Economics 7343

Macroeconomic Theory I

Dietrich Vollrath

Fall • 2017

Contents

CHAPTER 1

Preliminaries

To begin we need to establish a few accounting identities and definitions that will be used extensively in the course. The basis for the identities are the national income product accounts, and the various breakdowns that this provides for aggregate output.

1.1 Gross Domestic Product

Gross domestic product (GDP) is the total value of all goods and services produced within the borders of a given country in a given period (e.g. a year). We'll denote aggregate GDP as Y_t. If we have N_t people, then GDP per capita is denoted $y_t = Y_t/N_t$. Other variables will be defined similarly, with capital letters denoting aggregates and lower-case letters per-capita terms.

Now, there are various ways to break down GDP. Let's start with the income breakdown. National product accounts tell us that GDP can be split up into

$$Y_t = W_t + r_t K_t + \delta K_t. \tag{1.1}$$

The first element is compensation, or wages. We assume each person earns w_t in wages, so aggregate wages are $W_t = w_t N_t$. The second component, $r_t K_t$, is called "operating surplus" in the national accounts, and for us represents the total return to capital. r_t is the return to each unit of capital, and K_t is the amount of physical capital in the economy. The final term, δK_t, is depreciation. δ represents the fraction of the capital stock, K_t, that breaks down in period t. The inclusion of the depreciation term is why Y_t is known as "Gross" domestic product. "Net" domestic product would be $NDP_t = W_t + r_t K_t$.

An alternative way of breaking down GDP is the expenditure method. This will be familiar from your intermediate class in macro. Here we define

$$Y_t = C_t + I_t + G_t + X_t - X_t^M \tag{1.2}$$

where C_t is consumption spending, I_t is investment spending, G_t is government spending, and $X_t - X_t^M$ is net exports.

Finally, we will have a production definition of GDP. This, in your intermediate class, might have been referred to as the "value-added" approach. Here we will specify the production of GDP as

$$Y_t = F(K_t, N_t) \tag{1.3}$$

which says that GDP is a function (not surprisingly referred to as the "production function") of K_t and N_t, the capital stock and the population, respectively. Note that we do not specify individual value-added by firm, for example. Rather we provide a single aggregate production function that sweeps up all these value-added's together. The important point will be that GDP depends upon the factors of production - K_t and N_t - in the economy.

Note that each of these three are accounting identities. They hold by definition at all points in time. They are simply three different ways of conceiving of one single object, GDP. The are related in the circular flow diagram that you might remember from intermediate macro. GDP definitions, in various iterations, will form the budget constraints in our optimization problems.

1.2 Investment and Accumulation

One of the first things we'll study is the Solow model (and an extension called the Ramsey model) which involves the decision about how much to consume (C_t) and how much to invest (I_t). To keep our lives simple, we'll begin by assuming that $G_t = 0$ and $X_t - X_t^M = 0$.

The only reason you'd spend money on investment goods is because you expect to get something in return. Namely, future GDP that you can consume. So one of the first things we'll do is add the following dynamic equation to our arsenal

$$\Delta K_{t+1} \equiv K_{t+1} - K_t = I_t - \delta K_t \tag{1.4}$$

which tells us how capital tomorrow is determined by the capital stock today, as well as our investment decision today. Using the expenditure definition of GDP, we know that $I_t = Y_t - C_t$, so we can write down the following

$$\Delta K_{t+1} = Y_t - C_t - \delta K_t \tag{1.5}$$

which describes the economy-wide accumulation of capital. When we study the Solow model, we'll make some simple assumptions about how C_t is determined, while the Ramsey model will make the decision regarding C_t the outcome of an explicit utility maximization problem.

Recalling the production definition of GDP, $Y_t = F(K_t, N_t)$, we can write

$$\Delta K_{t+1} = F(K_t, N_t) - C_t - \delta K_t \tag{1.6}$$

which is a difference equation in K_t. In the Solow model we'll be attempting to solve this difference equation to find out how big the capital stock is at any given point in time.

1.3 Production, Wages, and the Return to Capital

As noted, one of the more important elements involved in our work is the production function, $F(K_t, N_t)$. We're going to use several assumptions regarding this function, as well as assumptions regarding the competitiveness of production, to figure out more explicitly the terms W_t and r_t.

First, we'll assume that the function F is constant returns to scale. Specifically, this means that

$$zF(K_t, N_t) = F(zK_t, zN_t) \tag{1.7}$$

or if we scale up both inputs by z, then output scales up by exactly z as well. To go further, take the derivative of the above with respect to z, which yields

$$F(K_t, N_t) = F_K K_t + F_N N_t \tag{1.8}$$

which, in effect, gives us another way of decomposing GDP. We can divide GDP up into the total product of capital ($F_K K_t$) and the total product of labor ($F_N N_t$).

Now, if the economy is perfectly competitive, then firms should be paying each unit of labor a wage equal to it's marginal product,

$$W_t = w_t N_t = F_N N_t. \tag{1.9}$$

Because of the constant returns to scale, we know that if firms pay out wages in this manner, then it must be that $r_t K_t + \delta K_t = F_K K_t$. This means that the return to capital is $r_t = F_K - \delta$. Owners of capital earn it's marginal product (similar to workers earning a wage equal to their marginal product), but their capital is depreciating at the rate δ, so their net return is $F_K - \delta$.

Let's put what we know together. Essentially, all we have are a set of different ways of defining GDP

$$\begin{aligned} Y_t &= F(K_t, N_t) && (1.10) \\ &= w_t N_t + r_t K_t + \delta K_t && (1.11) \\ &= w_t N_t + F_K K_t && (1.12) \\ &= C_t + I_t && (1.13) \\ &= C_t + \Delta K_{t+1} + \delta K_t && (1.14) \end{aligned}$$

and recall that all the items on the right-hand side are equal to each other as well. We know, for example, that

$$C_t + \Delta K_{t+1} = w_t N_t + r_t K_t \tag{1.15}$$

where the depreciation term has canceled. This enforces the same budget constraint as does the economy-wide accumulation equation in (1.6). The above, though, gives you a constraint that highlights how net income (wages plus returns to capital) is allocated to either consumption or an increase in the capital stock.

The most important point is that whether we are doing the Solow model, focusing on the accumulation equation in (1.6), or doing the Ramsey model, focusing on individual choices over consumption and accumulation using (1.15), we are using the same constraint.

1.4 The Consumption Problem

We'll spend some time focusing on people's decisions about what consumption should be. Their *individual* budget constraint will be

$$c_t + \Delta a_{t+1} = w_t + r_t a_t \tag{1.16}$$

which says that your income (wages plus the return on your assets, a_t) can be used for two things: consumption and changing your assets.

Note that this is really similar to the aggregate constraint above. If this is the individual constraint, and there are N_t individuals who are all identical, then in aggregate we'd have

$$N_t c_t + N_t \Delta a_{t+1} = N_t w_t + r_t N_t a_t \tag{1.17}$$

$$C_t + \Delta A_{t+1} = w_t N_t + r_t A_t \tag{1.18}$$

which looks even more like our aggregate constraint. Finally, if we assume that the economy is closed, then individuals assets have to be equal to the total amount of assets in the economy, which for now consists only of physical capital, so $A_t = K_t$. This leaves us with

$$C_t + \Delta K_{t+1} = w_t N_t + r_t K_t \tag{1.19}$$

which is just our aggregate constraint above. The point is that when we go through our individual problem we'll be able to use this to inform us about how aggregate consumption moves: *provided our assumption about identical individuals holds*. If not, we'd have more work to do.

1.5 Government Budgets

After we've covered the consumption and investment decisions in detail, we'll start adding back in the other elements of expenditure. First, consider government spending. For this, we can add an additional constraint that dictates how the governments finances evolve.

$$G_t + (1 + r_t^b) B_t = B_{t+1} + T_t \tag{1.20}$$

This says that government spending, G_t, plus the amount owed on government debt, B_t, which has a real interest rate of r_t^b, must be equal to government revenues. These are either taxes, T_t, or bonds that are due in the next period, B_{t+1}. For the moment, all bonds are only one-period bonds, so we aren't going to concern ourselves with things like 10 versus 30 year bonds.

Going back to our GDP definitions, we have now that

$$\begin{aligned} Y_t &= C_t + I_t + G_t && (1.21)\\ &= C_t + I_t + T_t + \Delta B_{t+1} - r_t^b B_t && (1.22)\\ &= C_t + \Delta K_{t+1} + \delta K_t + T_t + \Delta B_{t+1} - r_t^b B_t && (1.23)\\ &= w_t N_t + r_t K_t + \delta K_t && (1.24)\\ & && (1.25) \end{aligned}$$

which we can use to write

$$C_t + \Delta K_{t+1} + \Delta B_{t+1} = w_t N_t + r_t K_t + r_t^b B_t - T_t \tag{1.26}$$

and we have an expanded version of the constraint facing the economy. Now, there are three options for your net income: consumption, increasing capital, and increasing bond holdings. Your income is wages, returns to capital, returns to bond holdings, minus taxes.

1.6 Open Economies

We introduced government, so now let's add the last element of national accounts, net exports and imports. Once we do this, we'll need to distinguish GDP from GNP. We know what defines GDP. Gross National Product is

$$GNP_t = W_t + r_t A_t + \delta A_t \tag{1.27}$$

which looks a lot like the income definition of GDP, except that now we have assets (A_t) rather than the physical capital stock (K_t). The distinction matters because, in an open economy, total assets owned by citizens do not have to exactly equal the physical capital stock. (In a closed economy, by definition, $A_t = K_t$, so GDP = GNP). One other note is that I've implicitly assumed that assets earn an identical return of r_t no matter where they are located. You could be more refined and have the return earned on domestic assets be different from that earned on foreign. For now, let's keep them identical.

Let's define a new object, net foreign assets, F_t. This is

$$F_t = A_t - K_t. \tag{1.28}$$

If $F_t > 0$, then citizens have total assets higher than the total assets in the domestic economy (which is just the physical capital stock), so they must own foreign assets (i.e. physical capital located in other countries). If $F_t < 0$, then citizens have assets less than total assets in the domestic economy, and so some of the domestic capital stock must be owned by foreigners.

With this definition we get

$$\begin{aligned} GNP_t &= W_t + (r_t+\delta)(F_t+K_t) && (1.29)\\ &= W_t + (r_t+\delta)K_t + (r_t+\delta)F_t && (1.30)\\ &= Y_t + (r_t+\delta)F_t && (1.31) \end{aligned}$$

or GNP is just GDP adjusted by the return on net foreign assets.

The accumulation of physical capital in the domestic economy takes place according to the same mechanics: $\Delta K_{t+1} = I_t - \delta K_t$. However, now that assets and physical capital don't need to be identical for our citizens, we need an accumulation equation for assets as well.

$$\Delta A_{t+1} = GNP_t - C_t - G_t - \delta A_t \tag{1.32}$$

which says that you take your total income (GNP), subtract off all of the non-investment spending done on domestic goods (C_t and G_t), and subtract off the depreciation of your existing assets. The net result is the change in your asset holdings.

Using the prior decomposition of GNP, and noting that $\Delta A_{t+1} = \Delta F_{t+1} - \Delta K_{t+1}$, by definition, we get

$$\begin{aligned} \Delta A_{t+1} &= Y_t + (r_t+\delta)F_t - C_t - G_t - \delta(F_t+K_t) && (1.33)\\ \Delta A_{t+1} &= Y_t + r_tF_t - C_t - G_t - I_t - \delta K_t + I_t && (1.34)\\ \Delta A_{t+1} - (I_t - \delta K_t) &= Y_t - C_t - G_t - I_t + r_tF_t && (1.35)\\ \Delta A_{t+1} - \Delta K_{t+1} &= X_t - X_t^M + r_tF_t && (1.36)\\ \Delta F_{t+1} &= X_t - X_t^M + r_tF_t && (1.37) \end{aligned}$$

which is a big mess, but actually has some meaning. The right-hand side is called the *current account*. This is net exports plus the payments on net foreign assets (which if $F_t < 0$ could be negative). The term on the left is the *capital account*, the change in net foreign asset holdings.

This, again, is simply accounting. It tells us that you have to pay for what you get. If you are acquiring foreign assets ($\Delta F_{t+1} > 0$), then you must be paying for them with either exports (in an amount greater than you import), or with the earnings on your existing net foreign assets. If foreign countries are acquiring assets in your country on net ($\Delta F_{t+1} < 0$), then they are paying for them with imported goods or using the net payments they get from already owning assets in your country. Note, though, that this is simply another transformation of the accounting equations for GDP and GNP, along with the definitions of asset and capital accumulation. We haven't done any economics yet.

CHAPTER 2

The Solow Model

In this chapter we build up a basic Solow model of the economy. The dynamic nature of the economy arises because the amount of savings done today will influence the capital stock, and hence production, in the future. Thus the split of current output between consumption and savings is crucial. Within the Solow model we can (crudely) discuss the source of fluctuations in GDP around trend, the determination of the trend growth rate of GDP, openness to foreign capital flows, and the influence of taxation and government spending.

The mechanics of the Solow model sit at the heart of nearly every macroeconomic model. By itself, you can think of it as the most stripped down model of the macroeconomy that we can usefully put down on paper. This stripping down makes several heroic assumptions that we'll slowly relax over the course of the class. The assumptions aren't made because we think they are strictly true, but rather because we can't possibly hope to model the economy precisely, and we have to accept some simplifications to proceed. Hopefully our assumptions are relatively unimportant, and what we're left with in our model is useful in explaining why GDP moves the way it does.

2.1 Firms and Production

Let's begin with the determination of output. We will assume that all firms in the economy produce an identical good. That is, there is no specialization or differentiation between firms. There is thus no possibility of an individual firm having market power. We will also assume that there is free entry into production, so that the existing firms cannot establish any kind of cartel to gain market power. In short, we are assuming that there is perfect competition among firms in producing output, and hence zero profits.

These are going to be very restrictive assumptions, as it turns out. Without market power, prices will adjust instantly, and so there will be no way for nomianl shocks to generate fluctuations in the economy. Without market power and profits, there will be no incentives to innovate, and hence no deliberate productivity improvements. We'll relax those eventually, but this will form a good starting point.

Each firm is assumed to produce output according to the following production function

$$Y_i = F(K_i, N_i) \tag{2.1}$$

where K_i is the capital stock used by firm i and N_i is the labor used by firm i. The function $F(.,.)$ is identical for each firm, and this function is constant returns to scale (see the following boxed section for a definition of returns to scale).

Production Function Properties

One of the primary elements of any macro model is the production function. From a general perspective, there are several properties and terms worth understanding in more detail.

Write a production function in general as

$$Y = F(X_1, X_2, ...X_n) \tag{2.2}$$

so that output, Y, is produced by some combination of the factors of production denoted by X, of which there are n total. Typically, we'll use $n = 2$, and X_1 is equal to capital while X_2 is equal to labor. But we need not necessarily be that restrictive.

The first property to think about is returns to scale. Consider multiplying each factor of production by some factor z. Returns to scale refers to how this scaling affects output. In particular,

$$\text{Decreasing Returns} : F(zX_1, zX_2, ..., zX_n) < zY \tag{2.3}$$

$$\text{Increasing Returns} : F(zX_1, zX_2, ..., zX_n) > zY \tag{2.4}$$

$$\text{Constant Returns} : F(zX_1, zX_2, ..., zX_n) = zY. \tag{2.5}$$

Decreasing returns means, practically, that if you double inputs used, you get less than double the output. Increasing returns implies that you can more than double output by doubling inputs, and constant returns means output exactly doubles. In almost every case, we'll assume constant returns.

The marginal product of a factor of production is simply the derivative of the production function with respect to that factor.

$$MPX_i = \frac{\partial F(X_1, X_2, ...X_n)}{\partial X_i}. \tag{2.6}$$

Finally, we can think about substitution between factors of production. This is related to the concept of an isoquant. What combinations of factors of production produce an identical level *of output? This is most simply seen with a two-factor production function:* $Y = F(X_1, X_2)$. *The marginal rate of technical substitution between input 1 and 2 is, holding output constant*

$$MRTS_{1,2} = \frac{\partial X_2}{\partial X_1} = \frac{-MPX_1}{MPX_2} \tag{2.7}$$

which follows from the Implicit Function Theorem. This says that the isoquant is downward sloping (the negative slope) and depends on the relative marginal productivity of the two inputs. If the MRTS is small (in absolute value), then the marginal productivity of factor 2 is very large relative to 1. So to keep output constant, I would have to add a lot of factor 1 to replace the loss of a little of factor 2.

In addition, we assume that that $F(0,0) = 0$. Without inputs a firm can produce no outputs. The production function is presumed to exhibit diminishing returns to capital:

$$F_K(K_i, N_i) > 0 \text{ and } F_{KK}(K_i, N_i) \leq 0 \tag{2.8}$$

An increase in K_i increases output, but at a decreasing rate. Mathematically, we are assuming that the production function is concave in K_i. Similarly for labor we assume that

$$F_N(K_i, N_i) > 0 \text{ and } F_{NN}(K_i, N_i) \leq 0, \tag{2.9}$$

and we also assume that

$$F_{KN}(K_i, N_i) = F_{NK}(K_i, N_i) > 0. \tag{2.10}$$

This last assumption says that the marginal product of one factor is raised by the addition of the other. If we add labor, for example, then the marginal product of capital is higher.

A final set of assumptions we make regarding the production function is how it behaves as it approaches extreme values. Specifically

$$\lim_{K \to \infty} F_K(K_i, N_i) = 0 \text{ and } \lim_{K \to 0} F_K(K_i, N_i) = \infty. \tag{2.11}$$

These are known as the Inada (1964) conditions. They imply that when there is almost no capital in the firm, the marginal gain in output from a small amount of capital is infinite, but as they acquire large amounts of capital, the marginal product actually comes close to zero.

Given all this, our firms will be trying to maximize profits (π_i)

$$\pi_i = F(K_i, N_i) - RK_i - wN_i \tag{2.12}$$

where R is the rental rate that firms have to pay for capital and w is the wage they have to pay for labor. We are thus assuming that factor markets are perfectly competitive, and that firms have no monopsony power.

To maximize profits, firms will have the following first-order conditions

$$F_K(K_i, N_i) = R \text{ and } F_N(K_i, N_i) = w. \tag{2.13}$$

Given that every firm has an identical production function, and faces an identical R and w, every firm will make an identical choice regarding the amount of K_i and N_i to hire in.

Exactly how much capital and labor is that? That depends on the supplies of those factors of production. Let the aggregate stock of capital be K, and the aggregate stock of labor be N. We assume that both of these are supplied inelastically, meaning their supply curves are vertical, and do not respond at all to w or R.

The Cobb-Douglas Production Function and Income Distribution

The most common form of the production function that is used is the Cobb-Douglas, which has the following form:

$$Y = K^\alpha N^{1-\alpha} \tag{2.14}$$

where $\alpha \in (0,1)$. You can confirm easily that the Cobb-Douglas (CD) has constant returns to scale.

The marginal product of each factor in the CD is

$$MPK = \alpha K^{\alpha-1} N^{1-\alpha} \tag{2.15}$$

$$MPN = (1-\alpha) K^\alpha N^{-\alpha}. \tag{2.16}$$

Firms will set $MPK = R$, and $MPN = w$, giving us a way of working out exactly what the return to capital and wage will be in an economy with the CD production function in each firm (and perfect competition in output markets and factor markets).

We can consider the distribution of income between labor, capital, and profits. That is, what fraction of total income is made up of wages, or returns to capital, or profits? Start with labor and capital. Their shares of income can be written as

$$\frac{KR}{Y} = \frac{K \times MPK}{Y} = \frac{K \alpha K^{\alpha-1} N^{1-\alpha}}{Y} = \alpha \tag{2.17}$$

$$\frac{Nw}{Y} = \frac{N \times MPN}{Y} = \frac{(1-\alpha) K^\alpha N^{-\alpha}}{Y} = 1-\alpha. \tag{2.18}$$

What this shows is that exactly α of total income will be in the form of payments to capital, while the remaining $1-\alpha$ of income will be paid out as wages. As these shares add up to one, the share of income left to be paid out as profits is exactly zero (here we are talking about economic profits, not accounting profits).

This fits. We have started with an assumption of perfect competition between firms, and free entry. This should drive economic profits down to zero. There is nothing special about *the CD that ensures this is true. What the CD function adds is the fixed shares α and $1-\alpha$ for capital and labor. Regardless of the actual level of output, exactly α of output is used to pay for capital services, and the rest for labor. Cobb and Douglas developed this production function form precisely to match U.S. data that showed labor's share of output to hold steady at around 2/3 for several decades. This stability in labor's share is one of Kaldor's (1957) stylized facts regarding growth, although there is recent evidence from Loukas Karabarbounis and Brent Neiman (2013) that labor's share is actually declining in several countries.*

Across countries, Doug Gollin (2002) found that there is no tendency for labor's share to change as countries get richer. That does not mean labor's share is identical in every country - it varies between about 0.50 and 0.85. But there is not tendency for poor countries to have high labor shares (or vice versa).

Let there be M firms in the economy. M can be any positive finite number. One issue with assuming perfect competition is that we can't actually pin down the number of firms precisely. This means that we can easily assume that $M = 1$ if that is convenient. With free entry, even a single firm will act as a price-taker.

With M firms, it must be that

$$K_i M = K \text{ and } N_i M = N, \tag{2.19}$$

or the total demand equals total supply for both factors of production. These can obviously be rearranged to show that $K_i = K/M$ and $N_i = N/M$ are the amounts of capital and labor used by each firm.

Knowing this, we can say something about aggregate output. As we assumed that each firm produces an indentical good, we can write aggregate output as

$$Y = \sum_{i=1}^{M} Y_i = \sum_{i=1}^{M} F(K_i, N_i). \tag{2.20}$$

Here we can plug into the firm-level production function with what we know about K_i and N_i to reduce this to

$$Y = \sum_{i=1}^{M} F(K_i, N_i) = \sum_{i=1}^{M} F(K/M, N/M) = \sum_{i=1}^{M} \frac{F(K,N)}{M} = F(K,N). \tag{2.21}$$

In the above series of equations we can pull the $1/M$ out of the production function due to the constant returns to scale property of $F(.,.)$. Once we do this and sum over the M firms, the M will cancel and we're left with the aggregate production function

$$Y = F(K,N). \tag{2.22}$$

Often times we'll jump right to the aggregate production function, just assuming that $Y = F(K,N)$ exists. However, it is very useful to understand exactly what is going on behind that aggregate production function. As you can see, there are a whole series of assumptions built into the aggregate production function. In particular, perfect competition in output markets and factor markets. Having those assumptions gives us a nice aggregate production function to work with, but as mentioned above eliminates several potentially interesting aspects of the economy.

2.2 Accumulation and Dynamics

The aggregate production function says that output at time t is $Y_t = F(K_t, N_t)$. So to describe the time path of GDP we need to describe the time paths of K_t and N_t. If we know those we can plug through the production function and get Y_t. Note that we already have some answers here. In the baseline Solow model, the only possible source of fluctuations or growth are the stocks of capital and labor. If they fluctuate or grow, then so will output.

Let's begin with capital. We will describe it's dynamics as follows

$$K_{t+1} - K_t \equiv \Delta K_{t+1} = I_t - \delta K_t \tag{2.23}$$

where I_t is the amount of investment done in time t. This is simply the amount of output that is set aside for use as capital in period $t+1$. Recall that all output is identical, so there is no sense that we are producing machine tools (an investment good) that is differentiated from food (a consumption good). Some of the homogenous output is "invested", meaning simply that it is set aside.

The second term above is depreciation. A fracton δ of the existing stock of capital K_t just falls apart or breaks down every period. We assume that this rate is unaffected by anything, and is unavoidable.

What determines the level of investment? We will assume that it is described as follows

$$I_t = S_t, \tag{2.24}$$

or investment is exactly equal to the amount of output saved (S_t). Note that this is an assumption, and not an accounting identity. This assumes two major things. First, that all saved output is costlessly and perfectly translated into investment goods. If we want to imagine a financial sector taking in savings and loaning these out to firms, then the financial sector is doing this without charging any fees and without any loss of output. Secondly, there is no extra inflow of savings from abroad (or outflow of home savings to other countries). This economy is closed to capital flows.

This then leads to another assumption regarding the level of savings. We assume that it is a constant fraction of output (s)

$$S_t = sY_t. \tag{2.25}$$

In other words, the proportion saved is inelastic with respect to the rate of return on savings, the level of income, or anything else. In this economy it is always the case that a fraction s of output is set aside to be saved.

If we put together all of our various pieces, we are left with the following difference equation for capital

$$\Delta K_{t+1} = sF(K_t, N_t) - \delta K_t. \tag{2.26}$$

As one can see, this describes capital growth as a function of capital itself. This is a non-linear difference equation, because the function $F(.,.)$ is non-linear in capital.

The other moving part in this economy is the labor force. We assume that this grows at an exogenous rate, n

$$\Delta N_{t+1} = nN_t. \tag{2.27}$$

And that's it. We have an equation describing how capital changes over time, and one that describes how the labor force changes over time. Knowing K_t and N_t at any point in time, we can find Y_t from the production function.

The Solow model has several implications that we can compare to the data. It will turn out that the model does well in describing several economic facts we observe, but fails on others, which will be part of our motivation to expand on the model. To see these implications we will take the Solow model and write it in a more concise form.

In particular, we are generally interested in per capita GDP, rather than simply aggregate GDP. Per capita GDP is a good proxy of living standards, and our "big questions" are all about living standards.

Per capita GDP is simply Y_t/N_t, where we are now assuming that the labor force (N_t) is identical to the population. This is a relatively minor assumption, but one that can be relaxed for more realism. Regardless, this leads us to

$$y_t = \frac{Y_t}{N_t} = \frac{F(K_t, N_t)}{N_t} = F\left(\frac{K_t}{N_t}, 1\right) = F(k_t, 1) = f(k_t). \tag{2.28}$$

Let's walk though this chain of equations. The first is just a notational modification. Lower-case y_t refers to per-capita GDP to save us from having to write Y_t/N_t over and over again. This standard will be maintained throughout the class. Capital letters refer to aggregate values, while lower case letters refer to per-capita values.

In the second equation I've simply plugged in the production function. The third equation uses the constant returns to scale of $F(.,.)$ to divide through by N_t. In the fourth equation, I've replaced K_t/N_t with it's per-capita notation, k_t, capital per capita.

In the final equation I've made one more notational adjustment. Rather than write $F(k_t, 1)$ over and over again, with that bothersome 1 hanging around, I've replaced this with the simpler $f(k_t)$. This form of the production function is often called the *intensive form*, meaning it is written in per capita terms. The function $f(.)$ inherits all of the properties of $F(.,.)$, something you'll confirm for yourself in the homework problems. Specifically

$$f'(k_t) > 0 \text{ and } f''(k_t) < 0, \tag{2.29}$$

which says that output per capita is rising in capital per capita, but at a decreasing rate. The function $f(.)$ also satisfies the Inada conditions, with the marginal product of capital $f'(.)$ going to infinity as k_t goes to zero, and the marginal product going to zero as k_t goes to infinity.

Output per capita depends on capital per capita (I'll often also refer to this as capital per worker to avoid the awkward phrase). Hence we need to describe the determinants of capital per worker, and we can do that given our equations for capital and labor in (2.26) and (2.27). The growth rate of k_t can be approximated as follows,

$$\frac{\Delta k_{t+1}}{k_t} \approx \frac{\Delta K_{t+1}}{K_t} - \frac{\Delta N_{t+1}}{N_t}. \tag{2.30}$$

Note that this is an approximation. Strictly speaking, the growth rate of capital per worker is not exactly the growth rate of K minus the growth rate of N. However, so long as the growth rate of population is relatively small (and historically the yearly growth rate less than 2% in all but the most rapidly expanding population) this approximation is essentially an equality. We're using this discrete-time approximation to keep the notation consistent, as it will be useful as we continue in the course.

Now use equations (2.26) and (2.27) to plug into (2.30) and we have

$$\frac{\Delta k_{t+1}}{k_t} = \frac{sF(K_t, N_t)}{K_t} - \delta - n \tag{2.31}$$

$$= s\frac{F(K_t, N_t)/N_t}{K_t/N_t} - \delta - n \tag{2.32}$$

$$= s\frac{f(k_t)}{k_t} - \delta - n. \tag{2.33}$$

If we simply multiply through by k_t, then we have

$$\Delta k_{t+1} = sf(k_t) - (\delta + n)k_t. \tag{2.34}$$

The above equation is the "Solow equation". It describes how capital per worker evolves over time, and it constitutes our complete description of the economy. That is, for an economy with perfect competition among firms producing homogenous output, perfect factor markets, inelastically supplied labor and capital, closed to foreign capital flows, a constant savings rate, a costless financial sector, and constant population growth, this equation is all we need to describe how GDP evolves over time. With it, we can find the level of capital per worker at any period t, and from that find GDP per capita from the production function.

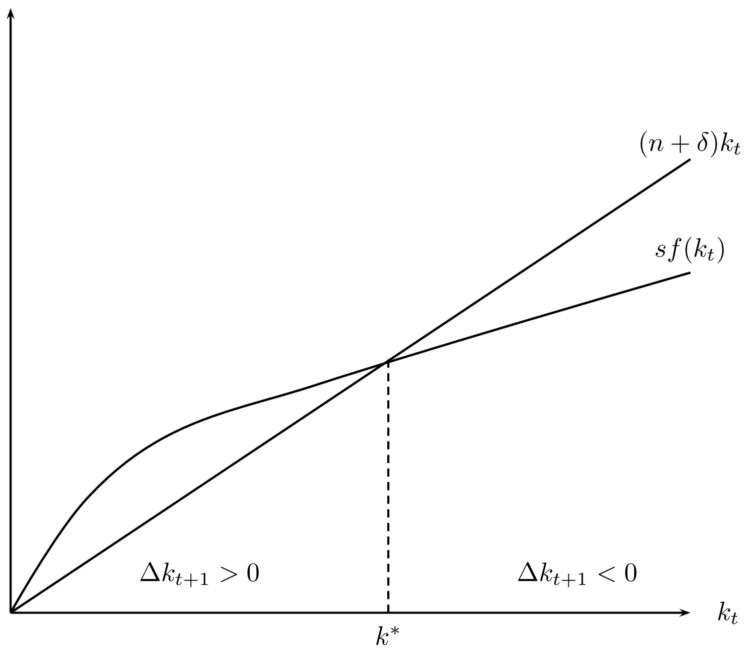


Figure 2.1: The Solow Steady State

Note: The steady state is the level of capital, k^*, at which the amount invested, $sf(k^*)$, is exactly equal to the amount of capital that is depreciating, $(n+\delta)k^*$. At any capital stock less than k^*, investment exceeds depreciation and the capital stock increases, so that over time k_t approaches k^*. For $k_t > k^*$, the opposite occurs. Because of these opposing effects, the steady state is stable.

2.3 Implications of the Solow Model

This model has specific predictions about how the economy will evolve over time, as well as predictions regarding the relationship of parameters (like the savings rate or the population growth rate) and output per capita. We can work through these to see how useful the Solow model will be.

Implication 1: The economy will eventually end up at a fixed level of capital per worker and stay there. This follows from the nature of the Solow equation. First we need to establish that there are fixed levels of capital per worker than, once reached, the economy never deviates from. We call these "steady states" of the Solow model. They occur wherever $\Delta k_{t+1} = 0$, meaning that capital worker is not growing, and that $k_{t+1} = k_t$. Using (2.34), we can solve for these steady states, which we denote k^*, as

$$sf(k^*) = (n+\delta)k^*. \tag{2.35}$$

One steady state is where $k^* = 0$. If the economy has no capital, then it cannot produce any output to invest in new capital, and it will stay at zero capital forever.

On the other hand, there is some value $k^* > 0$ that solves the above equation as well. To see why, it is easiest to look at a diagram. Figure 2.1 plots both terms on the right-hand side of equation (2.34) against the value k_t. The term $sf(k_t)$ is a concave function of k_t, given our assumptions that $f'(k_t) > 0$ and $f''(k_t) < 0$. The term $(n + \delta)k_t$ is linear in k_t. These curves cross where $sf(k_t) = (n + \delta)k_t$, which is precisely our defintion of the steady state. So that intersection tells us where k^* is found.

The second thing we need to establish is that the economy will head towards the steady state of its own accord. If the economy is *not* in steady state, then what this figure indicates is that the economy will in fact move towards steady state. To see this, note that for any $k_t < k^*$, it is the case that $sf(k_t) > (n + \delta)k_t$ and so $\Delta k_{t+1} > 0$, or capital per worker is growing. For $k_t > k^*$, it is the case that $sf(k_t) < (n + \delta)k_t$ and so $\Delta k_{t+1} < 0$, and capital per worker is shrinking. Hence, no matter where the economy begins, the capital stock tends to move towards steady state. The positive steady state is *stable*. If we deviate from the steady state, we will always return to it.

Solow with the Cobb-Douglas Production Function

Recall the Cobb-Douglas production function,

$$Y_t = K_t^{\alpha} N_t^{1-\alpha} \tag{2.36}$$

where $\alpha \in (0, 1)$. *The intensive form of the Cobb-Douglas is*

$$y_t = f(k_t) = k_t^{\alpha}. \tag{2.37}$$

Using this intensive form, we can write the accumulation equation as

$$\frac{\Delta k_{t+1}}{k_t} = s\frac{k_t^{\alpha}}{k_t} - \delta - n \tag{2.38}$$

and the steady state condition is that

$$sk^{*\alpha} = (n + \delta)k^*. \tag{2.39}$$

We can solve the steady state condition for

$$k^* = \left(\frac{s}{n + \delta}\right)^{1/(1-\alpha)} \tag{2.40}$$

and output per person in steady state is therefore

$$y^* = \left(\frac{s}{n + \delta}\right)^{\alpha/(1-\alpha)}. \tag{2.41}$$

From this it can easily be seen that output per person is not growing over time, as it is only a function of the fixed parameters s, n, *and* δ.

As output per capita is $y_t = f(k_t)$, we know that output per capita will be growing so long as $k_t < k^*$, shrinking if $k_t > k^*$, and constant if $k_t = k^*$. Output per capita hits a steady state just the same as capital per worker.

There's clearly something missing from the Solow model. We know that in the U.S. and other OECD countries there have been remarkably stable growth rates in output per capita over long periods of time. This version of the Solow model says that eventually growth in output per capita should have run down to zero. So our model cannot be precisely right at this point.

This doesn't mean that output (in aggregate) is not growing in steady state. Once the economy is in steady state, output in period $t+1$ is

$$Y_{t+1} = F(K_{t+1}, N_{t+1}) = F((1+n)K_t, (1+n)N_t) = (1+n)F(K_t, N_t) = (1+n)Y_t. \tag{2.42}$$

In other words, aggregate output is growing at the rate of population growth, n. The key step here is to note that $K_{t+1} = (1+n)K_t$ in steady state. Why? We know that k_t is constant in steady state, and as $k_t = K_t/N_t$, the only way for k_t to stay constant is for K_t to grow at exactly the same rate as N_t. So in steady state the aggregate economy is growing, but per capita output is stuck.

Implication 2: The steady state level of capital per worker, and income per capita, is positively related to s and negatively to n. This can be seen several ways. If you examine the steady state for the Cobb-Douglas situation, you'll see this immediately. From figure 2.1 one can see that if s rises, then the concave savings function shifts upwards, and the intersection with the depreciation line shifts to the right, indicating a higher steady state value of k^*. An increase in n rotates the depreciation line up, lowering the steady state.

More formally, the steady state condition in the Solow model can be written as

$$\frac{s}{n+\delta} = \frac{k^*}{f(k^*)}. \tag{2.43}$$

If s rises, then the right-hand side must rise as well. So what must happen to k^* in order to make the right-hand side increase? Take the derivative of the RHS with respect to k^* and you'll have

$$\frac{\partial k^*/f(k^*)}{\partial k^*} = \frac{f(k^*) - f'(k^*)k^*}{f(k^*)^2}. \tag{2.44}$$

This derivative is positive if $f(k^*) > f'(k^*)k^*$. Re-arranging, this derivative is positive if

$$\frac{f(k^*)}{k^*} > f'(k^*), \tag{2.45}$$

which says that the average product of capital is greater than the marginal product of capital. Is this true? It is given what we've assumed about the production function, particularly that it is concave.

Therefore, in equation (2.43), the right-hand side is increasing with k^*. So if s is higher, then to balance that equation k^* must increase. Similar logic shows that if n is higher, then it must be that k^* is lower.

These predictions can be evaluated against evidence from across countries. Remember that these predictions are about steady states. If countries are all at least close to their steady states then the influence of s and n should be apparent.

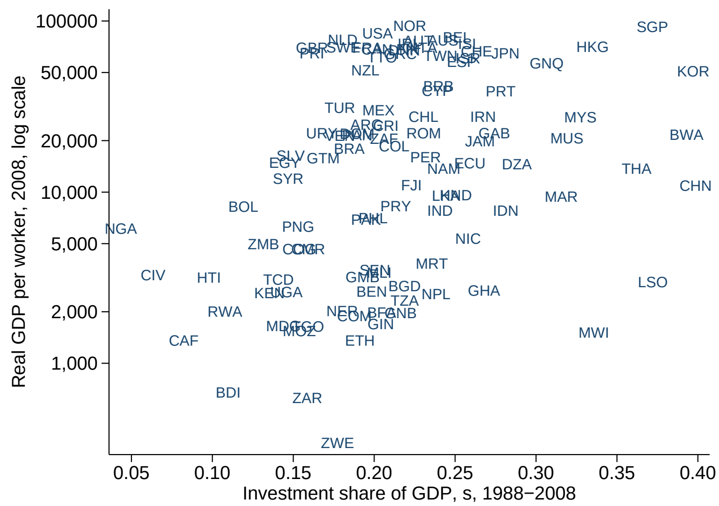


Figure 2.2: Relationship of savings s and income per capita

From figures 2.2 and 2.3 you can see that these rough relationships do hold in the data from the last 20 years. Countries that have a higher share of output saved (specifically, spent on investment goods) tend to have higher income per capita, while those with higher population growth rates tend to be poorer.

The relationships are not exact, and so describing differences in living standards between countries requires something more than just the savings rate and population growth rate. Another way of saying this is that a model that only has capital and labor is not sufficient to capture differences in output per capita across countries. We'll need to upgrade the Solow model to capture more accurately these living standards.

Implication 3: Out of steady state, the economy will grow faster the farther away from steady state it starts. To see this implication, start with the main Solow equation and divide through by k_t to put things in terms of growth rates,

$$\frac{\Delta k_{t+1}}{k_t} = s\frac{f(k_t)}{k_t} - n - \delta. \tag{2.46}$$

Now graph the growth rate against the level of k_t. To see what this graph should look like, first

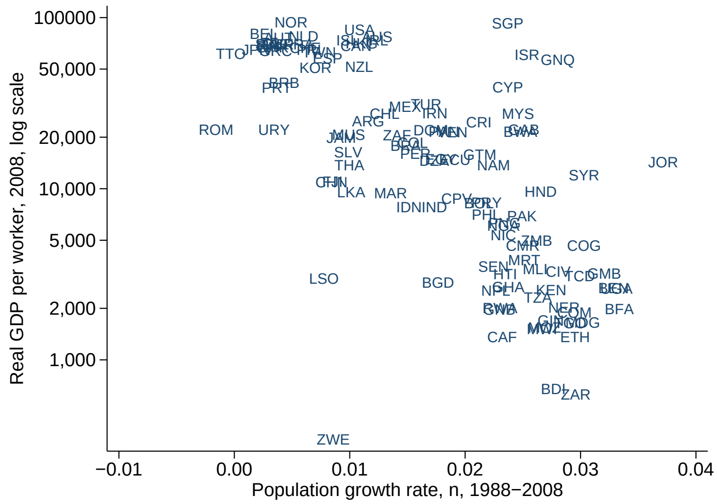


Figure 2.3: Relationship of population growth n and income per capita

consider how the growth rate acts as k_t goes to extreme values of zero and infinity:

$$\lim_{k\to 0} \frac{\Delta k_{t+1}}{k_t} = s\frac{f'(k_t)}{1} - \delta - n = \infty \tag{2.47}$$

$$\lim_{k\to \infty} \frac{\Delta k_{t+1}}{k_t} = s\frac{f'(k_t)}{1} - \delta - n = -\delta - n \tag{2.48}$$

where the first step in each situation relies on L'Hopital's rule. What we see is that as k_t goes to zero, the growth rate of capital per worker shoots up to infinity. As k_t goes to infinity, the growth rate of capital per worker becomes negative.

What happens between these extremes? For that we need to know how the growth rate of k_t responds to a change in k_t. Take the derivative of (2.46) with respect to k_t and you have

$$\frac{\partial \Delta k_{t+1}/k_t}{\partial k_t} = s\frac{f'(k_t)k_t - f(k_t)}{k_t^2}. \tag{2.49}$$

The sign of this term depends on $f'(k_t)k_t - f(k_t)$, must be negative, given that $f'(k_t) < f(k_t)/k_t$ as we established earlier.

So when k_t is close to zero, the growth rate approaches infinity, and as k_t increases the growth rate falls, until as k_t appreoaches infinity the growth rate approaches $-n-\delta$. Figure 3.1 plots this relationship. This figure identifies the stable steady-state of the Solow model as the point where capital per worker growth equals zero.

More interestingly, the figure shows that the farther away from steady state is capital per worker, the higher is the absolute value of the growth rate. A country with k_t very close to zero should grow

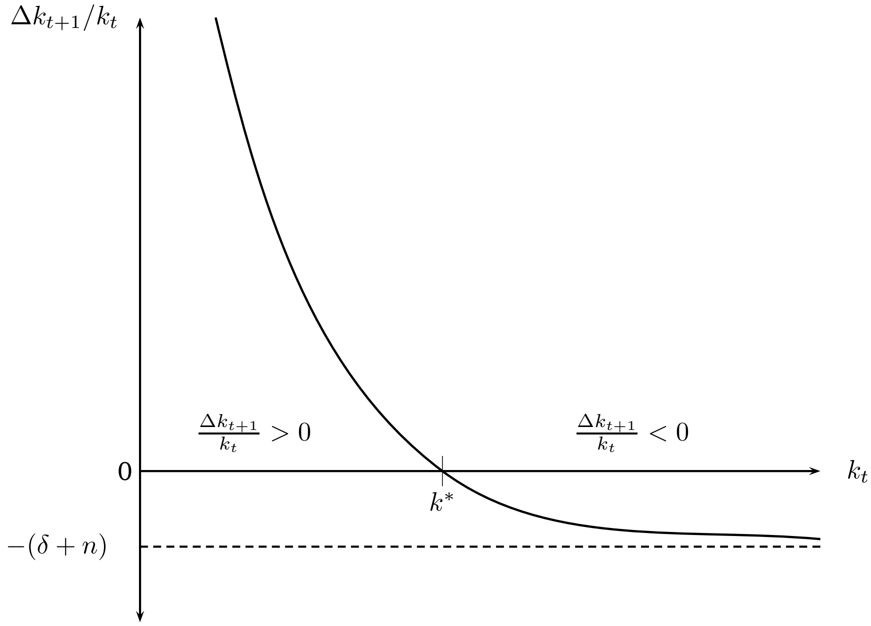


Figure 2.4: Growth Rates and k_t

faster than an economy with k_t close to the steady state. An economy with k_t *higher* than the steady state level should actually be shrinking. Overall, it predicts a negative relationship between growth rates in capital per worker and the level of k_t. Given that output per capita is simply a function of k_t, it predicts that growth in output per capita should be declining in the level of y_t.

Do we see this in the data? Yes and no. For a subset of currently rich countries, it most certainly does. Figure 2.5 plots growth rates against initial income levels for major European countries, Japan, the U.S., Canada, Australia, and New Zealand. Here, it's quite clear that income per capita grows faster the poorer a country starts out. Figure 2.6 shows the same kind of plot for all OECD countries starting in 1960. Again, the downward slope is obvious, as the Solow model would predict.

However, in figure 2.7 when all countries in the world are included the prediction falters. There is a mass of poor countries with low growth rates, many of them African, Asian, and Central American. There countries do not conform to the predictions of the simple Solow model, which would expect them to be growing as fast as Taiwain (TWN), Korea (KOR), and Botswana (BWA).

The importance of this relationship, and it's failure, has to do with what we call *convergence*. Going back to our stylized figure 3.1, note that what this implies is that eventually every country

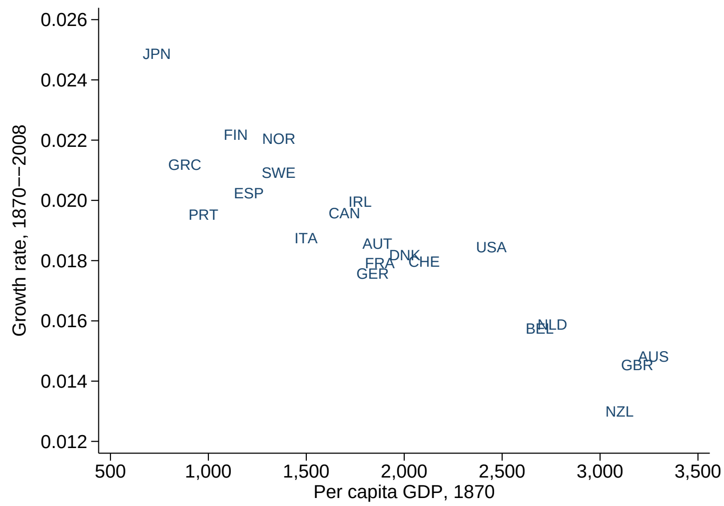


Figure 2.5: Growth rates and income levels, 1870-2008

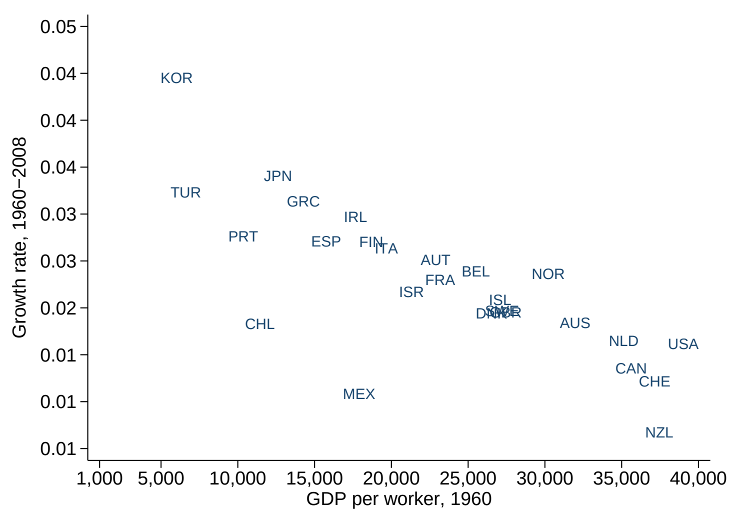


Figure 2.6: Growth rates and income levels, 1960-2008, OECD only

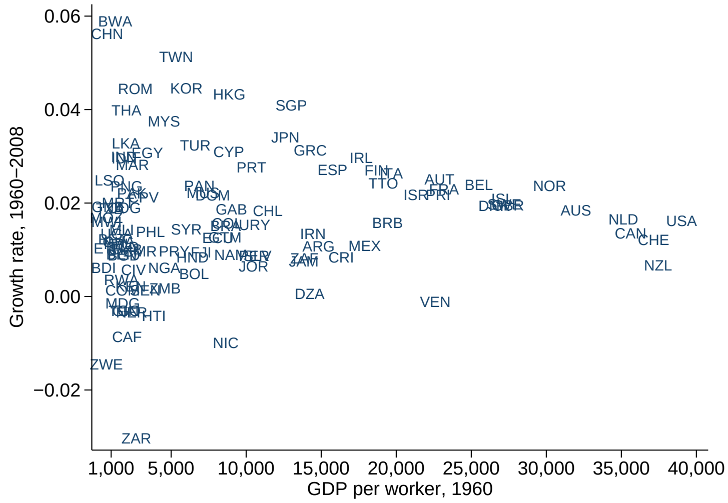


Figure 2.7: Growth rates and income levels, 1960-2008, all countries

should converge to a similar steady state. Poor countries grow faster than rich countries, and so they are catching up. Eventually, every economy should end up at k^*, which means they have identical levels of income per capita. This is what we appear to see in the OECD, with poor countries catching up. However, for many of the poorest countries in the world, this convergence is not occurring, as their growth rates are too low to catch-up with the rich countries. We'll explore some extensions of the Solow model that may help explain why that is, but at this point we just note that our model is incomplete.

An alternative way of seeing the connection of growth rates and income levels is to look at time-series from different countries. In particular, we'll look at countries that had a distinct drop in their k_t level, and see what happens afterwards. Germany and Japan both experienced bombing during World War II that wiped out most of their existing capital stock. According to the Solow model, this means that they should have grown more quickly than other countries (in particular the U.S., where capital was not destroyed) in the years after the war.

Figure 2.8 plots income per capita for the U.S., U.K., Germany, and Japan from 1870 to the present. As can be see, both Germany and Japan experience a distinct drop in income per capita immediately after the war. However, they both then experience a period of rapid growth (the slopes are much higher relative to the UK and US) until about 1970. By that point they have caught back up to the UK, and remain only slightly behind the U.S., as they were prior to the war. This is convergence in action. In this sense, the Solow model performs very well, predicting just such a catch-up. It fails, however, as we saw before, in that it would have predicted that growth rates fell

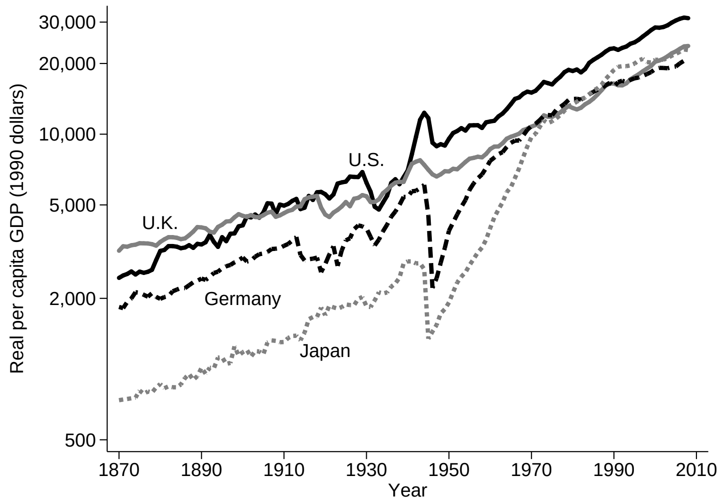


Figure 2.8: Income per capita over time, select countries

to zero eventually. We'll address that in the next section.[1]

Implication 4: The return on capital eventually ends up at a fixed level and stays there. The last implication of the Solow model worth emphasizing is how the return on capital acts over time. This was one of the motivating factors for Solow's original article. Economists had been wondering why the return to capital did get driven down to zero as rich countries accumulated larger and larger stocks of capital.

To see what happends to the return on capital, let's modify slightly the firm-level problem. The representative firm (recall that with perfect competition and free entry we can easily just work with one firm) has total profits of

$$\pi = F(K,N) - RK - wN = N\left(f(k) - Rk - w\right). \tag{2.50}$$

Maximize this alternative way of writing profits over N and k. This gives first-order conditions of

$$f(k) - Rk - w = 0 \tag{2.51}$$

$$f'(k) = R. \tag{2.52}$$

The second condition shows us that the return on capital in the economy is equal to the marginal product of capital per worker, $f'(k)$. We know already that in steady state k will be constant at k^*,

[1]This also highlights the importance of distinguishing between growth rates of y_t and levels of y_t. Living standards depend on the level, not the growth rate. Germany and Japan had high growth from 1950-1970, but it would have been demonstrably better from a material standpoint to have lived in the U.S. during that period.

and therefore the marginal product of capital per worker will be constant. That means that R is constant in steady state, and does not tend to decline to zero.

The key is that $F_{KN} > 0$. This means that as we add labor, the marginal product of capital actually goes up, raising R. By accumulating more capital we drive down the marginal product of capital, lowering R. In steady state the influence of capital accumulation on R just offsets the influence of adding labor, and therefore R stays constant even though in aggregate we keep accumulating K.

2.4 Consumption and Welfare

Up to this point we haven't actually said anything about welfare, or provided any way of measuring it. The Solow model doesn't have any optimizing individuals in it, trying to maximize their utility, so there really isn't any direct way to think about welfare.

We can examine consumption, though, which in our future work will be the item that individuals care about in their utility function. With our assumption of a constant savings rate, no trade, and no government sector to speak of at this point, it must be that

$$c_t = (1-s)y_t. \tag{2.53}$$

That is, consumption per capita is just a fixed fraction of output per capita. So therefore, consumption must reach a steady state just the same as output per capita, and remain constant thereafter.

This much follows directly from our prior work. However, note that consumption doesn't share the same striclty positive relationship with the savings rate, s, that output per capita has. To see this, note that s has two effects on consumption. If s goes up, then consumption goes down, as less output is left aside to consume. However, if s goes up, then output per capita rises, meaning higher consumption.

That means that there may be some level of s that delivers the maximum amount of consumption. But maximum over what time period? If the economy wants to maximize consumption in period t, then it could simply consume all of the output in time t, y_t. In fact, it could even consume all of the existing capital stock at time t, k_t, as recall that we've assumed capital is just unconsumed output. But if the economy did that, the capital stock in $t+1$ would be zero, and there would be no output at all, and no consumption.

Without an explicit utility function, we don't know whether that would be worth it or not. But we can think about an alternative, which is to ask what is the maximum level of consumption that the economy can sustain in steady state.

In steady state, consumption is

$$c^* = (1-s)f(k^*) = f(k^*) - sf(k^*) = f(k^*) - (n+\delta)k^* \tag{2.54}$$

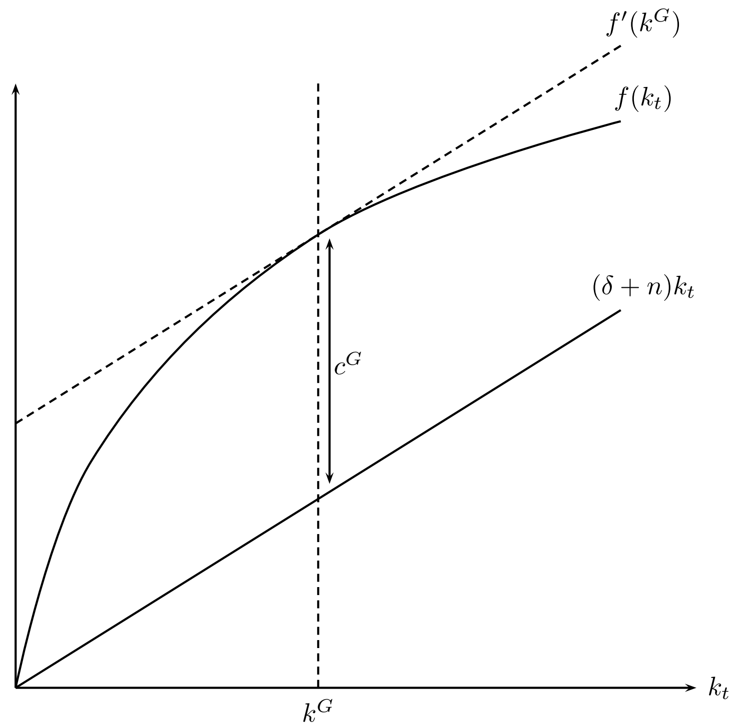


Figure 2.9: The golden rule capital stock

where the last equality follows from the Solow equation in steady state. Maximize c^* over k^* to find the steady state level of capital per worker that yields the highest consumption in steady state. The first-order condition is

$$f'(k^*) = n + \delta. \tag{2.55}$$

This says to set the marginal product of capital equal to population growth plus the depreciation rate. To see what is going on, consider figure 2.9. The top curve is the production function - not s times the production function. The depreciation line shows how much we need to be saving in steady state to keep the capital per worker constant. The difference between these curves is how much we can consume. The maximum amount of consumption comes where the production function is tangent to the depreciation line.

The Cobb-Douglas Golden Rule *If we presume that the production function is Cobb-Douglas, then how do we find the Golden Rule savings rate and consumption level?*

Let's begin with a production function of $Y_t = K_t^{\alpha} N_t^{1-\alpha}$. In intensive form, this is

$$y_t = k_t^{\alpha} \tag{2.56}$$

and we can write consumption in steady state as

$$c^* = k^{*\alpha} - (\delta + n)k^*. \tag{2.57}$$

Maximizing consumption over k^* yields the first order condition

$$\alpha k^{*(\alpha-1)} = \delta + n \tag{2.58}$$

which can be solved for the golden rule level of the capital stock

$$k^G = \left(\frac{\alpha}{\delta + n}\right)^{1/(1-\alpha)}. \tag{2.59}$$

How do we achieve this golden rule capital stock level? We have to set the savings rate so that our steady state capital stock per person is exactly equal to k^G. The steady state level of capital per worker is, as we saw previously,

$$k^* = \left(\frac{s}{\delta + n}\right)^{1/(1-\alpha)}. \tag{2.60}$$

Comparing this to the equation for k^G shows that we should set the savings rate to exactly $s = \alpha$ in order to reach the Golden Rule.

Note that if we looked at the steady state value for consumption from the Cobb-Douglas version, $c^* = (1-s)\left(\frac{s}{n+\delta}\right)^{\alpha/(1-\alpha)}$ and maximized this with respect to s we would get exactly the same answer.

There is thus a single k^G, or "Golden Rule" steady state level of capital per worker as Edmund Phelps termed it, that provides the maximum level of consumption in steady state. So is it possible for the economy to be at this Golden Rule steady state? It is, if the savings rate is set just right. We've already established that k^* rises smoothly with s. There must be some s such that the steady state $k^* = k^G$. By picking the right savings rate, the economy can achieve the Golden Rule level of consumption.

The final point to make regarding the Golden Rule is that while it is sustainable in the long run, this is only true so long as there are no disturbances to the economy. If a natural disaster, for example, were to destroy part of the capital stock, then $k_t < k^G$, and therefore $y_t < y^G$. To keep on consuming c^G would mean lowering the savings rate below s^G. But once the savings rate is below

s^G, the economy no longer has a steady state at k^G. The economy is not saving enough to remain at the Golden Rule level of consumption, and eventually consumption will have to go to zero as the economy uses up all the existing capital stock trying to maintain consumption at c^G.

The Golden Rule consumption level serves as a useful benchmark, but is not something that we expect a real economy to be able to achieve. An economy that has perfect competition, inelastic factor supplies, and that never experiences any disturbances can conceivably reach the Golden Rule level of consumption, but that kind of economy is most certainly not a good description of the real world.

2.5 Primitive Economic Growth

As we saw, in steady state there is no growth in output per capita in the Solow model. This runs counter to what we see in the data, with sustained growth in output per capita of roughly 1.2% per year in the OECD nations. Without having introduced any notions of technology improvement or productivity, is the Solow model capable of reproducing this fact? It is, but only with some very extreme assumptions. While it seems that these are unlikely to be true in the real world, examining the Solow model with these assumptions is useful as a benchmark.

2.5.1 The AK model

Growth runs down to zero in steady state in the Solow model because of our assumptions regarding the production function. Specifically, with $f'(k) > 0$ but $f''(k) < 0$, there are diminishing returns to accumulating more capital per worker. Eventually the additional output we get from a little more capital per worker is just enough to replace the capital per worker that is depreciating.

You can see this more clearly from the Solow accumulation equation

$$\frac{\Delta k_{t+1}}{k_t} = s\frac{f(k_t)}{k_t} - (n+\delta). \tag{2.61}$$

We established that the growth rate of k_t fell as k_t increased. That followed from the fact that $f'(k_t) < f(k_t)/k_t$, which in turn is a consequence of the assumption that $f''(k_t) < 0$, or that the intensive form of the production function is concave. From the accumulation equation, what this means is that as k_t rises, $f(k_t)$ rises more slowly that k_t, and so the first term declines towards zero.

The assumption that $f''(k_t) < 0$ is therefore driving the Solow model to have zero growth in steady state. So if the Solow model is to exhibit sustained growth in steady state, we will have to abandon this assumption. What does that mean? It means that $f(k_t)$ does not exhibit diminishing marginal returns, or as we accumulate k_t the marginal product of k_t does not decline. More precisely, we need $f''(k_t) = 0$.

A simple way to ensure that is to assume that the production function is linear. More precisely, if $F(K_t, N_t) = AK_t$, then $f(k_t) = Ak_t$. The A here is simply a scaling parameter, although one can also think of it as capturing the productivity of capital. The “AK” model in the title of this section comes from this form of the production function.

With that production function, the accumulation equation becomes

$$\frac{\Delta k_{t+1}}{k_t} = sA - (n + \delta). \tag{2.62}$$

Here, there is no relationship between the growth rate of capital and the size of k_t. Growth in capital per worker is constant at $sA - (n + \delta)$. So long as $sA > n + \delta$, then there will be positive growth in capital per worker and hence in output per capita. The AK model can replicate the steady growth rates that we see in the data.

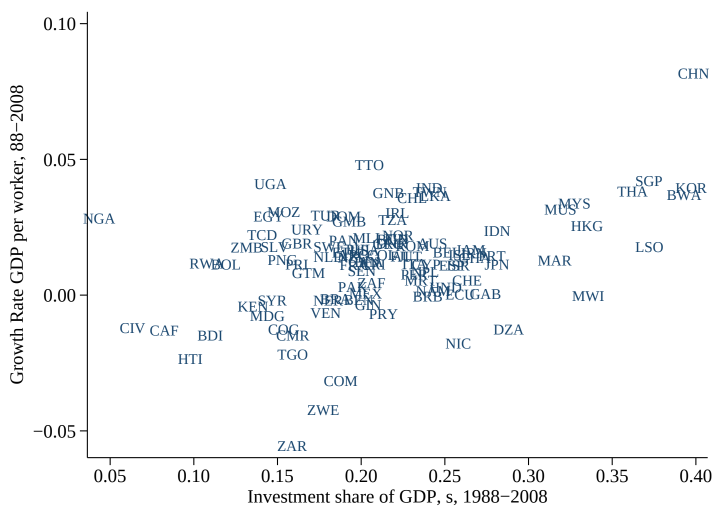


Figure 2.10: Growth rates and savings rates, 1988-2008

However, the AK model also says that *growth rates* are rising with s, and falling with n. Is this true in the data? Possibly. Figures 2.10 and 2.11 show growth rates of output per capita plotted against savings and population growth, respectively. As you can see, there is a tendency for growth rates to be higher for countries with higher savings rates, and lower for countries with higher population growth rates, just as the AK model would predict.

We need to be careful, though. The original Solow model would also predict a positive correlation of growth rates and savings rate *out of steady state*. So this is not definitive proof, as we don’t know whether countries are in or out of steady state. The AK model also says that growth rates should remain constant, regardless of the level of k_t. However, we saw in figures 2.5 and 2.6 that growth rates demonstrably slow down as countries get richer.

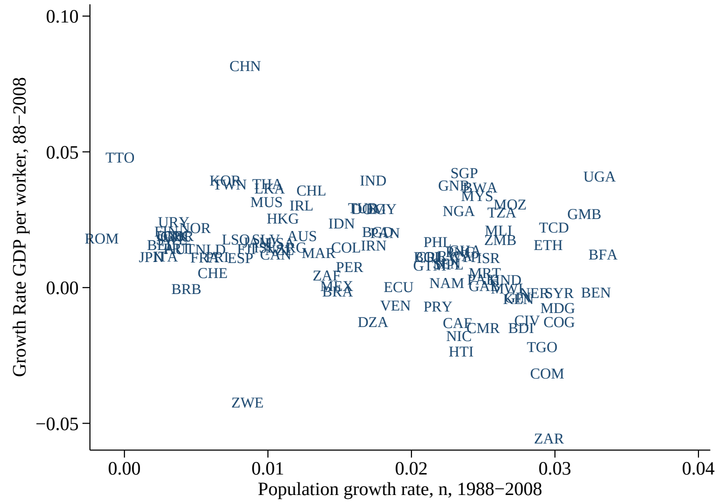


Figure 2.11: Growth rates and population growth rates, 1988-2008

Externalities and the AK Model *One way of introducing AK style production without having to assume that capital's share in output is $\alpha = 1$ is externalities. That is, capital will have spillover effects on productivity that firms and individuals don't account for directly.*

Let each perfectly competitive firm in the economy have a production function of

$$Y_i = BK_i^{\alpha} L_i^{1-\alpha} \tag{2.63}$$

where i indexes the firms, of which there are M. The terms B is a scaling term, think of it as an index of productivity, which the firm takes as given. Firms take the wage rate and rental rate as given. Each firm will thus pay out α of output to capital (as we see in the data and our typical model) and $1-\alpha$ to labor.

Each firm is identical, so each firm employs an identical amount of capital and labor. At the aggregate level, total production is

$$Y = \sum_i Y_i = BMK_i^{\alpha} L_i^{1-\alpha} = B(MK_i)^{\alpha}(ML_i)^{1-\alpha} \tag{2.64}$$

and for markets to clear it must be that the total demand for capital is equal to the total supply, $MK_i = K$ and total demand for labor is equal to total supply, $ML_i = L$. This means that aggregate output is just $Y = BK^{\alpha}L^{1-\alpha}$ and output per worker is $y = Bk^{\alpha}$.

Now, what about this term B? This productivity is presumed to depend on the physical capital per worker in use, on average, in the economy. Specifically, let $B = Ak^{\gamma}$. That is, as *firms use higher capital per worker, this has a positive productivity effect on all firms. Why? For many reasons one could imagine. Firms with high capital to labor ratios may spur competitors to innovate, or may demand suppliers meet certain technical standards or logistical requirements. Regardless, if the spill-over exists then output per worker is*

$$y = Ak^{\gamma+\alpha} \tag{2.65}$$

and we have a way of taking a production function with a relatively small α (say 1/3) and making it look more like an AK situation. If γ is large enough, say $\gamma = 1 - \alpha$, then $y = Ak$ and we have an exact AK model. If this is true, then we could have sustained growth just through capital accumulation.

This setup matches the observation that capital tends to earn a constant share of output, but still would have the prediction that growth rates are positively correlated with savings rates. As noted, this looks to hold across countries, but not within countries over time.

The AK model also, considering our underlying assumptions regarding firms and how they compenstate factors of production, has the implication that capital would earn 100% of output, and wages would be zero. Labor is unproductive, so there would be no need to compensate it. Obviously, we do not see that in the data, so the strict AK model seems unlikely to be true. Even if we relax the assumptions of perfect competition and perfect factor markets, we would have to believe that wages are simply rents that workers are able to extract from firms, and that seems a poor description of reality.

2.5.2 Multiple Capital Types

A different approach to embedding sustained growth in the Solow model is to allow for multiple types of capital; physical and human capital, for example. As it turns out, for these models to generate sustained growth, they will have to be essentially AK models, although with a more refined notion of what the "K" represents.

Let's use a concrete example to start. Let aggregate production be

$$Y_t = K_t^{\theta} H_t^{\phi} N_t^{1-\theta-\phi} \tag{2.66}$$

where H is human capital. In intensive form, the production function is

$$y_t = k_t^{\theta} h_t^{\phi}. \tag{2.67}$$

Let the two capital stocks accumulate in a Solow manner, treating human capital just like another type of capital

$$\frac{\Delta k_{t+1}}{k_t} = s_k \frac{y_t}{k_t} - (n+\delta) \tag{2.68}$$
$$\frac{\Delta h_{t+1}}{h_t} = s_h \frac{y_t}{h_t} - (n+\delta). \tag{2.69}$$

Take these equations, plug in for y_t, and you have

$$\frac{\Delta k_{t+1}}{k_t} = s_k \frac{(h_t/k_t)^{\phi}}{k_t^{1-\phi-\theta}} - (n+\delta) \tag{2.70}$$
$$\frac{\Delta h_{t+1}}{h_t} = s_h \frac{(k_t/h_t)^{\theta}}{h_t^{1-\phi-\theta}} - (n+\delta).$$

The growth rates of both k_t and h_t depend on the ratio of h/k. Similar to the original Solow model, we can look at the implications of these accumulation equations.

First, there will be a steady state *ratio* of k/h. This is easiest to see visually. Figure 2.12 plots the growth rates of the two capital stocks against the ratio k/h. For k, the growth rate is declining in k/h, going to infinity as k/h goes to zero, and to $-(n+\delta)$ as k/h goes to infinity. For h, the growth rate is rising in k/h, going to $-(n+\delta)$ as k/h goes to zero, and the rising concavely as k/h increases.

From this figure, we can work out that there is a stable steady state ratio of $(k/h)^*$, which is determined where the two curves cross. At that ratio, both k and h are growing at the same rate, and so the ratio remains steady. If the economy has $k_t/h_t < (k/h)^*$, then we see that k is growing faster than h, and hence k_t/h_t will rise over time. Similar logic shows that if the ratio is above the steady state, then the ratio will shrink. Hence $(k/h)^*$ is a stable steady state.

What is this steady state ratio k/h precisely? We know from the diagram that in steady state it must be that k and h grow at the same rate. To solve, set the growth rates from (2.71) equal to each other. Doing so results in

$$\left(\frac{k}{h}\right)^* = \frac{s_k}{s_h} \tag{2.71}$$

or the steady state ratio of physical to human capital depends on the relative savings rates, which seems intuitive.

Knowing the ratio, what is the growth rate of k and h at the steady state? To solve for this, simply take the steady state ratio of k/h and plug back into one of the accumulation equations. Using the accumulation equation for physical capital we find

$$\frac{\Delta k_{t+1}}{k_t} = \frac{s_k^{1-\phi} s_h^{\phi}}{k_t^{1-\phi-\theta}} - (n+\delta). \tag{2.72}$$

From this we see that the size of k_t potentially influences the growth rate of k_t in steady state. How do we make sense of that? It depends on what we assume about the size of ϕ and θ.

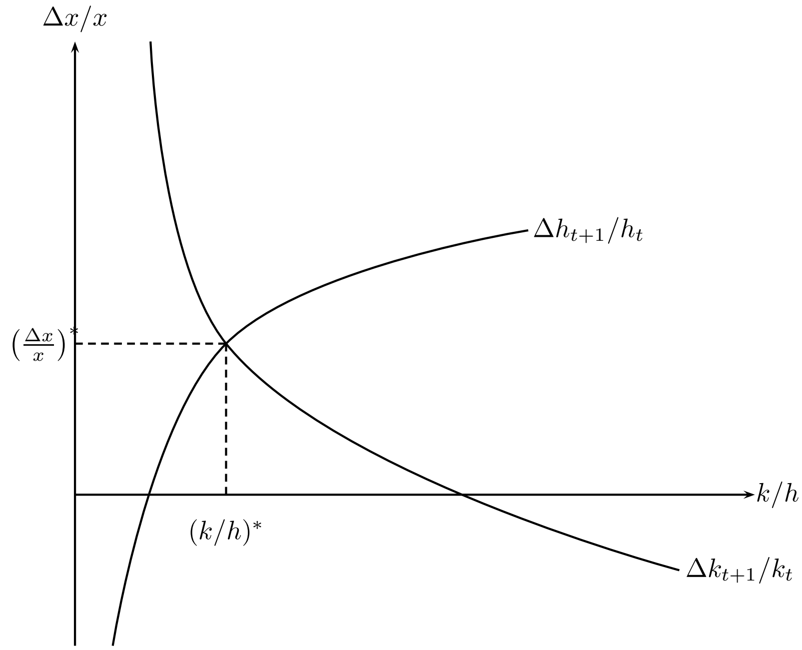


Figure 2.12: Growth with Two Capital Stocks

Note: The figure shows the growth rate of k and h relative to the ratio of their stocks. If $\theta + \phi < 1$, then the only possible intersection of these curves is along the axis - no growth in either. If $\theta + \phi = 1$, then the intersection can occur where growth in both is positive and the ratio stays constant. The steady state growth rate is denoted $\left(\frac{\Delta x}{x}\right)^*$.

Take the case where $\phi = 1 - \theta$. This implies that the aggregate production function is $Y = K^\theta H^{1-\theta}$, and raw labor ($N$) no longer matters. If this is true, then the $k_t^{1-\phi-\theta}$ in the denominator of the growth rate of physical capital reduces to one, and no longer matters for growth. We have an "AK" style result, where the growth rate of physical capital (and hence of human capital) equals $s_k^{1-\phi} s_h^{\phi} - (n + \delta)$. So long as the savings rates are high enough, k and h will continue to grow indefinitely, and hence so will output per capita.

What occurs in the case where $\phi + \theta < 1$? Then we are back to a typical Solow situation, where growth eventually has to wind down to zero. To see this, not that with $\phi + \theta < 1$, as k_t increases the growth rate of k_t continues to drop. Eventually it must be that the growth rate of k_t reaches zero. This will occur to human capital as well, with growth dropping to zero. The only possible way to be at the steady state ratio $(k/h)^*$ is for both capital stocks to remain constant. In the figure, the two curves shift until their intersection occurs exactly on the x-axis. Because growth in both k and h slows down to zero, so must growth in output per capita.

In general, what matters for sustained growth in the Solow model is not the number of capital varieties, but rather the degree of decreasing returns to capital in general.

The Lucas Model

A variant on the model including human capital comes from Lucas (1988), who specifies a production function

$$Y = K^{\alpha}(uhL)^{1-\alpha} \tag{2.73}$$

where L is labor, h is human capital per person, and u is the fraction of time spent working. In per capita terms

$$y = k^{\alpha}(uh)^{1-\alpha}. \tag{2.74}$$

Capital accumulates as $\Delta k = sy - (n+\delta)k$, as usual. Human capital accumulates as

$$\Delta h = \phi h(1-u) \tag{2.75}$$

which says that human capital is accumulated by spending time not working. So we have a tradeoff in output and accumulation. Notice that the growth rate of h is $\Delta h/h = \phi(1-u)$, a constant. Therefore human capital grows continuously, and so will output per capita.

In addition, as physical capital depends on output (which depends on h), physical capital accumulates continuously as well. Set the growth rates of k and h equal to each other, as in a steady state, and you will find that $\Delta k/k = \phi(1-u)$. Therefore the growth rate of output per capita is

$$\frac{\Delta y}{y} = \phi(1-u) \tag{2.76}$$

given that u is assumed to be a constant. Any change in u has two effects on the economy. First, u affects output immediately (if u goes down, we produce less by working less). Second, u affects output growth (if u goes down, growth goes up). So there is a tradeoff in selecting u. The optimal choice of u will depend on the time preference, as it depends on whether you want output today or in the future.

More importantly, the reason there is steady state growth in the Lucas model is not because there is some alternative assumption about human capital accumulation. The source of growth is the assumption that the aggregate production function has constant returns to scale in accumulable factors. To see this, write the total human capital stock as $H = hL$, and then the production function is $Y = K^{\alpha}H^{1-\alpha}u^{1-\alpha}$. With constant returns to generalized capital, there can be sustained growth.

Generally speaking, let the aggregate production function be

$$Y = F(K_1, K_2, ..., K_J, N) \tag{2.77}$$

where J is the number of types of capital used. If this function has constant returns to scale, then we can divide through by N and write

$$y = f(k_1, k_2, ..., k_J). \tag{2.78}$$

Totally differentiate this function

$$dy = f_1 dk_1 + f_2 dk_2 + ... + f_J dk_J, \tag{2.79}$$

where f_1 is simply the derivative of $f()$ with respect to k_1, and so on. This says that changes in output per worker are the sum of the marginal effect of each capital type times the change in that capital type. Divide this through by y, and multiple and divide each term on the right by the respective capital type

$$\frac{dy}{y} = \frac{f_1 k_1}{y}\frac{dk_1}{k_1} + ... + \frac{f_J k_J}{y}\frac{dk_J}{k_J}. \tag{2.80}$$

The growth rate of output per capita depends on the growth rate of each individual capital type, multiplied by that capital types share in output. We are still under the assumption that the economy has perfect factor markets, so that each capital type earns exactly its marginal product.

For each individual type of capital j, we would assume that the following holds: $f_j > 0$ and $f_{jj} < 0$. That is adding more k_j raises output, but at a decreasing rate. Given this series of assumptions, it must be that we reach a steady state where all of the capital types grow at the same rate, and output per capita grows at that same rate.

If it's the case that $dy/y = dk_j/k_j$ for all j types of capital, then equation (2.80) tells us what growth rates are permissible. Rewrite the equation as

$$\frac{dy}{y} = \frac{dy}{y}\left(\frac{f_1 k_1}{y} + ... + \frac{f_J k_J}{y}\right), \tag{2.81}$$

and consider the term in parentheses. Each individual term is the share of output paid to a type of capital. If those shares add up to one, then any value for dy/y is permissible. That is, if the shares add up to one, growth could possibly be non-zero. The exact growth rate would depend on the savings rates, population growth rates, and depreciation rates. When will the shares add up to one? Only if production function is constant returns to scale in accumulable factors. That is, $f(k_1, ..., k_J)$ is constant returns to scale. With constant returns, as we accumulate capital of one type, this pushes up the marginal product of all the other types of capital, allowing these marginal products to stay ahead of depreciation and population growth.

If $f(k_1, ..., k_J)$ has decreasing returns to scale, then the shares in the parentheses add up to something less than one. The only possible solution for the growth rate is then for $dy/y = 0$. In this case, we cannot accumulate capital fast enough to keep the marginal product of each capital type higher than it's depreciation rate, and growth eventually goes to zero.

Hence, the only way to generate sustained growth in the baseline Solow model is to assume that there are constant returns to accumulable factors in the economy. Perhaps we believe that is true. However, it implies that the share of output paid to raw labor is zero, and we do see unskilled labor earning *something*. It also implies that higher savings rates should lead to higher permanent growth rates. While across countries nations that save more have some tendency to grow faster, in the time series we see that countries growth rates slow down as they become richer (convergence).

CHAPTER 3

Productivity: Growth and Fluctuations

One of the serious faults of the baseline Solow model is that it does not feature any sustained growth in output per capita in steady state. Yet, as we know, trend growth in per capita income in the U.S. and other OECD nations averages about 1.2% per year.

By incorporating a way for the efficiency of production to increase even without raising inputs of capital or labor, we can create an updated model that captures this feature of the data. The important insight from the Solow model is that it's *only* through productivity improvements that living standards can rise in steady state. Accumulating more factors of production alone is not sufficient, as factors run into diminishing returns.

Having included productivity growth, we'll also have a mechanism for creating fluctuations in output that is not dependent on shocks to capital or labor supplies. We'll examine whether fluctuations in productivity themselves are capable of explaining fluctuations in output over time around the long trend growth.

Finally, we will attempt to distinguish how much of cross-country income differences are driven by factors of production (capital and labor) versus productivity differences. The end result will be that a sizable portion, perhaps as much as two-thirds, of the difference in living standards across countries is attibutable to productivity differences.

3.1 Technological Progress

We begin by maintaining all of the original assumptions of the Solow model regarding firms, competition and factor supplies. The one change is that production now takes the following form:

$$Y_t = F(K_t, E_t N_t) \tag{3.1}$$

where E_t is a term measuring the productivity of labor. An increase in E_t will increase output even without an increase in either capital or labor. I've jumped directly to the aggregate production function here, but in the background you can imagine lots of competitive firms using this technology.

Similar to before, we use the constant returns to scale property to redefine values in terms of *efficiency units*, E_tN_t, which is the total labor effort put into production. Dividing the production through by E_tN_t, we have

$$\tilde{y_t} = f(\tilde{k_t}) \equiv F\left(\frac{K_t}{E_tN_t}, 1\right) \tag{3.2}$$

where $\tilde{x}$ denotes a variable X_t that is scaled by E_tN_t.

Technological progress consists of a fixed growth rate in E_t, so that

$$E_{t+1} - E_t \equiv \Delta E_{t+1} = gE_t. \tag{3.3}$$

Evaluating the growth rate of $\tilde{k}$, we have

$$\frac{\Delta\tilde{k}}{\tilde{k}} = \frac{\Delta K_{t+1}}{K_t} - \frac{\Delta E_{t+1}}{E_t} - \frac{\Delta N_{t+1}}{N_t}. \tag{3.4}$$

As before, we can now use our knowledge of the different growth processes to describe this more fully.

$$\begin{aligned}
\frac{\Delta\tilde{k_{t+1}}}{\tilde{k_t}} &= \frac{I_t}{K_t} - \delta - n - g && (3.5)\\
&= s\frac{F(K_t, E_tN_t)}{K_t} - \delta - n - g && (3.6)\\
&= s\frac{F(K_t, N_t)/E_tN_t}{K_t/E_tN_t} - \delta - n - g && (3.7)\\
&= s\frac{f(\tilde{k_t})}{\tilde{k_t}} - \delta - n - g. && (3.8)
\end{aligned}$$

Here again we've used the assumption that a perfectly costless financial sector transforms savings into investment. Growth in capital per efficiency unit is driven by savings, of course, and "depreciates" with both population growth and technological progress. It is important to keep in mind that as E_t rises, $\tilde{k}_t$ falls, but this does not mean that capital is declining, just capital relative to the efficiency of labor.

Types of Technological Progress

In working with technological progress in the Solow model, I have assumed that technology enters the production function in a particular way, namely as "labor-augmenting", meaning that when E_t goes up, this adds to effective labor. This form of technological change is known as Harrod-neutral technological change, but it is not the only one available. What this means is that when E_t goes up, the isoquants of the production function shift inward (we need fewer *inputs to achieve the same level of output), but it also rotates the production function so that the MRTS goes up (meaning that it takes more additional capital to make up for lost labor).*

We could alternately write the production function as $Y_t = E_t F(K_t, N_t)$, so that technology augments all factors equally, which is called Hicks-neutral technological change. With Hicks-neutral changes, the isoquants shift inward, but remain identical in shape. In other words, the MRTS doesn't change, so the trade-off between capital and labor is not affected. Finally, Solow-neutral technical change is written as $Y_t = F(E_t K_t, N_t)$ and is similar to Harrod-neutral change, only it is "capital-augmenting".

A somewhat surprising results, due to Uzawa (1961), is that to have balanced growth in the long-run (meaning that $\Delta y/y = \Delta k/k$), we must be able to express technological change as Harrod-neutral. In other words, we cannot achieve balanced growth unless technological change acts in a labor-augmenting way.

For our purposes, this doesn't pose much of a problem because we are focusing mainly on Cobb-Douglas production functions. Recall that for the Cobb-Douglas, the EOS between capital and labor is always and exactly equal to one. What this means is that Harrod-, Hicks-, and Solow-neutral technological change are all simple transforms of each other, and therefore everything can be written in Harrod-neutral form.

To see this, take a Cobb-Douglas production function of the form $Y_t = K_t^\alpha N_t^{1-\alpha}$. Let E_t^{Hicks}, E_t^{Solow}, and E_t^{Harrod} be technological change of the three different types so that we have $Y_t = E_t^{Hicks}(E_t^{Solow} K_t)^\alpha (E_t^{Harrod} N_t)^{1-\alpha}$. We can simply write a simpler version of this as

$$Y_t = K_t^\alpha (E_t N_t)^{1-\alpha} \tag{3.9}$$

which is Harrod-neutral, so long as we define

$$E_t = E_t^{Harrod}(E_t^{Hicks})^{1/(1-\alpha)}(E_t^{Solow})^{\alpha/(1-\alpha)}. \tag{3.10}$$

Our Cobb-Douglas production function can be written as Harrod-neutral no matter how we'd like to actually define technological progress. For other production functions, this is not the case and therefore balanced growth may not be possible.

As before, an implication of this modified Solow model is that the economy will eventually come to a steady state, in this case with respect to the value of $\tilde{k}$. The steady state is defined as the point where $\Delta\tilde{k}_{t+1} = 0$, so that

$$sf(\tilde{k}^*) = (\delta + n + g)\tilde{k}^*. \tag{3.11}$$

Note, though, that we are generally not concerned with capital or output per efficiency unit, but rather with their per capita values. So what is capital per person?

$$k_t = E_t \tilde{k}_t \tag{3.12}$$

from which we can infer that even if $\tilde{k}$ is in steady state, k_t is growing over time at the same rate as E_t. This means that y_t is growing in steady state at the same rate as E_t as well. That growth rate is

$$\frac{\Delta y_{t+1}}{y_t} = \frac{\Delta E_{t+1}}{E_t} + \frac{\Delta \tilde{y^*}}{y^*} = g + 0 = g. \tag{3.13}$$

So the exogenously given growth in labor efficiency provides us with a mechanism by which we can generate trend growth in k and y at the rate g. We should stress the "mechanism" part, as this increased efficiency is simply assumed, and there is no economic reasoning given for why g exists or what value it may take. In later chapters we will address this point, and examine models that explain the growth rate g endogenously and as the result of explicit economic choices.

The other implications of the Solow model are kept intact. The return to capital will be constant in steady state. To see this, note that

$$F(K_t, E_t N_t) = E_t N_t f(\tilde{k}_t). \tag{3.14}$$

Take the derivative of both sides with respect to K. The marginal product of capital $F_K = f'(\tilde{k}_t)$. As $\tilde{k}_t$ is constant in steady state, it must be that F_K is constant as well. Steady state $\tilde{k}_t$ is still postively related to s, and negatively to n. In addition, it is negatively related to g. The faster that technology grows, the smaller will be the steady state level of capital per efficiency unit.

It is also true that the farther away from steady state an economy is, the faster it will grow, just as in the original Solow model. This particular implication will be used in the next section to look at how a Solow economy responds to shocks to the productivity term, E_t, which can ultimately be used as one explanation for why the eeconomy cycles around a long-run trend.

3.2 Dynamic Responses to Shocks

Whether we include population growth and technological progress or not, the steady state of the Solow model is stable (i.e. the non-zero steady state, to be accurate). In this section we consider more about the dynamics of the capital stock away from the steady state and what this can tell us about how the economy responds to shocks.

We know that $\tilde{k}^*$ is stable, but in addition the further away from the steady state, the more quickly the economy moves towards it. The logic of why the growth rate of $\tilde{k}$ falls as $\tilde{k}$ rises is identical to what we showed in the prior chapter

$$\lim_{\tilde{k}\to 0} \frac{\Delta \tilde{k}_{t+1}}{\tilde{k}_t} = s\frac{f'(\tilde{k}_t)}{1} - \delta - n - g = \infty \tag{3.15}$$

$$\lim_{\tilde{k}\to \infty} \frac{\Delta \tilde{k}_{t+1}}{\tilde{k}_t} = s\frac{f'(\tilde{k}_t)}{1} - \delta - n - g = -\delta - n - g \tag{3.16}$$

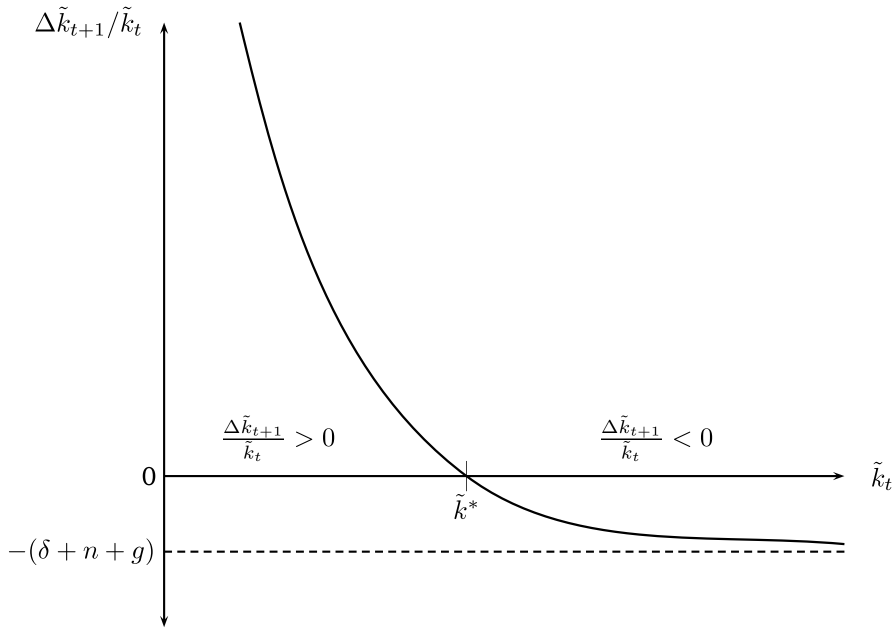


Figure 3.1: Growth Rates and $\tilde{k}_t$

where the first equality in each case is due to L'Hopital's rule and the second is due to the Inada conditions.

Between these extremes, the growth rate of the capital stock is decreasing, which can be seen by taking the derivative of (**??**) with respect to $\tilde{k}_t$:

$$\frac{\partial \Delta\tilde{k}_{t+1}/\tilde{k}_t}{\partial \tilde{k}_t} = s\frac{f'(\tilde{k}_t)\tilde{k}_t - f(\tilde{k}_t)}{\tilde{k}_t^2}. \tag{3.17}$$

The sign of this depends on the term $f'(\tilde{k}_t)\tilde{k}_t - f(\tilde{k}_t)$. This is negative when $f(\tilde{k}_t)/\tilde{k}_t > f'(\tilde{k}_t)$, or if the average product of capital is greater than the marginal product. Given our assumptions regarding the shape of $f(\cdot)$, it is always the case that this holds. Therefore, the growth rate of the capital stock is decreasing in $\tilde{k}_t$.

Using this information, we can plot the growth rate of $\tilde{k}_t$ against the level of $\tilde{k}_t$ in figure 3.1. Where this function crosses the x-axis, the growth rate is equal to zero and there is a steady state. We asserted before that the farther away $\tilde{k}_t$ is from the steady state, the faster it moves towards the steady state. We can see this in figure 3.1, where the absolute value of $\Delta\tilde{k}_{t+1}/\tilde{k}_t$ increases the farther away from $\tilde{k}^*$ one goes.

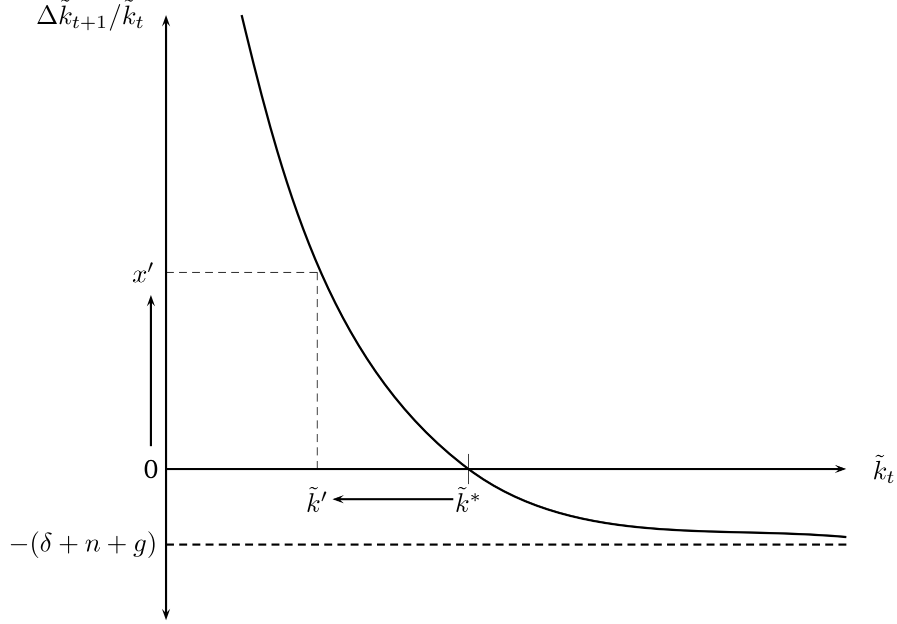


Figure 3.2: A Positive Shock to E_t

Note: The figure shows the effect of a one-time increase in productivity, E_t, on an economy initially in steady state. The increase in E_t lowers capital per efficiency unit to $\tilde{k}'$, and this increases the growth rate of capital per efficiency unit to x'. Over time $\tilde{k}$ will grow and reach the steady state of $\tilde{k}^*$ again.

How is growth of capital *per person* related to this? We know that by the definition of capital per efficiency unit

$$\frac{\Delta k_{t+1}}{k_t} = g + \frac{\Delta\tilde{k}_{t+1}}{\tilde{k}_t} \tag{3.18}$$

so that any change in the growth rate of $\tilde{k}$ translates directly into a change in the growth rate of capital per worker. Similarly, as output per worker can be written as

$$y_t = E_t\tilde{y} = E_t f(\tilde{k}_t) \tag{3.19}$$

the growth rate of output per person can be written as

$$\frac{\Delta y_{t+1}}{y_t} = g + \frac{\Delta\tilde{y}_{t+1}}{\tilde{y}_t}. \tag{3.20}$$

From this we can evaluate the effect of different shocks to the economy on the growth rate of capital per efficiency unit, and hence for capital per person and output per person. First, consider what would happen if the economy began in a steady state and there was a one-time increase in E_t, leaving the trend growth rate of efficiency still equal to g. At the time of the shock, capital per efficiency (and hence output per efficiency unit) *falls*. This drop in capital per efficiency unit implies that the growth rate of $\tilde{k}$ must increase. From (3.18) we know that this raises the growth rate of capital per worker as well. This can be seen clearly in figure 3.2. Over time, as $\tilde{k}$ accumulates the growth rate of capital per person falls towards the steady-state level of g. Output per worker growth follows exactly the same pattern. So this shock to E_t generates a boom, at least in terms of growth rates.

How does the *level* of output per person respond? Recall that $\tilde{y}$ has fallen because of the upward shock to E_t. Does this mean that output per person falls as well? No. Consider the derivative of output per worker with respect to E_t, given (3.19):

$$\frac{\partial y_t}{\partial E_t} = f(\tilde{k}_t) - E_t f'(\tilde{k}_t) \frac{k_t}{E_t^2} \tag{3.21}$$

which is positive if

$$\frac{f(\tilde{k}_t)}{\tilde{k}} > f'(\tilde{k}_t) \tag{3.22}$$

or the average product of capital per efficiency unit is larger than the marginal product. As noted before, this holds by our assumption regarding the concavity of the production function. Therefore a one-time shift in E_t increases income per person, even though income per efficiency unit has gone down.

So a shock to efficiency causes an immediate jump in output per person *and* increases the growth rate of output per person temporarily while the economy approaches a new steady state. If we reverse this shock, a negative shock to E_t will lower output per person immediately and then lower the growth rate of output per person temporarily until the steady state is reached again.

These immediate and dynamic responses to technological shocks thus form one potential explanation for fluctuations in economic performance over time. The real business cycle (RBC) literature takes precisely this view, arguing that real technological shocks hitting the economy are capable of explaining the origins of business cycles. Notice that these cycles occur without any reference to money or prices (hence the "real" moniker).

Solow in Continuous Time

We've done the model so far completely in discrete time. You can easily – and in fact probably more easily – write the Solow model in continuous time.

Using a $\dot{x}$ to denote the time derivative of a variable x, we can ask what the growth rate of *capital per efficiency unit is:*

$$\frac{\dot{\tilde{k}}}{\tilde{k}} = \frac{E(t)N(t)}{K(t)}\frac{\dot{K(t)}}{E(t)N(t)} \tag{3.23}$$

$$= \frac{E(t)N(t)}{K(t)}\left(\frac{\dot{K}}{E(t)N(t)} - \frac{K(t)\dot{E}}{E(t)^2N(t)} - \frac{K(t)\dot{N}}{E(t)N(t)^2}\right) \tag{3.24}$$

$$= \frac{\dot{K}}{K(t)} - \frac{\dot{E}}{E(t)} - \frac{\dot{N}}{N(t)} \tag{3.25}$$

$$= s\frac{Y(t)}{K(t)} - \delta - g - n \tag{3.26}$$

and this is essentially the identical expression to what we had in the discrete time situation.

The instantaneous growth rate of capital per efficiency unit depends on the growth rate of capital (which depends on savings and depreciation), the growth rate of technology (g) and the growth rate of population (n). The equation here is exact (as opposed to the approximate discrete time accumulation equation), but for all intents and purposes the discrete and continuous time versions are identical.

The addition of productivity growth has not changed the fundamental assumptions regarding perfect competition among final goods producing firms, with inelastic supplies of inputs. Hence there is no possibility of output varying due to aggregate demand shocks. A change in productivity, however, would shift the vertical aggregate supply curve, causing output to go up or down depending on the direction of the shock.

3.3 Productivity and Aggregate Fluctuations

The Solow model with productivity growth provides a mechanical way of understanding the source of long-run growth in output per capita. It also suggests one way of describing how output fluctuates about it's trend.

If productivity shocks are driving fluctuations, then this has implications for the time series properties of output. To see this clearly, we need to make one additional assumption. This is that depreciation is complete each period, or $\delta = 1$. This will highlight the role of productivity, but is not a particularly hard assumption to work around. We will also assume that $n = 0$ for simplicity, but again this is not terribly important.

It will also be helpful to describe efficiency at period $t + 1$ as

$$E_t = u_t E_0 (1 + g)^t \tag{3.27}$$

where E_0 is productivity in the initial period. We can set this to one if we like without changing anything. $(1+g)^t$ is just the accumulated growth in productivity between period 0 and period t. The term u_t is an idiosyncratic shock to E_t around it's trend. We assume that is is log-normally distributed, with $ln(u_t) \sim N(0, \sigma^2)$. Hence in expectation $u_t = 1$ and productivity would be at trend.

We can write output per capita as

$$y_t = E_t^{1-\alpha} k_t^{\alpha} \tag{3.28}$$

if we assume a Cobb-Douglas production technology for firms with $Y_i = K_i^{\alpha}(E_i N_i)^{1-\alpha}$. With complete depreciation and zero population growth, it must be the case that

$$k_t = s y_{t-1}. \tag{3.29}$$

Plugging this into the expression for output per capita, we have that

$$y_t = E_t^{1-\alpha}(s y_{t-1})^{\alpha}. \tag{3.30}$$

To work with this, insert our assumption regarding E_t and take logs of both sides. This leaves us with

$$\ln y_t = (1-\alpha)t\ln(1+g) + (1-\alpha)\ln E_0 + (1-\alpha)\ln u_t + \alpha \ln s + \alpha \ln y_{t-1}. \tag{3.31}$$

We can make several simplifications here. First, $\ln(1+g)$ is roughly equal to g. Second, assume that $E_0 = 1$. Now we can write output per capita in time t as

$$\ln y_t = (1-\alpha)tg + (1-\alpha)\ln u_t + \alpha \ln s + \alpha \ln y_{t-1}. \tag{3.32}$$

This says that output per capita is auto-regressive. That is, output per capita in the prior period y_{t-1} influences output per capita today. This is a simple outcome of the Solow model we set up. If anything lowers output in $t-1$, then this lowers the total amount saved in $t-1$, which therefore lowers the capital stock at time t, lowering output.

What the above representation of output per capita suggests is that fluctuations in y_t around its trend over time are due to fluctuations in u_t, the idiosyncratic shock to productivity. The auto-regressive portion of the expression says that when the economy does get shocked away from trend, it doesn't "snap back" right to trend next period. A negative u_t shock today will create a lingering effect on output per capita over the following periods. The coefficient in front of $\ln y_{t-1}$ dictates how long these shocks linger.

Our assumptions about perfect competition, inelastic factor supplies, and the Cobb-Douglas production function imply that this coefficient is exactly equal to α. Therefore, given our assumptions we'd expect this coefficient to be approximately 0.3–0.4, to match capital's share of output. However, if you actually regress output per capita on it's lag you'll find much higher coefficients. There is an exhaustingly long literature in time series macroeconometrics on exactly what this coefficient is. One of the arguments in that literature is that the coefficient is actually equal to one, meaning that shocks to output in $t-1$ permanently affect output in the future. Regardless, the time series evidence is not consistent with this basic set-up.

One of the issues with those time series regressions is that they do not have any way to measure u_t. If u_t has some auto-correlation of it's own, then perhaps the time series regressions are overestimating the coefficient on y_{t-1}. A different approach to evaluating this model is to look specifically at how u_t and y_t relate to each other. If productivity shocks are the most important driver of fluctuations in output per capita, then we should see u_t and y_t highly correlated.

To measure u_t we'll need to measure E_t. So how do we measure E_t? There is no data series on E_t. Therefore we have to back it out of the data. A common approach starts by writing output as follows:

$$Y_t = Res_t K_t^{1-s_N} N_t^{s_N} \tag{3.33}$$

where Res_t is the "residual". s_N is the share of labor in total revenues of firms. Practically speaking, one can back out the residual directly by simply re-arranging

$$Res_t = \frac{Y_t}{K_t^{1-s_N} N_t^{s_N}}. \tag{3.34}$$

This residual is often called "Solow's Residual", as he was the first to propose something like this in a separate paper from 1956.

To proceed, Solow made an important leap, one that is often glossed over in the literature. *Under the assumption that markets are perfect and production is constant returns to scale*, then it is the case that the revenue shares are exactly equal to the elasticity of output with respect to the factor of production. In the Cobb-Douglas case, this implies that $s_N = 1-\alpha$, but the principle holds regardless of the exact production technology.[1]

In practice, to obtain our estimate of s_N (and hence of α) we simply take total wages paid to workers (wN) and divide by total revenues, or GDP (Y). To obtain s_K, we are using the assumption on constant returns to get that $s_K = 1 - s_N$. With perfect competition, all revenues are divided up into either payments to labor or payments to capital.

Now because $s_N = 1 - \alpha$ and $s_K = \alpha$ under perfect competition and constant returns, the residual reduces to the following:

$$Res_t = E_t^{1-\alpha}. \tag{3.35}$$

In other words, under our assumption that the economy has perfect markets and constant returns, the residual gives us information about the level of productivity, E_t. So to measure E_t in this economy all I need to do is calculate the residual.

[1]To see this, note that the revenue share for labor is just $s_N = wN/Y$. With perfect competition and perfect factor markets, $w = F_N$, and therefore $s_N = F_N N/Y$. The right-hand side of this last equality is simply the definition of the elasticity of output with respect to N. With constant returns is has to be that $s_K = 1 - s_N$.

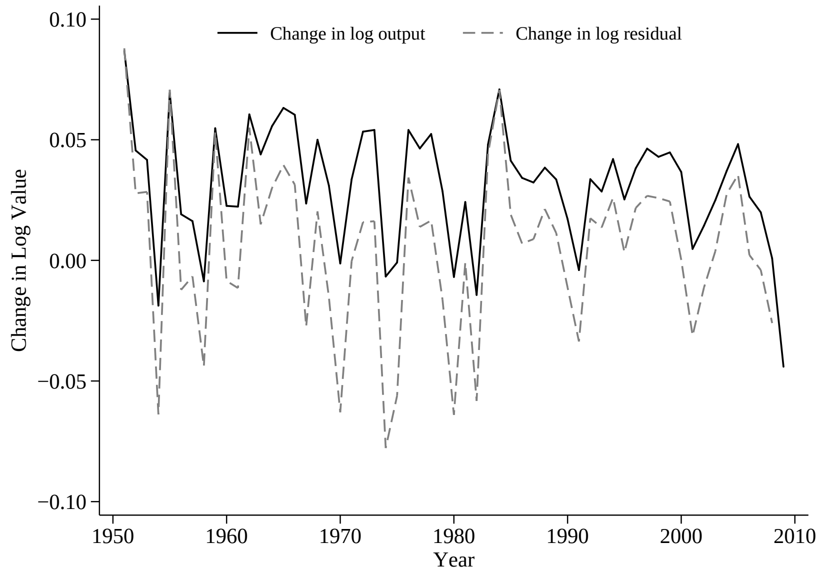


Figure 3.3: Output and Residual Co-movement in the U.S.

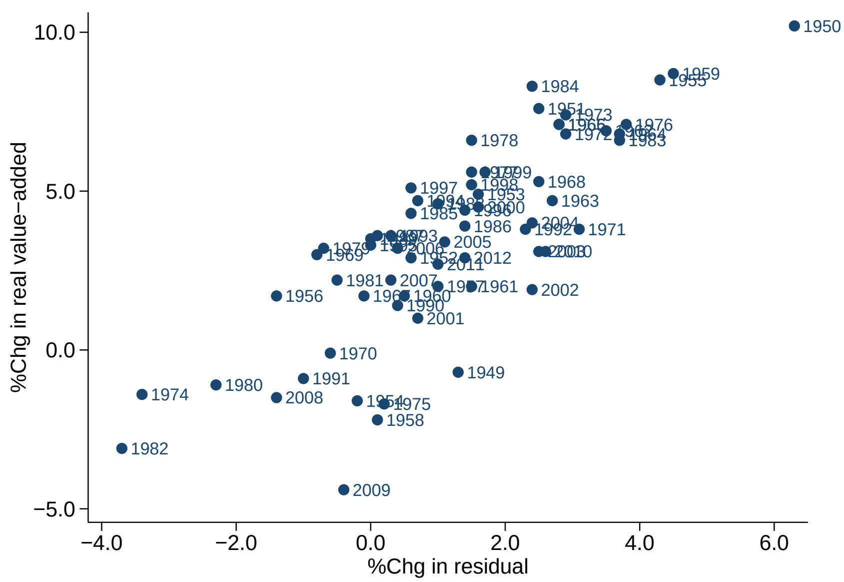


Figure 3.4: Output and Residual Growth Rates in the U.S.

For the moment, assume that our economy does have perfect competition and constant returns. We can calculate the residual for the U.S., which then tells us what E_t is for the U.S. We can plot this log of this E_t against the log of Y_t to see if in fact output and productivity are highly correlated as our Solow model suggests.

Figure 3.3 shows the results. As you can see, output and the residual move very similarly. The series are offset somewhat (the output series has a higher average change), and that gap represents the effect of changes in either K_t or N_t on output. From this figure, it appears that the residual is important in driving fluctuations in output. Perhaps a better way of seeing the correlation is to just plot the growth rates of output and the residual against each other, as in figure 3.4. Here you can clearly see that years in which the residual grows significantly, output also grows significantly.

Does this mean that it is real productivity shocks that drive fluctuations? Only if we assume that we have perfect markets and constant returns. With these assumptions the residual does pick up only fluctuations in E_t, and therefore we could conclude that productivity shocks appear to be very important in driving output fluctuations. However, that is only true under these very strict assumptions.

In a series of papers, Robert Hall (1988, 1990) showed that the residual Res_t does *not* necessarily measure E_t if there is imperfect competition among firms or increasing returns to scale. With imperfect competition, the share of labor does *not* equal the elasticity of output with respect to labor, and similarly for capital. What does this imply happens to our calculation of the residual?

Cost Minimization and Markups

To give us more flexibility in how we deal with firms, we can break the profit maximization problem up into two parts. First, minimize costs for producing any given level of output Y*. Second, maximize profits by selecting the optimal* Y *given market conditions.*

It turns out that for our purposes we can get a lot of traction simply by considering the first part of this: cost minimization. This will give us a very simply relationship that we can exploint regarding returns to scale, profits, and markups.

The firm's cost minimization problem is to minimize $wN + RK$ *such that* $F(K, N) = Y$*, where* Y *is an arbitrary output amount. We can write this as a Lagrangian*

$$\mathcal{L} = wN + RK + \theta(Y - F(K, N)) \tag{3.36}$$

where θ *is the Lagrange multiplier. In this problem, the multiplier has a very specific interpretation. It represents the additional cost of increasing* Y *incrementally. In other words,* θ *is the marginal cost of production.*

Taking first order conditions is straightforward, and yields that the firm will set $\theta F_N = w$ *and* $\theta F_K = R$*. Now, knowing these cost minimization conditions we can proceed to examine* *the firm's profits.*

$$\Pi = PY - wN - RK \quad (3.37)$$
$$= PF(K,N) - \theta F_N N - \theta F_K K, \quad (3.38)$$

where P is the price the firm charges for their output. Our assumptions regarding market power would inform us about how P is set (or not) by the firm, and hence how to maximize profits. However, we do not need to make any specific assumptions at this point to get a useful fact.

Re-arrange the profits of the firm as follows

$$1 = \frac{\theta}{P}\frac{F_N N}{F(K,N)} + \frac{\theta}{P}\frac{F_K K}{F(K,N)} + \frac{\Pi}{PF(K,N)} \quad (3.39)$$
$$= s_N + s_K + s_\pi \quad (3.40)$$

This says that the share of revenues paid to labor and capital - the first two terms - and the share of revenue paid out as profits must add up to one. The s_i terms are simply shorthand for the share of revenue.

Take the share paid to labor. This is equal to θ/P times the elasticity of output with respect to labor $F_N N/F(K,N)$. We will define the following as the ***markup****, μ:*

$$\mu = \frac{P}{\theta}. \quad (3.41)$$

The markup is the price charged for output relative to the marginal cost of that output, θ. If we assume the firm has market power, then it will be the case that $\mu > 1$, or they charge a price higher than marginal cost. If there is perfect competition, then $\mu = 1$, or firms charge exactly marginal cost. Hence the share of revenues going to labor, s_N, is equal to $(F_N N/F(K,N))/\mu$, or the elasticity of output with respect to labor divided by the markup.

Using this defintion of the markup, re-arrange again to get

$$\frac{F_N N}{F(K,N)} + \frac{F_K K}{F(K,N)} = \mu(1 - s_\pi). \quad (3.42)$$

Now, to complete the relationship, note what we have on the left-hand side. This is the elasticity of output with respect to labor plus the elasticity with respect to capital. Together, these two elasticities add up to the returns to scale in the production function. In other words, if those elasticities add up to one, then we have constant returns to scale. If those elasticities add up to more than one, we have increasing returns to scale, and if they are less than one, we have decreasing returns. You can confirm this yourself using Cobb-Douglas production functions quite readily, but the principle holds for any typical production function.

This relationship puts some structure on how firms profits, markups, and returns to scale can be related if they are cost-minimizing.

- *If production is constant returns to scale (so $F_N N/F(K,N) + F_K K/F(K,N) = 1$), then profits can be positive ($s_\pi > 0$) only if the markup is bigger than one. In other words, if there is market power for firms.*
- *If production is increasing returns to scale (so $F_N N/F(K,N) + F_K K/F(K,N) > 1$), then profits can be zero even if the markup is bigger than one. That is, even with no market power, firms will charge a markup over marginal cost if they have increasing returns to scale.*
- *If firms have market power ($s_\pi > 0$) but do not charge a markup ($\mu = 1$), then production must be decreasing returns to scale.*

First we have to understand how the shares of revenue, s_N and s_K, relate to the actual production elasticities. Hall uses the fact, and you'll find details for this in the boxed section, that these are related in the following manner

$$s_N = \frac{1-\alpha}{\mu} \tag{3.43}$$

where $\mu \geq 1$ is the *markup* over marginal cost that a firm charges. In general, if a firm has market power then it will charge a price for its output that is actually higher than the marginal cost of producing it. μ tells us the size of that markup. If $\mu = 2$, for example, then firms are charging a price that is twice marginal cost.

What you can see is that s_N is therefore *less* than the elasticity of output with respect to labor. So in calculating the residual from (3.34) we are under-weighting labor, and over-weighting capital. More importantly, if we plug in the relationship from (3.43) to the residual calculation, and reduce things we find that

$$Res_t = E_t^{1-\alpha} \left(\frac{N_t}{K_t}\right)^{s_N(\mu-1)}. \tag{3.44}$$

In other words, with imperfect competition it no longer is the case that the Solow residual is a measure of only efficiency, E_t. It now picks up any changes in N_t or K_t. The larger is the markup μ, the bigger the effect these factors have on the residual. If $\mu = 1$, or we have perfect competition, then the term involving N/K goes to zero and we're back to Solow's original version.

From this imperfect competition version of the residual, we can see that anything that makes N_t fall (unemployment) will drag down the residual. So the pattern seen in figure 3.3 may simply reflect the co-movement of employment and output, rather than actual productivity shocks. It's

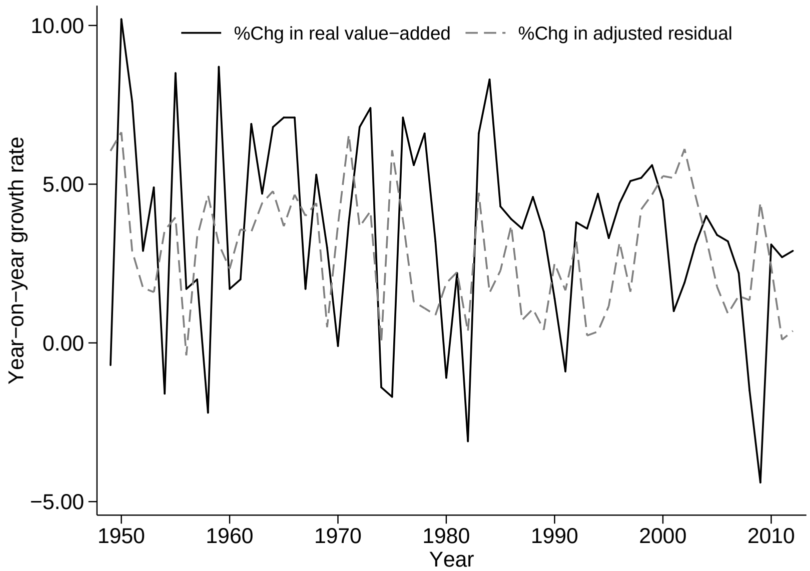


Figure 3.5: Output and the Adjusted Residual Co-movement in the U.S.

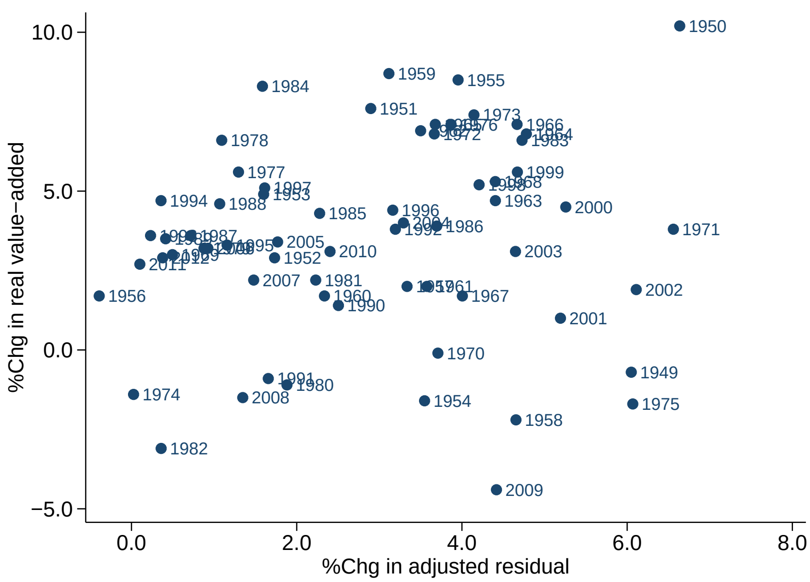


Figure 3.6: Output and the Adjusted Residual Growth Rates in the U.S.

possible that E_t is unchanging over time, or growing steadily, but the *residual* will vary with the business cycle. So we do not have any definitive evidence that productivity shocks drive the business cycle.

To see whether E_t is driving business cycles, we should back it out of equation (3.44). To do so, we need to make some assumptions about the size of μ. For the purposes here, I set it to $\mu = 2$, meaning that prices are twice marginal costs. One can argue for larger or smaller estimates, but this value illustrates the possible impact of imperfect competition. With μ, we can rearrange (3.44) and find

$$E_t^{1-\alpha} = Res_t \left(\frac{K_t}{N_t}\right)^{s_N(\mu-1)}. \tag{3.45}$$

In figure 3.5 the growth rate of $E_t^{1-\alpha}$ year-over-year is plotted against the growth rate of output, similar to figure 3.3. As can be seen, there is a less pronounced co-movement of the two series. Productivity - once we more accurately measure it under imperfect competition - does not appear to be as closely associated with business cycles. One can even see situations, such as the recessions in 2001 and 2009, where productivity appears to spike upwards when the economy is shrinking. A more useful way of seeing this is in figure 3.6, which plots the growth rate of output against the growth rate of this adjusted productivity measure. Here, unlike figure 3.4, there is essentially no relationship. It is not the case that years with large increases in productivity tend to be years in which there is positive output growth. Productivity does not appear to be correlated with business cycles.

This holds for the particular assumptions that we've implemented. Now we have imperfect competition with a markup of 2, which could be inaccurate. We've continued to assume that there are constant returns to scale (meaning that capital's share of costs is 1 minus labor's share). This may also be incorrect. And finally the figures are based on indicies of labor inputs and capital services, which may not accurately measure the size of K or N over time. Regardless, the point is that simply looking at the basic Solow residual does not necessarily provide us with evidence that productivity shocks are responsible for business cycle fluctuations. And if it is not productivity shocks that drive productivity shocks, then the points made in the prior chapter hold. There needs to be some upward-sloping aggregate supply curve and nominal shocks, or some kind of exogenous variation in labor and/or capital supplies.

3.4 Productivity and Long-Run Growth

With imperfect competition, the residual doesn't give us any precise information about productivity. That clouds the analysis of business cycles, but does it also affect our study of long-run growth? With perfect competition and constant returns, the residual would be a good way of measuring E_t, and we could track the growth in E_t over time, giving us a measure of trend growth in productivity, or g. This, recall, is the source of long-run growth in output per capita.

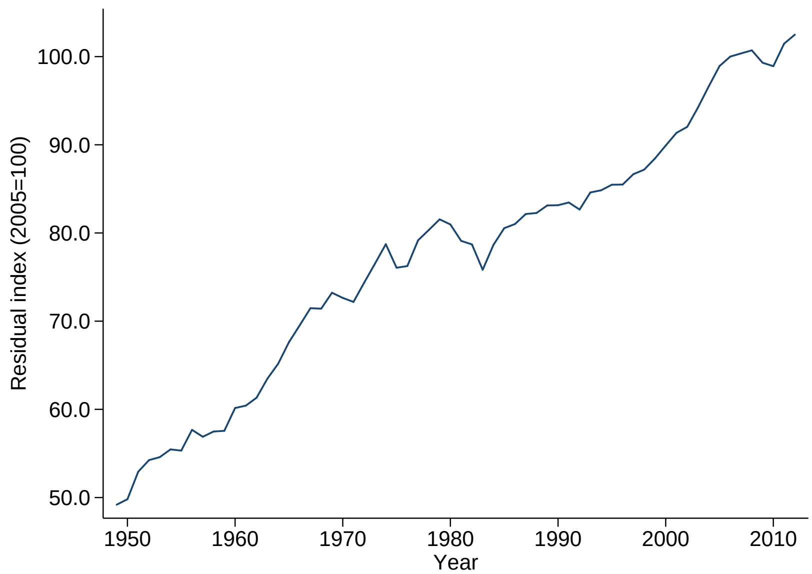


Figure 3.7: Residual Index in the U.S.

With imperfect competition, the residual can still provide useful information even though it isn't a pure measure of E_t. To see this, write the residual as follows

$$Res_t = \frac{E_t^{1-\alpha}}{E_t^{s_N(\mu-1)}} \left(\frac{N_t E_t}{K_t} \right)^{s_N(\mu-1)}. \tag{3.46}$$

The fraction in the parentheses is just the definition of $\tilde{k}_t$, and we can reduce the first fraction so that we get

$$Res_t = \frac{E_t^{s_N}}{\tilde{k}_t^{\ s_N(\mu-1)}}. \tag{3.47}$$

Now, the residual captures productivity E_t, scaled by the size of $\tilde{k}_t$. From the perspective of long-run growth, we think that $\tilde{k}_t$ will settle down to a steady state value of $\tilde{k}^*$, and so will be unchanging over time. If that is true, then the residual can give us a good indicator of the growth of E_t.

Figure 3.7 plots the Solow residual over time as an index with 2005 set to 100. As you can see, there has been steady upward growth in the residual, despite fluctuations around that trend. From this we can conclude that there has been steady upward growth in productivity, E_t, even if we are not precisely sure about the exact level of E_t at any given point in time.

One caveat to this; if the values of s_N or μ are changing over time, then it would be the case that the residual might not give us accurate information on E_t. If μ were climbing over time, then the residual would be rising too, and this would show up as a rising residual in the data. However, it seems unlikely that μ could grow sufficiently and over this long of a period to drive residual growth. Remember that μ represents the markup over marginal cost charged by firms, and this would have to reach ridiculous levels to explain the growth in the residual. Growth in s_N is also an unlikely candidate, as it is bounded between zero and one.

3.5 Cross-country Income Differences

In thinking about why some countries are rich and some are poor, a first step is to understand whether it is factors of production (capital and labor) or productivity that is driving the differences. In per-capita terms, if production can be written as

$$y_t = E_t^{1-\alpha} k_t^{\alpha}, \tag{3.48}$$

which in logs is just

$$\ln y_t = (1-\alpha)\ln E_t + \alpha \ln k_t. \tag{3.49}$$

We have lots of countries, and there is some variance across countries in their $\ln y_t$. We'd like to know how much of this variance is due to variation in k_t (factors), versus productivity (E_t). A simple way to think about doing this would be to simply regress $\ln y_t$ on $\ln k_t$. If we did that, we should get back a coefficient estimate of α. The R-squared from this regression would be

$$R^2 = \frac{\alpha^2 Var(\ln k_t)}{Var(\ln y_t)}. \tag{3.50}$$

This R-squared tells us precisely the fraction of variation in $\ln y_t$ that is accounted for by variation in $\ln k_t$.

However, running that regression will not give us a good estimate of α. That is because it is almost surely true that $\ln k_t$ is correlated with $\ln E_t$, and by not including $\ln E_t$ in the regression, we will get a biased estimate of α. In practice, if you did run this regression, you'd find a coefficient on log capital per worker of about 0.6–0.7. This seems high, given that capital's share of output is around 0.3–0.4.

If we cannot use the simple regression, then how do we determine something similar to the R-squared for the role of capital per worker? One option commonly used involves the residual again. Going back to the original definition of the residual in (3.33), we can re-write this in per-capita terms as

$$y_t = Res_t k_t^{1-s_n}, \tag{3.51}$$

and note that this looks similar to what we have in (3.48). As before, *if there is perfect competition and inelastic factor supplies* then $s_n = \alpha$ and $Res_t = E_t^{1-\alpha}$.

How is this useful? Consider the following ratio

$$\frac{Var(\ln k_t^{1-s_n})}{Var(\ln y_t)} \tag{3.52}$$

where we've taken the variance of log capital per worker across countries, and divided it by the variance of log output per capita across countries. The numerator gives us the amount of variation we'd expect to see in output per capita if the Res_t were identical across countries. If this ratio approaches one, then it suggests that capital per worker is sufficient to explain the variation in output per capita observed. If this ratio approaches zero, then it suggests that factors are unimportant in explaining variation in output per capita.

Caselli (2005) refers to this as the *success ratio*, and uses it as a measure of how important capital is in the variation in living standards across countries. If we apply some simple rules for logs and variances, then this success ratio is just

$$(1-s_n)^2 \frac{Var(\ln k_t}{Var(\ln y_t)} \tag{3.53}$$

and what Caselli is calculating is equivalent to the R-squared of a regression of log y_t on log k_t. The difference with the regression is that Caselli, and others doing similar work, simply make an assumption about the size of $1 - s_n$, rather than trying to estimate it from the data. In general, researchers have used the data on wages as a fraction of GDP to approximate s_n to be about 0.6. This is an average level. Gollin (2002) finds variation in s_n across countries, but it is not systematically related to output per capita.

One could calculate the success ratio using this equation along with data on capital stocks and output per capita. Using data from the Penn World Tables (PWT), version 8.0, on capital k_t and the output y_t, the success ratio is just 0.21. That is, variation in physical capital per worker only accounts for about 20% of the variation in income per capita across countries.

We touched briefly on human capital in the prior chapter. Here, we can add that additional capital type to the analysis. The PWT include a measure of human capital per worker in their data. If production is actually $y_t = k_t^{\alpha}(h_t E_t)^{1-\alpha}$, then we could try to back out a residual using

$$y_t = Res_t k_t^{1-s_n} h_t^{s_n}, \tag{3.54}$$

and now the residual will pick up any remaining differnce in output per worker that is not due to physical capital or human capital. The success ratio associated with this kind of analysis would be

$$\frac{Var(\ln k_t^{1-s_n} + \ln h_t^{s_n})}{Var(\ln y_t)}. \tag{3.55}$$

Using the human capital data, this success ratio is approximately 31%, or not quite one-third of the variation in output per worker is due to differences in capital stocks of both kinds. This leaves us with the conclusion that the residual is explaining far more of the gap between countries than accumulation of factors alone. Thus gaps in living standards are likely due mainly to levels of productivity, E_t.

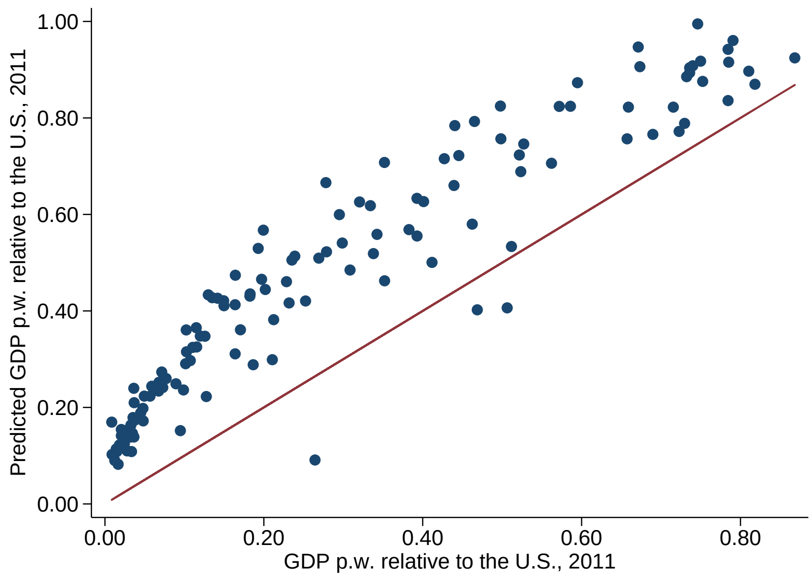


Figure 3.8: Predicted and Actual Output p.w. relative to the U.S., 2011

A different way to see the role of physical and human capital stocks is figure 3.8. The level of $k^{1-s_n}h^{s_n}$ relative to the U.S. is plotted against the observed level of y_t relative to the U.S. in 2011. If capital (physical and human) were the only driver of differences in output per worker between countries and the U.S., then the dots would line up along the 45 degree line. However, we see that essentially every country is above the line, indicating that capital stocks predict countries would be far richer relative to the U.S. than they actually are. Looking only at capital stocks, we'd expect poor countries to be better off than they are. The conclusion is that there must be some other factor that differs with the U.S., making them poorer. That factor is the residual, which we think is closely correlated with productivity, E_t.

Measuring Human Capital

How do we get a measure of human capital per worker? The norm in the literature at this point is to use a Mincerian setting. This means that we assume that human capital depends on years of schooling, and this is related to wages in a particular way. Specifically, we assume that human capital per worker can be written as

$$h_t = e^{\phi S_t} \tag{3.56}$$

where S_t is years of schooling for a worker, and ϕ is the percent increase in productivity that a worker gets for each year. It's not essential that we assume that ϕ is identical for each year, you can do more sophisticated versions that allow for different returns for primary, secondary, or tertiary education. However the actual empirics don't tend to change.

If this is the form that human capital takes, then wages can be written (assuming they are equal to marginal products) as

$$w_t = (1-\alpha)k_t^{\alpha}E_t^{1-\alpha}h_t \tag{3.57}$$

given our typical production function. Take logs of this and you have

$$\ln w_t = [\ln(1-\alpha)k_t^{\alpha}E_t^{1-\alpha}] + \phi S_t. \tag{3.58}$$

Log wages depend on an intercept term that involves the capital stock and productivity, and then a second term that depends on the years of schooling. The Mincerian name comes from the fact that an equation such as this is what Jacob Mincer proposed as the best fit for the relationship between wages and years of schooling when studying individual workers.

In practice, many researchers will assume that the value of ϕ is identical for each country, and then variation in human capital by country depends solely on S_t. There are departures from this, however, and depending on how one precisely models human capital one can get it to vary a lot or a little across countries. The fact is that we do not have a definitive answer for how to measure human capital in production.

3.6 Multiple Goods

The standard Solow model, or the expansions to involve multiple types of capital, all assume that there is a single homogenous good produced, "output". Another way of saying this is that the Solow model and variants assume that all the various types of goods and services produced in the world are perfect substitutes. This, or course, is not true. We can expand on the baseline Solow model to allow for the production of distinct types of goods, and each has their own individual productivity term. The mechanics of modeling this are very useful in a variety of contexts in economics. Here, we'll see that this does not change any of the significant predictions of the Solow model, or about how productivity affects the economy in the aggregate.

There are a few ways to set up and think about multiple goods, and while the semantics are quite different, they end up having the same consequences. Here are two main ways to think about this.

1. **Final and intermediate goods:** Individual have utility $U = f(Y)$, and $f(Y)$ has typical dminishing marginal utility properties, meaning individuals always want more. The Y itself is a "final good" that is produced by some perfectly competitive final goods firms with a production function like $Y = \Pi_i^N Y_i^{\theta_i}$, where Y_i are the N intermediate goods, and their profits are $\pi_i = Y - \sum_i P_i Y_i$. The price of the final good is the numeraire, so equals one. These intermediate goods are produced by separate firms/sectors with production functions like $Y_i = A_i K_i^{\alpha} L_i^{1-\alpha}$, and profits of $\pi_i = P_i Y_i - wL_i - RK_i$. There is a total of L and K in factor supplies. In this setting, all the action is about final goods firms buying intermediate goods. The consumer maximization problem is trivial.

2. **Consumer utility:** Individuals ahve utility $U = \Pi_i^N C_i^{\theta_i}$ and a budget constraint of $M = \sum_i P_i C_i$, where M is their income. There are firms/sectors that produce the consumption goods using a technology $Y_i = A_i K_i^{\alpha} L_i^{1-\alpha}$, and profits of $\pi_i = P_i Y_i - wL_i - RK_i$. Market clearing requires that $C_i = Y_i$ (supply equals demand), and national income accounting requires that $M = wL + RK$, where L and K are the total amount of labor of capital. In this setting, you solve the consumer maximization problem, solve the firm maximization problem, and then combine.

What I hope is obvious from those two descriptions is that the problems are effectively identical. It doesn't matter if it is final goods firms, or consumers, from which we derive the demand for the separate goods. Once we have that demand, everything works out identically. It's purely a question of labeling. You'll often see the final goods formulation setup used, because it tends to be more efficient in notation, but remember you could always go back and recast that as a consumer problem if you wanted to change the interpretation.

I made a very specific statement about the nature of utility or final goods production in the two examples. They are Cobb-Douglas, as in $Y = \Pi_i^N Y_i^{\theta_i}$. They are also constant returns to scale, meaning that $\sum_i^N \theta_i = 1$. This is a very restrictive formulation for demand, as it assumes the elasticity of substitution between goods is exactly one, and the final goods firm (or consumers if you did it that way) spend a constant fraction of their revenues (income) on each good. That's going to make our life easy in this section. In the next section, we'll relax that.

So let's proceed with the final goods type setup. The final goods firm wants to maximize profits, so it sets the marginal cost of each intermediate good equal to its price. That is, take the first-order condition with respect to each good i

$$\theta_i \Pi_i^N Y_i^{\theta_i} = Y_i P_i. \tag{3.59}$$

We assume the final goods sector is perfectly competitive, and so there are no net profits. This means that $\Pi_i^N Y_i^{\theta_i} = \sum_i^N P_i Y_i$, and thus we have that

$$P_i Y_i = \theta_i \sum_i^N P_i Y_i \tag{3.60}$$

or the expenditure on good i is a constant fraction θ_i of total expenditure. For now, that's all we need to know about the final good sectors demand for goods.

The output of each good is determined by

$$Y_i = A_i K_i^{\alpha} L_i^{1-\alpha} \tag{3.61}$$

which is just the familiar Cobb-Douglas. Note that we're assuming each good uses the same capital elasticity, α. This will make life much, much easier, but need not be a good description of the world. If we allowed α to differ by good, then we'd find later that the absolute size of the capital stock and/or labor supply would matter for which goods got produced.

We could dig down further, and assume that this production function is the result of a bunch of perfectly competitive firms all producing that good. Let's make that assumption, so that we can simply talk about the above production function as if it were for a single firm producing good i.

We're going to further assume that across the goods, the different firms all compete with one another, taking factor prices as given, and taking the price of their output, P_i, as given. If each good were produced by a single firm, we might think that firm had some market power to set price above marginal cost. We'll get to that. For now, the firms take prices as given.

Firms will maximize profits (although those profits will ultimately be zero) given by

$$\pi_i = P_i Y_i - w L_i - R K_i, \tag{3.62}$$

by choosing the amount of labor and capital to hire. This gives two first order conditions

$$\alpha \frac{P_i Y_i}{K_i} = R \tag{3.63}$$

$$(1-\alpha) \frac{P_i Y_i}{L_i} = w \tag{3.64}$$

(3.65)

which can be solved to show that

$$\frac{K_i}{L_i} = \frac{\alpha}{1-\alpha} \frac{w}{R}. \tag{3.66}$$

Note that the right-hand side is not specific to firm i. Every firm, producing any kind of output, all choose the same capital/labor ratio, regardless of their level of A_i. This is because the the values of α are identical, and this dictates the ratio of capital to labor that makes sense, given the factor prices.

Knowing this capital/labor ratio, go back and write the production function for each good as

$$Y_i = A_i L_i \left(\frac{\alpha}{1-\alpha} \frac{w}{R} \right)^{\alpha} \tag{3.67}$$

which shows that output of each sector can be written as linear in labor (we could do this to be linear in capital too, it isn't crucial that this is labor). Using this expression for Y_i, go back to the first-order condition for labor, and plug in to find that

$$P_i = \frac{1}{A_i} \left(\frac{R}{\alpha} \right)^{\alpha} \left(\frac{w}{1-\alpha} \right)^{1-\alpha}. \tag{3.68}$$

What's on the right is the marginal cost of producing good i. Note that as A_i goes up, P_i falls proportionately. Taking this expression for two goods, i, and j, it follows that

$$\frac{P_i}{P_j} = \frac{A_j}{A_i} \tag{3.69}$$

or the relative price of goods is directly proportional to the relative productivity between the two sectors.

To figure out how much labor, L_i, is used to produce each good (and hence how much capital is used, since we know the K/L ratio), we have to pull back in what we know about the final goods sector demand for intermediate goods. We know the expenditure on each intermediate good, and the relative expenditure on two goods i and j is

$$\frac{\theta_i}{\theta_j} = \frac{P_i Y_i}{P_j Y_j}. \tag{3.70}$$

Now, put in what we know about the relative price P_i/P_j, and then plug in what we know about how Y_i is determined, and we have

$$\frac{\theta_i}{\theta_j} = \frac{L_i}{L_j}. \tag{3.71}$$

The amount of labor employed in good i relative to good j is exactly proportional to the expenditure share on those goods. This, by the way, is going to be something of a general result. If the amount of expenditure on good i goes up, then there are only two ways for this to happen, either the price of i rises, or the amount of output of i rises. If the price rises, then this generates profits for the firm producing i, but given we assumed competition, this should pull in more firms, which would drive down the price while expanding production, which would involve hiring more workers. Thus the number of workers producing good i depends on the expenditure share.

To solve for exact values of L_i, we know that total labor must add up, or

$$\sum_i^N L_i = L. \tag{3.72}$$

Use the above expression for L_i/L_j in here, and we have

$$\sum_i^N L_j \frac{\theta_i}{\theta_j} = L \tag{3.73}$$

and note that the summation is not over j, but over i. So we can pull out the j terms to get $L_j \sum_i^M \theta_i = \theta_j L$. Now, the summation of the expenditure shares must, by definition, add up to one, so we have

$$L_j = \theta_j L \tag{3.74}$$

or the share of labor used in sector j is just the expenditure share. Remember that every sector uses exactly the same capital/labor ratio. It follows that therefore the share of capital employed in each sector is

$$K_j = \theta_j K \tag{3.75}$$

as well. It must be that the capital/labor ratio in any given firm/sector i is equal to the aggregate capital/labor ratio, $K_i/L_i = K/L$.

We can wrap all this up now by thinking about how to determine the production of final goods - which is what we wanted to know about in the first place. Simply plug back in to the final goods expression, remebering that the sum of the θ_i terms is equal to one.

$$\begin{aligned} Y &= \Pi_i^N Y_i^{\theta_i} && (3.76)\\ &= \Pi_i^N \left(A_i K_i^{\alpha} L_i^{1-\alpha}\right)^{\theta_i} && (3.77)\\ &= \Pi_i^N \left(A_i K^{\alpha} L^{1-\alpha}\theta_i\right)^{\theta_i} && (3.78)\\ &= K^{\alpha} L^{1-\alpha} \Pi_i^N \left(\theta_i A_i\right)^{\theta_i} && (3.79) \end{aligned}$$

What does this leave us with? Well, it's just a Cobb-Douglas production function with respect to total capital and labor. The last term involving the product of all those A_i terms is just total factor productivity. If you wanted to, you could write

$$E = \Pi_i^N \left(\theta_i A_i\right)^{\theta_i/(1-\alpha)} \tag{3.80}$$

and then write

$$Y = K^{\alpha}(EL)^{1-\alpha} \tag{3.81}$$

and we're right back where we started from with productivity. What we know from this section is that even if output is in fact the composition of multiple types of goods, this need not change anything about our overall analysis of the economy. It just tells us that the productivity term is an aggregate of the productivity terms of each individual firm/sector within the economy. Using a single aggregate production function, as we did with the Solow model and the initial part of the productivity chapter, is consistent with more complex structures of production.

You can think of improvements in productivity in any given firm, A_i, as having several outcomes. First, we know aggregate productivity will rise. How much it rises depends on how big θ_i is. The larger this is, the more powerful an increase in A_i will be in the aggregate. Think of the opposite, if $\theta_i = 0$, then no one buys the good, and so productivity increases in that firm are meaningless. But as A_i changes, it is also changing the relative price P_i, making it cheaper.

3.7 Complements and Substitutes

In the prior section, the final goods firm (and/or consumer preferences) were Cobb-Douglas, meaning that expenditure shares on different goods were constant. This was a result of the assumption that the elasticity of substitution between goods was exactly one. If a given firm has a productivity increase, then the price of its good falls in relative terms. By itself, this would lower the expenditure share on that good. But because people substitute towards that good due to the lower price, the expenditure share rises. With an elasticity of substitution of one, these two effects cancel perfectly, leaving the expenditure shares constant. A one percent fall in the relative price of good i is associated with a one percent rise in the relative consumption of that good, leaving expenditure unchanged.

That's fairly restrictive, and so by using an alternative structure for final good production (or consumer demand) we can get more interesting action on the demand for specific goods.

Let final goods be produced according to a CES production function as in the following

$$Y = \left(\sum_i^N \theta_i^{1/\sigma} Y_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \tag{3.82}$$

The value of σ here is the elasticity of substitution between goods. As this goes to zero, goods become perfect complements, and this would turn into a Leontief function based on the minimum consumption of the N goods. As σ goes to one, you could take the limit and find that this is Cobb-Douglas. As σ goes to infinity, this becomes the summation over the N goods, or perfect substitutes. The θ_i values serve the same purpose as in the Cobb-Douglas, they indicate some fixed relative importance of each good in final goods production.

Again, the final goods firm wants to maximize profits, and takes the price of each intermediate good as a given, P_i. What I'm going to change here is that we'll work with nominal prices, meaning I won't set the final goods price to be equal to one. This will just help you see that we can work out an aggregate price index, and that it is related to productivity in the end. So profits in the final goods sector are

$$\pi = PY - \sum_i^N P_i Y_i, \tag{3.83}$$

with P as the aggregate price index. Taking the first-order conditions with respect to any individual intermediate good i we have the following

$$P\theta_i^{1/\sigma}\left(\frac{Y}{Y_i}\right)^{1/\sigma} = P_i. \tag{3.84}$$

This is the inverse demand curve for good i from the final goods sector. You can confirm this is the right first-order condition by taking the derivative yourself and doing some simplification, where I've simply leapt to this solution. An alternate way to write this equation is as

$$P_i Y_i = \theta_i PY\left(\frac{P}{P_i}\right)^{\sigma-1}. \tag{3.85}$$

This shows the expenditure on good i is a fraction of total expenditure PY, and this looks a lot like the Cobb-Douglas case except for the extra term in the parentheses. What that extra terms says is that the expenditure share depends on the relative price of a good. If $\sigma > 1$, and goods are substitutes, then if P_i goes up relative to P, the expenditure on i falls, as people are willing to substitute away from an expensive good. If $\sigma < 1$, and goods are complements, then if P_i goes up relative to P, expenditure on i rises, as people are not willing to substitute, and so they have to spend more on i to keep consuming it. If $\sigma = 1$, you can see this reduces to the Cobb-Douglas case from the prior section.

Now, when we get to the intermediate goods producers, and we're going to treat them exactly like we did in the prior section. They are competitive, take prices and factor prices as given, and produce using those Cobb-Douglas production functions. They will end up with identical capital/labor ratios, and the relative price of goods will still be set by $P_i/P_j = A_j/A_i$. You can confirm that for yourself by going back and seeing that we didn't use anything about demand to find that relative price relationship.

We needed the expenditure shares when we went to try and solve for the actual allocation of labor and capital to each firm, and to solve for final output. So let's take the expenditure expression for two separate sectors, as given above, to write

$$\frac{P_i Y_i}{P_j Y_j} = \frac{\theta_i}{\theta_j}\left(\frac{P_j}{P_i}\right)^{\sigma-1}. \tag{3.86}$$

Now, plug in what we know about how relative prices and output work.

$$\frac{L_i}{L_j} = \frac{\theta_i}{\theta_j}\left(\frac{A_i}{A_j}\right)^{\sigma-1}. \tag{3.87}$$

We can stop here to see how productivity affects the allocation of labor to each sector. If A_i rises, then whether this results in L_i rising relative to L_j depends on σ. If $\sigma > 1$, then goods are substitutes, and as people shift their expenditures into good i, firm i expands the number of workers. The lower price associated with higher productivity expands the sector. The logic runs in reverse with complements. If productivity A_i rises, then because people don't want to raise their consumption of this good by much, fewer workers are necessary to meet their demand.

As above, you can solve for the L_i as a fraction of total labor, L. Use the adding up condition $\sum_i^N L_i = L$, and plug in L_i from the above relationship. Solve the resulting equation for L_j, which will show you that

$$L_j = \frac{\theta_j A_j^{\sigma-1}}{\sum_i^N \theta_i A_i^{\sigma-1}} L. \tag{3.88}$$

What matters here is the productivity relative to this aggregate productivity term in the denominator. The logic regarding substitutes and complements follows the above discussion. By similar logic to the prior section, that fraction is also the share of total capital used by the firm.

Taking this to the expression for final output, plug in to get

$$Y = \left(\sum_i^N \theta_i^{1/\sigma} \left(A_i K^\alpha L^{1-\alpha} \frac{\theta_i A_i^{\sigma-1}}{\sum_i^N \theta_i A_i^{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \tag{3.89}$$

This is a monster, but lots of things will cancel for us. First, pull out the capital and labor, as well as the aggregate productivity summation, and simplify the exponents on the A_i terms. This gives you

$$Y = \frac{K^\alpha L^{1-\alpha}}{\sum_i^N \theta_i A_i^{\sigma-1}} \left(\sum_i^N \theta_i A_i^{\sigma-1} \right)^{\frac{\sigma}{\sigma-1}}. \tag{3.90}$$

Now, notice that you've got the same summation in the denominator and numerator. You can work with the exponents on these overall summations to get

$$Y = K^\alpha L^{1-\alpha} \left(\sum_i^N \theta_i A_i^{\sigma-1} \right)^{\frac{1}{\sigma-1}}. \tag{3.91}$$

The term $\left(\sum_i^N \theta_i A_i^{\sigma-1}\right)^{\frac{1}{\sigma-1}}$ is productivity, and it is a generalized mean of the separate A_i terms. But similar to before, we have a more robust structure for goods and demand, but have not changed the fundamental nature of our aggregate output expression. All the analysis of the Solow model can go through, even with lots of firms producing multiple goods that are not perfect substitutes.

What we do have is more interesting results regarding how factors move between sectors in response to changes in productivity. We had that expression above regarding L_j, and how it relates to L. If productivity goes up in sector j, then this raises aggregate productivity, and this has an impact on the size of L_j. Changes in productivity will alter the structure of the economy, both in terms of where labor (and capital) are employed, and in terms of which goods are produced.

CHAPTER 4

Endogenous Growth

Everything in the last chapter showed the importance of productivity, both in explaining growth in GDP per capita, in explaining variation in GDP per capita across countries. But everything in the last chapter took the level or growth of productivity as a given. If productivity is so important, it seems worth thinking about the economics of what drives it. In this chapter I'll introduce some of the crucial elements of endogenous growth, where productivity improvements are the result of specific activities undertaken by firms and/or individuals.

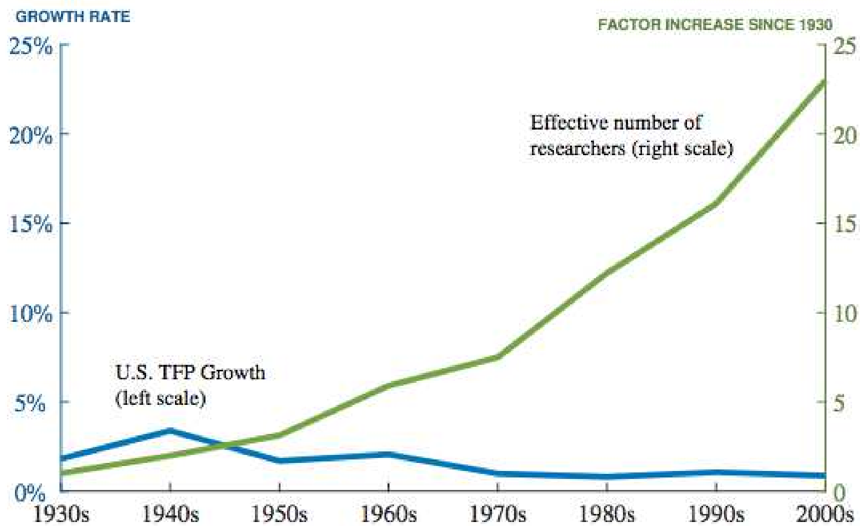
In developing theories of endogenous growth, we are going to be constrained by a few key facts. The first and most important is that the growth rate of GDP per capita, and of productivity, has been constant (or possibly falling slightly) over long stretches of time. To make sense the theories have to deliver constant growth in productivity. The second fact is that the efforts put into research and development, or innovation in general, have been rising steadily over time.

Figure 4.1 shows some simple facts from a recent paper by Nick Bloom, Chad Jones, John van Reenen, and Michael Webb (2017). What you can see here is that the growth rate of TFP (productivity) has been essentially flat over the last eighty years, while the absolute number of researchers (a proxy for the effort put into innovation) has been rising steadily over time.

Whatever your underlying model of research and development process, ultimately it has to accomodate these facts. Increases in the absolute amount of effort going into innovation have not demonstrably changed the growth rate of productivity over time. This will turn out to create some particular constraints on how any model of innovation will work.

Complicating matters is a last fact, which is that the efforts undertaken by firms or individuals to innovate are not done out of altruism. They hope to earn some kind of return for their efforts, otherwise why would firms keep putting all those researchers to work? Trying to combine a realistic model of firm innovation activity, which depends on the profits they can earn and hence the size of the market, with the restriction that overall productivity growth does not increase over time, turns

Figure 1: Aggregate Data on Growth and Research Effort



Note: The idea output measure is TFP growth, by decade (and for 2000-2015 for the latest observation). For the years since 1950, this measure is the BLS Private Business Sector multifactor productivity growth series. For the 1930s and 1940s, we use the measure from Robert Gordon (2016). The idea input measure is gross domestic investment in intellectual property products from the National Income and Product Accounts, deflated by a measure of the nominal wage for high-skilled workers.

Figure 4.1: Growth in TFP and the rise of researchers

out to be hard. There are a series of methods used accommodate these facts, and they have grown in complexity as researchers try to incorporate more realistic firm level models into growth theory.

4.1 The Basics of Endogenous Growth

Productivity is the crucial driver of long-run growth within countries, and a significant factor in income differences between countries. However, to this point we still have taken it as a completely exogenous process. That is, E_t simply goes up over time for some unexplained reason, at the rate g each period. Countries may vary in their g or in the baseline level of E_0, thus creating differences in their productivity levels.

How do we understand the source of E_t growth? We will address the specific incentives to actively pursue higher productivity by firms in a separate chapter, but here we can start to flesh out a more general model of E_t growth, and see that even once we endogenize the decision-making regarding innovation, we will still end up with something that looks like exogenous productivity growth.

Regardless of the underlying economics, any model of endogenous productivity growth has to specify the process through which productivity accumulates. In our baseline, the acummulation function for technology is

$$\Delta E_{t+1} = gE_t, \tag{4.1}$$

meaning that the absolute change in productivity between t and $t+1$ depends on the size of E_t itself. The higher is productivity, the greater the absolute jump in productivity next period. But there is no reason that productivity has to accumulate exactly in this manner. We can specify a more general accumulation process for productivity, as follows

$$\Delta E_{t+1} = \theta R_t^{\lambda} E_t^{\phi} \tag{4.2}$$

where R_t are the number of researchers working on improving productivity, and E_t is the current level of productivity. θ is a scaling parameter. We could think of R_t as the amount of resources spent on research instead (capital, labs, computers, etc..) but that won't change the point of what we're doing here. We're going to say that

$$R_t = s_R N_t \tag{4.3}$$

where s_R is the fraction of the labor force that is engaged in research. You could alternatively think of s_R as the amount of time that each person spends trying to innovate. Regardless, s_R is like the savings rate in the Solow model. The size of s_R will be chosen by profit-maximizing firms when we go into the economics of innovation, but for now just note that it is an input to accumulating new productivity.

Equation (4.2) is thus a general form for the accumulation of productivity, with our baseline model just being the special case where $\lambda = 1$ and $\phi = 1$. As it turns out, the size of those two parameters is going to be very important in determining how productivity growth works. The value of ϕ indicates how accumulation of productivity responds to the current productivity level. If $\phi > 0$, then we have that productivity accumulates faster as productivity rises. By discovering some ideas, we are more able to discover new ideas. If $\phi < 0$, then there is a "fishing out" effect, in which we acquire new ideas and that actually makes it harder to find the next new idea. If $\phi = 0$ then these exactly offset. There does not appear to be any particular reason to suspect that ϕ take on any specific value. One can come up with reasonable arguments for positive or negative values.

The value of λ captures the effect of adding new researchers to the accumulation process. Here, we generally assume that $0 < \lambda < 1$. That is, more researchers do make productivity more likely to grow, but there is some decreasing marginal return to having more researchers. They are likely to duplicate efforts unless there is perfect coordination of research activities, so that doubling the researchers does not necessarily double the accumulation of productivity.

So what does this more nuances formula imply about the growth rate of productivity? First, divide (4.2) through by E_t to get it in terms of growth rates.

$$\frac{\Delta E_{t+1}}{E_t} = \theta R_t^{\lambda} E_t^{\phi-1}. \tag{4.4}$$

Here, we have something similar to an accumulation equation for capital. The growth rate of E_t depends on E_t itself. For now, assume that $\phi < 1$ - it can be positive, but not equal to one (or greater). Then this implies that the growth rate of E_t falls as E_t rises. If we held R_t constant, then eventually the growth rate of productivity would get driven down to zero. Note that this doesn't mean that there is no innovation, just that the size of the jumps in E_t becomes vanishingly small compared to the level of E_t.

We don't see this in the actual data - recall the residual series for the U.S., capturing the upward trend in E_t over time. So why doesn't productivity growth slow down? Because R_t is in fact growing over time as well, counteracting the effect of $\phi < 1$. R_t, recall, depends on the population size. So holding s_R constant, R_t will be growing at a constant rate n over time.

An implication of (4.4) under the assumption that $\phi < 1$ is that the growth rate of E_t will become constant. To see why, consider the very particular case where $\phi = 0$ and $\lambda = 1$. Then the growth rate of E_t is simply

$$\frac{\Delta E_{t+1}}{E_t} = \theta \frac{R_t}{E_t}. \tag{4.5}$$

In figure 4.2 there is plotted the growth rate of E_t from this equation against the ratio R_t/E_t. I've also put in a horizontal line denoting the growth rate of population, n, which is also the growth rate of R_t, as $R_t = s_R N_t$. Because the growth rate of E_t is upward sloping in R_t/E_t, there is a stable steady state growth rate of E_t. Regardless of where the economy begins, the ratio of R_t/E_t will tend towards the value n/θ. At this point, the growth rate of E_t is exactly equal to n, the population growth rate. What is going on here? Even though every time E_t goes up, the growth rate of E_t falls, it is also the case that R_t is increasing over time, which counteracts the falling growth rate of productivity. Ultimately, the growth rate of productivity is tied to the growth rate of researchers. As they increase, the absolute increase in E_t increases, allowing the economy to continue to grow. Note that if population growth were to fall to zero then economic growth would fall to zero as well.

This holds more generally for accumulation of productivity described by (4.4). There will be a steady state growth rate of productivity that depends only on the population growth rate. To see this, note that we have the same kinds of forces at work, with R_t raising the growth rate while E_t is pulling it down. These forces will come into a balance. The balance point is where the growth rate of E_t is no longer changing. In other words, we need to think about the derivative of the growth

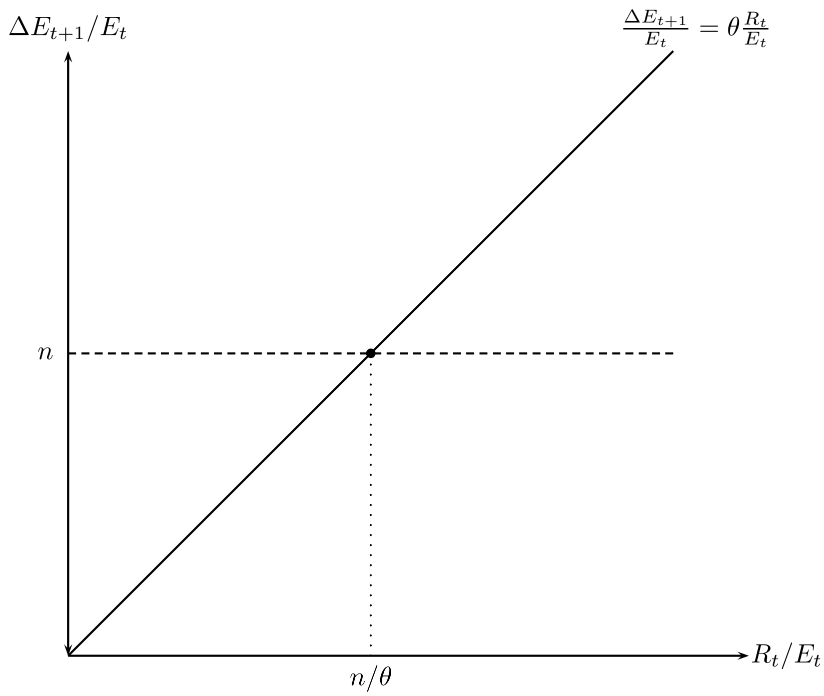


Figure 4.2: The Steady State Growth of Productivity

Note: The figure shows the determination of the steady state growth rate of E_t under the assumption that $\phi = 0$ and $\lambda = 1$. If R_t/E_t is below n/θ, then R_t is growing faster than E_t, and so R_t/E_t is rising. The opposite holds when R_t/E_t is above n/θ. Hence the value n/θ is a stable steady state value for the ratio R_t/E_t, and the growth rate of E_t is n in steady state.

rate of E_t with respect to time

$$\frac{\partial \Delta E_{t+1}/E_t}{\partial t} = \lambda \frac{\Delta R_{t+1}}{R_t} + (\phi - 1)\frac{\Delta E_{t+1}}{E_t} = 0. \tag{4.6}$$

The derivative of the growth rate w.r.t. time of E_t has to be equal to zero in the steady state - the growth rate is not changing. That means that $\lambda \Delta R_{t+1}/R_t$ must exactly offset $(\phi - 1)\Delta E_{t+1}/E_t$. If you rearrange the second equality you get that

$$\frac{\Delta E_{t+1}}{E_t} = \frac{\lambda}{1-\phi}\frac{\Delta R_{t+1}}{R_t} = \frac{n\lambda}{1-\phi}. \tag{4.7}$$

In steady state, the growth rate of productivity must be proportional to n. The only difference from above is that the proportion depends on the exact values of λ and $1 - \phi$. It is only at this growth rate of E_t that the growth rate of E_t will not change. Or, it is this growth rate that the growth rate of E_t will eventually reach no matter where it starts.

Does the idea that growth depends on population growth make sense? We saw in the Solow model that higher n led to lower living standards because it diluted capital. But note that the

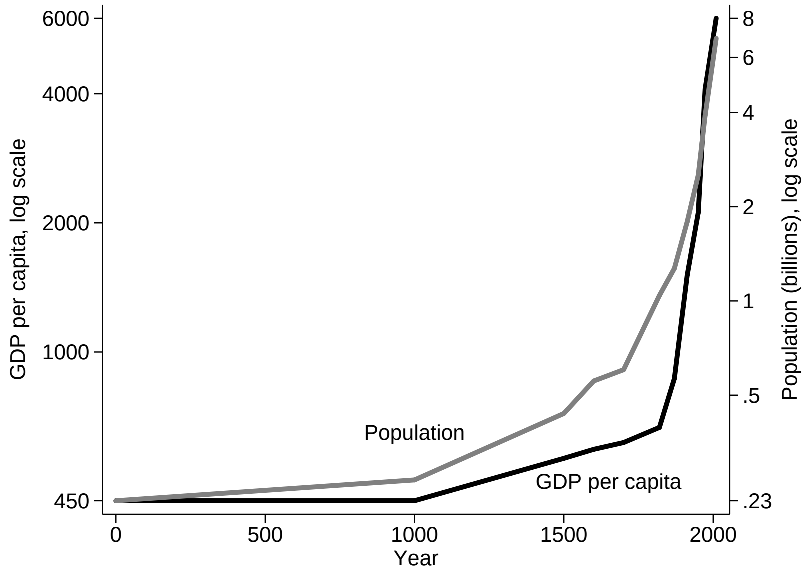


Figure 4.3: Output per capita and Population Levels Worldwide

effect of n in the Solow model was a level effect. The steady state level of k would be lower if n went up. However, we have here a counteracting effect of higher n, that it endgenders higher productivity growth through increasing the number of people producing new ideas. This may seem a prediction at odds with the evidence. On a country by country basis, this may not appear to work very well (why doesn't Belgium grow more slowly than Nigeria?), but from a global perspective this seems to be a good approximation. Figure 4.3 shows the levels of average GDP per capita and and population for the world over a long stretch of history. If we go back in time far enough, we see that both GDP per capita and population levels were quite low. It is only when population growth starts to explode that we have the rise of sustained growth in GDP per capita, around the time of the Industrial Revolution. The growth rate of income per capita is highly positively correlated over time for the whole world, even though an individual country may lower its living standards with rapid population growth. The reason is that higher population growth overall is able to produce greater productivity, through a greater stock of researchers coming up with productivity-enhancing innovations. Locally, though, increased population lowers capital per worker (or land per worker), hurting living standards. On net, the increased innovation has been the stronger force over history.

One thing missing from our steady state growth rate is s_R, the proportion of the population that is engaged in research. It would appear that this should matter, shouldn't it? If more researchers are important for innovation and long-run growth, then won't we grow faster if s_R is higher - that is, if more of the population engages in innovation? The issue here is that there are two offsetting effects. If s_R goes up, then indeed E_t will grow faster for a while, but eventually the increaed size of E_t will drag down the growth rate of innovation again. A one-time increase in s_R generates a level shift up in the number of researchers, but the growth rate of researchers remains n in the long-run. So s_R has no long-run impact on the growth rate, only on the level of productivity. You'll see this in a homework problem.

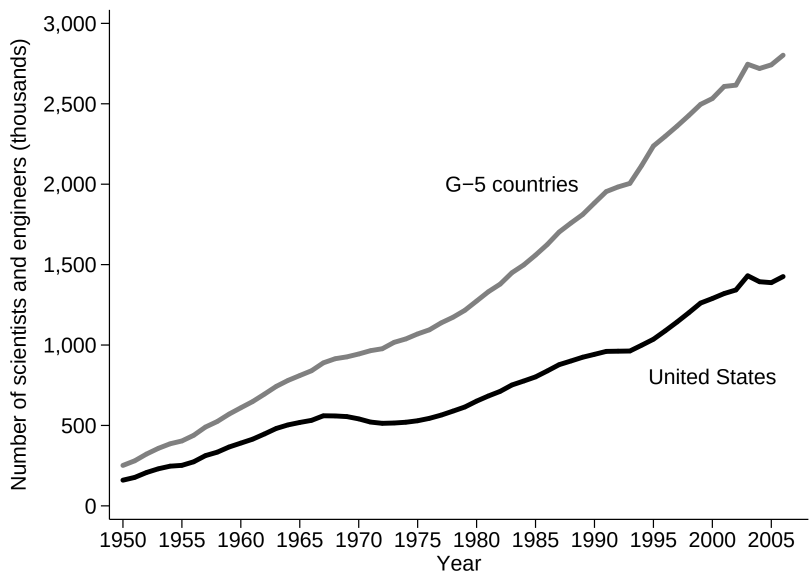


Figure 4.4: Absolute number of Scientists and Engineers, G-5 and U.S., 1950–2010

This is a consequence of assuming that $\phi < 1$. If we let $\phi = 1$, which was the original assumption in Romer's (1990) paper outlining endogenous growth, as well as $\lambda = 1$,then the the growth rate becomes

$$\frac{\Delta E_{t+1}}{E_t} = \theta R_t = \theta s_R N_t. \tag{4.8}$$

That is, the growth rate of E_t no longer is falling with E_t. This tells us that s_R has a permanent effect on the growth rate of E_t. However, it also tells us that the growth rate of productivity should be accelerating over time with the absolute size of N_t. We do not see either effect in the data. Growth rates of productivity have leveled off for all the richest countries in the world over the 20th century, despite populations growing regularly over that period. Additionally, there has been a distinct increase in the absolute number of researchers engaged in increasing productivity, and yet there has not been a change in the growth rate of E_t over the last 100 years. Extending the data on researchers in the OECD from Jones (1995), figure 4.4 shows that there has been a distinct increase in their numbers over time. However, the growth rate of productivity has not appreciably changed. The absolute number of researchers is not related to the growth rate of productivity. Rather, the growth rate of the number of researchers (related to n) is the key to the long-run growth rate.

Now, as mentioned at the outset of this section, this does not truly represent a model of endogenous productivity change. That is, we haven't specified why anyone would bother to become a researcher, or why firms undertake research and development at all. However, regardless of the exact models we do specify, we have to specify the accumulation technology for productivity, and once we've done that the implications for the long-run growth rate are fixed by the parameters we select for λ and ϕ. As there is no evidence consistent with $\phi = 1$, and rising productivity growth rates, it seems clear that we should stick with $\phi < 1$. In this case, we have what Jones (1995) calls "semi-endogenous" growth, meaning that productivity growth is tied to the population growth rate, and not to any parameters such as s_R that influence the intensity of research effort.

While s_R doesn't influence the growth rate, note that it influences the level of output per capita. To see this, let's go back to our basic Solow model, but account for the fact that if $s_R N_t$ people are doing research, they are not working. Output is therefore

$$Y_t = K_t^{\alpha}(E_t(1-s_R)N_t)^{1-\alpha} \tag{4.9}$$

in the Cobb-Douglas case.

In steady state, we know that $g = \lambda n/(1-\phi)$, and from prior work we can solve for

$$\tilde{y}^* = (1-s_R)\left(\frac{s_K}{n+g+\delta}\right)^{\alpha/(1-\alpha)} \tag{4.10}$$

where I used s_K to represent the savings rate for physical capital. The term $(1-s_R)$ enters now linearly, influencing the level of the steady state value of $\tilde{y}$. Knowing the definition of $\tilde{y}$, we can write this as

$$y_t = E_t(1-s_R)\left(\frac{s_K}{n+g+\delta}\right)^{\alpha/(1-\alpha)} \tag{4.11}$$

in steady state. This shows that output per capita grows with E_t, at the rate g. The rest of the terms on the right are all constant.

In steady state, we know that

$$g = \theta R_t^{\lambda} E_t^{\phi-1} \tag{4.12}$$

which means that along the steady state growth path,

$$E_t = \left(\frac{\theta(s_R N_t)^{\lambda}}{g}\right)^{1/(1-\phi)}. \tag{4.13}$$

Productivity is growing in proportion to the size of the population N_t, which is consistent with our finding on g.

Put this into our expression for output per capita in steady state and we have

$$y_t = \left(\frac{\theta(s_R N_t)^\lambda}{g}\right)^{1/(1-\phi)} (1 - s_R) \left(\frac{s_K}{n + g + \delta}\right)^{\alpha/(1-\alpha)}. \tag{4.14}$$

Ugly, yes. But note that s_R has two conflicting effects on output per capita. One, higher s_R means greater research effort, so the level of E_t will be higher, albeit growing at the same rate g regardless of what s_R may be. Two, a higher s_R will lower output because it takes people out of the final goods sector.

What this means is that higher s_R does not necessarily ensure the highest standard of living. There is some optimal amount of research effort to do that maximizes output per capita in steady state. In this sense, policies that influence s_R do have an important place in determining welfare. It's possible for a country to have too much or too little research effort.

4.2 A Model with Profits

The prior section laid out a model that could accomodate the first two facts mentioned in this chapter: growing research effort but constant growth. What that model did not establish was why anyone would bother to innovate at all. Why is s_R greater than zero? To answer this, we are going to need to allow for someone to earn some profits from an innovation.

There are some deeper concepts at work here. Innovations and/or ideas are often *non-rival* in nature, meaning that they can be copied or used by others without diminishing anyone's ability to use them. Think of these notes; your absorbtion of these ideas does not mean that I know less about growth. This separates many innovations from things like desks or pencils, which are *rival*. If you borrow my pencil, then I cannot use it.

It is the non-rival nature of productivity in general that allows for growth. But we still have that problem that no one would innovate to raise productivity without some kind of reward for doing so. We need the concept of *excludability*. Excludability is different than rivalry, and it is best to think of it in terms of institutions or property rights, rather than as an inherent feature of goods or ideas. My house is excludable, in the sense that I can legally deny other people the use of it, and when it gets sold, I have sole claim to the money that someone pays for it. Innovations may be non-rival, but they can be made excludable, even if by nature it may be easy to copy or use an innovation without paying the inventor. Things like patents or copyrights are examples of institutions that make an innovation excludable, even if it is non-rival.

So when we say that firms or individuals need some reward in order to innovate, we need to specify how innovations are *excludable*. That is, we need to specify a market structure or institution such that someone who owns or develops an idea can reap a reward from it. In this section, we're going to set up a model of monopolistic competition in which intermediate good firms are monopoly providers of their output to the final goods sector. The profits they earn will, in the next section, be the basis of the rewards for innovating, and create the incentive to undertake that activity. So for now, we need to set up the structure of those profits.

We will set things up in a static setting, taking the overall supply of capital and labor as fixed, for the time being. Final output (meaning GDP) is produced according to the a CES production function modified slightly

$$Y = N^{1-\alpha}\left(\sum_i^M Y_i^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\alpha\sigma}{\sigma-1}} \tag{4.15}$$

which looks like a bit of a mess. The CES combination of intermediate goods is put together with labor, N, to produce final output. The parameters α and $1-\alpha$ perform their typical functions here - this is just a Cobb-Douglas of labor and intermediate goods. Only here the intermediate goods are a CES function rather than simply a stock of capital.

This is a general form, but to develop the basic ideas of the monopolistic competition model we're going to make one particularly useful assumption. Namely that $\sigma = 1/(1-\alpha)$, or the elasticity of substitution is strictly related to the size of α. This is solely to make some things cancel nicely, and saves us from a lot of needless algebra. This assumption leads to the following production function

$$Y = N^{1-\alpha}\sum_i^M Y_i^{\alpha} \tag{4.16}$$

which has a very Cobb-Douglas-like look to it. Rather than simply capital raised to the α, it is the sum of the various intermediate goods each raised to the α that determines output.

For final goods, we assume that there is perfect competition among firms. That is, there is free entry into final goods production, each firm operating with the above function. However, we do need to know a little something about these firms and their demand for intermediate goods. Let profits for the final-goods sector be given by

$$\pi = Y - \sum_i^M P_i Y_i - wN, \tag{4.17}$$

where P_i is the price of an individual intermediate good, w is the wage rate, and the price of the final good is set to one as the numeraire. Maximizing profits requires that for each intermediate good used, the revenue marginal product equal the marginal cost, or

$$N^{1-\alpha}\alpha Y_i^{\alpha-1} = P_i. \tag{4.18}$$

This describes the inverse demand curve for good i from final goods firms. This will be used by the intermediate goods producers to determine how much to produce. Note that there is also a first order condition for labor setting the marginal product of labor equal to the wage, but for now we don't need to write that down explicitly.

Where do these intermediate goods comes from? This is where the imperfect competition comes in, and differs from the setup we used in the prior chapter. Each intermediate good is produced by a single firm, who owns the patent or holds the monopoly over production of that intermediate good. Each firm, though unique in it's output, has an identical production function

$$Y_i = A_i K_i \tag{4.19}$$

which is a simple linear production function in capital. Essentially, you can think of the intermediate goods as being capital services. A_i indicates the productivity of that particular type of capital in producing capital services.

Each of these firms is attempting to maximize profits, and because they hold monopolies on their output, we need to be explicit about their profit-maximization problem. Let profits be written as

$$\pi_i = P_i Y_i - R K_i, \tag{4.20}$$

which is the same as we've assumed all along for firms. As monopolists, the firms do not take P_i as given. Rather, they know the demand curve for their good, and as such will set their price to maximize profits. What is that demand curve? Demand for intermediate goods comes from the final goods producing firms, and we solved above for the inverse demand function for firm i's output. Putting that in yields

$$\pi_i = N^{1-\alpha} \alpha Y_i^{\alpha} - R K_i. \tag{4.21}$$

The firm then maximizes over K_i, which is incorporating the effect of producing more output on the price that they can charge. We assume that there are enough intermediate good producers (M is large) that no firm appreciates it's effect on Y. The first-order condition is

$$\alpha^2 N^{1-\alpha} A_i^{\alpha} K_i^{\alpha-1} = R. \tag{4.22}$$

which can be solved for

$$K_i = \left(\frac{\alpha^2 N^{1-\alpha} A_i^{\alpha}}{R} \right)^{1/(1-\alpha}. \tag{4.23}$$

and therefore

$$Y_i = \left(\frac{\alpha^2 A_i N^{1-\alpha}}{R} \right)^{1/(1-\alpha}. \tag{4.24}$$

Output of firm i rises with its productivity A_i. This allows the firm to produce capital services more cheaply, and so it can charge a lower price, which raises the demand form final good firms. If R rises relative to P, this raises marginal costs to the intermediate firm, and so output is lower, as it must charge a higher price to keep profits up. Note that the scale of the economy matters here. If N goes up, the final goods sector is larger, and this raises demand for the output of firm i.

From this determination of the output of firm i we can determine several outcomes of interest. First, consider the price charged by firm i for it's output. Putting (4.24) into (4.18) gives us the price charged by the firm

$$P_i = \frac{1}{\alpha}\frac{R}{A_i}. \tag{4.25}$$

The price is equal to the marginal cost of a unit of capital services (R/A_i) times a markup $1/\alpha$. You'll recall that we discussed markups previously, though without specifying where the markup comes from. Here, we have an explicit determination of the markup. The closer α is to zero, then higher the markup. When α is close to zero, the elasticity of substitution between intermediate goods is close to one, and the firm can essentially hold up the final goods firms. If α goes to one, then the elasticity of substitution goes to infinity and any individual firm has zero market power. Final goods firms can simply substitute to a different intermediate good, thus their price will be equal to the marginal cost. In other words, we're basically back at perfect competition.

What does this imply about aggregate output? Before we answer that, it is necessary to lay out the assumptions that we're going to impose to get the expression. We're assuming that the marginal cost of capital R is perfectly flexible. In other words, we have a perfect, frictionless financial market. This implies that no capital that is available in the economy is wasted. We take the aggregate stock of capital as given, K, and this amount is provided by savers (who we could model if we wished). The sum of the individual firm capital demands from (4.23) must add up to this total

$$K = \sum_i^M K_i = \sum_i^M \left(\frac{\alpha^2 N^{1-\alpha} A_i^{\alpha}}{R}\right)^{1/(1-\alpha)}. \tag{4.26}$$

This can be rearranged and solved for R as a function of the supplies of factors of production, N and K, and productivity levels, A_i.

$$R = \alpha^2 \left(\frac{N}{K}\right)^{1-\alpha} \left(\sum_i^M A_i^{\alpha/(1-\alpha)}\right)^{1-\alpha}. \tag{4.27}$$

This looks similar in form to a typical return on capital in a standard Solow model. That is, we have that the return depends inversely on the capital/labor ratio, K/N.

To solve for final output, it is necessary to put the solution for R back into the expression in (4.24) to get firm-level outputs, which are now

$$Y_i = A_i K \frac{A_i^{\alpha/(1-\alpha)}}{\sum_i^M A_i^{\alpha/(1-\alpha)}}. \tag{4.28}$$

I've written it this way to show that output of firm i consists of two parts. First is $A_i K$, the total amount that firm i could produce if it were the only intermediate good provider in the economy. The second is the ratio of $A_i^{\alpha/(1-\alpha)}$ to the sum of those terms across all firms. This dictates the share of capital that firm i actually is given access to through the market. The higher is A_i, the more of the capital stock that firm i will utilize, given it's higher productivity relative to the rest of the economy.

Finally, given Y_i, we can go back to the original aggregate production function and solve for

$$Y = N^{1-\alpha} \sum_i^M Y_i^\alpha = N^{1-\alpha} K^\alpha \left(\sum_i^M A_i^{\alpha/(1-\alpha)} \right)^{1-\alpha}. \tag{4.29}$$

This is somewhat cumbersome, but note that it is of exactly the same form as our standard Cobb-Douglas. Now, we happen to have some additional information about what constitutes the efficiency term. Here $E = \sum_i^M A_i^{\alpha/(1-\alpha)}$, a weighted sum of the individual productivities of firms. This makes some sense, aggregate productivity must depend on the firm-level productivity in the economy. This will be useful to us in the following sections where we talk about the optimal accumulation of new varieties (M) or increasing productivity at individual firms (A_i). But the overall economy works exactly like our standard Solow model from before.

Before going on, it will be useful to work through the distribution of income in this economy between wages, returns to capital, and profits. This gets back to the point raised in the prior section, which is that in order for there to be profits available to innovators, we are going to need to "underpay" factors of production. From the final goods sector profit maximization, we have a first order condition that wages are equal to the value of the marginal product of labor, or

$$w = (1-\alpha) N^{-\alpha} \sum_i^M Y_i^\alpha = (1-\alpha) \frac{Y}{N}. \tag{4.30}$$

This is similar to what we had for a standard Cobb-Douglas. Labor earns a fraction of final output. So the wage determination is standard.

This is helpful, as it means that α of final output remains to be paid out as returns to capital and profits. With perfect competition, we know that all of that will got to returns to captial, and none to profits. However, here we have monopolistically competitive firms who do earn profits.

Total payments to capital are RK, and if we take R from (4.27) we can write this as

$$RK = \alpha^2 N^{1-\alpha} K^\alpha \left(\sum_i^M A_i^{\alpha/(1-\alpha)} \right)^{1-\alpha} = \alpha^2 Y, \tag{4.31}$$

where the second equality follows from the formulation of output in (4.29). So capital earns not α of output, but only α^2. The payments to capital are lower than in perfect competition because firms are extracting some profits.

Finally, total profits of all the intermediate goods firms must add up to

$$\sum_i^M \pi_i = Y - wN - RK = \alpha(1-\alpha)Y, \tag{4.32}$$

or the remaining fraction of output is paid out as profits. For any individual firm, profits depend on their exact level of productivity A_i,

$$\pi_i = \alpha(1-\alpha)Y\frac{A_i^{\alpha/(1-\alpha)}}{\sum_i^M A_i^{\alpha/(1-\alpha)}}. \tag{4.33}$$

Similar to the expression for firm-level output, the ratio of productivity to the sum of productivities is important. Of the total profits in the economy, firm i takes a share based on their inherent produtivity.

This model of monopolistic competition delivers us a way of expressing profits for firms, and thus will allow us to express the value of a patent/idea which gives a firm the monopoly rights over a given product i. With that, we can actually talk about the motives for doing innovation, and what determines the level of effort put into research.

4.3 The Romer Growth Model

To go forward, we have to be explicit about exactly how someone can innovate. In the classic model of Romer (1990), innovation occurs through the introduction of new products (as opposed to making existing products more productive).

We use our baseline model of monopolistic competition to talk about innovation in the Romer model, as it features separate firms that each earn profits. To do this, we'll simplify the baseline model a little more. First, we'll assume that any firm that operates has $A_i = 1$. This means we can write aggregate output as

$$Y = K^\alpha (MN_Y)^{1-\alpha} \tag{4.34}$$

where N_Y is the number of people who work in the final goods sector. The remaining $N_R = N - N_Y$ individuals will work at innovating.The efficiency level of the economy is simply M, the number of varieties of intermediate goods. Any increase in that number will raise efficiency and hence output per capita.

The profits of any single intermediate good firm will be

$$\pi = \alpha(1-\alpha)\frac{Y}{M}. \tag{4.35}$$

Total profits are $\alpha(1-\alpha)Y$, and those profits are split up evenly among the M intermediate good firms, who are all identical now because $A_i = 1$. It's not indexed by i because this is the same for all firms.

So what do we mean by innovation? In short, if someone innovates, that means they have discovered a new variety of intermediate good. M potentially goes up by one, and that raises productivity in the economy. If they innovate, we'll assume that they receive some kind of patent that gives the holder exclusive rights to produce that particular intermediate good. There needs to be some kind of intellectual property rights to the innovation, otherwise it will be valueless, and no one will bother to innovate.

By figuring out the value of the patent for a new intermediate good, we'll be able to figure out how much effort people put into innovating. Let the value of a patent be V. Using a typical arbitrage argument, the value of the patent must be

$$rV_t = \pi_t + \Delta V_{t+1}. \tag{4.36}$$

The left-hand side tells us what one could earn by simply investing V_t in capital. The right-hand side tells me what I could earn by buying a patent for V_t dollars. I earn the flow of profits, plus whatever the change in value of the patent is over time. These must be equal. If they are not, there is an arbitrage opportunity.

Re-arranging this we have

$$r = \frac{\pi_t}{V_t} + \frac{\Delta V_{t+1}}{V_t}. \tag{4.37}$$

In steady state, it will have to be that r is constant. That is, the return on capital will be constant, which we know is true in a Solow model with efficiency growth. We know that in steady state, π_t will be growing at the rate n. Why? Because it is proportional to Y/M. Y grows at the rate $n+g$ in steady state, and M will be growing at g in steady state. So Y/M must grow at the rate n. The only way for the right-hand side to stay constant is for V_t to grow at the rate n as well, so π_t/V_t doesn't change. This means that $\Delta V_{t+1}/V_t = n$ in steady state. This implies that

$$V = \frac{\pi}{r-n} \tag{4.38}$$

in steady state. The value of a patent is equal to the flow of profits discounted at the rate $r-n$. This is simply the PDV of a flow of profits discounted at the rate $r-n$.

We know how valuable it is to innovate, and that needs to inform us about how much effort people put into innovation. For that we need some information on how people decide whether to work in the final goods sector or to innovate, as well as some information on how their labor effort at innovation translates into actual new intermediate goods.

If they work in the final goods sector, then they can earn

$$w_Y = (1-\alpha)\frac{Y}{N_Y}, \tag{4.39}$$

which is just the typical marginal product of labor. If they try to innovate, they will earn a "wage" by trying to innovate and possibly making a discover that they can patent. So their wage as a researcher is determined by the value of a patent, but depends upon how fast they can find new innovations to patent. The speed at which new ideas arrive for an individual researcher is assumed to be

$$\Delta M^{ind} = \theta N_R^{\lambda-1} M^{\phi}. \tag{4.40}$$

This accumulation equation for new varieties is of a similar form as our semi-endogenous growth model. This tells us the arrival of ideas for an individual researcher depends on N_R, the total number of researchers at work to begin with (remember, possible duplication of ideas) and M itself (the dilution of ideas). The reason that N_R is raised to $\lambda - 1$ is that this is the arrival rate for *one* researcher. The aggregate arrival of new intermediate good ideas is

$$\Delta M = \theta N_R^{\lambda-1} M^{\phi} \times N_R = \theta N_R^{\lambda} M^{\phi} \tag{4.41}$$

which is precisely the form of efficiency growth in our semi-endogenous model. No matter what, the growth rate of M is this model will end up being $\lambda n/(1-\phi)$. It is irrelevant what the exact incentives of the researchers are. Once we've posited this as the arrival rate of new ideas, the growth rate of M is pinned down in steady state.

Back to our individual researchers and workers. People will move back and forth from research to final-sector work until the expected payoffs are identical. That is, until

$$\Delta M^{ind} V = w_Y. \tag{4.42}$$

The left-hand side is the arrival rate of new innovations to a single researcher, multiplied by the value of one of those innovations, V, the steady state value of a patent. The left side is simply the outside option.

To continue, we're going to manipulate this expression to find the one thing we still don't know, how many people do research (N_R) versus how many work (N_Y). To find this, first multiple and divide the left-hand side by N_R, which gives us

$$\Delta M \frac{V}{N_R} = (1-\alpha)\frac{Y}{N_Y}. \tag{4.43}$$

We know that $V = [\alpha(1-\alpha)Y/M]/[r-n]$ from our analysis of the patent value and knowledge of the value of profits from an intermediate good producer. Putting this into the expression and canceling yields

$$\frac{\Delta M}{M}\frac{\alpha}{(r-n)N_R} = \frac{1}{N_Y} \tag{4.44}$$

which can be rearranged to

$$\frac{N_R}{N_Y} = \frac{\alpha}{r-n}\frac{\Delta M}{M} = \frac{\alpha}{r-n}\frac{\lambda n}{1-\phi}. \tag{4.45}$$

This tells us how the proportions of researchers to final-goods sector workers will be determined. If you like, this tells us the value of s_R in the semi-endogenous growth model.

The faster that innovations accumulate (higher λ or higher ϕ), the higher the proportion of people that will do research. This makes some sense. The more quickly that patentable ideas arrive, the higher the value of doing research. Those profits are discounted at the rate $r-n$, so the higher this discount rate, the less research is done. Note that n is positive for research. Faster population growth means faster growth in profits. Why? In the imperfect competition economy, profits depend on the scale of the economy - I can make more money by selling my intermediate good to a larger final goods sector. So anything that makes the aggregate economy grow faster will raise my incentive to do research and innovate.

If you think back to our generic semi-endogenous growth model, recall that s_R was the proportion of workers that did research. So $s_R = N_R/(N_R + N_Y)$, and equation (4.45) defines s_R implicitly for us. Also recall that there is some optimal level of s_R in these growth models. There is nothing in the solution in (4.45) that ensures we are at that optimal level. If n is very high, or r very low, then it is quite possible that a country could have too much research effort as innovators chase profits, but leave few people left to work on final goods. The opposite situation occurs if r is too high or n is too low, there will be a ratio s_R less than the optimal level. Just because we have profit-maximizing firms and innovators does not guarantee we get the best possible outcome.

CHAPTER 5

Savings and the Supply of Capital

One of the key presumptions we've made to this point is that the supply of capital and labor is fixed and inelastic with respect to the rate of return or wage, respectively. This does not appear to be a good description of reality, as both investment spending (the supply of new capital) and employment (the supply of workers) fluctuate with the business cycle. In addition, investment spending and employment vary across countries, and there are long-run trends in these supplies within countries that we would like to understand at some more fundamental level.

To discuss what drives these supplies, we need to specify who is supplying them, and what their motives are. In other words, we're going to need to explicitly write down a model of individuals that provide savings and labor to firms. These individuals will have some kind of utility function that they are trying to maximize, and they'll take the rental rate for capital and the wage rate as given when they make their decisions. Their optimization will result in elastic supplies of capital and labor, providing one means through which output may fluctuate or grow over time.

We'll begin by talking about the savings decision alone, and then incorporate an explicit labor supply decision once we have the basic framework in place. Once we've described how individuals act to supply these factors, we can fold these decisions into the Solow model and see how this changes the predictions of that model.

5.1 The Fisher Model

To begin describing savings, we need to understand that it is fundametally a problem in maximizing utility subject to a budget constraint. Rather than choosing how to allocate their income between beer and pizza, as in a typical intermediate micro problem, people here are going to decide how to allocate their income between buying goods (of whatever kind) *today* versus buying them *in the future*. The "goods" that individuals will care about here are distinct in time, rather than in physical characteristics. People will save so that they can enjoy more consumption later in life. Their budget will depend on the income they earn, but will now also depend on *when* they earn it. The goods they consume will be identical in all periods of time, but the relative price of consumption goods today compared to consumption goods tomorrow will vary with the interest rate.

The easiest setting in which to describe the savings decision is the Fisher model, which posits that people live for only two periods. They have a utility function that looks like this

$$V = U(c_1) + \beta U(c_2) \tag{5.1}$$

where V is total lifetime utility, and $U(\cdot)$ is sometimes called the "felicity" function, or the utility that one gets from consuming in any particular period. Note that we assume that $U(\cdot)$ is identical in both period 1 and period 2. The term β is the time preference parameter. It scale the value of consumption in period 2 by some amount compared to period 1. If $\beta < 1$ then people are impatient, and would prefer to consume today rather than in period 2. If $\beta > 1$ then people are willing to wait.[1]

The utility function described in (5.1) has a very important property, *additive separability*. This means that the marginal rate of substitution between any two periods (say t and $t+1$) is independent of any other periods. Another way of saying this is that the marginal utility of consumption in period t is independent of consumption in any other period. Therefore, what you consume yesterday has no influence on your utility today. What you consume today has no influence on your utility tomorrow. More nuanced models will discard this property (i.e. with durable goods that give utility in multiple periods) but for now we retain it for the simplicity it brings.

In addition to additive separability, we make assumptions about the form of the period-specific utility function $U(\cdot)$. Specifically, we assume that:

$$U'(c_t) > 0 \text{ and } U''(c_t) < 0 \tag{5.2}$$

which says that utility goes up as c_t goes up. However, the marginal utility of consumption is falling as c_t goes up. The per-period utility function is concave in consumption.

Constant Relative Risk Aversion

A specific type of utility function that we will use frequently is the constant relative risk aversion (CRRA) one. This has the form

$$U(c) = \frac{c^{1-\sigma}}{1-\sigma} \text{ where } \sigma > 0 \tag{5.3}$$

where σ is the coefficient of relative risk aversion. If $\sigma = 0$ then utility is just linear in consumption. If $\sigma > 1$, then notice that utility actually has a negative value; this is immaterial, as

[1] It's also common to see β written in terms of a *discount rate*, θ, where $\beta = 1/(1+\theta)$. At times it is convenient to flip back and forth in terms of solving problems.

we'll see the function still has the proper derivatives. Finally, if $\sigma = 1$ then the utility function reduces to $U(c) = \ln c$.

Taking the derivatives yields

$$U'(c) = c^{-\sigma} > 0 \tag{5.4}$$
$$U''(c) = -\sigma c^{-\sigma-1} < 0 \tag{5.5}$$

which matches the assumptions we made regarding utility functions.

The name of the function comes from the definition of risk aversion. Risk aversion measures the elasticity of marginal utility with respect to consumption. In this case

$$-\frac{\partial U'(c)}{\partial c}\frac{c}{U'(c)} = -c\frac{U''(c)}{U'(c)} = \sigma \tag{5.6}$$

which is an indication of the curvature of the utility function. The larger is σ, the more the utility function "curves down" as c increases. In other words, it tells us how severely average utility goes down relative to the utility of an average. We'll see exactly how this fits in with risk later on in this chapter.

How does the utility function reflect consumption smoothing? In other words, what is the elasticity of consumption with respect to a change in the marginal utility of consumption? How much will a person move consumption across time when the marginal utility of this consumption changes. We call this the elasticity of inter-temporal substitution (EIS). It is calculated as

$$-\frac{\partial c}{\partial U'(c)}\frac{U'(c)}{c} = -\frac{U'(c)}{cU''(c)} = \frac{1}{\sigma} \tag{5.7}$$

and we can see that it is simply the inverse of the coefficient of relative risk aversion. So the parameter σ measures the degree to which we are willing to substitute consumption across time, as well as our risk aversion. If σ is large, the EIS is small, and we will not respond much to changes in marginal utility.

An implication of the assumption of concavity is that individuals would like *smooth consumption*. That is, individuals would rather consume similar amounts in periods 1 and 2 than having a big jump (or drop) in consumption. More concretely, an individual will prefer to consume $c_1 = 15$ and $c_2 = 15$ to consuming $c_1 = 5$ and $c_2 = 25$, even though the total consumption is identical. A second implication of this concavity is *risk aversion*. Individuals will actually give up some consumption to avoid having to face a lottery. Specifically, an individual might prefer to consume $c_1 = 15$ rather than facing a coin flip between $c_1 = 10$ and $c_1 = 30$, even though the coin flip has a higher expected value of consumption.

To go along with the utility function, we need some kind of budget constraint. We'll assume that the individual earns w_1 and w_2 in wages in the two periods of life, respectively. In the first period of life they consume an amount c_1, which means they have savings of

$$s_1 = w_1 - c_1. \tag{5.8}$$

In their second period of life, their consumption depends not only on their wage, but on the savings they set aside in period one. We assume that they earn a return of r on their savings between periods 1 and 2, so that

$$c_2 = w_2 + (1+r)s_1. \tag{5.9}$$

We can combine those two conditions easily, employing one additional assumption. That is that the *individual can borrow and lend freely at the rate r*. If they'd like to borrow in the first period (so that $s_1 < 0$), they can. If they do borrow, then in the second period the $(1+r)s_1$ term is capturing how much they have to pay back to the lender. We'll address what changes if they cannot borrow or lend freely later on.

Putting those two equations together and re-arranging we have

$$c_1 + \frac{c_2}{1+r} = w_1 + \frac{w_2}{1+r} \tag{5.10}$$

which is sometimes referred to as the lifetime budget constraint of the individual. The left-hand side is the present discounted value of all consumption, while the right-hand side is the present discounted value of all income.

This is now a simple constrained optimization problem. We could simply plug the budget constraint into the utility funtion and solve, but it will be useful to practice using a Lagrangian to solve it. Form the Lagrangian as

$$\mathcal{L} = U(c_1) + \beta U(c_2) + \lambda \left(w_1 + \frac{w_2}{1+r} - c_1 - \frac{c_2}{1+r} \right) \tag{5.11}$$

and maximize with respect to c_1, c_2, and λ, the Lagrange multiplier. This yields the following first order conditions:

$$U'(c_1) - \lambda = 0 \tag{5.12}$$

$$\beta U'(c_2) - \frac{\lambda}{1+r} = 0 \tag{5.13}$$

$$w_1 + \frac{w_2}{1+r} - c_1 - \frac{c_2}{1+r} = 0 \tag{5.14}$$

which we can then use to solve for the optimal consumption path. Take the first two conditions and solve them together for the following:

$$\frac{U'(c_1)}{\beta U'(c_2)} = (1+r). \tag{5.15}$$

This is simply the ratio of marginal utilities equal to a ratio of marginal costs. The interest rate $(1 + r)$ is the relative cost of c_1 in terms of c_2. That is, I'd have to give up $(1 + r)$ units of c_2 in order to afford an additional unit of c_1. Why? Because if I consume an extra unit in period 1, then I will not be able to earn interest on it in period 2. Likewise, the relative price of c_2 in terms of c_1 is $1/(1+r)$. I have to give up $1/(1+r)$ units of consumption today in order to afford an additional unit of consumption tomorrow.

Note that the relative price of consumption in period c_1 is strictly higher than in period c_2, assuming that $r > 0$. That is, consumption today is more expensive than consumption in the future. Why? Because you would be foregoing interest payments.

So what keeps the individual from only buying c_2? The marginal utility of c_1 relative to c_2. As c_1 drops the marginal utility rises, given our assumptions, so at some point it will be worth buying c_1. In addition, note that the marginal utility of c_2 is scaled down by β, so the individual will be less inclined to want to purchase that good.

It can be helpful to write this as

$$\frac{U'(c_1)}{U'(c_2)} = \beta(1+r), \tag{5.16}$$

and either this form or the prior one are often referred to as the *Euler equation*. This refers to the solution to an optimal dynamic programming model, which we are doing in a simplified setting. The Euler equation describes how the control variable (consumption) is related over time.

From the Euler equation, note that if $\beta(1 + r) > 1$, then it must be that $U'(c_1) > U'(c_2)$, or $c_1 < c_2$. In other words, consumption should be rising over time if $\beta(1 + r) > 1$. When will this be true? When either β is very large (and so the individual values the future a lot) or when $(1 + r)$ is large (and so the cost of consuming is very high). The inverse holds as well: if $\beta(1 + r) < 1$, then it will be the case that $c_1 > c_2$, or the person has falling consumption over time. If $\beta(1 + r) = 1$ exactly, then $c_1 = c_2$ and the person chooses to have the same amount of consumption in both periods.

The Euler equation yields an answer on how c_1 and c_2 should be related over time, but does not describe their exact values. To find this we need to employ the final first-order condition, the budget constraint. Visually, this is no different than a typical intermediate microeconomics problem. Figure 5.1 presents the optimal choice problem. The budget constraint is fixed by the location of w_1, w_2, which represent the endowment of the person. They look for the point at which their indifference curve is tangent to this budget line. That is their optimal choice for consumption, c_1^*, c_2^*. In the figure, this leaves them with positive net savings in period 1, and those savings are used to increase consumption in period 2 above their actual wage.

An Exact Solution with CRRA Utility

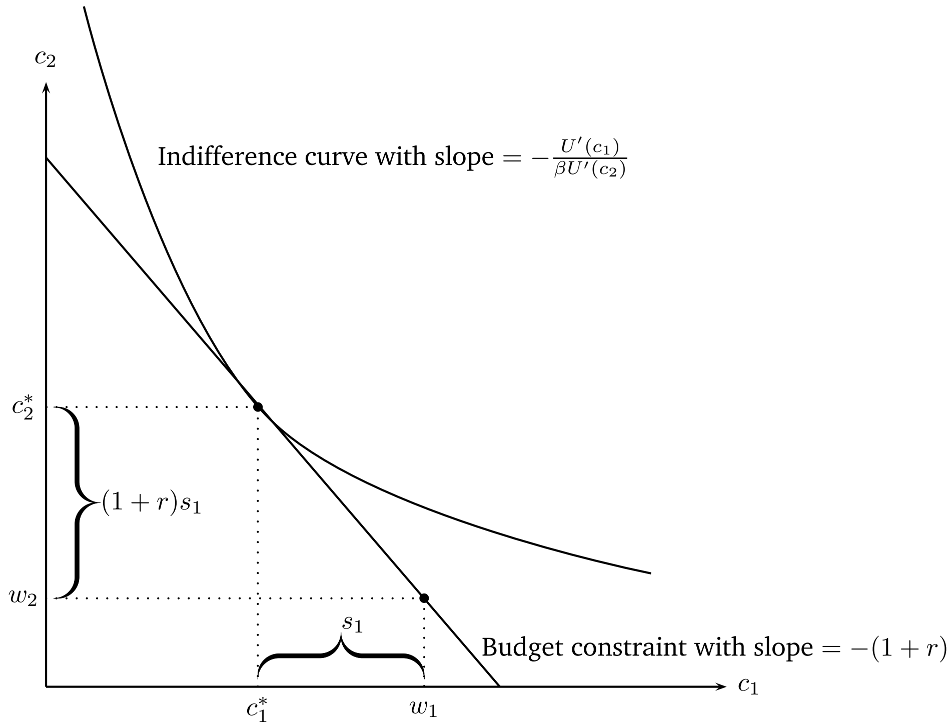


Figure 5.1: Optimal Consumption Decision

Note: The figure shows the optimal decision for an individual with $V = U(c_1) + \beta U(c_2)$ *and a budget constraint of* $c_1 + c_2/(1+r) = w_1 + w_2/(1+r)$*. The individual has an endowment of* (w_1, w_2)*, which dictates where the budget line must go. The slope of the budget line is* $-(1+r)$*, or the relative cost of* c_1 *with respect to* c_2*. The optimal point is where the indifference curve is tangent to the budget constraint, which occurs at the point* (c_1^*, c_2^*)*. The individual is thus a net saver.*

Using CRRA utility the felicity function is

$$U(c) = \frac{c^{1-\sigma}}{1-\sigma} \quad \text{where} \quad \sigma > 0 \tag{5.17}$$

where σ *is the coefficient of relative risk aversion. Using this the Euler equation becomes*

$$\frac{c_1^{-\sigma}}{\beta c_2^{-\sigma}} = (1+r). \tag{5.18}$$

This solves to

$$c_2 = c_1[\beta(1+r)]^{1/\sigma}. \tag{5.19}$$

Note that this preserves the idea that as $\beta(1+r)$ *rises, the amount of* c_2 *relative to* c_1 *rises. However, here we can see that an important factor is the intertemporal elasticity of substitution,*

$1/\sigma$. If the EIS gets very large (σ is small), then the person is willing to substitute consumption between periods, and so any small change in $\beta(1+r)$ will have a large impact on relative consumption. On the other hand, if the EIS is small (σ is large), then c_2 will tend to be very similar to c_1, regardless of the value of $\beta(1+r)$. The individual is unwilling to substitute between periods even if the interest rate or time preference rates are pointing in that direction.

We can use this relationship in the budget constraint to find

$$c_1 + \frac{c_1[\beta(1+r)]^{1/\sigma}}{(1+r)} = w_1 + \frac{w_2}{(1+r)} \tag{5.20}$$

which can be solved for

$$c_1^* = \frac{1}{1+\beta^{1/\sigma}(1+r)^{1/\sigma-1}}\left(w_1 + \frac{w_2}{(1+r)}\right). \tag{5.21}$$

There are two parts to this solution. First is a fraction telling us what proportion of the value in parentheses the individual should consume in the first period. The object in parentheses is the PDV of lifetime wealth, so the optimal solution is phrased as a proportion of lifetime wealth.

The fraction depends on β and $(1+r)$, as you would expect. We can simplify somewhat if we consider different extreme cases. If $1/\sigma = 1$, which would be the case if the person had log utility, $U = ln(c_t)$, then the fraction is just $1/(1+\beta)$. The fraction of lifetime wealth that should be consumed depends only on the time perference parameter, β. The higher it is, the less they consume in period 1.

If σ goes to infinity, then the EIS $1/\sigma$ goes to zero. In this case the fraction reduces to $(1+r)/(2+r)$, which is very close to 1/2. In other words, if they do not like to substitute between periods, then individuals will consume almost exactly half of their lifetime wealth in each period.

What about savings? They are

$$s_1 = w_1 - c_1^* = \left(\frac{\beta^{1/\sigma}(1+r)^{1/\sigma-1}}{1+\beta^{1/\sigma}(1+r)^{1/\sigma-1}}\right)w_1 - \left(\frac{1}{1+\beta^{1/\sigma}(1+r)^{1/\sigma-1}}\right)\frac{w_2}{(1+r)}. \tag{5.22}$$

This is very ugly. However, notice that whether savings are positive or negative depends in part on the relative size of w_1 and w_2. If w_1 is very large, and $w_2 = 0$, then savings will be positive for sure. The individual must save something. If $w_1 = 0$ and w_2 is large, then they must borrow ($s_1 < 0$) for sure.

With this optimal consumption/savings problem in place, we can think about how savings (and hence the supply of investment in the economy) responds to changes in the return on savings, $1+r$. Imagine that there is an increase in $1+r$. Does this increase savings? That depends on three separate effects:

- *The Substitution Effect:* The increase in $1+r$ is like an increase in the price of consumption in the initial period. This induces the individual to consume less c_1 and more c_2. Their savings, defined as $s_1 = w_1 - c_1$, would therefore rise, all else being equal (which it is not).

- *The Income Effect:* An increase in $1+r$ also allows the individual to have higher consumption in the last period for any given amount of savings. This effect essentially expands the set of consumption combinations lying within their budget constraint, and therefore may actually increase their consumption in the initial period. Whether this income effect outweighs the substitution effect depends on the intertemporal elasticity of substitution ($1/\sigma$ in the CRRA). If $EIS < 1$, then the substitutability of consumption across periods is low, the individual is unwilling to lower consumption much in the initial period, and the income effect dominates. c_1 goes up, and savings actually decline. Alternately, if $EIS > 1$, individuals do not mind moving consumption between periods, the substitution effect dominates, and savings increase in response to $1+r$ rising.

- *The Wealth Effect:* In addition to the effect of $1+r$ on how individuals allocate their wealth between periods, an increase in $1+r$ actually lowers their wealth by decreasing the present discounted value of their income. Because of this, the wealth effect acts to lower consumption in the initial period, and therefore savings tend to go up.

What this means is that savings does not respond monotonically to the rate of return. It depends on the nature of the utility function (through the EIS) as well as individuals initial position as a saver or borrower. As part of the homework you will work out the derivative of savings with respect to $1+r$, but the general pattern that holds once we assume that $EIS < 1$ can be seen in figure 5.2.

Here, savings always rise with the rate of return if the individual is a borrower. To gain some intutition for this, consider someone who has $w_1 = 0$ and $w_2 > 0$, or all their wealth is in the second period of life. They must borrow to consume in period 1. If the rate of return goes up, this means a big decrease in their lifetime wealth ($w_1 + w_2/(1+r)$). Therefore they will consume less in period 1, and if they are doing less consumption in period 1 they need to borrow less (save more).

In contrast, now consider someone who has $w_1 > 0$ and $w_2 = 0$, so that they are always a saver. Now, if $1+r$ goes up, their lifetime wealth is unaffected. Hence the income effect dominates and they want to consume more in period 1. This necessarily means they save less. For this individual, the increase in the rate of return means they can save less but still enjoy a large amount of consumption in the future.

Finally, it is worth considering the log form of utility, meaning that $\sigma = 1$ in the CRRA, and the EIS is precisely one. In this case the income and substitution effects exactly offset each other. The only effect of the rate of return on savings is through the wealth effect. In this case, savings

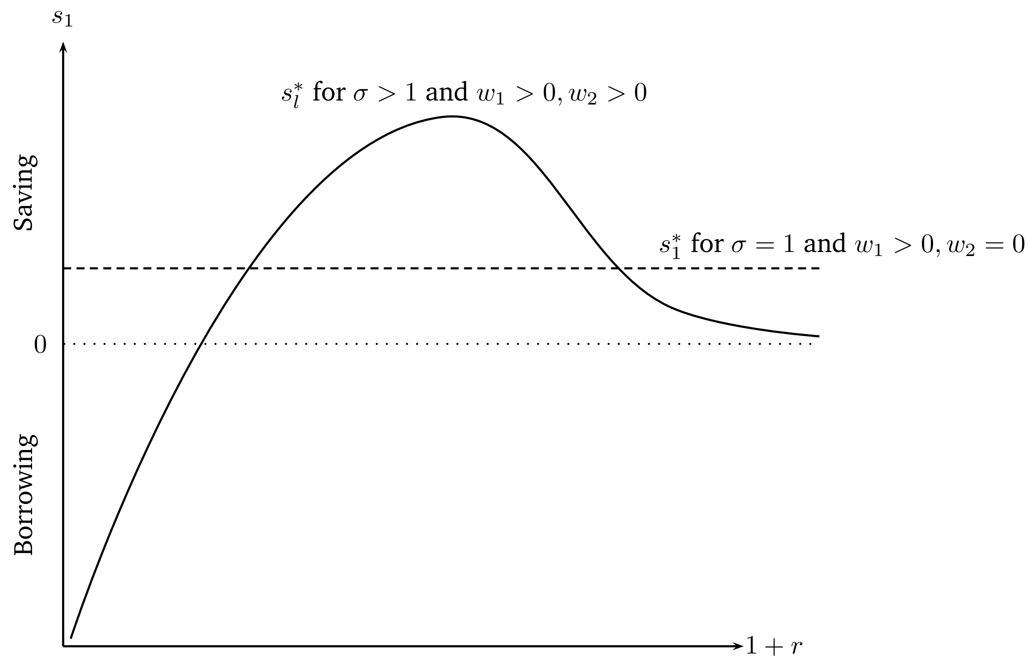


Figure 5.2: Savings and the Rate of Return

Note: The figure shows the optimal savings for an individual with an intertemporal elasticity of substitution of less than one. Savings rise monotonically at low levels of $1 + r$, as the wealth effect dominates the income effect. Once savings are already large, however, the income effect dominates and savings actually decrease with further increases in rates of return.

are always increasing with the rate of return, and the savings curve in figure 5.2 never reaches a point where it dips. In fact, individuals who are always savers, with $w_1 > 0$ and $w_2 = 0$, will save a specifc amount of their wealth, and this will not change regardless of what the interest rate may be. In other words, if $\sigma = 1$ and $w_1 > 0$ while $w_2 = 0$, then the savings curve is just the horizontal line noted in figure 5.2.

5.2 The Over-lapping Generations Model

We now have enough knowledge to incorporate endogenous savings into the original Solow model of accumulation we developed. Recall that in the Solow model the return on savings was irrelevant to the amount of investment, and hence to the capital stock. The Fisher model gives us a way of letting investment spending depend on the return on savings. Hence, if firms will pay a higher rate of return, individuals will adjust their savings behavior.

The over-lapping generations (OLG) model is a specific (and not the only) way of setting up accumulation and savings based on assuming that each period there is a new generation of Fisher-ian consumers born. In period t, then, there are two generations of consumers alive. Young consumers, for whom period t is their period 1. And old consumers, for whom period t is their period 2. In period $t+1$, the young consumers from period t are now old, and it is their period 2. Then there is a new generation of young people in period $t+1$ for whom $t+1$ is their period one.

To keep things relatively simple, we typically assume that depreciation is complete, or $\delta = 1$. This means that capital falls apart completely every period. We will additionally assume that each generation is $1+n$ times larger than the prior one. Given these assumptions, we can describe the accumulation of capital as simply.

$$K_{t+1} = I_t \tag{5.23}$$

In per-worker terms, we have

$$\frac{K_{t+1}}{N_{t+1}} = \frac{I_t}{N_t}\frac{N_t}{N_{t+1}} \tag{5.24}$$

$$k_{t+1} = \frac{i_t}{1+n}. \tag{5.25}$$

This doesn't look exactly like the Solow model of accumulation because of the odd way in which we've modeled population growth. k_{t+1} refers to the capital per young person in period $t+1$, while i_t is the investment per young person in period t. The division by $1+n$ captures the diluation of the investment across a larger number of young people.

Accounting for capital per young person is done because of a second assumption that we impose to keep the problem manageable. We assume that only young individuals provide labor. Old individuals cannot work, and will only be able to consume by saving some of their wages as young people. In terms of our Fisher model, we're assuming that $w_1 > 0$ but $w_2 = 0$. Hence every young individual will save something for their old age.

To go further we need to specify the utility function for individuals. And here's where we run into a particularly annoying problem. If we assume a general CRRA form for the $U(c_t)$ function, then what will occur given our assumption that $w_1 > 0$ and $w_2 = 0$? It will be the case that there is no wealth effect. And given that we are fairly certain that $\sigma > 1$, this means that the income effect outweighs the substitution effect for these individuals, and any increase in $1+r$ will actually *lower* savings. The supply of savings, so to speak, slopes downward.

Generally, we get around this by assuming that utility is of the form $U(c_t) = \ln c_t$. Along with the assumption that people only have wages in period one, this means that savings do not change with respect to the rate of return. This will work for finding a solution, but note that we've now assumed away the reason for endogenizing savings in the first place - to let it depend on the rate of return. Regardless, it's worth solving out the simple OLG model, as it forms the basis for a number of models that are not particularly interested in savings rates *per se*. This includes models of fertility and bubbles involving nominal assets.

In the simple OLG, lifetime utility is $V = \ln c_{1t} + \beta \ln c_{2,t+1}$, where I've indicated both the time period from the perspective of the individual (1 or 2) and from the perspective of the economy as a whole (t and $t+1$). We already noted that their lifetime budget constraint is $c_{1t}+c_{2,t+1}/(1+r_{t+1}) = w_t$. You can solve this model out for optimal savings of a young person in time t, which are

$$s^*_{1t} = \frac{\beta}{1+\beta} w_t. \tag{5.26}$$

As noted above, accumulation of capital is driven by the equation

$$k_{t+1} = \frac{i_t}{1+n}. \tag{5.27}$$

So to finish off the model we need to specify how savings and investment are related.

Typically, we'd just assume that $i_t = s^*_{1t}$, or savings are costlessly transformed directly into investment goods. However, that need not necessarily be the case. First, the collection of savings and loaning of them to firms to do investment is something that is handled (often) by the financial sector. So we could add some detail on how the financial market works. Alternatively, we could propose that for every unit of output that is saved, only some fraction of it can be actually translated into investment.

For now, let's continue with the presumption that total savings are costlessly translated into new investment. This implies that

$$k_{t+1} = \frac{s_{1t}}{1+n} = \frac{\beta}{1+\beta} \frac{w_t}{1+n}. \tag{5.28}$$

To complete the model, we need to describe how wages are determined. If we use a typical Cobb-Douglas production function, $y_t = k_t^\alpha$, then we have that

$$w_t = (1-\alpha) k_t^\alpha. \tag{5.29}$$

Putting this into the prior equation yields

$$k_{t+1} = \frac{\beta}{1+\beta} \frac{(1-\alpha) k_t^\alpha}{1+n}. \tag{5.30}$$

Here we have a difference equation in k. Capital per worker in the future depends on capital per worker today. We can work through the implications of this simple OLG model similar to what we did with the Solow model.

First, the economy will reach a steady state with a constant level of capital per worker, k^*. This is easy to see visually in figure 5.3. The level of k_{t+1} is mapped against k_t. From equation (5.30) you can see that the relationship is concave. At any level of k_t less than k^*, $k_{t+1} > k_t$, and the capital stock is increasing, as shown at point A in the figure. The capital per worker continues to grow until it reaches k^*. Similar logic explains that if $k_t > k^*$ then the capital stock will fall towards the steady state. Hence it is stable.[2].

[2]As with the Solow model, $k = 0$ is a steady state as well

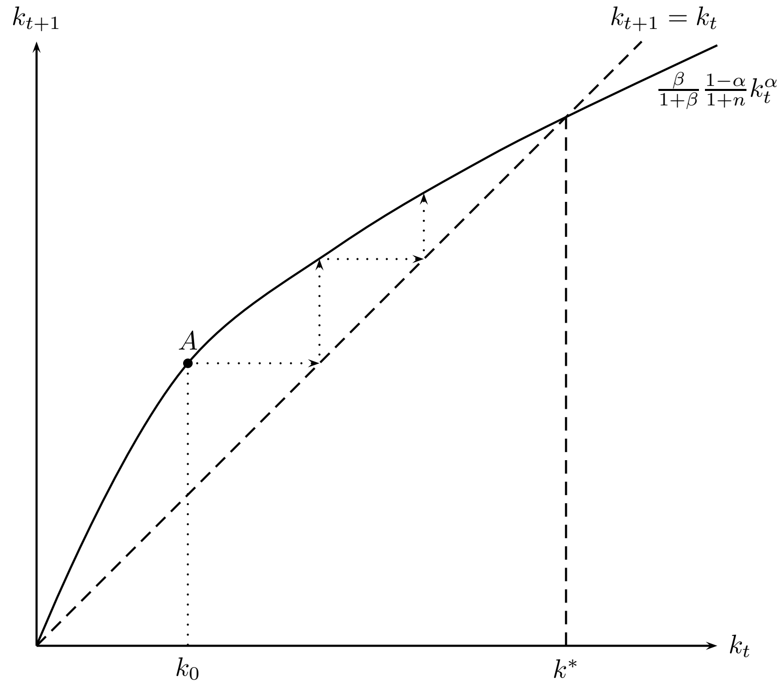


Figure 5.3: Dynamics in the OLG Model

Note: The figure shows the evolution of k_{t+1} given the value of k_t. Where this curve crosses the line of equality, $k_{t+1} = k_t$, there is a steady state. From an original endowment of capital of k_0, one can follow the evolution of the capital stock period by period as it climbs towards the steady state.

Solving for the steady state from (5.30), set $k_{t+1} = k_t = k^*$. Then we have

$$k^* = \left(\frac{\beta}{1+\beta}\frac{(1-\alpha)}{1+n}\right)^{1/(1-\alpha)}. \tag{5.31}$$

This let's us consider the second set of implications of the Solow model. The steady state is increasing in the time preference rate β, and decreasing in the population growth rate. As $\beta/(1+\beta)$ is the savings rate, this is a similar conclusion that we drew from the Solow model.

What's the Difference between Solow and OLG?

There isn't much of one. The OLG assumes total depreciation ($\delta = 1$) and this over-lapping means of population growth, and those combined end up giving us a slightly different way of tracking what happends to capital. But it's simply an accounting difference.

Recall from the Solow model that I said one could approximate the growth rate of capital *per worker as*

$$\frac{\Delta k_{t+1}}{k_t} \approx \frac{\Delta K_{t+1}}{K_t} - \frac{\Delta N_{t+1}}{N_t}. \tag{5.32}$$

However, the exact relationship is

$$\frac{\Delta k_{t+1}}{k_t} = \frac{N_t}{N_{t+1}} \left(\frac{\Delta K_{t+1}}{K_t} - \frac{\Delta N_{t+1}}{N_t} \right). \tag{5.33}$$

If we start from this exact equation in the Solow model and apply the OLG assumptions, we will get the same relationship of k_{t+1} to k_t that we have in the OLG.

First, by N_t we now mean just the young generation of workers. And we know that $N_{t+1} = (1+n)N_t$. Using that and re-arranging slightly the above equation we have

$$\frac{\Delta k_{t+1}}{k_t} = \frac{1}{(1+n)} \frac{K_{t+1}}{K_t} - 1. \tag{5.34}$$

Multiple both sides of this equation by k_t and you have

$$k_{t+1} - k_t = \frac{K_{t+1}/N_t}{(1+n)} - k_t. \tag{5.35}$$

You can drop the k_t from both sides. We also know that $K_{t+1} = I_t$ in the OLG, because we've assumed that $\delta = 1$. Finally, if we assume that savings are instantly translated into investment, then $I_t = s_t N_t$ in the OLG. So we have

$$k_{t+1} = \frac{s_t}{1+n}, \tag{5.36}$$

as stated before. In the OLG, we know that $s_t = \frac{\beta}{1+\beta}(1-\alpha)y_t$. So from the perspective of the Solow model, the fixed savings rate s is just $s = \frac{\beta}{1+\beta}(1-\alpha)$. But the two models are essentially identical, with the OLG a special case of the Solow.

5.3 Infinitely-lived Savers

With our intuition from the Fisher model, we can turn to an expanded savings problem, one where individuals live for an infinite number of periods. The infinite number of periods is non-sensical on the face of it, but is useful as a reference. Essentially, we are interested in individuals with long time horizons, or individuals that are concerned about the future beyond their own lives (for example, they care about their children's consumption).

Infinite lives just expands on the original Fisher model set up. The utility function can be written as:

$$V_0 = \sum_{t=0}^{\infty} \beta^t U(c_t). \tag{5.37}$$

where β is again the time preference rate, but note that now it is to the t power, so that at any period t, the time preference rate for $t+1$ is β. From the perspective of time zero, the preference for consumption in period t is β^t. Note that I'm also running time from period zero forward, which is going to make accounting for the infinite sums easier.

One somewhat tedious point is that we often will find it easier to refer to a person's "time discount rate", rather than the preference parameter β. Let θ be the rate at which people discount future utility. You can think of this as the percentage that utility increases by consuming today rather than tomorrow. There is nothing different in thinking of preferences this way. Very simply

$$\beta = \frac{1}{1+\theta} \tag{5.38}$$

and thus β and θ capture the same effect - we enjoy consumption today rather than consumption tomorrow. Sometimes it will be easier, notationally, to use β, while other times θ is useful to keep equations clean. You will simply need to get comfortable translating back and forth.

We have a utility function, now we require a budget constraint. We'll be a little more specific here in writing the budget constraint in two different forms. The *dynamic budget constraint* states that:

$$a_{t+1} + c_t = w_t + (1+r)a_t \tag{5.39}$$

where a_{t+1} are our stock of assets tomorrow, c_t is the amount we consume today, w_t is our income today, and a_t is our existing stock of assets. The value $(1+r)$ is the rate of return on our assets, and for now we will assume that this is constant to simplify the mathematics. For any given periods t and $t+1$, this equation describes how our assets evolve based on our consumption decision. The only other thing we require given the dynamic budget constraint is an initial asset level, a_1. That tells us how much we have available prior to making any choices.

This dynamic budget constraint is equivalent to the *lifetime budget constraint*, which is what we worked with in the Fisher model. For infinitely lived individuals the lifetime budget constraint is

$$\sum_{t=0}^{\infty} \frac{c_t}{(1+r)^t} = a_0 + \sum_{t=0}^{\infty} \frac{w_t}{(1+r)^t}. \tag{5.40}$$

This lifetime budget is arrived at by applying the dynamic budget constraint over and over again until one has an expression for a_T, and asserting that $a_T/(1+r)^T$ goes to zero as T goes to infinity. In other words, the present discounted value of assets at infinity must be zero. One cannot accumulate an infinite supply of assets, nor can one accumulate a non-zero amount of debt ($a_T < 0$).

To maximize utility subject to this lifetime budget constraint we can set up a very large Lagrangian, as in

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t U(c_t) + \lambda \left(a_1 + \sum_{t=0}^{\infty} \frac{w_t}{(1+r)^t} - \sum_{t=0}^{\infty} \frac{c_t}{(1+r)^t} \right). \tag{5.41}$$

This is fine, but has an infinite number of first order conditions. Needless to say, this makes actually solving this problem somewhat difficult. However, we do not really need all the first-order conditions to understand what the optimal solution should look like.

Take the first-order condition for time t and time $t+1$. These are

$$\beta^t U'(c_t) - \lambda \frac{1}{(1+r)^t} = 0 \tag{5.42}$$

$$\beta^{t+1} U'(c_{t+1}) - \lambda \frac{1}{(1+r)^{t+1}} = 0 \tag{5.43}$$

and solving these together to eliminate λ we have

$$\frac{U'(c_t)}{U'(c_{t+1})} = \beta(1+r) \tag{5.44}$$

which is simply the Euler equation we identified before, only written with respect to any two adjacent periods, t and $t+1$, rather than for periods 1 and 2 specifically. What this says is that for any adjacent periods, the relative marginal utility must be equal to $\beta(1+r)$. The Euler equation essentially gives us a rule for how consumption should evolve from period to period. Note that this condition does not depend at all on the budget constraint. In other words, it does not matter how large is our total lifetime wealth, the pattern of consumption over time will be identical. Just as with the Fisher model, whether consumption is growing or falling over time depends on whether $\beta(1+r)$ is greater than, less than, or equal to one.

If we wanted to solve for the exact path of consumption (e.g. a specific value for c_{14} or c_{1456}), then we need to incorporate the budget constraint. While the Euler equation tells us what the pattern of consumption is over time, the budget constraint will tell us the level. This can be seen in figure 5.4, which depicts the optimal consumption path for two individuals. Person A has high lifetime wealth, while person B has low wealth. Person A obviously gets to enjoy higher consumption in every period, but notice that the *shape* of the the consumption path is identical, only the height differs. This is because the Euler equation, which dictates the slope of the consumption path at every point t, is independent of the budget itself.

A Specific Solution for Infinitely-Lived Consumption

Assume that we have CRRA utility, so that $U(c_t) = c_t^{1-\sigma}/(1-\sigma)$. If this is the case, then the Euler equation is

$$\frac{c_{t+1}}{c_t} = [\beta(1+r)]^{1/\sigma}. \tag{5.45}$$

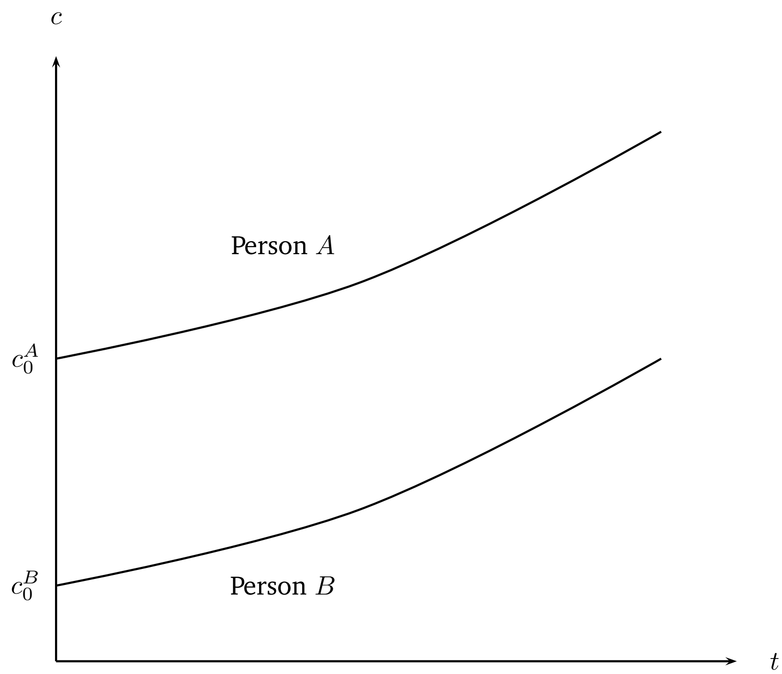


Figure 5.4: Consumption Paths

Note: Both person A and person B face identical interest rates and have identical discount rates. Therefore the slope of their consumption path through time is identical. However, person A has a higher lifetime wealth than person B, and so consumption is higher at every point.

Knowing this general Euler equation holds between any two given periods, it must be the case that

$$c_t = c_0[\beta(1+r)]^{t/\sigma} \tag{5.46}$$

and we can put this into the budget constraint from (5.40) to solve for the actual consumption path. Before doing so, define the following term as lifetime wealth, W:

$$W \equiv a_0 + \sum_{t=0}^{\infty} \frac{w_t}{(1+r)^t} \tag{5.47}$$

and therefore we can write the budget constraint as

$$\sum_{t=0}^{\infty} \frac{c_t}{(1+r)^t} = W \tag{5.48}$$

$$\sum_{t=0}^{\infty} \frac{c_0(\beta(1+r))^{t/\sigma}}{(1+r)^t} = W \tag{5.49}$$

$$c_0\left(\frac{1}{1-(\beta(1+r))^{1/\sigma-1}}\right) = W \tag{5.50}$$

$$c_0 = \frac{(1+r)-(\beta(1+r))^{1/\sigma}}{(1+r)} W \tag{5.51}$$

and initial consumption is a fraction of total lifetime wealth.

This determines the level of initial consumption, while the Euler equation determines the path over time after c_0, as in figure 5.4. Note that if $\beta(1+r) > 1$ then consumption is growing over time, and the fraction of lifetime wealth consumed at time zero is small. In contrast, when $\beta(1+r) < 1$ consumption is falling over time and the fraction of wealth consumed initially will be large. Only when $\beta(1+r) = 1$ will consumption be perfectly smooth over time. Note that in this last case, initial consumption reduces to the simpler function of

$$c_0 = \frac{r}{1+r} W. \tag{5.52}$$

We now have a description of consumption in period zero, can we describe consumption in any given period t? We can, using the relationship in (5.46) that tells us what c_t is given c_0. However, this gives us an expression for c_t as a function of W, total lifetime wealth. Can we describe c_t in terms of remaining lifetime wealth at time t itself? Note that in any given period t, we could describe remaining lifetime wealth as

$$W_t \equiv a_t + \sum_{s=0}^{\infty} \frac{w_{s+t}}{(1+r)^s} \tag{5.53}$$

and we can describe consumption in all future periods with respect to consumption in period t as

$$c_s = c_t(\beta(1+r))^{s/\sigma} \tag{5.54}$$

by using the Euler equation. We can solve the problem the same as before, finding that

$$c_t = \frac{(1+r)-(\beta(1+r))^{1/\sigma}}{(1+r)} W_t \tag{5.55}$$

or consumption in time t should be a fraction of lifetime wealth. No matter what period we are in, we should always consume the same fraction of our wealth. Equation (5.55) can be referred to as the consumption function, or the rule for determining how much to consume in any given period.

5.4 The Continuous Time Problem

We can reconsider the whole question of dynamic optimization of consumption behavior, except now we can look at a continuous time version. That is, people don't have discrete periods of life, but evolve, well, continuously. This is primarily just a change in notation and mathematical technique, but all the same intuitions still apply.

Now we have

$$\max_{c} \int_{0}^{\infty} e^{-\theta t} U(c)\, dt \tag{5.56}$$

subject to the constraint that

$$\dot{a} = ra + w - c \tag{5.57}$$

where a is now the instantaneous level of assets, w is the instantaneous wage rate, c is the instantaneous level of consumption, and r is the constant rate of interest. In the utility function, θ represents the discount rate, only it is set up in continuous time. The level of a is the state variable, meaning that it does not jump around, while c is the control variable, meaning that it can.[3]. For all variables, I've dropped the time subscript for convenience.

We again want to eliminate the possibility of unlimited borrowing and infinite consumption, so we have the present value budget constraint of

$$a_0 + \int_{0}^{\infty} e^{-rt} w dt = \int_{0}^{\infty} e^{-rt} c dt \tag{5.58}$$

You could set up a Lagrangian, but then we'd need an infinite number of multipliers. Instead, we'll set up a Hamiltonian, which is written as

$$H = \max_{c} \left\{ e^{-\theta t} U(c) + \mu (ra + w - c) \right\} \tag{5.59}$$

which looks a lot like a Lagrangian. However, the multiplier μ is the instantaneous shadow value of assets, and is time varying. Now, to solve this problem you need to apply several conditions to the Hamiltonian. This method is called optimal control theory, and we will use this theory without explicitly discussing the origins of it.

First, you maximize H with respect to c, as written

$$U'(c)\, e^{-\theta t} - \mu = 0 \tag{5.60}$$

and this gives you something that looks like the FOC from the Lagrangian. It's telling us that we have to balance out the marginal gain in utility from consumption against the marginal cost, which is given by μ and represents the shadow value of assets at any given point in time.

[3] A state variable is like your weight, while a control variable is like your calorie intake for the day. You can vary your calorie intake daily, jumping from 100 to 1000 to 5000 calories or back. But your weight will only adjust slowly in reaction to changes in the control.

Next, you recover the constraint by taking $\partial H/\partial \mu = \dot{a}$ or

$$\dot{a} = ra + w - c \tag{5.61}$$

which is just ensuring that we meet the constraint on how assets accumulate. The rate of change of the shadow value of assets, $\dot{\mu} = -\partial H/\partial a$ or

$$\dot{\mu} = -\mu r \tag{5.62}$$

and this describes how the shadow value of assets evolves over time. Finally, we have a transversality condition, which keeps the problem from "blowing up" and having infinite consumption later in life. This condition is that

$$\lim_{t\to\infty} \mu a = 0 \tag{5.63}$$

You solve (5.60) , (5.61), and (5.62) together in order to find the solution. First, take (5.60) and take the derivative with respect to time

$$\dot{c}U''(c)\, e^{-\theta t} - \theta U'(c)\, e^{-\theta t} - \dot{\mu} = 0$$

which gives us another expression for $\dot{\mu}$. Plug (5.62) into the above equation to get

$$\dot{c}U''(c)\, e^{-\theta t} - \theta U'(c)\, e^{-\theta t} = -\mu r$$

Now notice from (5.60) that $\mu = U'(c)\, e^{-\theta t}$ and plug that in to get

$$\dot{c}U''(c)\, e^{-\theta t} - \theta U'(c)\, e^{-\theta t} = -U'(c)\, e^{-\theta t} r$$

Start going crazy with the algebra and you can get the following statement

$$\frac{\dot{c}}{c} = (r - \theta)\,\frac{U'(c)}{cU''(c)} \tag{5.64}$$

which describes the growth of consumption over time. Assume for the moment that the $\frac{U'(c)}{cU''(c)}$ term is constant. Then whether consumption is growing or falling depends on the relative size of r and θ, or exactly what we saw in the Fisher model. If r is larger, then consumption is rising as people save their incomes, and if θ is larger then consumption is falling as people discount the future a lot.

The second term on the right hand side of (5.64) should be familiar. It's just the inverse of the coefficient of relative risk aversion. What it says is that your consumption growth will be slower if your risk aversion (smoothing preference) is higher.

In the CRRA case, we know exactly what the risk aversion is, σ. So that means that if preferences are CRRA, the optimal consumption growth is

$$\frac{\dot{c}}{c} = (r - \theta)\,\frac{1}{\sigma}$$

Again, solving explicitly for the level of consumption is possible. Equation (5.64) is a first order differential equation with a simple form and has the solution that

$$c_t = c_0 e^{\frac{1}{\sigma}(r-\theta)t}$$

which gives us a nice way to describe consumption in any period. Now we need the budget constraint, which was

$$a_0 + \int_0^\infty e^{-rt} w dt = \int_0^\infty e^{-rt} c dt$$

and we can plug in our formula for c_t, play with some algebra and get

$$a_0 + \int_0^\infty e^{-rt} w dt = c_0 \int_0^\infty e^{\frac{1}{\sigma}(r(1-\sigma)-\theta)t} dt$$

The integral on the right hand side can be evaluated to be a positive, finite number if

$$(1-\sigma)\, r < \theta$$

which is just like the condition we saw in discrete time. Now, evaluating the integral and rearranging we get that

$$c_0 = \frac{1}{\sigma}\left(\theta - (1-\sigma)\, r\right)\left[a_0 + \int_0^\infty e^{-rt} w dt\right].$$

This form of the consumption function yields essentially identical conclusions to the discrete time model. At times the continuous time model proves more convenient to use in further analysis, and at others the discrete model works. It doesn't really matter which form of the problem we use, but for the purposes of this class we'll generally stick with the discrete model.

5.5 The Ramsey Model

Recall that the Solow model told us how capital accumulated, given a savings rate. The amount of capital in the economy determines the rate of return on savings. Consumption theory tells us that the amount we save depends on the rate of return to savings. The Ramsey model puts these two concepts together: the capital stock depends on the consumption decision, and the consumption decision depends on the capital stock. We'll see that the steady state solution to this model has many of the properties of the original Solow model, but predictions about how the economy acts out of the steady state differ.

5.5.1 The Centralized Solution

To begin, we'll consider the centralized problem. We can think of a benevolent dictator or social planner attempting to maximize discounted utility for the economy as a whole. Alternatively, we could imagine that this is an economy populated by a single individual. The important part is that that we are solving for the optimal solution, meaning that the maximization is done with full knowledge of the ramifications of one's decisions.

Maximization

The problem is to maximize utility

$$V_0 = \sum_{t=0}^{\infty} \beta^t U(c_t) \tag{5.65}$$

subject to the following dynamic budget constraint

$$k_{t+1} = (1-\delta)k_t + f(k_t) - c_t \tag{5.66}$$

where k_t is the capital stock, $f(\cdot)$ is the production function, δ is the depreciation of capital. In addition, a transversality condition will dictate that utility cannot become infinite

$$\lim_{t\to\infty} \beta^t U'(c_t)k_t = 0. \tag{5.67}$$

Two notes about the dynamic budget constraint. First, the dynamic budget constraint can be written as follows

$$\frac{\Delta k_{t+1}}{k_t} = \frac{f(k_t) - c_t}{k_t} - \delta \tag{5.68}$$

which is simply the equation of motion from the Solow model written more generally. In the Solow model we presumed that $c_t = (1-s)f(k_t)$, but now c_t will be chosen optimally to maximize utility. Second, the dynamic budget constraint here is the same as the dynamic constraint in the consumption problem. We are assuming in the Solow that the only assets in the economy are capital, k_t, so it must be that $a_t = k_t$. Second, we are assuming that the centralized decision-maker earns all the income in the economy, which means she has $f(k_t) - \delta k_t$ available to spend. In other words, her w_t is equal $f(k_t) - \delta k_t$.

We do not know the full time path of income, though, as we did in the consumption problem. That is because the decisions the central planner makes today will influence income in the future through the capital stock. This means we cannot roll up some lifetime budget constraint for the planner. We'll need to optimize using the dynamic budget constraint directly. At time zero, the Lagrangian for this can be written as

$$\mathcal{L}_0 = \sum_{t=0}^{\infty} \left(\beta^t U(c_t) + \lambda_t \left[(1-\delta)k_t + f(k_t) - c_t - k_{t+1}\right]\right) \tag{5.69}$$

where the Lagrangian has an infinite number of multipliers, one for each period off into the future, and so we denote them λ_t. We want to maximize this Lagrangian with respect to c_t, k_{t+1}, and λ_t

for each of the infinite number of periods. However, we can examine only the first-order conditions from two adjacent periods to find the general rules the economy must follow.

For any given period, the value of k_t is given, so we do not have to maximize with respect to it. But for period $t+1$, we can optimize over k_{t+1}, because we have the option about what to set this to. So for periods t and $t+1$, we have the following first order conditions:

$$\frac{\partial \mathcal{L}}{\partial c_t} = \beta^t U'(c_t) - \lambda_t = 0 \tag{5.70}$$

$$\frac{\partial \mathcal{L}}{\partial c_{t+1}} = \beta^{t+1} U'(c_{t+1}) - \lambda_{t+1} = 0 \tag{5.71}$$

$$\frac{\partial \mathcal{L}}{\partial k_{t+1}} = \lambda_{t+1}\left[f'(k_{t+1}) + (1-\delta)\right] - \lambda_t = 0. \tag{5.72}$$

The form of the final first-order condition follows because the term k_{t+1} shows up in two different constraints: in the period t constraint, k_{t+1} is one of the options on how to spend our current income, in the period $t+1$, k_{t+1} dictates the stock of capital available and output.

Notice that the first two first-order conditions are similar to the typical conditions from the optimal consumption problem. However, note that now the multipliers are not identical. Why not? In the optimal consumption problem, recall that the first order conditions were

$$\frac{\partial \mathcal{L}}{\partial c_t} = \beta^t U'(c_t) - \lambda = 0 \tag{5.73}$$

$$\frac{\partial \mathcal{L}}{\partial c_{t+1}} = \beta^{t+1} U'(c_{t+1}) - \frac{\lambda}{1+r} = 0 \tag{5.74}$$

or the marginal value of wealth in period t was equal to λ, and the marginal value of wealth in period $t+1$ was equal to $\lambda/(1+r)$. So the marginal value of wealth was different across periods, but because the rate of return was given, we had a simple way or relating these marginal values.

Now, in the Ramsey model, the rate of return depends on our consumption decision, and so we cannot make a simplification regarding the marginal value of wealth between periods. Instead, we have to appeal to the third first-order condition that relates λ_t to λ_{t+1}.

Solving the first-order conditions from the Ramsey problem to eliminate the multipliers, we end up with

$$\frac{U'(c_t)}{U'(c_{t+1})} = \beta\left[f'(k_{t+1}) + 1 - \delta\right] \tag{5.75}$$

which is a generalized form of the Euler equation. It has the same interpretation as before, the ratio of marginal utilities must equal the ratio of marginal costs for consumption across periods. What the Ramsey model tells us is that the cost to one extra dollar of consumption today is equal to $1-\delta+f'(k_{t+1})$, while the cost of an extra unit of consumption in the future is still $1/\beta$.

The Euler equation has to be combined with the resource constraint $k_{t+1} = (1-\delta-n)k_t + f(k_t) - c_t$ to solve for the optimal time path of consumption as well as the steady state solution.

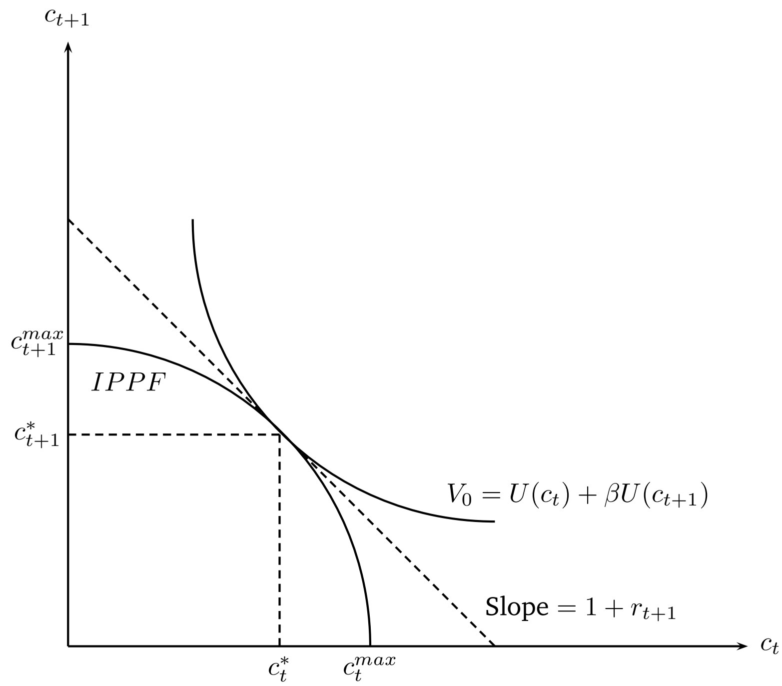


Figure 5.5: Inter-temporal Optimization

Visualizing the Solution

To understand what the central planner or individual is doing between any two periods, we can examine the maximization in a simple graph.

First, consider just periods t and $t+1$, holding constant consumption in every other period of time, as well as the capital stock in every other period. We can write lifetime utility as

$$V_0 = U(c_t) + \beta U(c_{t+1}) + \Omega \tag{5.76}$$

where Ω represents the utility obtained in all other periods in life, and is constant.

Holding V_0, constant, marginal rate of substitution between c_{t+1} and c_t is

$$\frac{\partial c_{t+1}}{\partial c_t} = -\frac{U'(c_t)}{\beta U'(c_{t+1})} \tag{5.77}$$

which gives us the slope of the indifference curves for this economy.

From the production side, the resource constraints tell us that

$$c_{t+1} = f(k_{t+1}) - k_{t+2} + (1-\delta)k_{t+1} \tag{5.78}$$

$$= f\left(f(k_t) - c_t + (1-\delta)k_t\right) - k_{t+2} + (1-\delta)\left[f(k_t) - c_t + (1-\delta)k_t\right] \tag{5.79}$$

which relates consumption in period c_{t+1} to consumption in period c_t, holding constant k_t and k_{t+2}. This equation describes an inter-temporal production possibility frontier (IPPF), showing all the different combinations of output in the two periods that satisfy the constraints imposed by the fixed values of k_t and k_{t+2}.

The slope of the IPPF is

$$\frac{\partial c_{t+1}}{\partial c_t} = -\left(f'(k_{t+1}) + (1-\delta)\right) \tag{5.80}$$

and the second derivative of this relationship is

$$\frac{\partial^2 c_{t+1}}{\partial c_t^2} = f''(k_{t+1}) < 0 \tag{5.81}$$

meaning that the IPPF is a concave function.

The individual is trying to maximize utility, subject to the IPPF. As in a typical economics problem, we look for the point at which the indifference curve is tangent to the budget constraint. Figure 5.5 displays this optimal consumption decision.

As the slopes of the IPPF and indifference curve must be equal at this tangency, we have

$$\frac{\partial c_{t+1}}{\partial c_t}\,|_{Util} = -\frac{U'(c_t)}{\beta U'(c_{t+1})} = -\left(f'(k_{t+1}) + (1-\delta)\right) = \frac{\partial c_{t+1}}{\partial c_t}\,|_{IPPF} \tag{5.82}$$

which you'll see is just recovering the Euler equation relating consumption across periods.

From this representation, we can also ask ourselves an important question that links this problem to the decentralized problem we'll explore later. What is the price of consumption in period t relative to consumption in period $t+1$ that is consistent with this equilibrium? That price, which we have called $(1 + r_{t+1})$ in our consumption chapter, is simply the slope of the IPPF and indifference curve at the optimal solution.

This slope is

$$1 + r_{t+1} = f'(k_{t+1}) + (1-\delta). \tag{5.83}$$

If we were to save one extra dollar, what would we receive? We'd get the marginal return on this new capital, but some of this dollar would be lost to depreciation, so the net return on the savings is $f'(k_{t+1}) - \delta$.

Steady State Solution

As in the Solow model, we can look for steady states in which consumption, capital, and output are all constant. What this steady state implies is that $\Delta k_{t+1} = 0$ and $\Delta c_t = 0$. Let c^* and k^* be the long-run steady state equilibrium levels of consumption and capital.

From the Euler equation, we have in steady state that

$$\frac{U'(c^*)}{U'(c^*)} = \beta\left[f'(k^*) + 1 - \delta\right] \tag{5.84}$$

which implies that

$$f'(k^*) = \frac{1}{\beta} + \delta - 1 \tag{5.85}$$

and if we have an explicit form for the intensive production function we could solve this for an exact value of k^*. An explicit form of the utility function would yield exact solutions for steady state consumption.

Recall from the previous section that we determined that the return to savings, $1+r$, is equal to $f'(k) + 1 - \delta$. In steady state, then, the return to savings is

$$1 + r^* = 1 + f'(k^*) - \delta = 1 + \frac{1}{\beta} + \delta - 1 - \delta = \frac{1}{\beta} \tag{5.86}$$

which tells us that the steady state return to savings is a function only of the time discount rate. This has to hold, because the only way for a steady state to even exist is if consumption is constant over time. Consumption is constant over time only when $(1+r) = 1/\beta$.

This is one point where the usefulness of talking about discount rates (θ) is cleaner than talking about β. Recall that $\beta = 1/(1+\theta)$. Given that, the steady state return to capital is

$$r^* = \theta \tag{5.87}$$

which tells us that the steady state interest rate is equivalent to the rate at which individuals discount the future. This provides a little intuition for how patience influences the return to capital and hence the capital stock. Impatient people have a high discount rate, and to get them to save enough to replenish the capital stock every period we must offer them a large return to their savings. The only way for the return to savings to be large is to have a small capital stock. Hence impatient people lead to small steady state capital stocks.

In addition, we can compare this steady state solution to the Golden Rule steady state of the Solow model. Recall that the Golden Rule steady state was achieved when we set $f'(k^{GR}) = \delta$ (ignoring population growth). In contrast, the Ramsey model tells us that the steady state marginal product of capital is $f'(k^*) = 1/\beta + \delta - 1$. As $1/\beta - 1 > 0$, it must be the case that the marginal product of capital is higher in the Ramsey steady state than under the Golden Rule. This implies that $k^* < k^{GR}$. In other words, the Ramsey model delivers a steady state capital stock *below* the Golden Rule level.

Why does this arise? In the Ramsey model, we have incorporated a time discount rate, β, into the utility function, and therefore the short-run costs of achieving the Golden Rule (high savings and low consumption) are factored against the discounted future gains (high consumption). Because of the discounting, it becomes optimal to forego the Golden Rule consumption level in the long run in return for some higher consumption in the present.

An Explicit Ramsey Model

Let us use a CRRA function for per-period utility, so that

$$U(c_t) = \frac{c_t^{1-\sigma}}{1-\sigma} \tag{5.88}$$

and the production function is Cobb-Douglas

$$y_t = k_t^{\alpha}. \tag{5.89}$$

In steady state, we determined that it must be that $f'(k^*) = \theta + \delta$, which given the production function implies that

$$k^* = \left(\frac{\alpha}{\delta+\theta}\right)^{1/(1-\alpha)} \tag{5.90}$$

and note the similarities to the Solow model steady state with the Cobb-Douglas function. In place of the savings rate, we have α, which you'll recall is actually the Golden rule savings rate. However, in addition to dividing by the depreciation rate, we also must divide by a term related to the time discount rate.

In steady state, the dynamic budget constraint is

$$k^* = k^{*\alpha} + (1-\delta)k^* - c^* \tag{5.91}$$

which can be solved for, using (5.90),

$$c^* = \left(\frac{1}{\delta+\theta}\right)^{1-\alpha} \frac{(1-\alpha)\delta+\theta}{\alpha^{\alpha}} \tag{5.92}$$

Dynamics

To understand what is happening *out* of the steady state, we can construct a phase diagram of the Ramsey model. This is built off of the two difference equations that the optimization delivers: the Euler equation and the dynamic budget constraint. They are:

$$\frac{U'(c_t)}{U'(c_{t+1})} = \beta\left[f'(k_{t+1}) + 1 - \delta\right] \tag{5.93}$$

$$\Delta k_{t+1} = f(k_t) - c_t - \delta k_t \tag{5.94}$$

and both are non-linear difference equations. The phase diagram will look at how each variable, consumption and the capital stock, changes based on given values of c_t and k_t.

Taking consumption first, note that the Euler equation implies a growth rate in consumption, Δc_{t+1}, between periods t and $t+1$. If the capital stock is at steady state, and $f'(k_{t+1}) = 1/\beta - 1 + \delta$,

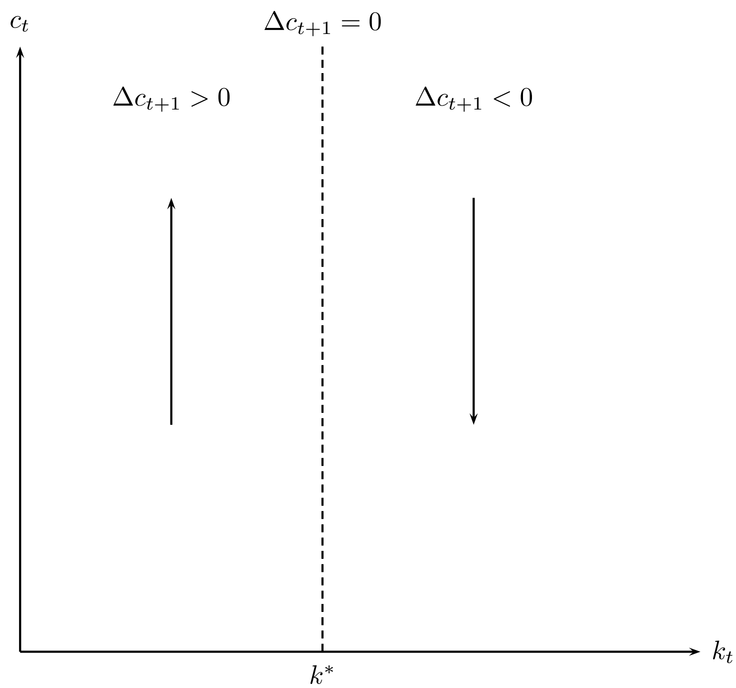


Figure 5.6: Consumption Dynamics

then it must be that $c_t = c_{t+1}$ and $\Delta c_{t+1} = 0$. Graphing this, we have a vertical line at k^* in figure 5.6. Regardless of the level of consumption, if $k_t = k^*$ consumption is not changing.

Now what if capital is *not* at the steady state? Consider $k_t < k^*$. If the capital stock is smaller than k^*, the marginal product of capital is relatively large, and so the term $\beta\left[f'(k_{t+1}) + 1 - \delta\right]$ is greater than one. If this ratio is greater than one, then the present is "expensive" relative to the future, and so $U'(c_t) > U'(c_{t+1})$. Thus $c_t < c_{t+1}$ and $\Delta c_{t+1} > 0$. So for any level of $k_t < k^*$ it must be that consumption is growing over time.

A similar analysis reveals that for any values of $k_t > k^*$, it must be the case that $\Delta c_{t+1} < 0$, as the present is "cheap" relative to the future. We cannot say exactly what Δc_{t+1} is, but we do know that it is negative in this region.

Turning to the capital stock, we can perform a similar analysis. Let's begin at steady state, when $\Delta k_{t+1} = 0$. This tells us that $c_t = f(k_t) - \delta k_t$. Graphing $f(k_t)$ and δk_t together as in the Solow diagram, we see that consumption (the difference between these two curves), must be hump-shaped, with a maximum at the Golden Rule level of $f'(k_t) = \delta$.

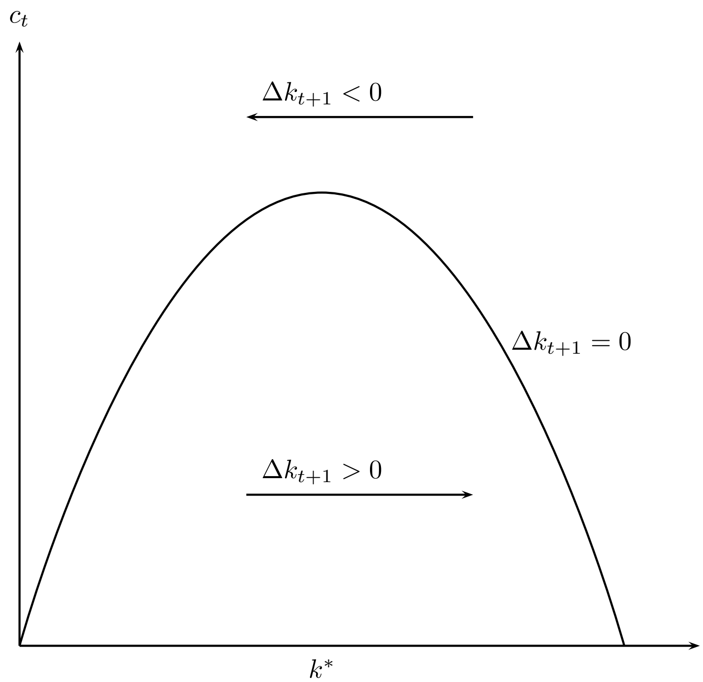


Figure 5.7: Capital Dynamics

Mathematically, this follows from examining the derivatives of c_t with respect to k_t.

$$\frac{\partial c_t}{\partial k_t} = f'(k_t) - \delta \tag{5.95}$$

$$\frac{\partial^2 c_t}{\partial k_t^2} = f''(k_t) < 0 \tag{5.96}$$

which indicates that the relationship between consumption and capital is convex with a maximum at the Golden Rule. Graphing this in figure 5.7, the solid line represents all the points at which $\Delta k_{t+1} = 0$. If the economy finds itself with a combination of consumption and capital that put it on this line, the capital stock will not be changing.

Off of this line, what is happening? Below the curve, $c_t < f(k_t) - \delta k_t$, and so $\Delta k_{t+1} > 0$, or the capital stock is growing. This should make intuitive sense. If the economy is consuming less than the steady state amount, there must be extra savings available to increase the capital stock. A similar analysis suggests that everywhere that c_t lies above the curve, $\Delta k_{t+1} < 0$.

Now, if we combine diagrams 5.6 and 5.7, we have figure 5.8. This shows the phase diagram of the Ramsey model, indicating the direction of change in both consumption and the capital stock in each of the four distinct regions defined by the two steady state curves.

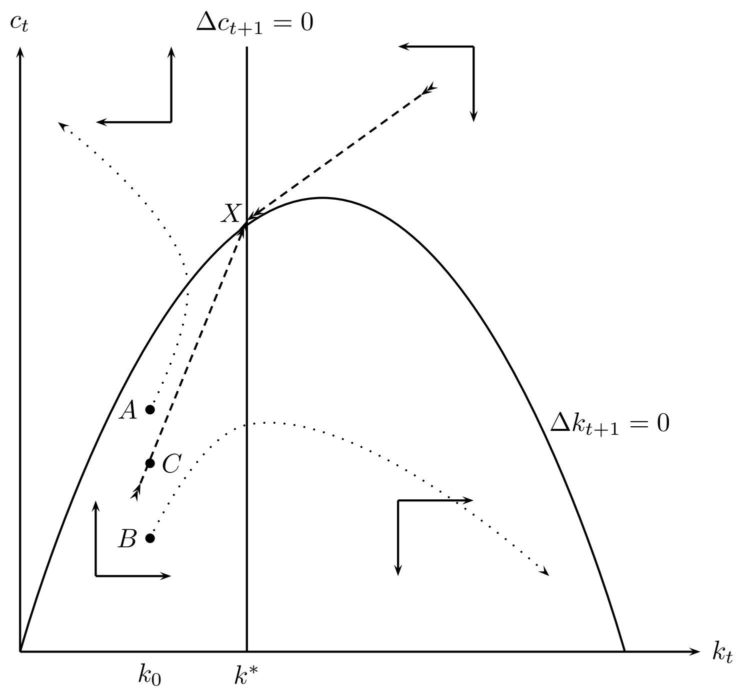


Figure 5.8: The Ramsey Diagram

Note: The Ramsey diagram combines the dynamics of consumption and capital. The point X is the steady state. The arrows indicate the dynamics governing consumption and capital in each of the four regions. In only the southwest and northeast regions are the dynamics pointing towards the steady state. In those regions, there is only one path of consumption and capital, indicated by the dashed lines, that leads to the steady state in the long run. Given an initial stock of k_0, only the consumption choice corresponding to C puts the economy on the path to the steady state. A and B violate the transversality condition and are non-optimal, respectively.

The intersection of the curves at point X is the steady state, the point at which neither consumption nor capital changes. This point is stable, in that once we reach it, we will not leave this point. But we are not necessarily at the steady state, and the phase diagram allows us to analyze how the economy will act outside of the steady state.

Consider an economy with the capital stock of k_0, less than k^*. The economy cannot simply jump to k^*, they must save over time to acquire the steady state capital stock. The decision facing the economy (and built in implicitly in our optimization problem) is to determine what the right level of consumption is in period zero that will put the economy *on the trajectory* towards the steady state.

If the economy selects a very large consumption amount (point A), it will put itself in the northwest quadrant, and the dynamics in this region tell us that the capital stock will start falling and consumption will start rising. This solution is unsustainable because it implies infinite consumption and a zero capital stock in the long run, violating the transversality condition.

On the other hand, if the economy chooses a very low level of consumption (so that savings are large), it will begin to acquire capital and consumption will rise, but eventually the economy will spill over into the southeast quadrant of the diagram and consumption begins to fall while the capital stock accumulates to enormous proportions. This cannot be optimal, because consumption goes to zero.

There is one point, C, that if we select this initial level of consumption, will allow for the capital stock to grow as well as consumption. Over time, the economy will approach the steady state point of X and will reach a consumption level that can be sustained indefinitely.

This path is called the “stable arm”, and a similar analysis yields a similar path for the northeast quadrant in which we begin with more capital than in the steady state.

The point of the stable arms is that there is a single optimal consumption level for any given capital stock. Only one consumption level leaves the economy in a position to reach the steady state and sustain consumption into infinity. We can use this phase diagram to understand how the economy responds to shocks and changes in technology, but prior to addressing those issues we will establish that the same solutions we derived here can be obtained by a decentralized economy in which individuals and firms are acting independently.

5.5.2 The Decentralized Solution

In the previous section we assumed that a single optimizing agent made the decisions for the economy, internalizing the effect of consumption decisions on the accumulation of the capital stock. Now we separate the decision-making and consider a) consumers, who are trying to optimize their lifetime utility by making savings decisions and b) firms, who are trying to optimize their profits by employing factors of production.

To fit this all together, we will need to introduce a financial market that links the savings of individuals with the capital requirements of firms.

Individuals

Here we have nothing more than the infinitely-lived consumption problem of the previous chapter. Individuals maximize utility,

$$V_0 = \sum_{t=0}^{\infty} \beta^t U(c_t) \tag{5.97}$$

subject to the dynamic budget constraint

$$a_{t+1} = (1 + r_t)a_t + w_t + x_t - c_t \tag{5.98}$$

where a_t are the assets of individuals, w_t is labor income, x_t is additional income (dividends, etc.), and c_t is their consumption. The value r_t is the interest rate that individuals can earn in the financial market in period t. The individual takes the time path of r_t, w_t, and x_t as given.

We know that the solution for the individuals is to have their consumption be dictated by the Euler equation, which says that

$$\frac{U'(c_t)}{U'(c_{t+1})} = \beta(1 + r_{t+1}). \tag{5.99}$$

This gives us our first piece of the solution.

Firms

As discussed in the chapter on the Solow model, we consider a perfectly competitive economy. For simplicity, we assume that all firms exist for only one period, so they do not have long-run financing considerations.

Firms require both capital and labor to operate. To obtain capital for use in period t, a firm borrows money from the financial market, paying a rate R_t for those funds. The firm's demand for these funds depends on it's profit maximization decision. Profits in period t will be

$$\Pi_t = F(K_t, N_t) - w_t N_t - R_t K_t \tag{5.100}$$

where w_t is the wage rate the firm takes as given, and the return R_t is their net cost of capital. Maximization yields that

$$F_N(K_t, N_t) = w_t \tag{5.101}$$

$$F_K(K_t, N_t) = R_t. \tag{5.102}$$

To cast these results in per capita terms so that they align with our individual problem, consider writing the firms profits as

$$\Pi_t = N_t \left[f(k_t) - w_t - R_t k_t\right] \tag{5.103}$$

and maximizing over N_t yields

$$[f(k_t) - w_t - R_t k_t] + N_t \left[\frac{-f'(k_t)k_t}{N_t} + R_t k_t \frac{1}{N_t}\right] = 0 \tag{5.104}$$

$$[f(k_t) - w_t - R_t k_t] + [-f'(k_t)k_t + R_t k_t] = 0 \tag{5.105}$$

$$w_t = f(k_t) - f'(k_t)k_t \tag{5.106}$$

and the maximization over capital yields the simple

$$f'(k_t) = R_t. \tag{5.107}$$

As firms are perfectly competitive and the production function is constant returns to scale, it must be that

$$F(K_t, N_t) = F_N(K_t, N_t)N_t + F_K(K_t, N_t)K_t \tag{5.108}$$

and given the firms profit-maximizing decisions, it must be the case that profits (x_t) are equal to zero.

Market Clearing and Equilibrium

The financial market we have in mind here works with zero frictions. This market received deposits of a_t from N_t individuals, and must repay those individuals an amount $N_t(1 + r_t)a_t$. The financial market loaned an amount K_t to firms to use in production, and from them it receives an amount $(1 + R_t - \delta)K_t$, where note that we have incorporated the fact that some of the capital depreciates in use, and so the total amount returned to the financial market is smaller by this amount.

To clear the market it must be that these amounts are equal

$$N_t(1 + r_t)a_t = (1 + R_t - \delta)K_t \tag{5.109}$$

and as this is a closed economy, the total assets in the economy must be equal to the total capital stock, or $N_t a_t = K_t$. What this implies is that

$$(1 + r_t) = (1 + R_t - \delta) \tag{5.110}$$

$$r_t = R_t - \delta. \tag{5.111}$$

From the firm maximization problem, we know that they will employ capital until $R_t = f'(k_t)$, so that in the end we have

$$r_t = f'(k_t) - \delta. \tag{5.112}$$

Using this last result in the Euler equation for individuals yields

$$\frac{U'(c_t)}{U'(c_{t+1})} = \beta(1 + f'(k_{t+1}) - \delta) \tag{5.113}$$

which is identical to the Euler equation derived for the centralized economy.

Now consider the individuals dynamic budget constraint, $a_{t+1} = (1 + r_t)a_t + w_t + x_t - c_t$. The value x_t represents other income to the individual, but as all of output is paid to either labor or capital (i.e. profits are zero), there is no other income to pay out, and $x_t = 0$. We know from market clearing that $a_{t+1} = k_{t+1}$ and $a_t = k_t$. Finally, we know from (5.112) what the return on assets is, and we know from (5.106) how to describe the wage rate.

Putting all this information together in the dynamic budget constraint, we have

$$k_{t+1} = (1 + f'(k_t) - \delta)k_t + f(k_t) - f'(k_t)k_t - c_t \tag{5.114}$$

which reduces to

$$k_{t+1} = f(k_t) + (1-\delta)k_t - c_t \tag{5.115}$$

and this is identical to the original accumulation function for capital we described in the centralized model.

As the equations describing the motion of consumption and capital over time are identical in the de-centralized and centralized economy, they must have the exact same solution for the steady state and the dynamics. Note that the centralized solution was an optimal solution, meaning that all costs and benefits were internalized by the optimizing agent. What we've seen here is that a decentralized solution delivers exactly the same result. This is an example of First Welfare Theorem in action. We have complete markets in all goods (consumption in every period), and so the market solution is Pareto Optimal. This is useful because we can concentrate on working with the centralized problem, knowing that the market can (under our assumptions about how firms and markets work) deliver the same results.

5.5.3 Population, Technology, and Growth

As we did previously, we can consider how the model changes when we incorporate growth in the number of people and growth in productivity.

Population Growth

To add in population growth, we assume that the total population (or alternately, all of the households) are growing at the rate n per period. We can write the growth rate of assets per capita as

$$\frac{\Delta a_{t+1}}{a_t} = \frac{\Delta A_{t+1}}{A_t} - \frac{\Delta N_{t+1}}{N_t} \tag{5.116}$$

which we can evaluate by determining the growth rate of aggregate assets, A_t. Given the dynamic budget constraint, it must be that

$$A_{t+1} = (1+r_t)A_t + w_t N_t + x_t N_t - c_t N_t \tag{5.117}$$

so that the growth rate of assets per capita can be written as

$$\frac{\Delta a_{t+1}}{a_t} = r_t + \frac{w_t}{a_t} + \frac{x_t}{a_t} - \frac{c_t}{a_t} - n \tag{5.118}$$

which can be rearranged into the following constraint

$$a_{t+1} = (1 + r_t - n)a_t + w_t + x_t - c_t. \tag{5.119}$$

Given this budget constraint, the Ramsey model solution is not functionally different. As we did before, the financial market sets assets equal to the capital stock, $r_t = f'(k_t) - \delta$, $w_t = f(k_t) -$

$f'(k_t)k_t$, and $x_t = 0$ which leads to

$$k_{t+1} = f(k_t) + (1 - \delta - n)k_t - c_t \tag{5.120}$$

and this is exactly what we derived in the Solow model.

Solving with this dynamic constraint, we have an Euler equation

$$\frac{U'(c_t)}{U'(c_{t+1})} = \beta \left[f'(k_{t+1}) + 1 - \delta - n\right] \tag{5.121}$$

which then leads to a steady state value of k^* that solves

$$f'(k^*) = \frac{1}{\beta} + \delta + n - 1 \tag{5.122}$$

and notice that the addition of population growth has increased the marginal product of capital in steady state. Therefore, the steady state level of capital is lower the larger population growth is. This follows somewhat mechanically, as increasing the number of individuals each period makes maintaining a large capital stock per person more difficult.

Note that we did not alter the utility function in this analysis. That is, utility is the discounted sum of *per capita* consumption. An alternate version of utility would be the following,

$$V_0 = \sum_{t=0}^{\infty} \beta^t N_0 (1+n)^t U(c_t) \tag{5.123}$$

and utility is the discounted sum of total utility in the economy. For this formulation of the problem to make sense, it has to be that

$$\beta(1+n) < 1 \tag{5.124}$$

or the growth rate of population is smaller than $1/\beta - 1$, which you'll recall is the steady state value of r^*. So if we have utility over total utility of each generation, we need the population to grow more slowly than assets. If we did not, then utility (which is based on assets per person) could grow to infinity.

Technological Change

Here we can again adapt the model as we did with the Solow version, allowing the production function that incorporates technological change,

$$Y_t = F(K_t, E_t N_t) \tag{5.125}$$

where the efficiency term, E_t, grows exogenously at the rate g.

How does this alter the problem? Given the dynamic budget constraint, it must be that

$$A_{t+1} = (1 + r_t)A_t + \tilde{w}_t E_t N_t + \tilde{x}_t E_t N_t - \tilde{c}_t E_t N_t \tag{5.126}$$

where the $\tilde{\cdot}$ indicates a variable written in per-efficiency unit terms. We can evaluate the growth rate of assets per efficiency unit using this equation, knowing that $E_{t+1} = (1+g)E_t$ and $N_{t+1} = (1+n)N_t$,

$$\frac{\Delta \tilde{a}_{t+1}}{\tilde{a}_t} = r_t + \frac{\tilde{w}_t}{\tilde{a}_t} + \frac{\tilde{x}_t}{\tilde{a}_t} + \frac{\tilde{c}_t}{\tilde{a}_t} - n - g \tag{5.127}$$

and this gives us a dynamic constraint for assets per efficiency unit of

$$\tilde{a}_{t+1} = (1 + r_t - n - g)\tilde{a}_t + \tilde{w}_t + \tilde{x}_t - \tilde{c}_t. \tag{5.128}$$

The financial sector again ensures that capital and assets are equal to each other, and therefore $\tilde{k}_t = \tilde{a}_t$. Applying this, as well as noting that $\tilde{w}_t = f(\tilde{k}_t) - f'(\tilde{k}_t)\tilde{k}_t$, $\tilde{x}_t = 0$, and $r_t = f'(\tilde{k}_t) - \delta$, we have that

$$\tilde{k}_{t+1} = (1 - \delta - n - g)\tilde{k}_t + f(\tilde{k}_t) - \tilde{c}_t \tag{5.129}$$

which is identical to the accumulation equation for the Solow model with technological change.

Utility maximization takes place as before, but note that people only care about consumption per person, not consumption per efficiency unit. Therefore, you'll see in the following Lagrangian that we have modified the term inside the felicity function to be $\tilde{c}_t E_t$, which is simply consumption per person.

$$\mathcal{L} = \sum_{t=0}^{\infty} \left(\beta^t U(\tilde{c}_t E_t) + \lambda_t \left[(1 - \delta - n - g)\tilde{k}_t + f(\tilde{k}_t) - \tilde{c}_t - \tilde{k}_{t+1} \right] \right) \tag{5.130}$$

and the first order conditions for two adjacent periods t and $t+1$ are

$$\frac{\partial \mathcal{L}}{\partial \tilde{c}_t} = \beta^t U'(\tilde{c}_t E_t)E_t - \lambda_t = 0 \tag{5.131}$$

$$\frac{\partial \mathcal{L}}{\partial \tilde{c}_{t+1}} = \beta^{t+1} U'(\tilde{c}_{t+1} E_{t+1})E_{t+1} - \lambda_{t+1} = 0 \tag{5.132}$$

$$\frac{\partial \mathcal{L}}{\partial \tilde{k}_{t+1}} = \lambda_{t+1} \left[f'(\tilde{k}_{t+1}) + (1 - \delta - n - g) \right] - \lambda_t = 0. \tag{5.133}$$

Solving these all together to eliminate the multipliers, and noting that $E_{t+1} = (1+g)E_t$ we have an Euler equation of

$$\frac{U'(c_t)}{U'(c_{t+1})} = \beta(1+g) \left[f'(\tilde{k}_{t+1}) + (1 - \delta - n - g) \right]. \tag{5.134}$$

This tells us that the optimal growth rate of consumption now depends not only on the discount rate, β, and the interest rate $f'(\tilde{k}_{t+1}) + (1 - \delta - n - g)$, but also on the growth rate of technology itself. Technological growth gives the economy a built-in increase in consumption over time, and this lowers the marginal utility of consumption in the future.

In the steady state, just as in the Solow model, it will have to be $\tilde{k}$ and $\tilde{c}$ that are constant. This tells us that consumption and capital per person must grow at the rate g in steady state.

Without functional forms for utility and production, it isn't possible to be more explicit about the solution. We cannot even draw much information from the Euler equation, as this involves the ratio of marginal utilities of consumption per person, not per efficiency unit. So it is not the case that $U'(c_t) = U'(c_{t+1})$. We do know that consumption is growing, though, so that $U'(c_t) > U'(c_{t+1})$. This implies that in steady state it must be the case that

$$f'(\tilde{k}^*) > \frac{1}{\beta(1+g)} - 1 + \delta + n + g \tag{5.135}$$

or the marginal product of capital has to be larger than the value that would make consumption constant over time. This means that $\tilde{k}^*$ has to be lower. So with technological change, the marginal product of capital will be relatively high.

An Explicit Solution with Technological Change

Using a standard Cobb-Douglas production function, so that $\tilde{y}_t = \tilde{k}_t^{\alpha}$, and a CRRA utility function, $U(c_t) = c_t^{1-\sigma}/(1-\sigma)$, we can provide more details on the solution to the Ramsey model.

The Euler equation can be written as

$$\left(\frac{c_{t+1}}{c_t}\right)^{\sigma} = \beta(1+g)\left[\alpha\tilde{k}_{t+1}^{\alpha-1} + (1-\delta-n-g)\right] \tag{5.136}$$

and in steady state we know that $c_{t+1} = (1+g)c_t$, so this can be solved for

$$\tilde{k}^* = \left(\frac{\alpha}{\frac{(1+g)^{\sigma-1}}{\beta} - 1 + \delta + n + g}\right)^{1/(1-\alpha)} \tag{5.137}$$

which is complex, but similar in form to the original solution without technological progress. The difference is that the future is discounted now by not only β, but also by the growth rate of technology. The larger this growth rate, the more the economy ignores the future and consumes today. Why? Because the growth of technology allows the economy to enjoy large consumption in the future without having to save for it.

5.6 Fluctuations and Savings

Having developed a whole structure of the Ramsey model to endogenize savings, what do we find? First, that fundamentally savings are still exogenous - now they depend on β, as opposed to s, as in the Solow model. The Ramsey model is more nuanced, however, in explaining that savings rates (specifically s_t/y_t) do respond to the rate of return on savings. This means that savings rates will change dynamically as the economy accumulates capital. There are some homework problems

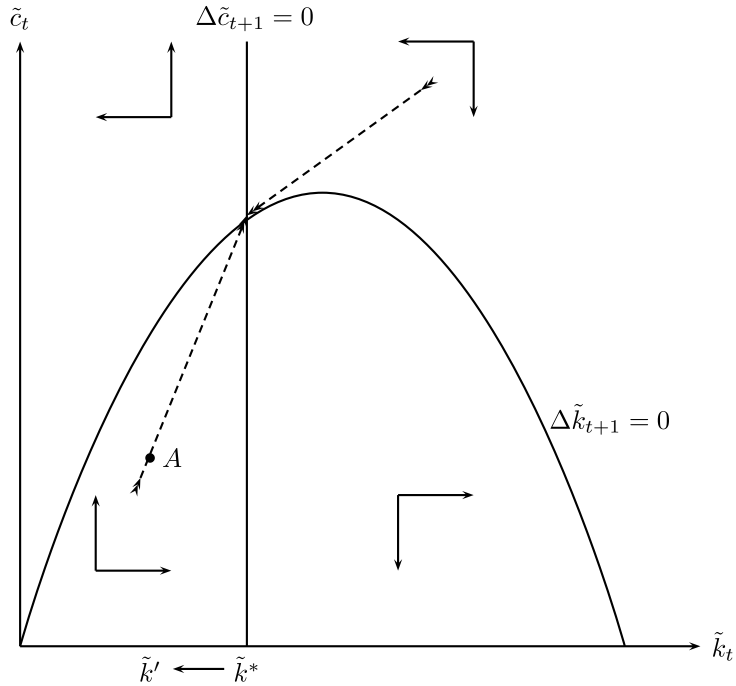


Figure 5.9: A Positive Shock to E_t

Note: Beginning in steady state, a positive increase in E_t lowers the capital per efficiency unit from $\tilde{k}^*$ to $\tilde{k}'$. The optimal response to this is for consumption per efficiency unit to fall to point A. From there, the economy will evolve towards the steady state.

asking you to think about how the savings rate responds to $1 + r$, and what that means for how savings rates respond to economic growth.

In terms of fluctuations, does this model suggest anything different than what we've already seen in the Solow model? Yes and no. Yes, in response to a productivity shock, the response of the economy is more nuanced in the sense that savings will respond endogenously. No, in the sense that the general pattern of response to a productivity shock will look awfully similar in the two types of economies. That is, following a positive productivity shock both models predict that economic growth will increase temporarily until the economy returns to steady state.

One additional feature that the Ramsey model can handle that the Solow cannot is the distinction between anticipated and unanticipated changes. We can start by thinking about how the Ramsey model reacts to an unanticipated increase in efficiency E_t.

Consider a *permanent* upward shock to E_t. This immediately lowers the value of $\tilde{k}_t$. But note that the steady state values of $\tilde{k}^*$ and $\tilde{c}^*$ are unchanged, as they only depend on the growth rate of technology, g, not the level.

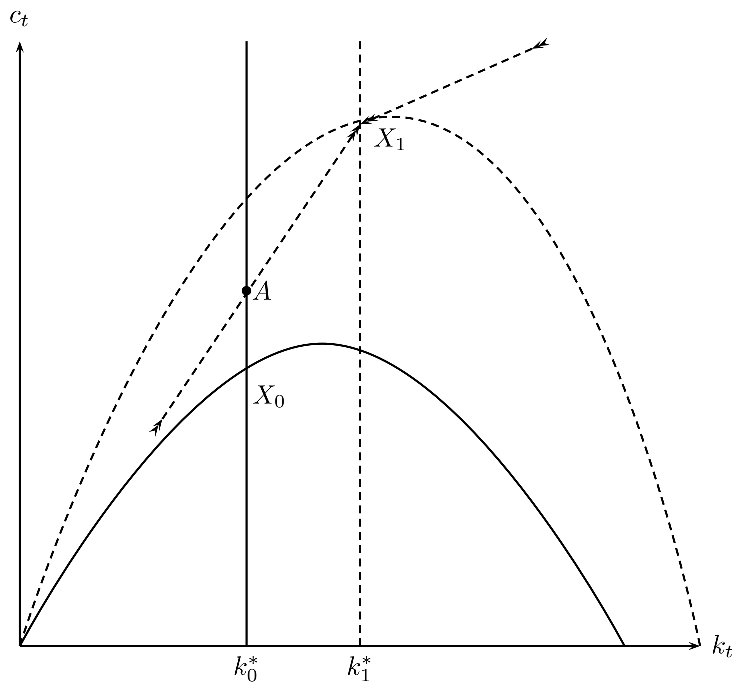


Figure 5.10: A Positive Shock to E_t, in per capita terms

Note: Beginning at steady state X_0, a positive increase in E_t has two effects. First, it increases capital per person in the steady state. Second, it increases the consumption possible at any given level of capital. These two changes create a new steady state at the point X_1. The optimal response to this shock, given that the capital stock is given by k_0^* at the time of the shock, is to increase consumption to the point A so that the economy will be on the stable arm to the new steady state levels of consumption.

A drop in $\tilde{k}_t$ implies an increase in the marginal product of capital, and hence an increase in the rate of return to savings. This induces a drop in $\tilde{c}_t$, and from that point forward the economy will approach the steady state. This can be seen in figure 5.9, which shows how the drop in $\tilde{k}_t$ due to the increase in E_t is handled by the economy. Essentially, we simply move down the stable arm, and then follow it back to steady state.

This is concise, but perhaps not very informative on what happens to per capita variables. To see how these react, we can consider a Ramsey diagram of this economy done in per capita terms rather than per efficiency unit terms. We can think of this diagram as showing us how the economy responds to shocks *relative to a baseline*.

Figure 5.10 shows how the economy responds to the shock to E_t in per capita terms. Here, the increased productivity has two effects. First, it raises the marginal product of capital, and therefore it takes a larger amount of capital (k_1^*) to set $\Delta c_{t+1} = 0$ from the Euler equation. This shifts the

$\Delta c_{t+1} = 0$ line to the right. Second, an increase in productivity increases consumption for any given amount of capital, and hence the $\Delta k_{t+1} = 0$ curve shifts upwards as well.

This results in a new steady state, and a new set of stable arms. From the existing capital stock, k_0^*, the optimal response is to increase consumption immediately, and then let both consumption and capital grow over time towards the new steady state.

Note that this is all relative to the baseline, or the per-efficiency unit steady state, in which consumption and capital per person are growing at the rate g every period. So what figure 5.10 indicates is that consumption growth becomes higher than g for a period of time following the shock, which conforms with the results of figure 5.9 that suggest $\tilde{c}_t$ grows at a rate greater than zero for a period of time before reaching the new steady state.

The dynamics of consumption, capital, and the interest rate are plotted in 5.11. Both consumption and capital are growing at the rate g prior to the shock. Once E_t increases, consumption jumps discretely and continues to grow more quickly than g for a while before it asymptotes to a growth rate of g again as the economy approaches the new steady state. In the long run, consumption growth is still g, but at a higher level than before.

Capital, in contrast, cannot jump. Once the increase in E_t hits, though, and increases the marginal return to capital, accumulation accelerates, and k_t grows at a rate higher than g until the economy reaches steady state again. The time path of r_{t+1} is consistent with this pattern; the initial increase induces the relatively fast growth in the capital stock and as capital accumulates the rate of return falls back towards its original level.

Compared to this unanticipated jump, consider what happens if there is an *anticipated* jump in E_t? That is, if people are informed today that there will be a distinct jump in E_t at some point in the future. Would the economy just stay at the original steady state, as if nothing had changed, until the point that E_t actually jumps? At that point the economy would find itself at point A in figure 5.9, and would grow relatively fast until it was back at steady state.

Why would individuals choose to do this? Their per capita consumption would be constant until the time of the shock, and then it would jump up discretely. That is fine, but we do not think people like distinct jumps in consumption. Knowing that E_t is going to jump, why not increase consumption a little now? If individuals did so, their consumption would jump upon learning the information about E_t, but only a little. They'll spread the anticipated new income out over their whole infinite lifetime.

In terms of the Ramsey diagram, figure 5.12 shows you the what is happening. The jump in consumption is to point A, which is *below* where the new stable arm will be in the future. As such, the dynamics at point A are still dictated by the old steady state, which tells us that k falls and c rises. This is the path from point A to point B in the figure. So individuals respond to the new information about productivity by instantly adjusting their consumption behavior, but only slightly. However, this slight adjustment is sufficient to get them to the new steady state once E_t changes.

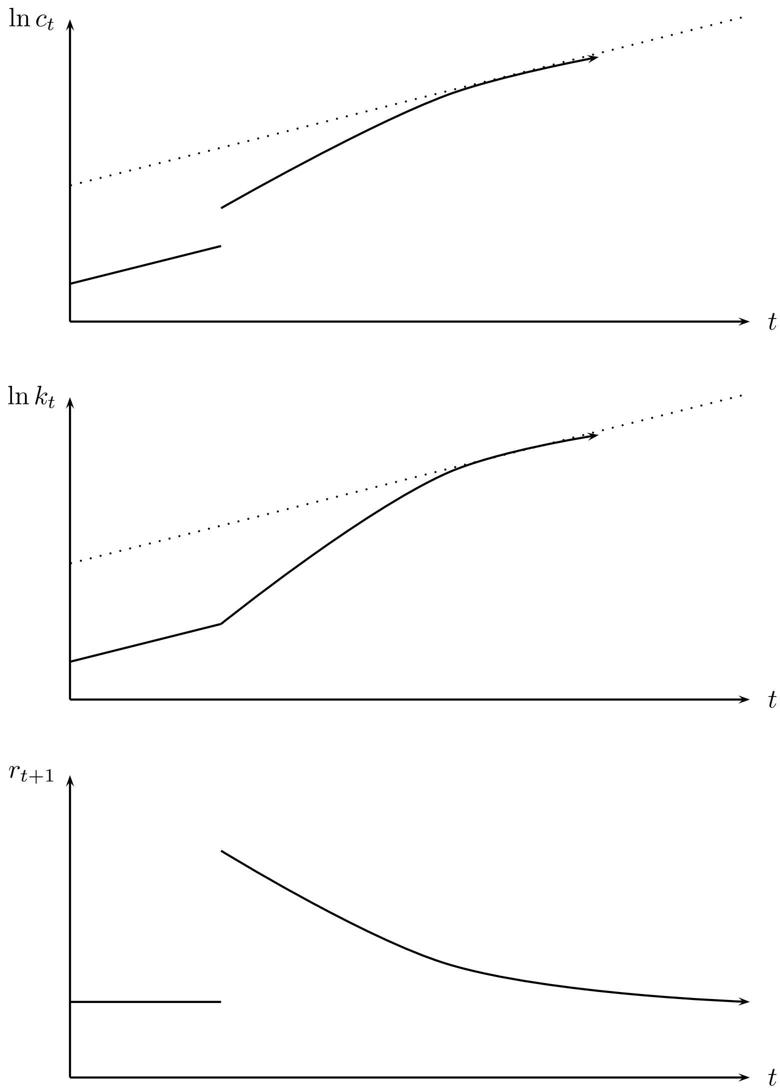


Figure 5.11: The Dynamic Response to an Unanticipated Shock in E_t

Note: The three graphs show how log consumption per capita, log capital per capita, and the interest rate respond to a positive shock in E_t. Consumption jumps discretely due to the unanticipated shock, grows at a rate faster than g, and eventually asymptotes back to a growth rate of g, although at a higher level. Capital does not jump, but begins growing at a rate faster than g for a period of time before returning to growth at g. Finally, the interest rate discretely jumps up in response, and then returns over time to the original level.

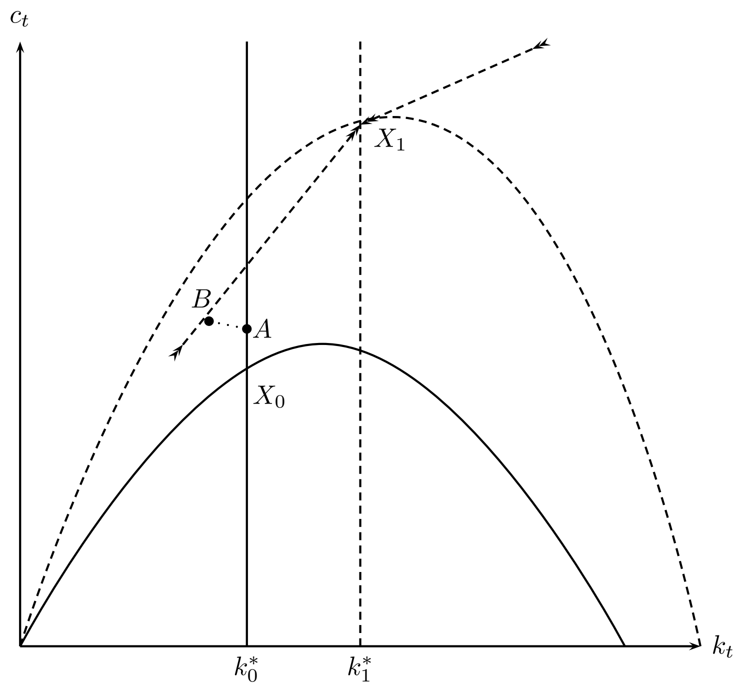


Figure 5.12: An Anticipated Positive Shock to E_t, in per capita terms

Note: Beginning at steady state X_0, the anticipated positive increase in E_t will have two effects. First, it increases capital per person in the steady state. Second, it increases the consumption possible at any given level of capital. These two changes create a new steady state at the point X_1. The optimal response to this anticipated shock, given that the capital stock is given by k_0^* at the time of the shock, is to increase consumption to the point A so that the economy will increase consumption and decrease k in the time between learning about the jump in E_t, and the actual jump. The jump in c is made so that when E_t actually jumps, individuals will be right on the new stable arm, at point B.

Following along the paths of c, k, and therefore r, we have figure 5.13. Here, we see that the path for consumption looks quite similar to before, but note two differences. First, the jump in consumption is at the announcement, not at the actual change in E_t. Second, the actual jump in consumption is smaller.

Similarly, the jump in r_{t+1} is at the actual change in E_t, not at the announcement. The interest rate rises gradually after the announcement, in line with the gradual decline in capital. Once efficiency actually rises, capital begins to rise again and interest rates fall towards steady state.

The important implication is that simple news about productivity changes is sufficient to cause fluctuations in output. Announcements regarding productivity will cause individuals to adjust behavior immediately, creating the possibilities that fluctuations could be driven in part by anticipated changes in the future.

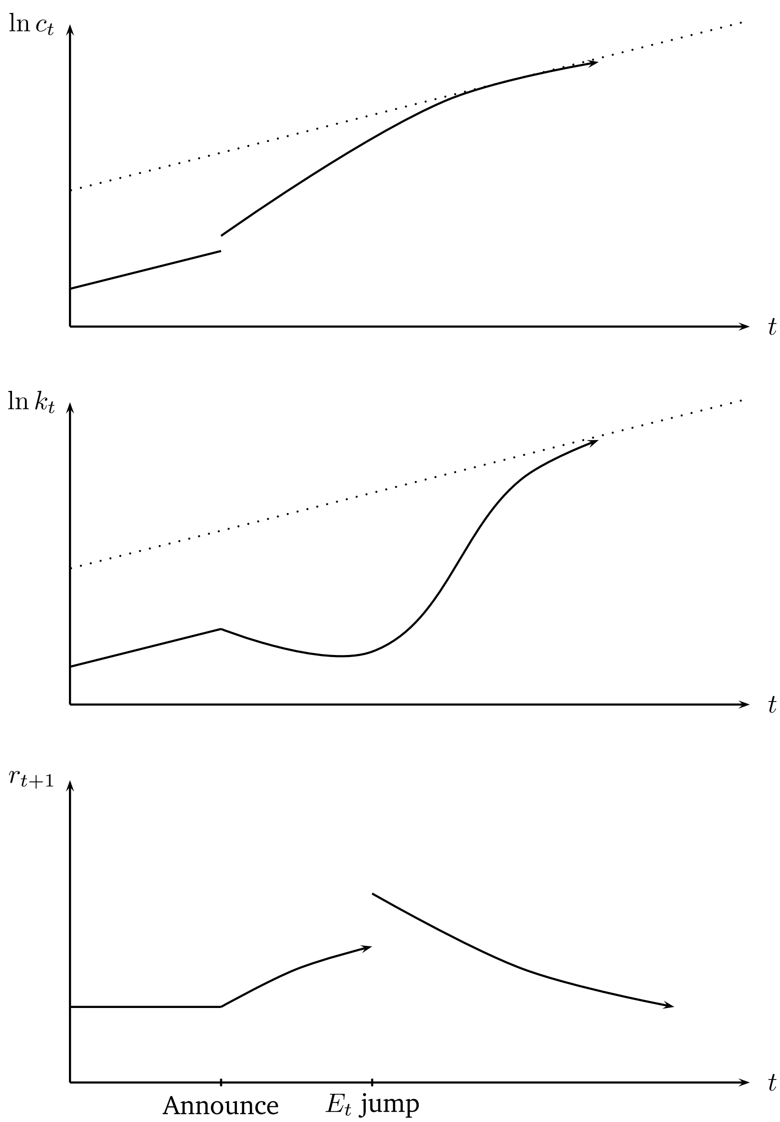


Figure 5.13: The Dynamic Response to an Anticipated Shock in E_t

Note: The three graphs show how log consumption per capita, log capital per capita, and the interest rate respond to a positive shock in E_t. Consumption jumps discretely due to the unanticipated shock, grows at a rate faster than g, and eventually asymptotes back to a growth rate of g, although at a higher level. Capital does not jump, but begins growing at a rate faster than g for a period of time before returning to growth at g. Finally, the interest rate discretely jumps up in response, and then returns over time to the original level.

However, note there is nothing about having endogenous savings that changes the conclusions we made before regarding the role of nominal demand in fluctuations. Even though the supply of capital is endogenous, there is still perfect competition and perfect price flexibility, so the economy will always be operating at full capacity. So a change in nominal demand will change nominal quantities, such as prices, but not real things like output or employment. We will still need to introduce some kind of market imperfection or nominal rigidity to have any impact of nominal demand on output.

Chapter 6

Population and Growth

At this point, we have some methods of describing where steady state growth comes from. This is useful for describing current trend growth, perhaps, but fails to account for the longer-run experience of economic development.

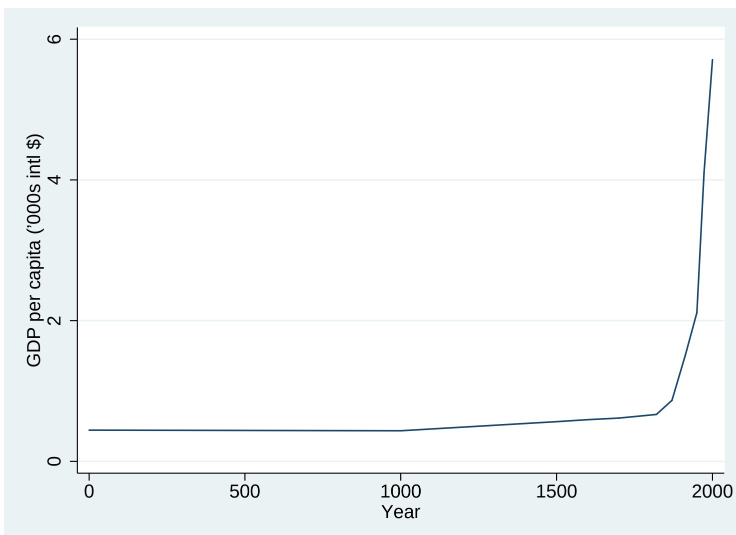


Figure 6.1: Output per capita over the very long run

Source: Calculations using data on GDP and population from Maddison (2001)

Between the year 1 and the year 1000, growth in output per capita was essentially zero. Between 1000 and 1820 growth in output per capita was only 0.19% per year. Following 1820, world growth rates started to rise to about 0.5% per year and increasing to close to 3% per year by the 1960's.

Growth has not been a constant feature of economic history, and so our models are useful for modern economies only. In addition, many developing countries have very low growth rates, when our existing models would predict very high growth rates (due to steady state convergence). How do we capture this transition from stagnation to sustained growth?

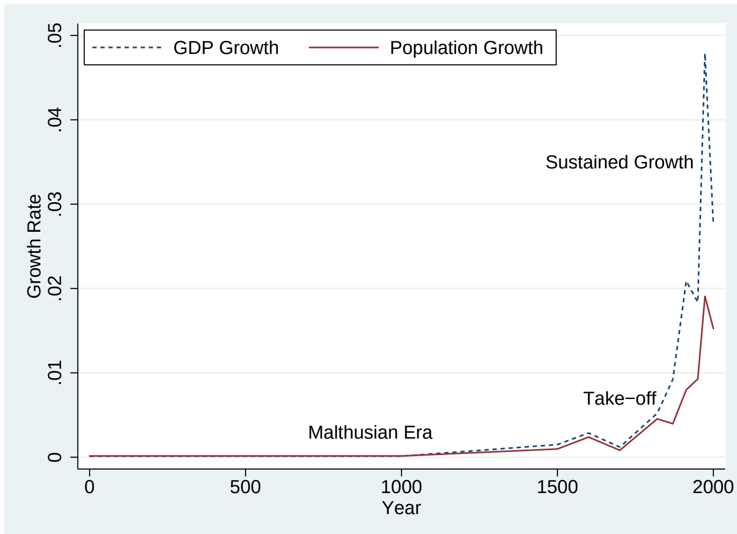


Figure 6.2: Growth Rates of Output and Population over the Long Run

Source: Data on both GDP and population are from Maddison (2001)

We're going to start with a simple Malthusian model, useful for understanding the period of time prior to the onset of sustained growth. The key elements of this are a fixed factor of production (non-accumulable) and a positive response of population growth to income per capita. After that, we'll see how to get a little more sophisticated with respect to the fertility decision, allowing for quality and quantity.

6.1 Malthusian Economics

To begin with, we will incorporate another aspect of behavior into the model explicitly: fertility. The growth rate of the population will no longer be exogenous, but will depend on economic conditions. We'll model this in a Malthusian sense, named after Thomas Malthus, who proposed the idea that fertility responds positively to income. This response will yield a model that has a steady state level of income per person, as any increase in income over this value generates more people and drives down incomes.

A key feature of the Malthusian model is a production function that has a fixed factor of production - usually land, denoted X. Equivalently, we could have included physical capital, but assumed that the production function had decreasing returns to scale. Let output be

$$Y_t = A_t X^{\beta} L_t^{1-\beta} \tag{6.1}$$

and therefore output per capita is

$$y_t = A_t X^{\beta} L_t^{-\beta} \tag{6.2}$$

so that standards of living are positively related to land and negatively related to the size of the population.

From an individuals perspective, the best way to approach optimal fertility is in an OLG framework. Let utility of individuals be

$$U = \ln c_t + \gamma \ln n_t \tag{6.3}$$

where c_t is consumption and n_t is the number of children they have in the second period of life. In the first period of life, each individual is just a kid and makes no choices and consumes nothing.

The budget constraint for an individual involves the cost of children. Each child is assumed to take up an amount of resources, θ, so that the constraint is

$$y_t = c_t + \theta n_t \tag{6.4}$$

which says that increasing the number of children I have reduces the amount of consumption I can do. Income for the individual, y_t, is the same as total output per capita. This is like saying that individuals earn both their wage and the rental income on land. Land is passed down from one generation to the next automatically - there is no accumulation process to worry about.

The optimal response for an individual to to choose

$$n_t = \frac{\gamma}{1-\gamma} \frac{y_t}{\theta} \tag{6.5}$$

and fertility is increasing in output. Now, going back to the per-capita income, we can plug this into the optimal fertility expression and find that

$$n_t = \frac{\gamma}{1+\gamma} \frac{A_t X^{\beta}}{\theta L_t^{\beta}} \tag{6.6}$$

and notice now that fertility goes down when the population goes up.

Given this expression, and the fact that $L_{t+1} = n_t L_t$, the dynamics of this economy will automatically yield a fixed steady state population of

$$L_{ss} = X \left(\frac{\gamma}{1+\gamma} \frac{A_t}{\theta} \right)^{1/\beta} \tag{6.7}$$

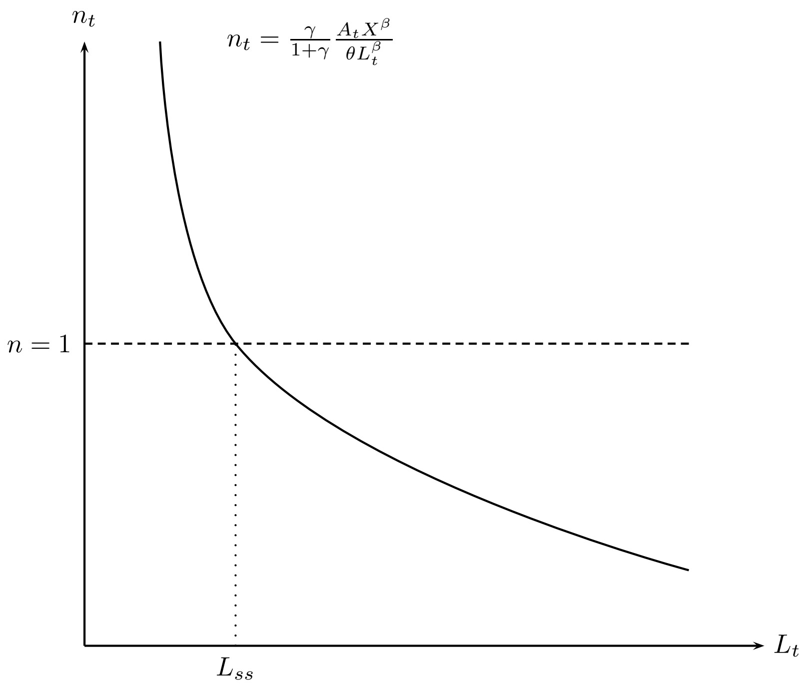


Figure 6.3: The Malthusian Steady State

Note: Optimal fertility is declining in the size of the population as this lowers income per capita. At levels of population less than L_{ss}*, fertility is greater than one, and the population size is increasing. At population greater than* L_{ss}*, fertility is below one and the population shrinks.*

which can be seen more clearly in figure 6.3. When population is low, output per capita is large and so is fertility - population grows. If population is large, output per capita is low and so people have few children - population falls. The size of the steady state population gets larger with more X, or with higher productivity A_t.

What happens to output per capita? Using the steady state population level, output per capita becomes

$$y_{ss} = \theta\frac{1+\gamma}{\gamma} \tag{6.8}$$

and notice that this does not depend at all on productivity or land. In other words, output per capita is dictated completely by the preference parameters and the cost of children. The more valuable children are (γ is larger), the lower is output per capita, as people tend to have lots of children. The more costly children are (θ is larger), the higher is income per capita, as fertility is lower.

Productivity doesn't matter here because any increase in productivity just increases fertility and, *because we have a fixed factor of production*, this lowers income per person again. The crucial point

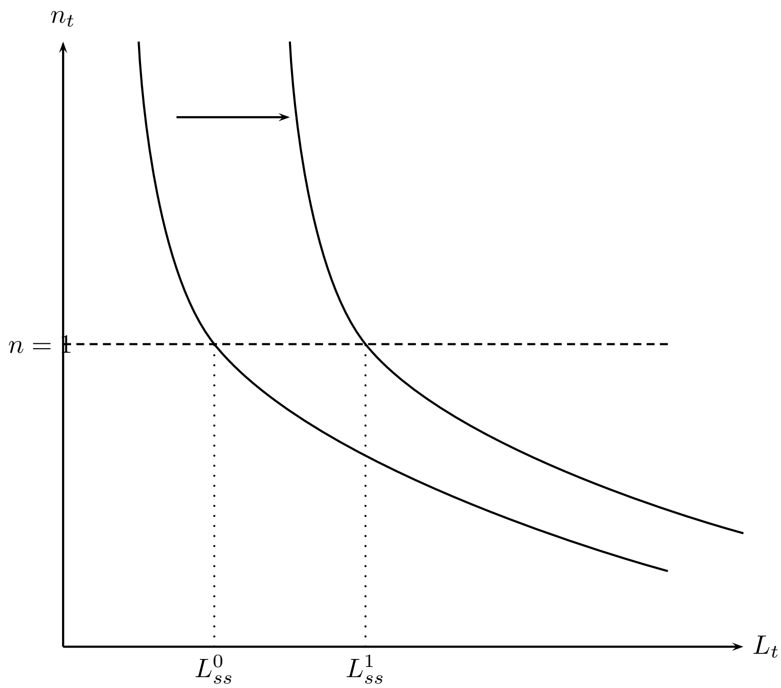


Figure 6.4: A Malthusian Productivity Increase

Note: Following an increase in A, the entire fertility curve shifts to the right. Immediately, fertility increases and the population grows to L^1_{ss}. The initial effect on income per capita is positive, but this eventually dissipates.

is that there is a fixed factor. The Solow model delivered a constant output per capita - but one that increased with productivity. The Malthusian model gives us a constant output per capita, but cannot even match the Solow model because there is nothing here to accumulate to keep up with population growth.

6.2 Quantity and Quality

To introduce fertility in a manner that will matter for the dynamics of the aggregate economy, we need children to act in part like any other inter-temporal asset. That is, having children must provide some utility in the future that we otherwise could not have.

We could treat kids as another kind of consumption good that matters to us just today. Then the relative price of kids compared to other consumption would determine the fertility rate, and we could sink that fertility rate into our OLG model above and see how the economy responds. What we would like to do instead is to incorporate the fertility decision as an explicit inter-temporal choice - that is, I want to take into account how my decisions about children today influence the economy tomorrow.

There are different possible ways to do this. One could be to make my children's utility explicitly part of my utility (I'm altruistic). Because my kids will care about my grandkids, then I care about my grandkids too (because it affects my kids utility). Keep this chain going, and you have that my fertility decision depends on every future generation off to infinity. Under this type of scenario, the problem reduces to a standard infinitely-lived individual problem. I make a decision about consumption and fertility today that I know will lead to certain decisions by my kids tomorrow, and so forth.

We'll take a slightly different approach that doesn't end up resolving itself into so neat of a package. We'll assume that parents care about both the quantity and the quality (to be defined more precisely) of kids. The important difference is that the quality is solely a function of things I can control today (e.g. schooling and bequests to kids), while their quality does not depend at all on what they do with their children. So I won't end up with this recursive structure that leads to an infinitely-lived problem.

More specifically, let's define utility for a parent as follows:

$$V = \ln c_t + \beta \ln n_t y_{t+1} \tag{6.9}$$

where c_t is parent consumption in period t, n_t is the number of children they have, and y_{t+1} is the income of their children in the next period. Here "quality" is defined solely by income. One can be fancier about how quality is defined, but for this all to work quality must be something that depends in part on some kind of economic sacrifice by the parent.

We've got a very specific form for utility here. It gives us that the marginal utility of children is unaffected by their quality. That is, raising quality does not make me want more (or less) kids. This is identical to the effect that using log utility has on the interest rate – income and substitution effects offset. With a more general utility function kids and quality can be either complements or substitutes, and exactly how they respond to each other will depend on how we specify utility. For now we stick with this very simple specification.

Here we'll also be very straightforward about life-spans as well. Each generation lives for one period, and one period only. In their one period of life a generation earns income y_t (which consists of all the returns to both labor and capital), consumes, has kids, and then dies. Quite the life. What is the budget constraint for this boring individual?

$$y_t = c_t + n_t(\phi + b_t) \tag{6.10}$$

where ϕ is the cost of having a child (regardless of their quality), and b_t is a bequest that I leave to each child. This b_t is the way in which I can give my child "quality" - you could think of it as schooling, or simply a lump sum of capital handed down to the child.

Finally, where does income come from? We'll keep it very simple again and say

$$y_t = k_t^\alpha = b_{t-1}^\alpha \tag{6.11}$$

which just says that production is our standard Cobb-Douglas function in per worker terms, and the second equality presumes that the capital stock is exactly equal to the bequest left to a generation by its parents. Again, you could still interpret this as human capital if you liked. We've also assumed that depreciation is exactly equal to one - which also fits with b_t being human capital as it depreciates rather immediately upon the death of a given generation.

The problem is thus to maximize a generations utility, subject to the constraint given on their income, but knowing how their choice of b_t will influence their children's quality. Because generations don't over-lap, but do pass on bequests to each other, this is sometimes called an "over-lapping bequest" or OLB model.

The solution is relatively simple. Plug in the budget constraint to the utility function, and use the production function to resolve the y_{t+1} term. Yes, we could set up a Lagrangian, but the problem is simple enough to dispense with that. We've now got

$$V = \ln[y_t - n_t(\phi + b_t)] + \beta \ln[n_t b_t^\alpha] \tag{6.12}$$

and we can maximize this over both n_t and b_t.

The first-order conditions are

$$\frac{\phi + b_t}{y_t - n_t(\phi + b_t)} = \frac{\beta}{n_t} \tag{6.13}$$

$$\frac{n_t}{y_t - n_t(\phi + b_t)} = \frac{\beta\alpha}{b_t} \tag{6.14}$$

and if you work these two together you can end up with the following

$$\frac{\phi + b_t}{n_t} = \frac{\beta/n_t}{\beta\alpha/b_t} \tag{6.15}$$

where I haven't canceled everything so that the interpretation is easier. This combined FOC tells us very standard information about how marginal utilities must equal to marginal costs. The left-hand side is the ratio of marginal costs. $\phi + b_t$ is the cost of having one additional child - you need to pay ϕ and you also will have to bequeath them b_t. n_t is the cost of raising your bequest per child, b_t by one dollar - you hand over an additional dollar to each child. On the right-hand side the value β/n_t is the marginal utility of one additional child, while $\beta\alpha/b_t$ is the marginal utility of raising the bequest per child by one dollar.

Notice that income has nothing to do with this (for now). That is, the trade-off between quantity and quality does not depend on income. We'll see how that changes as we continue with the problem.

Using the marginal utility/marginal cost condition, we can go back and solve the FOC for the following

$$b_t = \frac{\alpha}{1-\alpha}\phi \tag{6.16}$$

$$n_t = \frac{\beta}{1+\beta}\frac{y_t(1-\alpha)}{\phi} \tag{6.17}$$

which tells us several interesting things. First, child quality is fixed with respect to income. All the dynamics, then, arise through numbers of kids rather than their bequests. This will have pretty severe consequences.

To see this, consider the dynamics of income per capita. We know that

$$y_{t+1} = b_t^{\alpha} = \left(\frac{\alpha}{1-\alpha}\phi\right)^{\alpha} \tag{6.18}$$

which tells us that income per capita is *constant*. This is not a steady state condition - this tells us that no matter what, income per capita is always exactly equal to the above value. There really are no dynamics at all in this model. Any increase in income per capita is matched by an increase in the number of children, which spreads the increased income out over a larger number of bequests, making every child's bequest fall exactly back to $b_t = \alpha\phi/(1-\alpha)$.