studied.

3.6. Summary

This chapter began our study of *economic growth*, defined as annual rates of change in income or capital (total or per-capita). We observed that growth rates of 2% or more per-capita are a relatively modern phenomenon, systematically attained first by the United States, and only in the 20th century. Developing economies are way behind in per-capita income relative to their developed counterparts, and faster growth rates are consequently high on their priority list.

Our study of growth theory began with the basic notion of macroeconomic balance. Savings equals investment: abstention from current consumption paves the way for increases in capital equipment. And more capital creates more output. Thus two parameters are immediately relevant: the *savings rate*, which tells us how much an economy abstains from consumption, and the *production function*, which tells us how the resulting increase in capital translates into output. This allows us to derive the fundamental growth equation (3.7).

The Solow model builds on this equation. It emphasizes diminishing returns to capital and labor, and in particular diminishing returns in per-capita output as a function of the capital-labor ratio. If capital grows faster than the labor force, then each unit of capital has less labor to work with, so that average output per unit of capital falls. That implies that savings — which are proportional to output — fall *relative to the capital stock*, which slows down the rate of growth of capital. (Exactly the opposite happens if capital is growing too slowly relative to labor.) This mechanism ensures that in the long run, capital and working population grow exactly at the same rate, and per-capita growth ultimately vanishes. Capital and labor maintain a constant long-run balance at the *steady-state* capital stock (per-capita).

From this perspective, growth dies out because there is no technical progress. We introduced this next, thinking of such progress as ongoing growth in knowledge that continually increases the productivity of labor. It then became important to distinguish between the working population and *effective labor*, which is the working population multiplied by (the changing level of) productivity. Thus effective labor grows as the sum of population growth and technical progress. With this amendment, the Solow arguments apply exactly as before, with all per-capita magnitudes re-expressed per unit of effective labor. This means, for instance, that while the long-run capital stock *relative to effective labor* settles down to a steady-state ratio, the capital stock per person keeps growing and it does so at the rate of technical progress. Likewise, per-capita income keeps increasing in the long run precisely at the rate of technical progress. In this sense, the Solow model is one of *exogenous growth*, as long as we view the rate of technical progress as some non-behavioral parameter.

Finally, we introduced two variants on the Solow framework. We noted that if production is entirely dependent on physical capital, then sustained output growth can be maintained at all per-capita income levels. Now behavioral parameters such as savings rates and population growth rates can be seen to affect not just the levels of output but also the rate of per-capita income growth. The resulting model, due to Harrod and Domar, can therefore be thought of as a prototypical model of *endogenous growth*, in the sense that human behavior can directly and persistently affect the long run rate of growth.

Our second extension included human capital along with physical capital, and we studied a model in which both inputs could be freely accumulated. This, too, could give rise to a model of endogenous growth provided that there is constant returns to the joint pair of physical and human capital inputs. There are two important takeaways from this second extension. The first is that the responsiveness of per-capita steady state income to various parameters increases as the share of *accumulable* factors in production increases. Physical capital is one of those factors, but education or human capital can be another. This is a topic to which we shall return in Chapter 4.

In particular, endogenous growth is not limited to a setting in which there is constant returns to the physical capital input alone, as in the Harrod-Domar model. It could occur whenever there is constant (or even increasing) returns to the entire complex of accumulable inputs. In the particular variant we studied, that complex was physical and human capital, but it could also include other factors, such as technical knowledge if deliberately accumulated — think of knowledge as an input like any other. If all those inputs are not sufficient to generate constant or increasing returns together, then we are back to the Solow model (or some close relative of it), where all growth must perforce be exogenous.

3.7. Appendix

Returns to Scate for General Production Functions. Recall that for two inputs, say capital and labor, a production function can be written as

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

Suppose that we scale all inputs by some common factor $\lambda > 0$. (Note that this implies an increase in all inputs if $\lambda > 1$ and a decrease in all inputs if $\lambda < 1$.) We say that there is *constant returns to scale* if

$$F(\lambda K, \lambda L) = \lambda F(K, L) \text{ for every } \lambda > 0. \tag{3.28}$$

The left hand side of this equality is the new output $F(\lambda K, \lambda L)$ produced when each input is scaled by λ. The right hand side is the old output, which is just $F(K, L)$, scaled by λ. The definition states that the new output equals the old output scaled by λ.

The definitions of increasing and decreasing returns to scale are written in similar fashion but we must be careful about the direction of scaling. We say that there is *increasing returns to scale* if

$$F(\lambda K, \lambda L) > \lambda F(K, L) \text{ for every } \lambda > 1,$$

which means that the new output climbs by *more* than λ when inputs are each scaled upward by $\lambda > 1$. Likewise, there is *decreasing returns to scale* if

$$F(\lambda K, \lambda L) < \lambda F(K, L) \text{ for every } \lambda > 1,$$

which means that the new output climbs by *less* than λ when inputs are each scaled upward by $\lambda > 1$. In the special case of a Cobb-Douglas production function, we have

$$Y = AK^a L^b,$$

for constants $(A, a, b) \gg 0$. You should verify that constant returns to scale imply $a + b = 1$, increasing returns to scale imply $a + b > 1$, and decreasing returns to scale imply $a + b < 1$.

Per-Capita Version of the Production Function. Production functions with constant returns to scale have a nice property. They allow us to relate *per-capita* input to *per-capita* output. To see this, choose $\lambda = 1/L$ in equation (3.28). Then

$$\frac{F(K, L)}{L} = F\left(\frac{K}{L}, 1\right).$$

But $Y = F(K, L)$, so we can conclude that

$$\frac{Y}{L} = F\left(\frac{K}{L}, 1\right), \text{ or } y = F(k, 1), \tag{3.29}$$

where as in the main text, $y = Y/L$ and $k = K/L$. Now we can simply rewrite $F(k,1)$ in more compact form as $f(k)$ for some function f. Then equation (3.29) turns into

$$y = f(k), \tag{3.30}$$

which is the relationship are looking for. In the main text, $f(k)$ corresponds to the function Ak^a; see equation (3.3).

A warning to those who are seeing this for the first time: equation (3.30) is deceptive in its simplicity. It represents production relationships between per-capita levels *only*; that is, between per-capita capital k and per-capita output y. You *cannot* express y as a function of k alone when returns to scale are not constant. The scale of operations will matter. As an exercise, try going through the same process with the Cobb-Douglas production function, both when returns to scale are constant ($a + b = 1$) and when they are not ($a + b \neq 1$), and appreciate why your steps must fail in the latter case.

Assume there is constant returns to scale, The following facts about $f(k)$ can be proved, and you should try and establish these.

[1] If $F(K,L)$ is increasing in each of its arguments, then $f(k)$ is increasing in k.

[2] If $F(K,L)$ exhibits diminishing returns to each of its inputs, then $f(k)$ exhibits diminishing returns in k. Hint: a little calculus goes a long way; recall that $f(k) = F(k,1)$ and differentiate twice.

Point [2] is worth noting carefully. It says that if there are diminishing returns to both factors in the production function, then the *per-capita* production function displays diminishing returns in per-capita capital used in production. In the main text, we used constant returns to scale to arrive at the per-capita description of a production function, and then used diminishing returns to each input *separately* to assert that per-capita output exhibits diminishing returns with respect to per-capita capital. This particular shape lay at the heart of the predictions of the Solow model.

More on the Endogenous Growth Model with Human Capital. Recall the extended production function with three inputs, reproduced here:

$$Y = AK^a(eL)^b H^c. \tag{3.31}$$

Written this way without any further interpretation of (or restriction on) H, it should be clear that H and c have no real *separate* meaning: mathematically, we could always create a new way of measuring H, say, by raising it to any power we please, and c would have to adjust accordingly.[13] The "units" of H (and therefore c) only has meaning once we describe how it changes or how it's acquired. In the main text, we took the view that H is human capital, but viewed in this way H could just as easily be a model of endogenous TFP. For instance, first shut down exogenous technical progress by setting $e = 1$ in (4.7), and then define a new variable

$$z = H^{c/b}, \text{ so that as a consequence, } Y = AK^a(zL)^b,$$

which means that z plays exactly the same role as e, but in addition it is *endogenously* accumulated. (The savings generated by s_h change the level of d.) So, variants of our setup, while similar in their algebra, add convenient flexibility to our interpretations.

[13] For instance, measure productivity as $D \equiv \sqrt{H}$. In that case we would replace H^c in (3.19) by the equally kosher formulation D^{2c}.

In the main text we focused on the case $a + c < 1$, and then described the steady state $(\hat{k}^*, \hat{h}^*, \hat{y}^*)$. We also discussed how the system changed into an endogenous growth model once every input is accumulable, which in the constant returns to scale setting implies $a + c = 1$. In this Appendix we show how to calculate the endogenous rate of growth in this case, and also provide a convenient formula for the long run *ratio* of human to physical capital.

Define r to be the ratio H/K, which of course is just the same as $\hat{h}/\hat{k}$. Because $a + c$ is equal to 1, in what follows we will simply replace c by $1 - a$. The production function (3.22) in its per-capita form can be rewritten to see that

$$\frac{\hat{y}}{\hat{k}} = r^{1-a} \text{ and } \frac{\hat{y}}{\hat{h}} = r^{-a}, \tag{3.32}$$

In the main text, we assumed that depreciation rates of physical and human capital, as well as the rates of technical progress and population growth, are all zero (but our method of analysis doesn't depend on any of these assumptions). Then, combining (3.20) and (3.32), we see that

$$(1+n)\frac{\hat{k}(t+1)}{\hat{k}(t)} = s_k \frac{\hat{y}(t)}{\hat{k}(t)} + 1 = s_k r(t)^{1-a} + 1, \tag{3.33}$$

and likewise combining (3.21) and (3.32),

$$(1+n)\frac{\hat{h}(t+1)}{\hat{h}(t)} = s_h \frac{\hat{y}(t)}{\hat{h}(t)} + 1 = s_h r(t)^{-a} + 1. \tag{3.34}$$

Equations (3.33) and (3.34) together yield

$$\frac{r(t+1)}{r(t)} = \frac{s_h r(t)^{-a} + 1}{s_k r(t)^{1-a} + 1},$$

and manipulating this a bit gives us the two equivalent forms

$$r(t+1) = \frac{s_h}{s_k}\frac{1 + [r(t)^a/s_h]}{1 + [r(t)^{a-1}/s_k]} = r(t)\frac{[s_h/r(t)] + r(t)^{a-1}}{s_k + r(t)^{a-1}}. \tag{3.35}$$

Now we can use (3.35) to establish the following two Observations:

[1] If at any date t, we have $r(t) > s_h/s_k$, then it must be that $r(t) > r(t+1) > s_h/s + k$.

[2] If at any date t, we have $r(t) < s_h/s_k$, then it must be that $r(t) < r(t+1) < s_h/s_k$.

These two Observations, coupled with the statement (see main text) that once $r(t)$ *equals* s_h/s_k it will stay right there, constitute the proof of convergence.

We will prove Observation 1. (The proof of Observation 2 is analogous.) Imagine that at some date t, we have $r(t) > s_h/s_k$. Then it is easy to see that

$$\frac{1 + [r(t)^a/s_h]}{1 + [r(t)^{a-1}/s_k]} > 1,$$

and combining this with the first equality in (3.35), we see that $r(t+1) > s_h/s_k$.

Likewise, it is easy to see that

$$\frac{[s_h/(r(t)] + r(t)^{a-1}}{s_k + r(t)^{a-1}} < 1,$$

and combining this observation with the second equality in (3.35), we conclude that $r(t) > r(t+1)$. This proves Observation 1, and our proof of convergence is complete.