Lecture Notes in Economic Growth

Christian Groth

February 3, 2014

Contents

Preface

This is a collection of earlier separate lecture notes in Economic Growth. The notes have been used in recent years in the course Economic Growth within the Master's Program in Economics at the Department of Economics, University of Copenhagen.

Compared with the earlier versions of the lecture notes some chapters have been extended and in some cases divided into several chapters. In addition, discovered typos and similar have been corrected. In some of the chapters a terminal list of references is at present lacking.

The lecture notes are in no way intended as a substitute for the textbook: D. Acemoglu, *Introduction to Modern Economic Growth*, Princeton University Press, 2009. The lecture notes are meant to be read along with the textbook. Some parts of the lecture notes are alternative presentations of stuff also covered by the textbook, while many other parts are complementary in the sense of presenting additional material. Sections marked by an asterisk, *, are cursory reading.

For constructive criticism I thank Niklas Brønager, class instructor since 2012, and plenty of earlier students. No doubt, obscurities remain. Hence, I very much welcome comments and suggestions of any kind relating to these lecture notes.

February 2014

Christian Groth

Chapter 1

Introduction to economic growth

This introductory lecture is a refresher on basic concepts.

Section 1.1 defines Economic Growth as a field of economics. In Section 1.2 formulas for calculation of compound average growth rates in discrete and continuous time are presented. Section 1.3 briefly presents two sets of stylized facts. Finally, Section 1.4 discusses, in an informal way, the different concepts of cross-country income convergence. In his introductory Chapter 1, §1.5, Acemoglu briefly touches upon these concepts.

1.1 The field

Economic growth analysis is the study of what factors and mechanisms determine the time path of *productivity* (a simple index of productivity is output per unit of labor). The focus is on

- productivity levels and
- productivity growth.

1.1.1 Economic growth theory

Economic growth theory endogenizes productivity growth via considering human capital accumulation (formal education as well as learning-by-doing) and endogenous research and development. Also the conditioning role of geography and juridical, political, and cultural institutions is taken into account.

Although for practical reasons, economic growth theory is often stated in terms of easily measurable variables like per capita GDP, the term “economic growth” may be interpreted as referring to something deeper. We could think of “economic growth” as the widening of the opportunities of human beings to lead freer and more worthwhile lives.

To make our complex economic environment accessible for theoretical analysis we use economic models. What *is* an economic model? It is a way of organizing one’s thoughts about the economic functioning of a society. A more specific answer is to define an economic model as a conceptual structure based on a set of mathematically formulated assumptions which have an economic interpretation and from which empirically testable predictions can be derived. In particular, an economic growth model is an economic model concerned with productivity issues. The union of connected and non-contradictory models dealing with economic growth and the theorems derived from these constitute an *economic growth theory*. Occasionally, intense controversies about the validity of different growth theories take place.

The terms “New Growth Theory” and “endogenous growth theory” refer to theory and models which attempt at explaining sustained per capita growth as an outcome of internal mechanisms in the model rather than just a reflection of exogenous technical progress as in “Old Growth Theory”.

Among the themes addressed in this course are:

- How is the world income distribution evolving?
- Why do living standards differ so much across countries and regions? Why are some countries 50 times richer than others?
- Why do per capita growth rates differ over long periods?
- What are the roles of human capital and technology innovation in economic growth? Getting the questions right.
- Catching-up and increased speed of communication and technology diffusion.
- Economic growth, natural resources, and the environment (including the climate). What are the limits to growth?
- Policies to ignite and sustain productivity growth.
- The prospects of growth in the future.

The course concentrates on *mechanisms* behind the evolution of productivity in the industrialized world. We study these mechanisms as integral parts of dynamic general equilibrium models. The exam is a test of the extent to which the student has acquired understanding of these models, is able to evaluate them, from both a theoretical and empirical perspective, and is able to use them to analyze specific economic questions. The course is calculus intensive.

1.1.2 Some long-run data

Let Y denote real GDP (per year) and let N be population size. Then Y/N is GDP per capita. Further, let g_Y denote the average (compound) growth rate of Y per year since 1870 and let $g_{Y/N}$ denote the average (compound) growth rate of Y/N per year since 1870. Table 1.1 gives these growth rates for four countries.

	g_Y	$g_{Y/N}$
Denmark	2,67	1,87
UK	1,96	1,46
USA	3,40	1,89
Japan	3,54	2,54

Table 1.1: Average annual growth rate of GDP and GDP per capita in percent, 1870–2006. Discrete compounding. Source: Maddison, A: The World Economy: Historical Statistics, 2006, Table 1b, 1c and 5c.

Figure 1.1 displays the time path of annual GDP and GDP per capita in Denmark 1870-2006 along with regression lines estimated by OLS (logarithmic scale on the vertical axis). Figure 1.2 displays the time path of GDP per capita in UK, USA, and Japan 1870-2006. In both figures the average annual growth rates are reported. In spite of being based on exactly the same data as Table 1.1, the numbers are slightly different. Indeed, the numbers in the figures are slightly lower than those in the table. The reason is that discrete compounding is used in Table 1.1 while continuous compounding is used in the two figures. These two alternative methods of calculation are explained in the next section.

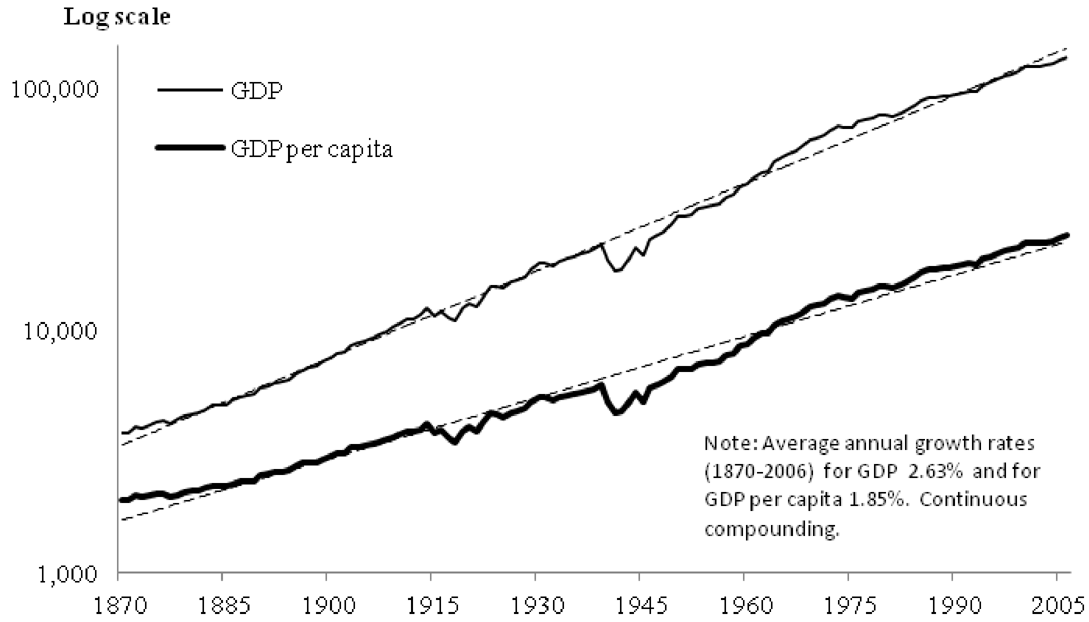


Figure 1.1: GDP and GDP per capita (1990 International Geary-Khamis dollars) in Denmark, 1870-2006. Source: Maddison, A. (2009). Statistics on World Population, GDP and Per Capita GDP, 1-2006 AD, www.ggdc.net/maddison.

1.2 Calculation of the average growth rate

1.2.1 Discrete compounding

Let y denote aggregate labor productivity, i.e., $y \equiv Y/L$, where L is employment. The average growth rate of y from period 0 to period t, with discrete compounding, is that G which satisfies

$$\begin{aligned} y_t &= y_0(1+G)^t, \qquad t = 1, 2, ... , \qquad \text{or} \qquad (1.1) \\ 1+G &= (\frac{y_t}{y_0})^{1/t}, \text{ i.e.,} \\ G &= (\frac{y_t}{y_0})^{1/t} - 1. \qquad (1.2) \end{aligned}$$

"Compounding" means adding the one-period "net return" to the "principal" before adding next period's "net return" (like with interest on interest, also called "compound interest"). Obviously, G will generally be quite different from the arithmetic average of the period-by-period growth rates. To

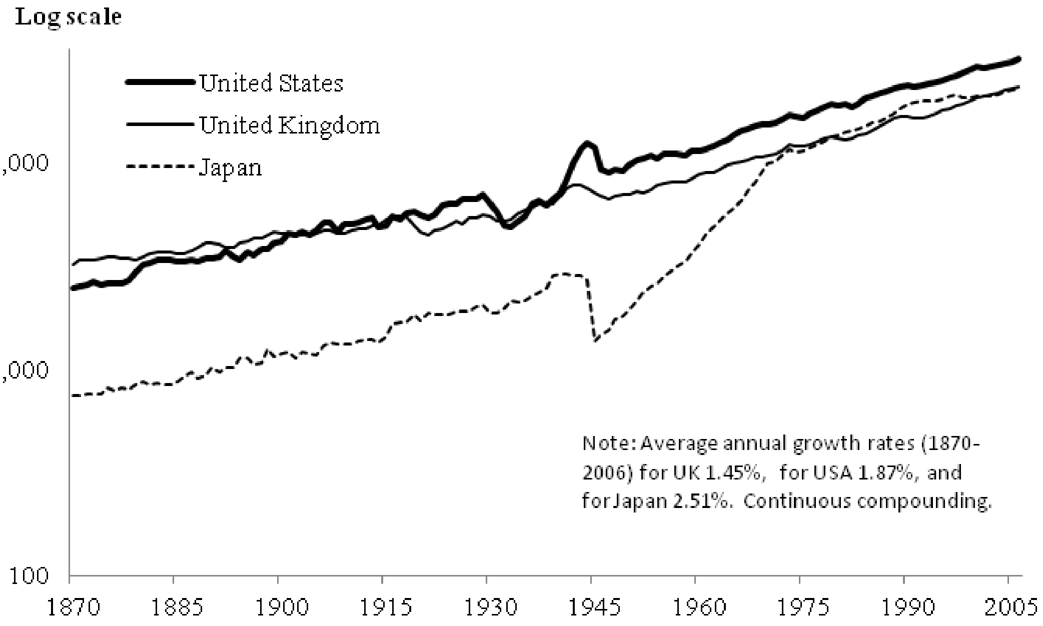


Figure 1.2: GDP per capita (1990 International Geary-Khamis dollars) in UK, USA and Japan, 1870-2006. Source: Maddison, A. (2009). Statistics on World Population, GDP and Per Capita GDP, 1-2006 AD, www.ggdc.net/maddison.

underline this, G is sometimes called the "average compound growth rate" or the "geometric average growth rate".

Using a pocket calculator, the following steps in the calculation of G may be convenient. Take logs on both sides of (1.1) to get

$$\begin{aligned} \ln\frac{y_t}{y_0} &= t\ln(1+G) \;\Rightarrow \\ \ln(1+G) &= \frac{\ln\frac{y_t}{y_0}}{t} \Rightarrow & (1.3) \\ G &= \text{antilog}\Big(\frac{\ln\frac{y_t}{y_0}}{t}\Big) - 1. & (1.4) \end{aligned}$$

Note that t in the formulas (1.2) and (1.4) equals the number of periods *minus 1.*

1.2.2 Continuous compounding

The average growth rate of y, with continuous compounding, is that g which satisfies

$$y_t = y_0 e^{gt}, \tag{1.5}$$

where e denotes the Euler number, i.e., the base of the natural logarithm.[1] Solving for g gives

$$g = \frac{\ln \frac{y_t}{y_0}}{t} = \frac{\ln y_t - \ln y_0}{t}. \tag{1.6}$$

The first formula in (1.6) is convenient for calculation with a pocket calculator, whereas the second formula is perhaps closer to intuition. Another name for g is the "exponential average growth rate".

Again, the t in the formula equals the number of periods minus 1.

Comparing with (1.3) we see that $g = \ln(1 + G) < G$ for $G > 0$. Yet, by a first-order Taylor approximation about $G = 0$ we have

$$g = \ln(1 + G) \approx G \quad \text{for } G \text{ "small"}. \tag{1.7}$$

For a given data set the G calculated from (1.2) will be slightly above the g calculated from (1.6), cf. the mentioned difference between the growth rates in Table 1.1 and those in Figure 1.1 and Figure 1.2. The reason is that a given growth force is more powerful when compounding is continuous rather than discrete. Anyway, the difference between G and g is usually unimportant. If for example G refers to the annual GDP growth rate, it will be a small number, and the difference between G and g immaterial. For example, to $G = 0.040$ corresponds $g \approx 0.039$. Even if $G = 0.10$, the corresponding g is 0.0953. But if G stands for the inflation rate and there is high inflation, the difference between G and g will be substantial. During hyperinflation the monthly inflation rate may be, say, $G = 100\%$, but the corresponding g will be only 69%.

Which method, discrete or continuous compounding, is preferable? To some extent it is a matter of taste or convenience. In period analysis discrete compounding is most common and in continuous time analysis continuous compounding is most common.

For calculation with a pocket calculator the continuous compounding formula, (1.6), is slightly easier to use than the discrete compounding formulas, whether (1.2) or (1.4).

To avoid too much sensitiveness to the initial and terminal observations, which may involve measurement error or depend on the state of the business

[1] Unless otherwise specified, whenever we write $\ln x$ or $\log x$, the *natural* logarithm is understood.

cycle, one can use an OLS approach to the trend coefficient, g, in the following regression:

$$\ln Y_t = \alpha + gt + \varepsilon_t.$$

This is in fact what is done in Fig. 1.1.

1.2.3 Doubling time

How long time does it take for y to double if the growth rate with discrete compounding is G? Knowing G, we rewrite the formula (1.3):

$$t = \frac{\ln \frac{y_t}{y_0}}{\ln(1+G)} = \frac{\ln 2}{\ln(1+G)} \approx \frac{0.6931}{\ln(1+G)}.$$

With $G = 0.0187$, cf. Table 1.1, we find

$$t \approx 37.4 \text{ years},$$

meaning that productivity doubles every 37.4 years.

How long time does it take for y to double if the growth rate with continuous compounding is g? The answer is based on rewriting the formula (1.6):

$$t = \frac{\ln \frac{y_t}{y_0}}{g} = \frac{\ln 2}{g} \approx \frac{0.6931}{g}.$$

Maintaining the value 0.0187 also for g, we find

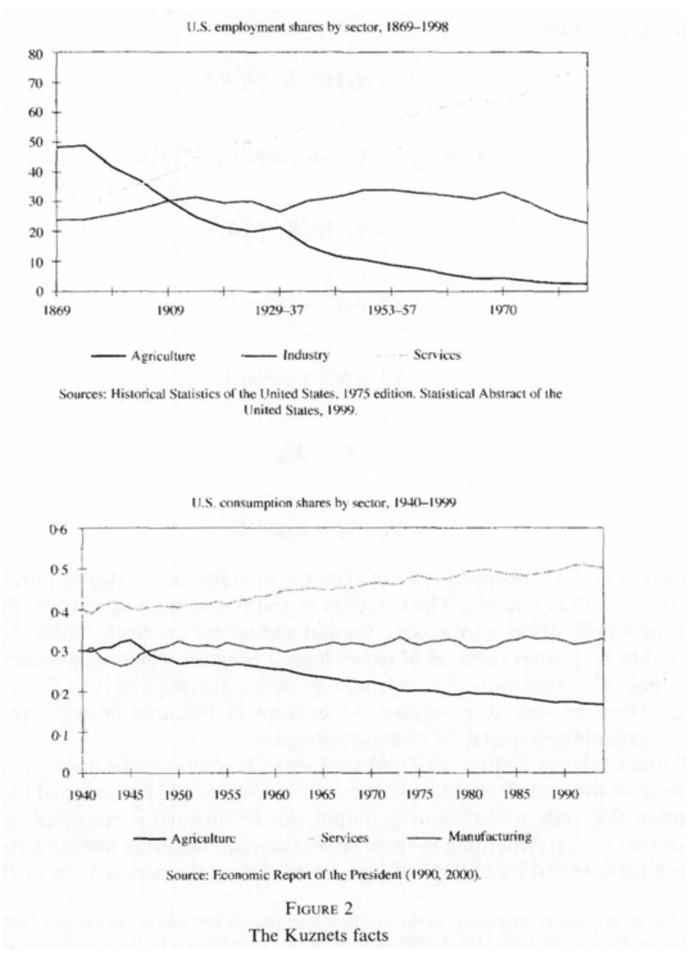
$$t \approx \frac{0.6931}{0.0187} \approx 37.1 \text{ years}.$$

Again, with a pocket calculator the continuous compounding formula is slightly easier to use. With a lower g, say $g = 0.01$, we find doubling time equal to 69.1 years. With $g = 0.07$ (think of China since the 1970's), doubling time is about 10 years! Owing to the compounding exponential growth is extremely powerful.

1.3 Some stylized facts of economic growth

1.3.1 The Kuznets facts

A well-known characteristic of modern economic growth is structural change: unbalanced sectorial growth. There is a massive reallocation of labor from agriculture into industry (manufacturing, construction, and mining) and further into services (including transport and communication). The shares of



Sources: Historical Statistics of the United States, 1975 edition. Statistical Abstract of the United States, 1999.

Source: Economic Report of the President (1990, 2000).

FIGURE 2
The Kuznets facts

Figure 1.3: The Kuznets facts. Source: Kongsamut et al., Beyond Balanced Growth, Review of Economic Studies, vol. 68, Oct. 2001, 869-82.

total consumption expenditure going to these three sectors have moved similarly. Differences in the demand elasticities with respect to income seem the main explanation. These observations are often referred to as the *Kuznets facts* (after Simon Kuznets, 1901-85, see, e.g., Kuznets 1957).

The two graphs in Figure 1.3 illustrate the Kuznets facts.

1.3.2 Kaldor's stylized facts

Surprisingly, in spite of the Kuznets facts, the evolution at the *aggregate* level in developed countries is by many economists seen as roughly described by what is called Kaldor's "stylized facts" (after the Hungarian-British economist Nicholas Kaldor, 1908-1986, see, e.g., Kaldor 1957, 1961)[2]:

1. Real output per man-hour grows at a more or less constant rate over fairly long periods of time. (Of course, there are short-run fluctuations superposed around this trend.)

2. The stock of physical capital per man-hour grows at a more or less constant rate over fairly long periods of time.

3. The ratio of output to capital shows no systematic trend.

4. The rate of return to capital shows no systematic trend.

5. The income shares of labor and capital (in the national accounting sense, i.e., including land and other natural resources), respectively, are nearly constant.

6. The growth rate of output per man-hour differs substantially across countries.

These claimed regularities do certainly not fit all developed countries equally well. Although Solow's growth model (Solow, 1956) can be seen as the first successful attempt at building a model consistent with Kaldor's "stylized facts", Solow once remarked about them: "There is no doubt that they are stylized, though it is possible to question whether they are facts" (Solow, 1970). But the Kaldor "facts" do at least seem to fit the US and UK quite well, see, e.g., Attfield and Temple (2010). The sixth Kaldor fact is, of course, well documented empirically (a nice summary is contained in Pritchett, 1997).

Kaldor also proposed hypotheses about the links between growth in the different sectors (see, e.g., Kaldor 1967):

a. Productivity growth in the manufacturing and construction sectors is enhanced by output growth in these sectors (this is also known as Verdoorn's Law). Increasing returns to scale and learning by doing are the main factors behind this.

b. Productivity growth in agriculture and services is enhanced by output growth in the manufacturing and construction sectors.

1.4 Concepts of income convergence

The two most popular across-country income convergence concepts are "β convergence" and "σ convergence".

[2] Kaldor presented his six regularities as "a stylised view of the facts".

1.4.1 β convergence vs. σ convergence

Definition 1 *We say that β convergence occurs for a given selection of countries if there is a tendency for the poor (those with low income per capita or low output per worker) to subsequently grow faster than the rich.*

By "grow faster" is meant that the growth rate of per capita income (or per worker output) is systematically higher.

In many contexts, a more appropriate convergence concept is the following:

Definition 2 *We say that σ convergence, with respect to a given measure of dispersion, occurs for a given collection of countries if this measure of dispersion, applied to income per capita or output per worker across the countries, declines systematically over time. On the other hand, σ divergence occurs, if the dispersion increases systematically over time.*

The reason that σ convergence must be considered the more appropriate concept is the following. In the end, it is the question of increasing or decreasing dispersion across countries that we are interested in. From a superficial point of view one might think that β convergence implies decreasing dispersion and vice versa, so that β convergence and σ convergence are more or less equivalent concepts. But since the world is not deterministic, but stochastic, this is not true. Indeed, β convergence is only a necessary, not a sufficient condition for σ convergence. This is because over time some reshuffling among the countries is always taking place, and this implies that there will always be some extreme countries (those initially far away from the mean) that move closer to the mean, thus creating a negative correlation between initial level and subsequent growth, in spite of equally many countries moving from a middle position toward one of the extremes.[3] In this way β convergence may be observed at the same time as there is no σ convergence; the mere presence of random measurement errors implies a bias in this direction because a growth rate depends negatively on the initial measurement and positively on the later measurement. In fact, β convergence may be consistent with *σ divergence* (for a formal proof of this claim, see Barro and Sala-i-Martin, 2004, pp. 50-51 and 462 ff.; see also Valdés, 1999, p. 49-50, and Romer, 2001, p. 32-34).

[3] As an intuitive analogy, think of the ordinal rankings of the sports teams in a league. The dispersion of rankings is constant by definition. Yet, no doubt there will allways be some tendency for weak teams to rebound toward the mean and of champions to revert to mediocrity. (This example is taken from the first edition of Barro and Sala-i-Martin, *Economic Growth*, 1995; I do not know why, but the example has been deleted in the second edition from 2004.)

Hence, it is wrong to conclude from β convergence (poor countries tend to grow faster than rich ones) to σ convergence (reduced dispersion of per capita income) without any further investigation. The mistake is called "regression towards the mean" or "Galton's fallacy". Francis Galton was an anthropologist (and a cousin of Darwin), who in the late nineteenth century observed that tall fathers tended to have not as tall sons and small fathers tended to have taller sons. From this he falsely concluded that there was a tendency to averaging out of the differences in height in the population. Indeed, being a true aristocrat, Galton found this tendency pitiable. But since his conclusion was mistaken, he did not really have to worry.

Since σ convergence comes closer to what we are ultimately looking for, from now, when we speak of just "income convergence", σ convergence is understood.

In the above definitions of σ convergence and β convergence, respectively, we were vague as to what kind of selection of countries is considered. In principle we would like it to be a representative sample of the "population" of countries that we are interested in. The population could be all countries in the world. Or it could be the countries that a century ago had obtained a certain level of development.

One should be aware that historical GDP data are constructed retrospectively. Long time series data have only been constructed for those countries that became relatively rich during the after-WWII period. Thus, if we as our sample select the countries for which long data series exist, a so-called *selection bias* is involved which generates a spurious convergence. A country which was poor a century ago will only appear in the sample if it grew rapidly over the next 100 years. A country which was relatively rich a century ago will appear in the sample unconditionally. This selection bias problem was pointed out by DeLong (1988) in a criticism of widespread false interpretations of Maddison's long data series (Maddison 1982).

1.4.2 Measures of dispersion

Our next problem is: *what* measure of dispersion is to be used as a useful descriptive statistics for σ convergence? Here there are different possibilities. To be precise about this we need some notation. Let

$$\begin{aligned} y &\equiv \frac{Y}{L}, \qquad \text{and} \\ q &\equiv \frac{Y}{N}, \end{aligned}$$

where Y = real GDP, L = employment, and N = population. If the focus is on living standards, Y/N, is the relevant variable.[4] But if the focus is on (labor) productivity, it is Y/L, that is relevant. Since most growth models focus on Y/L rather than Y/N, let os take y as our example.

One might think that the standard deviation of y could be a relevant measure of dispersion when discussing whether σ convergence is present or not. The *standard deviation* of y across n countries in a given year is

$$\sigma_y \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{1.8}$$

where

$$\bar{y} \equiv \frac{\sum_i y_i}{n}, \tag{1.9}$$

i.e., $\bar{y}$ is the average output per worker. However, if this measure were used, it would be hard to find *any* group of countries for which there is income convergence. This is because y tends to grow over time for most countries, and then there is an inherent tendency for the variance also to grow; hence also the square root of the variance, σ_y, tends to grow. Indeed, suppose that for all countries, y is doubled from time t_1 to time t_2. Then, automatically, σ_y is also doubled. But hardly anyone would interpret this as an increase in the income inequality across the countries.

Hence, it is more adequate to look at the standard deviation of *relative* income levels:

$$\sigma_{y/\bar{y}} \equiv \sqrt{\frac{1}{n}\sum_i(\frac{y_i}{\bar{y}} - 1)^2}. \tag{1.10}$$

This measure is the same as what is called the *coefficient of variation*, CV_y, usually defined as

$$CV_y \equiv \frac{\sigma_y}{\bar{y}}, \tag{1.11}$$

that is, the standard deviation of y standardized by the mean. That the two measures are identical can be seen in this way:

$$\frac{\sigma_y}{\bar{y}} \equiv \frac{\sqrt{\frac{1}{n}\sum_i(y_i - \bar{y})^2}}{\bar{y}} = \sqrt{\frac{1}{n}\sum_i(\frac{y_i - \bar{y}}{\bar{y}})^2} = \sqrt{\frac{1}{n}\sum_i(\frac{y_i}{\bar{y}} - 1)^2} \equiv \sigma_{y/\bar{y}}.$$

[4] Or perhaps better, Q/N, where $Q \equiv GNP \equiv GDP - rD - wF$. Here, rD, denotes net interest payments on foreign debt and wF denotes net labor income of foreign workers in the country.

The point is that the coefficient of variation is "scale free", which the standard deviation itself is not.

Instead of the coefficient of variation, another scale free measure is often used, namely the standard deviation of $\ln y$, i.e.,

$$\sigma_{\ln y} \equiv \sqrt{\frac{1}{n}\sum_i (\ln y_i - \ln y^*)^2}, \tag{1.12}$$

where

$$\ln y^* \equiv \frac{\sum_i \ln y_i}{n}. \tag{1.13}$$

Note that y^* is the geometric average, i.e., $y^* \equiv \sqrt[n]{y_1 y_2 \cdots y_n}$. Now, by a first-order Taylor approximation of $\ln y$ around $y = \bar{y}$, we have

$$\ln y \approx \ln \bar{y} + \frac{1}{\bar{y}}(y - \bar{y})$$

Hence, as a very rough approximation we have $\sigma_{\ln y} \approx \sigma_{y/\bar{y}} = CV_y$, though this approximation can be quite poor (cf. Dalgaard and Vastrup, 2001). It may be possible, however, to defend the use of $\sigma_{\ln y}$ in its own right to the extent that y tends to be approximately lognormally distributed across countries.

Yet another possible measure of income dispersion across countries is the *Gini index* (see for example Cowell, 1995).

1.4.3 Weighting by size of population

Another important issue is whether the applied dispersion measure is based on a *weighting of the countries by size of population.* For the world as a whole, when no weighting by size of population is used, then there is a slight tendency to income divergence according to the $\sigma_{\ln q}$ criterion (Acemoglu, 2009, p. 4), where q is per capita income ($\equiv Y/N$). As seen by Fig. 4 below, this tendency is not so clear according to the CV_q criterion. Anyway, when there *is* weighting by size of population, then in the last twenty years there has been a tendency to income convergence at the global level (Sala-i-Martin 2006; Acemoglu, 2009, p. 6). With weighting by size of population (1.12) is modified to

$$\sigma^w_{\ln q} \equiv \sqrt{\sum_i w_i (\ln q_i - \ln q^*)^2},$$

where

$$w_i = \frac{N_i}{N} \quad \text{and} \quad \ln q^* \equiv \sum_i w_i \ln q_i.$$

1.4.4 Unconditional vs. conditional convergence

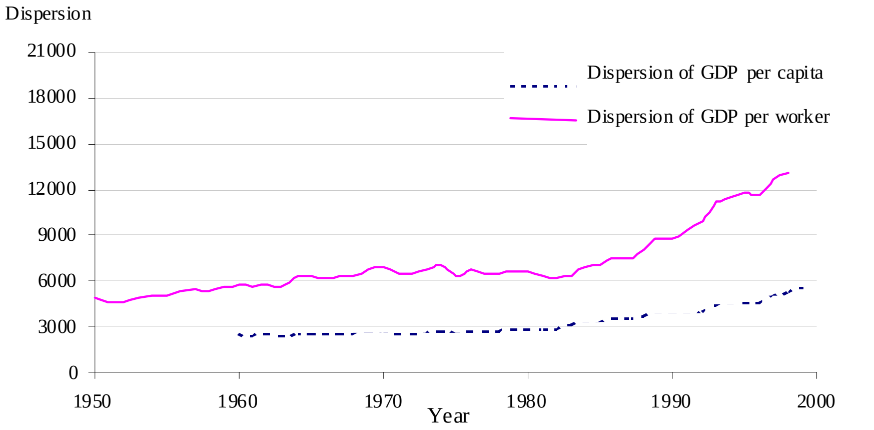
Yet another distinction in the study of income convergence is that between unconditional (or absolute) and conditional convergence. We say that a large heterogeneous group of countries (say the countries in the world) show *unconditional* income convergence if income convergence occurs for the whole group without conditioning on specific characteristics of the countries. If income convergence occurs only for a subgroup of the countries, namely those countries that in advance share the same "structural characteristics", then we say there is *conditional* income convergence. As noted earlier, when we speak of just income "convergence", income "σ convergence" is understood. If in a given context there might be doubt, one should of course be explicit and speak of unconditional or conditional σ convergence. Similarly, if the focus for some reason is on β convergence, we should distinguish between unconditional and conditional β convergence.

What the precise meaning of "structural characteristics" is, will depend on what model of the countries the researcher has in mind. According to the Solow model, a set of relevant "structural characteristics" are: the aggregate production function, the initial level of technology, the rate of technical progress, the capital depreciation rate, the saving rate, and the population growth rate. But the Solow model, as well as its extension with human capital (Mankiw et al., 1992), is a model of a closed economy with exogenous technical progress. The model deals with "within-country" convergence in the sense that the model predicts that a closed economy being initially below or above its steady state path, will over time converge towards its steady state path. It is far from obvious that this kind of model is a good model of cross-country convergence in a globalized world where capital mobility and to some extent also labor mobility are important and some countries are pushing the technological frontier further out, while others try to imitate and catch up.

1.4.5 A bird's-eye view of the data

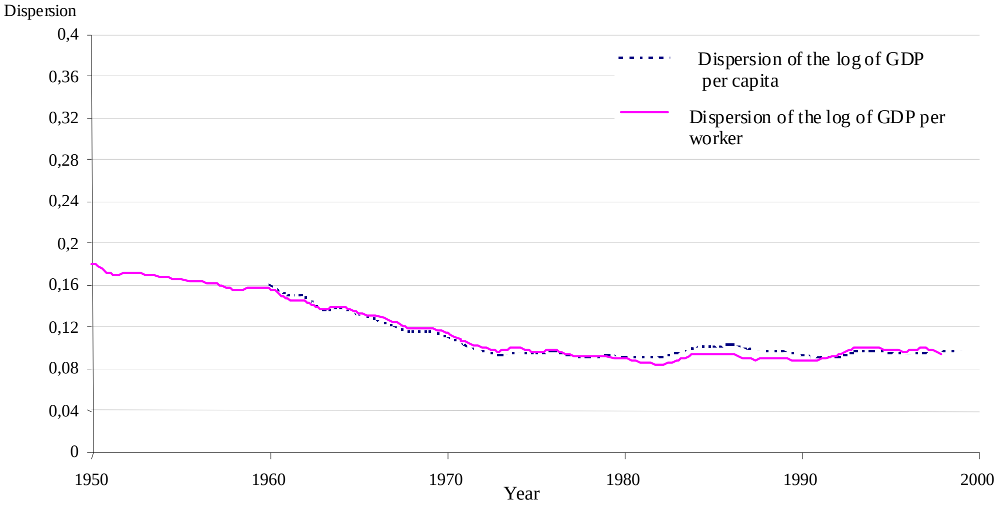
In the following no serious econometrics is attempted. We use the term "trend" in an admittedly loose sense.

Figure 1.4 shows the time profile for the standard deviation of y itself for 12 EU countries, whereas Figure 1.5 and Figure 1.6 show the time profile of the standard deviation of $\log y$ and the time profile of the coefficient of variation, respectively. Comparing the upward trend in Figure 1.4 with the downward trend in the two other figures, we have an illustration of the fact that the movement of the standard deviation of y itself does not capture



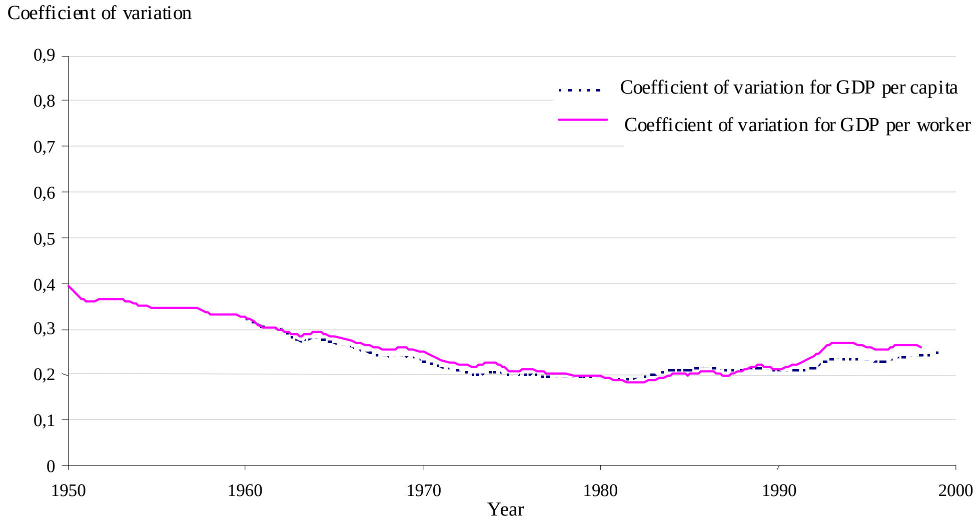
Remarks: Germany is not included in GDP per worker. GDP per worker is missing for Sweden and Greece in 1950, and for Portugal in 1998. The EU comprises Belgium, Denmark, Finland, France, Greece, Holland, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden, Germany, the UK and Austria.
Source: Pwt6, OECD Economic Outlook No. 65 1999 via Eco Win and World Bank Global Development Network Growth Database.

Figure 1.4: Standard deviation of GDP per capita and per worker across 12 EU countries, 1950-1998.



Remarks: Germany is not included in GDP per worker. GDP per worker is missing for Sweden and Greece in 1950, and for Portugal in 1998. The EU comprises Belgium, Denmark, Finland, France, Greece, Holland, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden, Germany, the UK and Austria.
Source: Pwt6, OECD Economic Outlook No. 65 1999 via Eco Win and World Bank Global Development Network Growth Database.

Figure 1.5: Standard deviation of the log of GDP per capita and per worker across 12 EU countries, 1950-1998.



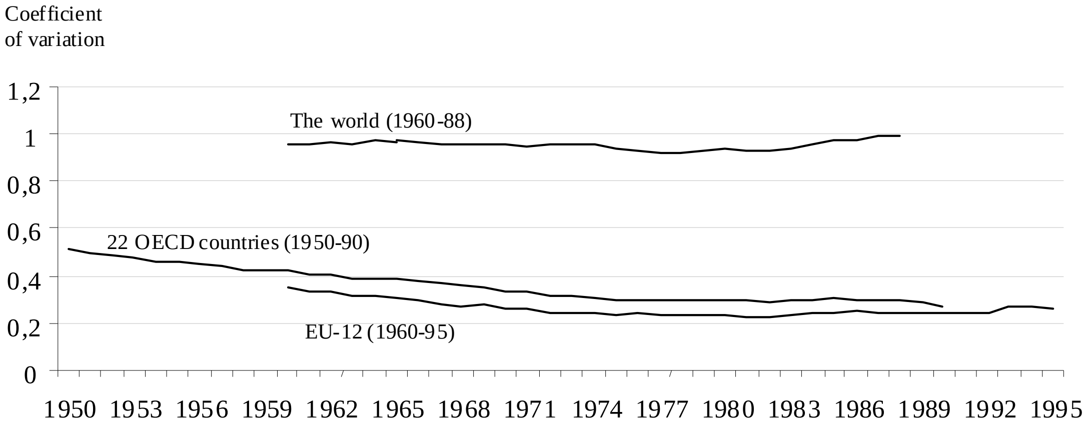
Remarks: Germany is not included in GDP per worker. GDP per worker is missing for Sweden and Greece in 1950, and for Portugal in 1998. The EU comprises Belgium, Denmark, Finland, France, Greece, Holland, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden, Germany, the UK and Austria.
Source: Pwt6, OECD Economic Outlook No. 65 1999 via Eco Win and World Bank Global Development Network Growth Database.

Figure 1.6: Coefficient of variation of GDP per capita and GDP per worker across 12 EU countries, 1950-1998.

income convergence. To put it another way: although there seems to be conditional income convergence with respect to the two scale-free measures, Figure 1.4 shows that this tendency to convergence is *not* so strong as to produce a narrowing of the absolute distance between the EU countries.[5]

Figure 1.7 shows the time path of the coefficient of variation across 121 countries in the world, 22 OECD countries and 12 EU countries, respectively. We see the lack of unconditional income convergence, but the presence of conditional income convergence. One should not over-interpret the observation of convergence for the 22 OECD countries over the period 1950-1990. It is likely that this observation suffer from the selection bias problem mentioned in Section 1.4.1. A country that was poor in 1950 will typically have become a member of OECD only if it grew relatively fast afterwards.

[5] Unfortunately, sometimes misleading graphs or texts to graphs about across-country income convergence are published. In the collection of exercises, Chapter 1, you are asked to discuss some examples of this.



Remarks: 'The world' comprises 121 countries (not weighed by size) where complete time series for GDP per capita exist.
The OECD countries exclude South Korea, Hungary, Poland, Iceland, Czech Rep., Luxembourg and Mexico.
EU-12 comprises: Benelux, Germany, France, Italy, Denmark, Ireland, UK, Spain, Portugal og Greece.
Source: Penn World Table 5.6 and OECD Economic Outlook, Statistics on Microcomputer Disc, December 1998.

Figure 1.7: Coefficient of variation of income per capita across different sets of countries.

1.4.6 Other convergence concepts

Of course, just considering the time profile of the first and second moments of a distribution may sometimes be a poor characterization of the evolution of the distribution. For example, there are signs that the distribution has polarized into *twin peaks* of rich and poor countries (Quah, 1996a; Jones, 1997). Related to this observation is the notion of club convergence. If income convergence occurs *only* among a subgroup of the countries that to some extent share the same initial conditions, then we say there is *club-convergence.* This concept is relevant in a setting where there are *multiple* steady states toward which countries can converge. At least at the theoretical level multiple steady states can easily arise in overlapping generations models. Then the initial condition for a given country matters for which of these steady states this country is heading to. Similarly, we may say that *conditional club-convergence* is present, if income convergence occurs *only* for a subgroup of the countries, namely countries sharing similar structural characteristics (this may to some extent be true for the OECD countries) *and*, within an interval, similar initial conditions.

Instead of focusing on income convergence, one could study *TFP conver-*

gence at aggregate or industry level.[6] Sometimes the less demanding concept of *growth rate convergence* is the focus.

The above considerations are only of a very elementary nature and are only about descriptive statistics. The reader is referred to the large existing literature on concepts and econometric methods of relevance for characterizing the evolution of world income distribution (see Quah, 1996b, 1996c, 1997, and for a survey, see Islam 2003).

1.5 Literature

Acemoglu, D., 2009, *Introduction to Modern Economic Growth*, Princeton University Press: Oxford.

Attfield, C., and J.R.W. Temple, 2010, Balanced growth and the great ratios: New evidence for the US and UK, *Journal of Macroeconomics*, vol. 32, 937-956.

Barro, R. J., and X. Sala-i-Martin, 1995, *Economic Growth,* MIT Press, New York. Second edition, 2004.

Bernard, A.B., and C.I. Jones, 1996a, ..., *Economic Journal.*

- , 1996b, Comparing Apples to Oranges: Productivity Convergence and Measurement Across Industries and Countries, *American Economic Review*, vol. 86 (5), 1216-1238.

Cowell, Frank A., 1995, *Measuring Inequality. 2. ed.,* London.

Dalgaard, C.-J., and J. Vastrup, 2001, On the measurement of σ-convergence, *Economics letters,* vol. 70, 283-87.

Dansk økonomi. Efterår 2001, (Det økonomiske Råds formandskab) Kbh. 2001.

Deininger, K., and L. Squire, 1996, A new data set measuring income inequality, *The World Bank Economic Review*, 10, 3.

Delong, B., 1988, ... *American Economic Review.*

Handbook of Economic Growth, vol. 1A and 1B, ed. by S. N. Durlauf and P. Aghion, Amsterdam 2005.

[6] See, for instance, Bernard and Jones 1996a and 1996b.

Handbook of Income Distribution, vol. 1, ed. by A.B. Atkinson and F. Bourguignon, Amsterdam 2000.

Islam, N., 2003, What have we learnt from the convergence debate? *Journal of Economic Surveys 17*, 3, 309-62.

Kaldor, N., 1957, A model of economic growth, *The Economic Journal,* vol. 67, pp. 591-624.

- , 1961, "Capital Accumulation and Economic Growth". In: F. Lutz, ed., *Theory of Capital,* London: MacMillan.

- , 1967, *Strategic Factors in Economic Development*, New York State School of Industrial and Labor Relations, Cornell University.

Kongsamut et al., 2001, Beyond Balanced Growth, *Review of Economic Studies*, vol. 68, 869-882.

Kuznets, S., 1957, Quantitative aspects of economic growth of nations: II, *Economic Development and Cultural Change*, Supplement to vol. 5, 3-111.

Maddison, A., 1982,

Mankiw, N.G., D. Romer, and D.N. Weil, 1992,

Pritchett, L., 1997, Divergence – big time, *Journal of Economic Perspectives*, vol. 11, no. 3.

Quah, D., 1996a, Twin peaks ..., *Economic Journal,* vol. 106*,* 1045-1055.

-, 1996b, Empirics for growth and convergence, *European Economic Review*, vol. 40 (6).

-, 1996c, Convergence empirics ..., *J. of Ec. Growth,* vol. 1 (1).

-, 1997, Searching for prosperity: A comment, *Carnegie-Rochester Conferende Series on Public Policy,* vol. 55, 305-319.

Romer, D., 2012, *Advanced Macroeconomics*, 4th ed., McGraw-Hill: New York.

Sala-i-Martin, X., 2006, The World Distribution of Income, *Quarterly Journal of Economics 121*, No. 2.

Solow, R.M., 1970, *Growth theory. An exposition,* Clarendon Press: Oxford. Second enlarged edition, 2000.

Valdés, B., 1999, *Economic Growth. Theory, Empirics, and Policy*, Edward Elgar.

On measurement problems, see: http://www.worldbank.org/poverty/inequal/methods/index.htm

Chapter 2

Review of technology

The aim of this chapter is, first, to introduce the terminology concerning firms' technology and technological change used in the lectures and exercises of this course. At a few points I deviate somewhat from definitions in Acemoglu's book. Section 1.3 can be used as a formula manual for the case of CRS.

Second, the chapter contains a brief discussion of the somewhat controversial notions of a representative firm and an aggregate production function.

Regarding the distinction between discrete and continuous time analysis, most of the definitions contained in this chapter are applicable to both.

2.1 The production technology

Consider a two-factor production function given by

$$Y = F(K, L), \tag{2.1}$$

where Y is output (value added) per time unit, K is capital input per time unit, and L is labor input per time unit ($K \geq 0$, $L \geq 0$). We may think of (2.1) as describing the output of a firm, a sector, or the economy as a whole. It is in any case a very simplified description, ignoring the heterogeneity of output, capital, and labor. Yet, for many macroeconomic questions it may be a useful first approach. Note that in (2.1) not only Y but also K and L represent *flows,* that is, quantities per unit of time. If the time unit is one year, we think of K as measured in machine hours per year. Similarly, we think of L as measured in labor hours per year. Unless otherwise specified, it is understood that the rate of utilization of the production factors is constant over time and normalized to one for each production factor. As explained in Chapter 1, we can then use the same symbol, K, for the *flow* of capital services as for the *stock* of capital. Similarly with L.

2.1.1 A neoclassical production function

By definition, K and L are non-negative. It is generally understood that a production function, $Y = F(K, L)$, is *continuous* and that $F(0, 0) = 0$ (no input, no output). Sometimes, when specific functional forms are used to represent a production function, that function may not be defined at points where $K = 0$ or $L = 0$ or both. In such a case we adopt the convention that the domain of the function is understood extended to include such boundary points whenever it is possible to assign function values to them such that continuity is maintained. For instance the function $F(K, L) = \alpha L + \beta KL/(K + L)$, where $\alpha > 0$ and $\beta > 0$, is not defined at $(K, L) = (0, 0)$. But by assigning the function value 0 to the point $(0, 0)$, we maintain both continuity and the "no input, no output" property, cf. Exercise 2.4.

We call the production function *neoclassical* if for all (K, L), with $K > 0$ and $L > 0$, the following additional conditions are satisfied:

(a) $F(K, L)$ has continuous first- and second-order partial derivatives satisfying:

$$\begin{aligned} F_K &> 0, \quad F_L > 0, && (2.2) \\ F_{KK} &< 0, \quad F_{LL} < 0. && (2.3) \end{aligned}$$

(b) $F(K, L)$ is strictly quasiconcave (i.e., the level curves, also called isoquants, are strictly convex to the origin).

In words: (a) says that a neoclassical production function has continuous substitution possibilities between K and L and the *marginal productivities* are positive, but diminishing in own factor. Thus, for a given number of machines, adding one more unit of labor, adds to output, but less so, the higher is already the labor input. And (b) says that every isoquant, $F(K, L) = \bar{Y}$, has a strictly convex form qualitatively similar to that shown in Figure 2.1.[1] When we speak of for example F_L as the marginal *productivity* of labor, it is because the "pure" partial derivative, $\partial Y/\partial L = F_L$, has the denomination of a productivity (output units/yr)/(man-yrs/yr). It is quite common, however, to refer to F_L as the marginal *product* of labor. Then a unit marginal increase in the labor input is understood: $\Delta Y \approx (\partial Y/\partial L)\Delta L = \partial Y/\partial L$ when $\Delta L = 1$. Similarly, F_K can be interpreted as the marginal *productivity* of capital or as the marginal *product* of capital. In the latter case it is understood that $\Delta K = 1$, so that $\Delta Y \approx (\partial Y/\partial K)\Delta K = \partial Y/\partial K$.

[1] For any fixed $\bar{Y} \geq 0$, the associated *isoquant* is the level set $\{(K, L) \in \mathbb{R}_+ | \ F(K, L) = \bar{Y}\}$.

The definition of a neoclassical production function can be extended to the case of n inputs. Let the input quantities be $X_1, X_2, \ldots, X_n$ and consider a production function $Y = F(X_1, X_2, \ldots, X_n)$. Then F is called neoclassical if all the marginal productivities are positive, but diminishing, and F is strictly quasiconcave (i.e., the upper contour sets are strictly convex, cf. Appendix A).

Returning to the two-factor case, since $F(K, L)$ presumably depends on the level of technical knowledge and this level depends on time, t, we might want to replace (2.1) by

$$Y_t = F^t(K_t, L_t), \tag{2.4}$$

where the superscript on F indicates that the production function may shift over time, due to changes in technology. We then say that $F^t(\cdot)$ is a neoclassical production function if it satisfies the conditions (a) and (b) for all pairs (K_t, L_t). *Technological progress* can then be said to occur when, for K_t and L_t held constant, output increases with t.

For convenience, to begin with we skip the explicit reference to time and level of technology.

The marginal rate of substitution Given a neoclassical production function F, we consider the isoquant defined by $F(K, L) = \bar{Y}$, where $\bar{Y}$ is a positive constant. The *marginal rate of substitution*, MRS_{KL}, of K for L at the point (K, L) is defined as the absolute slope of the isoquant at that point, cf. Figure 2.1. The equation $F(K, L) = \bar{Y}$ defines K as an implicit function of L. By implicit differentiation we find $F_K(K, L)dK/dL + F_L(K, L) = 0$, from which follows

$$MRS_{KL} \equiv -\frac{dK}{dL}_{|Y=\bar{Y}} = \frac{F_L(K, L)}{F_K(K, L)} > 0. \tag{2.5}$$

That is, MRS_{KL} measures the amount of K that can be saved (approximately) by applying an extra unit of labor. In turn, this equals the ratio of the marginal productivities of labor and capital, respectively.[2] Since F is neoclassical, by definition F is strictly quasi-concave and so the marginal rate of substitution is diminishing as substitution proceeds, i.e., as the labor input is further increased along a given isoquant. Notice that this feature characterizes the marginal rate of substitution for any neoclassical production function, whatever the returns to scale (see below).

[2] The subscript $|Y = \bar{Y}$ in (2.5) indicates that we are moving along a given isoquant, $F(K, L) = \bar{Y}$. Expressions like, e.g., $F_L(K, L)$ or $F_2(K, L)$ mean the partial derivative of F w.r.t. the second argument, evaluated at the point (K, L).

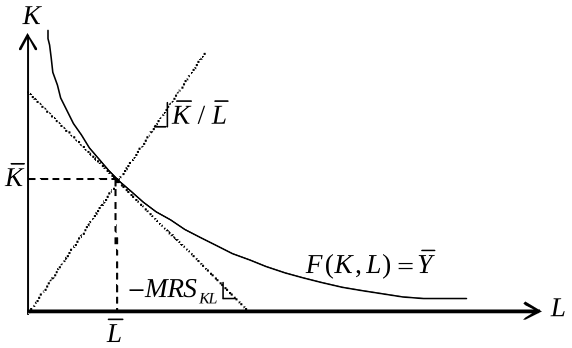


Figure 2.1: MRS_{KL} as the absolute slope of the isoquant.

When we want to draw attention to the dependency of the marginal rate of substitution on the factor combination considered, we write $MRS_{KL}(K, L)$. Sometimes in the literature, the marginal rate of substitution between two production factors, K and L, is called the *technical* rate of substitution (or the technical rate of transformation) in order to distinguish from a consumer's marginal rate of substitution between two consumption goods.

As is well-known from microeconomics, a firm that minimizes production costs for a given output level and given factor prices, will choose a factor combination such that MRS_{KL} equals the ratio of the factor prices. If $F(K, L)$ is homogeneous of degree q, then the marginal rate of substitution depends only on the factor proportion and is thus the same at any point on the ray $K = (\bar{K}/\bar{L})L$. That is, in this case the expansion path is a straight line.

The Inada conditions A continuously differentiable production function is said to satisfy the *Inada conditions*[3] if

$$\lim_{K\to 0} F_K(K,L) = \infty, \lim_{K\to\infty} F_K(K,L) = 0, \tag{2.6}$$
$$\lim_{L\to 0} F_L(K,L) = \infty, \lim_{L\to\infty} F_L(K,L) = 0. \tag{2.7}$$

In this case, the marginal productivity of either production factor has no upper bound when the input of the factor becomes infinitely small. And the marginal productivity is gradually vanishing when the input of the factor increases without bound. Actually, (2.6) and (2.7) express *four* conditions, which it is preferable to consider separately and label one by one. In (2.6) we have two *Inada conditions for MPK* (the marginal productivity of capital), the first being a *lower*, the second an *upper* Inada condition for MPK. And

[3] After the Japanese economist Ken-Ichi Inada, 1925-2002.

in (2.7) we have two *Inada conditions for MPL* (the marginal productivity of labor), the first being a *lower*, the second an *upper* Inada condition for MPL. In the literature, when a sentence like "the Inada conditions are assumed" appears, it is sometimes not made clear which, and how many, of the four are meant. Unless it is evident from the context, it is better to be explicit about what is meant.

The definition of a neoclassical production function we gave above is quite common in macroeconomic journal articles and convenient because of its flexibility. There are textbooks that define a neoclassical production function more narrowly by including the Inada conditions as a requirement for calling the production function neoclassical. In contrast, in this course, when in a given context we need one or another Inada condition, we state it explicitly as an additional assumption.

2.1.2 Returns to scale

If all the inputs are multiplied by some factor, is output then multiplied by the same factor? There may be different answers to this question, depending on circumstances. We consider a production function $F(K, L)$ where $K > 0$ and $L > 0$. Then F is said to have *constant returns to scale* (CRS for short) if it is homogeneous of degree one, i.e., if for all (K, L) and all $\lambda > 0$,

$$F(\lambda K, \lambda L) = \lambda F(K, L).$$

As all inputs are scaled up or down by some factor > 1, output is scaled up or down by the same factor.[4] The assumption of CRS is often defended by the *replication argument.* Before discussing this argument, lets us define the two alternative "pure" cases.

The production function $F(K, L)$ is said to have *increasing returns to scale* (IRS for short) if, for all (K, L) and all $\lambda > 1$,

$$F(\lambda K, \lambda L) > \lambda F(K, L).$$

That is, IRS is present if, when all inputs are scaled up by some factor $>$ 1, output is scaled up by *more* than this factor. The existence of gains by specialization and division of labor, synergy effects, etc. sometimes speak in support of this assumption, at least up to a certain level of production. The assumption is also called the *economies of scale* assumption.

[4] In their definition of a neoclassical production function some textbooks add constant returns to scale as a requirement besides (a) and (b). This course follows the alternative terminology where, if in a given context an assumption of constant returns to scale is needed, this is stated as an additional assumption.

Another possibility is *decreasing returns to scale* (DRS). This is said to occur when for all (K, L) and all $\lambda > 1$,

$$F(\lambda K, \lambda L) < \lambda F(K, L).$$

That is, DRS is present if, when all inputs are scaled up by some factor, output is scaled up by *less* than this factor. This assumption is also called the *diseconomies of scale* assumption. The underlying hypothesis may be that control and coordination problems confine the expansion of size. Or, considering the "replication argument" below, DRS may simply reflect that behind the scene there is an additional production factor, for example land or a irreplaceable quality of management, which is tacitly held fixed, when the factors of production are varied.

EXAMPLE 1 The production function

$$Y = AK^{\alpha}L^{\beta}, \qquad A > 0, 0 < \alpha < 1, 0 < \beta < 1, \tag{2.8}$$

where A, α, and β are given parameters, is called a *Cobb-Douglas production function*. The parameter A depends on the choice of measurement units; for a given such choice it reflects "efficiency", also called the "total factor productivity". Exercise 2.2 asks the reader to verify that (2.8) satisfies (a) and (b) above and is therefore a neoclassical production function. The function is homogeneous of degree $\alpha + \beta$. If $\alpha + \beta = 1$, there are CRS. If $\alpha + \beta < 1$, there are DRS, and if $\alpha + \beta > 1$, there are IRS. Note that α and β must be less than 1 in order not to violate the diminishing marginal productivity condition. □

EXAMPLE 2 The production function

$$Y = \min(AK, BL), \qquad A > 0, B > 0, \tag{2.9}$$

where A and B are given parameters, is called a *Leontief production function* or a *fixed-coefficients production function;* A and B are called the *technical coefficients*. The function is not neoclassical, since the conditions (a) and (b) are not satisfied. Indeed, with this production function the production factors are not substitutable at all. This case is also known as the case of *perfect complementarity* between the production factors. The interpretation is that already installed production equipment requires a fixed number of workers to operate it. The inverse of the parameters A and B indicate the required capital input per unit of output and the required labor input per unit of output, respectively. Extended to many inputs, this type of production function is often used in multi-sector input-output models (also called Leontief models).

In aggregate analysis neoclassical production functions, allowing substitution between capital and labor, are more popular than Leontief functions. But sometimes the latter are preferred, in particular in short-run analysis with focus on the use of already installed equipment where the substitution possibilities are limited.[5] As (2.9) reads, the function has CRS. A generalized form of the Leontief function is $Y = \min(AK^\gamma, BL^\gamma)$, where $\gamma > 0$. When $\gamma < 1$, there are DRS, and when $\gamma > 1$, there are IRS. □

The replication argument The assumption of CRS is widely used in macroeconomics. The model builder may appeal to the *replication argument.* To explain the content of this argument we have to first clarify the distinction between rival and nonrival inputs or more generally the distinction between rival and nonrival goods. A good is *rival* if its character is such that one agent's use of it inhibits other agents' use of it at the same time. A pencil is thus rival. Many production inputs like raw materials, machines, labor etc. have this property. In contrast, however, technical knowledge like a farmaceutical formula or an engineering principle is *nonrival.* An unbounded number of factories can simultaneously use the same farmaceutical formula.

The replication argument now says that by, conceptually, doubling all the rival inputs, we should always be able to double the output, since we just "replicate" what we are already doing. One should be aware that the CRS assumption is about *technology* in the sense of functions linking inputs to outputs − limits to the *availability* of input resources is an entirely different matter. The fact that for example managerial talent may be in limited supply does not preclude the thought experiment that *if* a firm could double all its inputs, including the number of talented managers, then the output level could also be doubled.

The replication argument presupposes, first, that *all* the relevant inputs are explicit as arguments in the production function; second, that these are changed equiproportionately. This, however, exhibits the weakness of the replication argument as a defence for assuming CRS of our present production function, $F(\cdot)$. One could easily make the case that besides capital and labor, also land is a necessary input and should appear as a separate argument.[6] If an industrial firm decides to duplicate what it has been doing, it needs a piece of land to build another plant like the first. Then, on the basis of the replication argument we should in fact expect DRS w.r.t. capital and labor alone. In manufacturing and services, empirically, this and other possible

[5] Cf. Section 2.4.

[6] We think of "capital" as producible means of production, whereas "land" refers to non-producible natural resources, including for example building sites.

sources for departure from CRS may be minor and so many macroeconomists feel comfortable enough with assuming CRS w.r.t. K and L alone, at least as a first approximation. This approximation is, however, less applicable to poor countries, where natural resources may be a quantitatively important production factor.

There is a further problem with the replication argument. Strictly speaking, the CRS claim is that by changing all the inputs equiproportionately by *any* positive factor, λ, which does not have to be an integer, the firm should be able to get output changed by the same factor. Hence, the replication argument requires that indivisibilities are negligible, which is certainly not always the case. In fact, the replication argument is more an argument *against* DRS than *for* CRS in particular. The argument does not rule out IRS due to synergy effects as size is increased.

Sometimes the replication line of reasoning is given a more subtle form. This builds on a useful *local* measure of returns to scale, named the *elasticity of scale.*

The elasticity of scale* To allow for indivisibilities and mixed cases (for example IRS at low levels of production and CRS or DRS at higher levels), we need a local measure of returns to scale. One defines the *elasticity of scale*, $\eta(K, L)$, of F at the point (K, L), where $F(K, L) > 0$, as

$$\eta(K,L) = \frac{\lambda}{F(K,L)}\frac{dF(\lambda K, \lambda L)}{d\lambda} \approx \frac{\Delta F(\lambda K, \lambda L)/F(K,L)}{\Delta\lambda/\lambda}, \text{ evaluated at } \lambda = 1. \tag{2.10}$$

So the elasticity of scale at a point (K, L) indicates the (approximate) percentage increase in output when both inputs are increased by 1 percent. We say that

$$\text{if } \eta(K,L) \begin{cases} > 1, & \text{then there are locally } \textit{IRS,} \\ = 1, & \text{then there are locally } \textit{CRS,} \\ < 1, & \text{then there are locally } \textit{DRS.} \end{cases} \tag{2.11}$$

The production function *may* have the same elasticity of scale everywhere. This is the case if and only if the production function is homogeneous. If F is homogeneous of degree h, then $\eta(K, L) = h$ and h is called the *elasticity of scale parameter.*

Note that the elasticity of scale at a point (K, L) will always equal the sum of the partial output elasticities at that point:

$$\eta(K,L) = \frac{F_K(K,L)K}{F(K,L)} + \frac{F_L(K,L)L}{F(K,L)}. \tag{2.12}$$

This follows from the definition in (2.10) by taking into account that

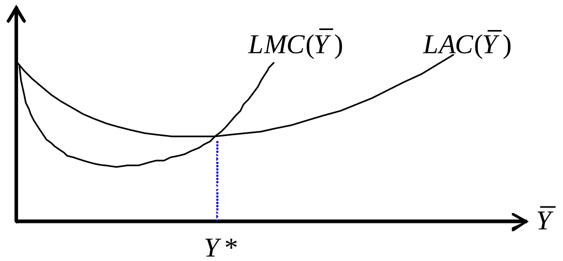


Figure 2.2: Locally CRS at optimal plant size.

$$\begin{aligned}\frac{dF(\lambda K,\lambda L)}{d\lambda} &= F_K(\lambda K,\lambda L)K + F_L(\lambda K,\lambda L)L \\ &= F_K(K,L)K + F_L(K,L)L, \text{ when evaluated at } \lambda = 1.\end{aligned}$$

Figure 2.2 illustrates a popular case from introductory economics, an average cost curve which from the perspective of the individual firm (or plant) is U-shaped: at low levels of output there are falling average costs (thus IRS), at higher levels rising average costs (thus DRS).[7] Given the input prices, w_K and w_L, and a specified output level, $\bar{Y}$, we know that the cost minimizing factor combination $(\bar{K},\bar{L})$ is such that $F_L(\bar{K},\bar{L})/F_K(\bar{K},\bar{L}) = w_L/w_K$. It is shown in Appendix A that the elasticity of scale at $(\bar{K},\bar{L})$ will satisfy:

$$\eta(\bar{K},\bar{L}) = \frac{LAC(\bar{Y})}{LMC(\bar{Y})}, \tag{2.13}$$

where $LAC(\bar{Y})$ is average costs (the minimum unit cost associated with producing $\bar{Y}$) and $LMC(\bar{Y})$ is marginal costs at the output level $\bar{Y}$. The L in LAC and LMC stands for "long-run", indicating that both capital and labor are considered variable production factors within the period considered. At the optimal plant size, Y^*, there is equality between LAC and LMC, implying a unit elasticity of scale, that is, locally we have CRS. That the long-run average costs are here portrayed as rising for $\bar{Y} > Y^*$, is not essential for the argument but may reflect either that coordination difficulties are inevitable or that some additional production factor, say the building site of the plant, is tacitly held fixed.

Anyway, we have here a more subtle replication argument for CRS w.r.t. K and L at the aggregate level. Even though technologies may differ across plants, the surviving plants in a competitive market will have the same average costs at the optimal plant size. In the medium and long run, changes in

[7] By a "firm" is generally meant the company as a whole. A company may have several "manufacturing plants" placed at different locations.

aggregate output will take place primarily by entry and exit of optimal-size plants. Then, with a large number of relatively small plants, each producing at approximately constant unit costs for small output variations, we can without substantial error assume constant returns to scale at the aggregate level. So the argument goes. Notice, however, that even in this form the replication argument is not entirely convincing since the question of indivisibility remains. The optimal plant size may be large relative to the market − and is in fact so in many industries. Besides, in this case also the perfect competition premise breaks down.

2.1.3 Properties of the production function under CRS

The empirical evidence concerning returns to scale is mixed. Notwithstanding the theoretical and empirical ambiguities, the assumption of CRS w.r.t. capital and labor has a prominent role in macroeconomics. In many contexts it is regarded as an acceptable approximation and a convenient simple background for studying the question at hand.

Expedient inferences of the CRS assumption include:

(i) marginal costs are constant and equal to average costs (so the right-hand side of (2.13) equals unity);

(ii) if production factors are paid according to their marginal productivities, factor payments exactly exhaust total output so that pure profits are neither positive nor negative (so the right-hand side of (2.12) equals unity);

(iii) a production function known to exhibit CRS and satisfy property (a) from the definition of a neoclassical production function above, will automatically satisfy also property (b) and consequently *be* neoclassical;

(iv) a neoclassical two-factor production function with CRS has always $F_{KL} > 0$, i.e., it exhibits "direct complementarity" between K and L;

(v) a two-factor production function known to have CRS and to be twice continuously differentiable with positive marginal productivity of each factor everywhere in such a way that all isoquants are strictly convex to the origin, *must* have *diminishing* marginal productivities everywhere.[8]

[8] Proofs of these claims can be found in intermediate microeconomics textbooks and in the Appendix to Chapter 2 of my Lecture Notes in Macroeconomics.

A principal implication of the CRS assumption is that it allows a reduction of dimensionality. Considering a neoclassical production function, $Y = F(K, L)$ with $L > 0$, we can under CRS write $F(K, L) = LF(K/L, 1) \equiv Lf(k)$, where $k \equiv K/L$ is called the *capital-labor ratio* (sometimes the *capital intensity*) and $f(k)$ is the *production function in intensive form* (sometimes named the per capita production function). Thus output per unit of labor depends only on the capital intensity:

$$y \equiv \frac{Y}{L} = f(k).$$

When the original production function F is neoclassical, under CRS the expression for the marginal productivity of capital simplifies:

$$F_K(K, L) = \frac{\partial Y}{\partial K} = \frac{\partial \left[Lf(k)\right]}{\partial K} = Lf'(k)\frac{\partial k}{\partial K} = f'(k). \tag{2.14}$$

And the marginal productivity of labor can be written

$$\begin{aligned} F_L(K, L) &= \frac{\partial Y}{\partial L} = \frac{\partial \left[Lf(k)\right]}{\partial L} = f(k) + Lf'(k)\frac{\partial k}{\partial L} \\ &= f(k) + Lf'(k)K(-L^{-2}) = f(k) - f'(k)k. \end{aligned} \tag{2.15}$$

A neoclassical CRS production function in intensive form always has a positive first derivative and a negative second derivative, i.e., $f' > 0$ and $f'' < 0$. The property $f' > 0$ follows from (2.14) and (2.2). And the property $f'' < 0$ follows from (2.3) combined with

$$F_{KK}(K, L) = \frac{\partial f'(k)}{\partial K} = f''(k)\frac{\partial k}{\partial K} = f''(k)\frac{1}{L}.$$

For a neoclassical production function with CRS, we also have

$$f(k) - f'(k)k > 0 \text{ for all } k > 0, \tag{2.16}$$

in view of $f(0) \geq 0$ and $f'' < 0$. Moreover,

$$\lim_{k \to 0} \left[f(k) - f'(k)k\right] = f(0). \tag{2.17}$$

Indeed, from the mean value theorem[9] we know there exists a number $a \in (0, 1)$ such that for any given $k > 0$ we have $f(k) - f(0) = f'(ak)k$. From this follows $f(k) - f'(ak)k = f(0) < f(k) - f'(k)k$, since $f'(ak) > f'(k)$ by $f'' < 0$.

[9] This theorem says that if f is continuous in $[\alpha, \beta]$ and differentiable in (α, β), then there exists at least one point γ in (α, β) such that $f'(\gamma) = (f(\beta) - f(\alpha))/(\beta - \alpha)$.

In view of $f(0) \geq 0$, this establishes (2.16). And from $f(k) > f(k) - f'(k)k > f(0)$ and continuity of f follows (2.17).

Under CRS the Inada conditions for MPK can be written

$$\lim_{k \to 0} f'(k) = \infty, \qquad \lim_{k \to \infty} f'(k) = 0. \tag{2.18}$$

In this case standard parlance is just to say that "f satisfies the Inada conditions".

An input which must be positive for positive output to arise is called an *essential input*; an input which is not essential is called an *inessential input.* The second part of (2.18), representing the upper Inada condition for MPK under CRS, has the implication that *labor* is an essential input; but capital need not be, as the production function $f(k) = a + bk/(1+k)$, $a > 0, b > 0$, illustrates. Similarly, under CRS the upper Inada condition for MPL implies that *capital* is an essential input. These claims are proved in Appendix C. Combining these results, when *both* the upper Inada conditions hold and CRS obtain, then both capital and labor are essential inputs.[10]

Figure 2.3 is drawn to provide an intuitive understanding of a neoclassical CRS production function and at the same time illustrate that the lower Inada conditions are more questionable than the upper Inada conditions. The left panel of Figure 2.3 shows output per unit of labor for a *CRS neoclassical production function* satisfying the Inada conditions for MPK. The $f(k)$ in the diagram could for instance represent the Cobb-Douglas function in Example 1 with $\beta = 1 - \alpha$, i.e., $f(k) = Ak^{\alpha}$. The right panel of Figure 2.3 shows a non-neoclassical case where only two alternative *Leontief techniques* are available, technique 1: $y = \min(A_1 k, B_1)$, and technique 2: $y = \min(A_2 k, B_2)$. In the exposed case it is assumed that $B_2 > B_1$ and $A_2 < A_1$ (if $A_2 \geq A_1$ at the same time as $B_2 > B_1$, technique 1 would not be efficient, because the same output could be obtained with less input of at least one of the factors by shifting to technique 2). If the available K and L are such that $k < B_1/A_1$ or $k > B_2/A_2$, some of either L or K, respectively, is idle. If, however, the available K and L are such that $B_1/A_1 < k < B_2/A_2$, it is efficient to *combine* the two techniques and use the fraction μ of K and L in technique 1 and the remainder in technique 2, where $\mu = (B_2/A_2 - k)/(B_2/A_2 - B_1/A_1)$. In this way we get the "labor productivity curve" OPQR (the envelope of the two techniques) in Figure 2.3. Note that for $k \to 0$, MPK stays equal to $A_1 < \infty$, whereas for all $k > B_2/A_2$, $MPK = 0$. A similar feature remains true, when we consider *many,* say n, alternative efficient Leontief techniques available. Assuming these techniques cover a considerable range w.r.t. the B/A ratios,

[10] Given a Cobb-Douglas production function, both production factors are essential whether we have DRS, CRS, or IRS.

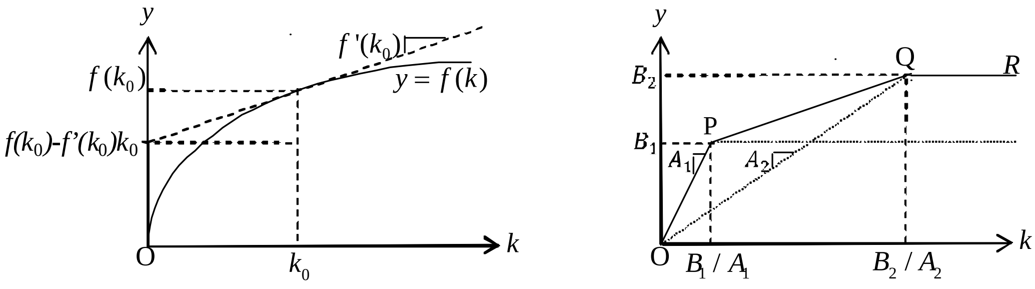


Figure 2.3: Two labor productivity curves based on CRS technologies. Left: neoclassical technology with Inada conditions for MPK satisfied; the graphical representation of MPK and MPL at $k = k_0$.as $f'(k_0)$ and $f(k_0)- f'(k_0)k_0$ are indicated. Right: a combination of two efficient Leontief techniques.

we get a labor productivity curve looking more like that of a neoclassical CRS production function. On the one hand, this gives some intuition of what lies behind the assumption of a neoclassical CRS production function. On the other hand, it remains true that for all $k > B_n/A_n$, $MPK = 0$,[11] whereas for $k \to 0$, MPK stays equal to $A_1 < \infty$, thus questioning the lower Inada condition.

The implausibility of the lower Inada conditions is also underlined if we look at their implication in combination with the more reasonable upper Inada conditions. Indeed, the four Inada conditions taken *together* imply, under CRS, that output has no upper bound when either input goes to infinity for fixed amount of the other input (see Appendix C).

2.2 Technological change

When considering the movement over time of the economy, we shall often take into account the existence of *technological change*. When technological change occurs, the production function becomes time-dependent. Over time the production factors tend to become more productive: more output for given inputs. To put it differently: the isoquants move inward. When this is the case, we say that the technological change displays *technological progress*.

[11] Here we assume the techniques are numbered according to ranking with respect to the size of B.

Concepts of neutral technological change

A first step in taking technological change into account is to replace (2.1) by (2.4). Empirical studies typically specialize (2.4) by assuming that technological change take a form known as *factor-augmenting* technological change:

$$Y_t = F(a_t K_t, b_t L_t), \tag{2.19}$$

where F is a (time-independent) neoclassical production function, Y_t, K_t, and L_t are output, capital, and labor input, respectively, at time t, while a_t and b_t are time-dependent efficiencies of capital and labor, respectively, reflecting technological change. In macroeconomics an even more specific form is often assumed, namely the form of *Harrod-neutral technological change.*[12] This amounts to assuming that a_t in (2.19) is a constant (which we can then normalize to one). So only b_t, which we will then denote T_t, is changing over time, and we have

$$Y_t = F(K_t, T_t L_t). \tag{2.20}$$

The efficiency of labor, T_t, is then said to indicate the *technology level.* Although one can imagine natural disasters implying a fall in T_t, generally T_t tends to rise over time and then we say that (2.20) represents *Harrod-neutral technological progress.* An alternative name for this is *labor-augmenting* technological progress (technological change acts *as if* the labor input were augmented).

If the function F in (2.20) is homogeneous of degree one (so that the technology exhibits CRS w.r.t. capital and labor), we may write

$$\tilde{y}_t \equiv \frac{Y_t}{T_t L_t} = F(\frac{K_t}{T_t L_t}, 1) = F(\tilde{k}_t, 1) \equiv f(\tilde{k}_t), \qquad f' > 0, f'' < 0.$$

where $\tilde{k}_t \equiv K_t/(T_t L_t) \equiv k_t/T_t$ (habitually called the "effective" capital intensity or, if there is no risk of confusion, just the capital intensity). In rough accordance with a general trend in aggregate productivity data for industrialized countries we often assume that T grows at a constant rate, g, so that in discrete time $T_t = T_0(1+g)^t$ and in continuous time $T_t = T_0 e^{gt}$, where $g > 0$. The popularity in macroeconomics of the hypothesis of labor-augmenting technological progress derives from its consistency with Kaldor's "stylized facts", cf. Chapter 4.

There exists two alternative concepts of neutral technological progress. *Hicks-neutral* technological progress is said to occur if technological development is such that the production function can be written in the form

$$Y_t = T_t F(K_t, L_t), \tag{2.21}$$

[12] The name refers to the English economist Roy F. Harrod, 1900−1978.

where, again, F is a (time-independent) neoclassical production function, while T_t is the growing technology level.[13] The assumption of Hicks-neutrality has been used more in microeconomics and partial equilibrium analysis than in macroeconomics. If F has CRS, we can write (2.21) as $Y_t = F(T_t K_t, T_t L_t)$. Comparing with (2.19), we see that in this case Hicks-neutrality is equivalent with $a_t = b_t$ in (2.19), whereby technological change is said to be *equally factor-augmenting.*

Finally, in a kind of symmetric analogy with (2.20), *Solow-neutral* technological progress[14] is often in textbooks presented by a formula like:

$$Y_t = F(T_t K_t, L_t). \tag{2.22}$$

Another name for the same is *capital-augmenting* technological progress (because here technological change acts as if the capital input were augmented). Solow's original concept[15] of neutral technological change is not well portrayed this way, however, since it is related to the notion of *embodied* technological change and capital of *different vintages,* see below.

It is easily shown (Exercise 2.5) that the Cobb-Douglas production function (2.8) satisfies all three neutrality criteria at the same time, if it satisfies one of them (which it does if technological change does not affect α and β). It can also be shown that within the class of neoclassical CRS production functions the Cobb-Douglas function is the only one with this property (see Exercise 4.? in Chapter 4).

Note that the neutrality concepts do not say anything about the *source* of technological progress, only about the quantitative form in which it materializes. For instance, the occurrence of Harrod-neutrality should not be interpreted as indicating that the technological change emanates specifically from the labor input in some sense. Harrod-neutrality only means that technological innovations predominantly are such that not only do labor and capital in combination become more productive, but this happens to *manifest itself* in the form (2.20). Similarly, if indeed an improvement in the quality of the labor input occurs, this "labor-specific" improvement may be manifested in a higher A_t, B_t, or both.

Before proceeding, we briefly comment on how the capital stock, K_t, is typically measured. While data on gross investment, I_t, is available in national income and product accounts, data on K_t usually is not. One ap-

[13] The name refers to the English economist and Nobel Prize laureate John R. Hicks, 1904–1989.

[14] The name refers to the American economist and Nobel Prize laureate Robert Solow (1924–).

[15] Solow (1960).

proach to the measurement of K_t is the *perpetual inventory method* which builds upon the accounting relationship

$$K_t = I_{t-1} + (1-\delta)K_{t-1}. \tag{2.23}$$

Assuming a constant capital depreciation rate δ, backward substitution gives

$$K_t = I_{t-1}+(1-\delta)\left[I_{t-2} + (1-\delta)K_{t-2}\right] = \ldots = \sum_{i=1}^{T}(1-\delta)^{i-1}I_{t-i}+(1-\delta)^{T}K_{t-T}. \tag{2.24}$$

Based on a long time series for I and an estimate of δ, one can insert these observed values in the formula and calculate K_t, starting from a rough conjecture about the initial value K_{t-T}. The result will not be very sensitive to this conjecture since for large T the last term in (2.24) becomes very small.

Embodied vs. disembodied technological progress

There exists an additional taxonomy of technological change. We say that technological change is *embodied*, if taking advantage of new technical knowledge requires construction of new investment goods. The new technology is incorporated in the design of newly produced equipment, but this equipment will not participate in subsequent technological progress. An example: only the most recent vintage of a computer series incorporates the most recent advance in information technology. Then investment goods produced later (investment goods of a later "vintage") have higher productivity than investment goods produced earlier at the same resource cost. Thus investment becomes an important driving force in productivity increases.

We way formalize embodied technological progress by writing capital accumulation in the following way:

$$K_{t+1} - K_t = q_t I_t - \delta K_t, \tag{2.25}$$

where I_t is gross investment in period t, i.e., $I_t = Y_t - C_t$, and q_t measures the "quality" (productivity) of newly produced investment goods. The rising level of technology implies rising q so that a given level of investment gives rise to a greater and greater addition to the capital stock, K, measured in *efficiency units*. In aggregate models C and I are produced with the same technology, the aggregate production function. From this together with (2.25) follows that q capital goods can be produced at the same minimum cost as one consumption good. Hence, the equilibrium price, p, of capital goods in terms of the consumption good must equal the inverse of q, i.e., $p = 1/q$. The output-capital ratio in value terms is $Y/(pK) = QY/K$.

Note that even if technological change does not directly appear in the production function, that is, even if for instance (2.20) is replaced by $Y_t = F(K_t, L_t)$, the economy may experience a rising standard of living when q is growing over time.

In contrast, *disembodied technological change* occurs when new technical and organizational knowledge increases the combined productivity of the production factors independently of when they were constructed or educated. If the K_t appearing in (2.20), (2.21), and (2.22) above refers to the total, historically accumulated capital stock as calculated by (2.24), then the evolution of T in these expressions can be seen as representing disembodied technological change. All vintages of the capital equipment benefit from a rise in the technology level T_t. No new investment is needed to benefit.

Based on data for the U.S. 1950-1990, and taking quality improvements into account, Greenwood et al. (1997) estimate that embodied technological progress explains about 60% of the growth in output per man hour. So, empirically, *embodied* technological progress seems to play a dominant role. As this tends not to be fully incorporated in national income accounting at fixed prices, there is a need to adjust the investment levels in (2.24) to better take estimated quality improvements into account. Otherwise the resulting K will not indicate the capital stock measured in efficiency units.

2.3 The concepts of a representative firm and an aggregate production function

Many macroeconomic models make use of the simplifying notion of a *representative firm.* By this is meant a fictional firm whose production "represents" aggregate production (value added) in a sector or in society as a whole.

Suppose there are n firms in the sector considered or in society as a whole. Let F^i be the production function for firm i so that $Y_i = F^i(K_i, L_i)$, where Y_i, K_i, and L_i are output, capital input, and labor input, respectively, $i = 1, 2, \ldots, n$. Further, let $Y = \Sigma_{i=1}^n Y_i$, $K = \Sigma_{i=1}^n K_i$, and $L = \Sigma_{i=1}^n L_i$. Ignoring technological change, suppose these aggregate variables in a given society turn out to be related through some production function, $F^*(\cdot)$, in the following way:

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

Then $F^*(K, L)$ is called the *aggregate production function* or the production function of the *representative* firm. It is *as if* aggregate production is the result of the behavior of such a single firm.

A simple example where the aggregate production function is well-defined is the following. Suppose that all firms have the *same* production function, i.e., $F^i(\cdot) = F(\cdot)$, so that $Y_i = F(K_i, L_i)$, $i = 1, 2, \ldots, n$. If in addition F has CRS, we then have

$$Y_i = F(K_i, L_i) = L_i F(k_i, 1) \equiv L_i f(k_i),$$

where $k_i \equiv K_i/L_i$. Hence, facing given factor prices, cost minimizing firms will choose the same capital intensity $k_i = k$, for all i. From $K_i = kL_i$ then follows $\sum_i K_i = k \sum_i L_i$ so that $k = K/L$. Thence,

$$Y \equiv \sum Y_i = \sum L_i f(k_i) = f(k) \sum L_i = f(k)L = F(k, 1)L = F(K, L).$$

In this (trivial) case the aggregate production function is well-defined and turns out to be exactly the same as the identical CRS production functions of the individual firms.

Allowing for the existence of *different* production functions at firm level, we may define the aggregate production function as

$$\begin{aligned} F(K, L) &= \max_{(K_1, L_1, \ldots, K_n, L_n) \geq 0} F^1(K_1, L_1) + \cdots + F^n(K_n, L_n) \\ \text{s.t. } \sum_i K_i &\leq K, \quad \sum_i L_i \leq L. \end{aligned}$$

Allowing also the existence of different output goods, different capital goods, and different types of labor makes the issue more intricate, of course. Yet, if firms are price taking profit maximizers and there are nonincreasing returns to scale, we at least know that the aggregate outcome is *as if,* for given prices, the firms jointly maximize aggregate profit on the basis of their combined production technology. The problem is, however, that the conditions needed for this to imply existence of an aggregate production function which is *well-behaved* (in the sense of inheriting simple qualitative properties from its constituent parts) are restrictive.

Nevertheless macroeconomics often treats aggregate output as a single homogeneous good and capital and labor as being two single and homogeneous inputs. There was in the 1960s a heated debate about the problems involved in this, with particular emphasis on the aggregation of different kinds of equipment into one variable, the capital stock "K". The debate is known as the "Cambridge controversy" because the dispute was between a group of economists from Cambridge University, UK, and a group from Massachusetts Institute of Technology (MIT), which is located in Cambridge, USA. The former group questioned the theoretical robustness of several of the neoclassical tenets, including the proposition that rising aggregate capital intensity tends to be associated with a falling rate of interest. Starting at the disaggregate level, an association of this sort is not a logical necessity because, with different production functions across the industries, the relative prices of produced inputs tend to change, when the interest rate changes. While acknowledging the possibility of "paradoxical" relationships, the latter group maintained that in a macroeconomic context they are likely to cause devastating problems only under exceptional circumstances. In the end this is a matter of empirical assessment.[16]

To avoid complexity and because, for many important issues in growth theory, there is today no well-tried alternative, we shall in this course most of the time use aggregate constructs like "Y", "K", and "L" as simplifying devices, hopefully acceptable in a first approximation. There are cases, however, where some disaggregation is pertinent. When for example the role of imperfect competition is in focus, we shall be ready to disaggregate the production side of the economy into several product lines, each producing its own differentiated product. We shall also touch upon a type of growth models where a key ingredient is the phenomenon of "creative destruction" meaning that an incumbent technological leader is competed out by an entrant with a qualitatively new technology.

Like the representative firm, the *representative household* and the *aggregate consumption function* are simplifying notions that should be applied only when they do not get in the way of the issue to be studied. The importance of budget constraints may make it even more difficult to aggregate over households than over firms. Yet, *if* (and that is a big if) all households have the *same constant* propensity to consume out of income, aggregation is straightforward and the representative household is a meaningful concept. On the other hand, if we aim at understanding, say, the *interaction* between lending and borrowing households, perhaps via financial intermediaries, the representative household is not a useful starting point. Similarly, if the theme is conflicts of interests between firm owners and employees, the existence of *different* types of households should be taken into account.

[16] In his review of the Cambridge controversy Mas-Colell (1989) concluded that: "What the 'paradoxical' comparative statics [of disaggregate capital theory] has taught us is simply that modelling the world as having a single capital good is not *a priori* justified. So be it."

2.4 Long-run vs. short-run production functions*

Is the substitutability between capital and labor the same "ex ante" and "ex post"? By ex ante is meant "when plant and machinery are to be decided upon" and by ex post is meant "after the equipment is designed and constructed". In the standard neoclassical competitive setup there is a presumption that also after the construction and installation of the equipment in the firm, the ratio of the factor inputs can be fully adjusted to a change in the relative factor price. In practice, however, when some machinery has been constructed and installed, its functioning will often require a more or less fixed number of machine operators. What can be varied is just the *degree of utilization* of the machinery. That is, after construction and installation of the machinery, the choice opportunities are no longer described by the neoclassical production function but by a Leontief production function,

$$Y = \min(Au\bar{K}, BL), \qquad A > 0, B > 0, \tag{2.26}$$

where $\bar{K}$ is the size of the installed machinery (a fixed factor in the short run) measured in efficiency units, u is its utilization rate ($0 \leq u \leq 1$), and A and B are given technical coefficients measuring efficiency.

So in the short run the choice variables are u and L. In fact, essentially only u is a choice variable since efficient production trivially requires $L = Au\bar{K}/B$. Under "full capacity utilization" we have $u = 1$ (each machine is used 24 hours per day seven days per week). "Capacity" is given as $A\bar{K}$ per week. Producing efficiently at capacity requires $L = A\bar{K}/B$ and the marginal product by increasing labor input is here nil. But if demand, Y^d, is *less* than capacity, satisfying this demand efficiently requires $u = Y^d/(A\bar{K}) < 1$ and $L = Y^d/B$. As long as $u < 1$, the marginal productivity of labor is a *constant*, B.

The various efficient input proportions that are possible *ex ante* may be approximately described by a neoclassical CRS production function. Let this function on intensive form be denoted $y = f(k)$. When investment is decided upon and undertaken, there is thus a choice between alternative efficient pairs of the technical coefficients A and B in (2.26). These pairs satisfy

$$f(k) = Ak = B. \tag{2.27}$$

So, for an increasing sequence of k's, $k_1, k_2, \ldots, k_i, \ldots$, the corresponding pairs are $(A_i, B_i) = (f(k_i)/k_i, f(k_i))$, $i = 1, 2, \ldots$[17] We say that ex ante,

[17] The points P and Q in the right-hand panel of Fig. 2.3 can be interpreted as con-

depending on the relative factor prices as they are "now" and are expected to evolve in the future, a suitable technique, (A_i, B_i), is chosen from an opportunity set described by the given neoclassical production function. But ex post, i.e., when the equipment corresponding to this technique is installed, the production opportunities are described by a Leontief production function with $(A, B) = (A_i, B_i)$.

In the picturesque language of Phelps (1963), technology is in this case *putty-clay.* Ex ante the technology involves capital which is "putty" in the sense of being in a malleable state which can be transformed into a range of various machinery requiring capital-labor ratios of different magnitude. But once the machinery is constructed, it enters a "hardened" state and becomes "clay". Then factor substitution is no longer possible; the capital-labor ratio at full capacity utilization is fixed at the level $k = B_i/A_i$, as in (2.26). Following the terminology of Johansen (1972), we say that a putty-clay technology involves a "long-run production function" which is neoclassical and a "short-run production function" which is Leontief.

In contrast, the standard neoclassical setup assumes the same range of substitutability between capital and labor ex ante and ex post. Then the technology is called *putty-putty.* This term may also be used if ex post there is at least *some* substitutability although less than ex ante. At the opposite pole of putty-putty we may consider a technology which is *clay-clay.* Here neither ex ante nor ex post is factor substitution possible. Table 2.1 gives an overview of the alternative cases.

Table 2.1. Technologies classified according to factor substitutability ex ante and ex post

	Ex post substitution	
Ex ante substitution	possible	impossible
possible	putty-putty	putty-clay
impossible		clay-clay

The putty-clay case is generally considered the realistic case. As time proceeds, technological progress occurs. To take this into account, we may replace (2.27) and (2.26) by $f(k_t, t) = A_t k_t = B_t$ and $Y_t = \min(A_t u_t \bar{K}_t, B_t L_t)$, respectively. If a new pair of Leontief coefficients, (A_{t_2}, B_{t_2}), efficiency-dominates its predecessor (by satisfying $A_{t_2} \geq A_{t_1}$ and $B_{t_2} \geq B_{t_1}$ with at

structed this way from the neoclassical production function in the left-hand panel of the figure.

least one strict equality), it may pay the firm to invest in the new technology at the same time as some old machinery is scrapped. Real wages tend to rise along with technological progress and the scrapping occurs because the revenue from using the old machinery in production no longer covers the associated labor costs.

The clay property ex-post of many technologies is important for short-run analysis. It implies that there may be non-decreasing marginal productivity of labor up to a certain point. It also implies that in its investment decision the firm will have to take expected future technologies and future factor prices into account. For many issues in long-run analysis the clay property ex-post may be less important, since over time adjustment takes place through new investment.

2.5 Literature notes

As to the question of the empirical validity of the constant returns to scale assumption, Malinvaud (1998) offers an account of the econometric difficulties associated with estimating production functions. Studies by Basu (1996) and Basu and Fernald (1997) suggest returns to scale are about constant or decreasing. Studies by Hall (1990), Caballero and Lyons (1992), Harris and Lau (1992), Antweiler and Treffler (2002), and Harrison (2003) suggest there are quantitatively significant increasing returns, either internal or external. On this background it is not surprising that the case of IRS (at least at industry level), together with market forms different from perfect competition, has in recent years received more attention in macroeconomics and in the theory of economic growth.

Macroeconomists' use of the value-laden term "technological progress" in connection with technological change may seem suspect. But the term should be interpreted as merely a label for certain types of shifts of isoquants in an abstract universe. At a more concrete and disaggregate level analysts of course make use of more refined notions about technological change, recognizing for example not only benefits of new technologies, but also the risks, including risk of fundamental mistakes (think of the introduction and later abandonment of asbestos in the construction industry).

An informative history of technology is ...

Embodied technological progress, sometimes called investment-specific technological progress, is explored in, for instance, Solow (1960), Greenwood et al. (1997), and Groth and Wendner (2014). Hulten (2001) surveys the literature and issues related to measurement of the direct contribution of capital accumulation and technological change, respectively, to productivity growth.

Conditions ensuring that a representative household is admitted and the concept of Gorman preferences are discussed in Acemoglu (2009). Another useful source, also concerning the conditions for the representative firm to be a meaningful notion, is Mas-Colell et al. (1995). For general discussions of the limitations of representative agent approaches, see Kirman (1992) and Gallegati and Kirman (1999). Reviews of the "Cambridge Controversy" are contained in Mas-Colell (1989) and Felipe and Fisher (2003). The last-mentioned authors find the conditions required for the well-behavedness of these constructs so stringent that it is difficult to believe that actual economies are in any sense close to satisfy them. For a less distrustful view, see for instance Ferguson (1969), Johansen (1972), Malinvaud (1998), Jorgenson et al. (2005), and Jones (2005).

It is often assumed that capital depreciation can be described as geometric (in continuous time exponential) evaporation of the capital stock. This formula is popular in macroeconomics, more so because of its simplicity than its realism. An introduction to more general approaches to depreciation is contained in, e.g., Nickell (1978).

2.6 References

(incomplete)

Chapter 3

Continuous time analysis

Because dynamic analysis is generally easier in continuous time, growth models are often stated in continuous time. This chapter gives an account of the conceptual aspects of continuous time analysis. Appendix A considers simple growth arithmetic in continuous time. And Appendix B provides solution formulas for linear first-order differential equations.

3.1 The transition from discrete time to continuous time

We start from a discrete time framework. The run of time is divided into successive periods of equal length, taken as the time-unit. Let us here index the periods by $i = 0, 1, 2, ...$ Thus financial wealth accumulates according to

$$a_{i+1} - a_i = s_i, \qquad a_0 \text{ given,}$$

where s_i is (net) saving in period i.

3.1.1 Multiple compounding per year

With time flowing continuously, we let $a(t)$ refer to financial wealth at time t. Similarly, $a(t+\Delta t)$ refers to financial wealth at time $t+\Delta t$. To begin with, let Δt equal one time unit. Then $a(i\Delta t)$ equals $a(i)$ and is of the same value as a_i. Consider the *forward* first difference in a, $\Delta a(t) \equiv a(t+\Delta t) - a(t)$. It makes sense to consider this change in a in relation to the length of the time interval involved, that is, to consider the *ratio* $\Delta a(t)/\Delta t$. As long as $\Delta t = 1$, with $t = i\Delta t$ we have $\Delta a(t)/\Delta t = (a_{i+1} - a_i)/1 = a_{i+1} - a_i$. Now, keep the time unit unchanged, but let the length of the time interval $[t, t+\Delta t)$

approach zero, i.e., let $\Delta t \to 0$. When $a(\cdot)$ is a differentiable function, we have

$$\lim_{\Delta t \to 0} \frac{\Delta a(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{a(t+\Delta t) - a(t)}{\Delta t} = \frac{da(t)}{dt},$$

where $da(t)/dt$, often written $\dot{a}(t)$, is known as the *derivative of* $a(\cdot)$ at the point t. Wealth accumulation in continuous time can then be written

$$\dot{a}(t) = s(t), \qquad a(0) = a_0 \text{ given}, \tag{3.1}$$

where $s(t)$ is the saving flow at time t. For Δt "small" we have the approximation $\Delta a(t) \approx \dot{a}(t)\Delta t = s(t)\Delta t$. In particular, for $\Delta t = 1$ we have $\Delta a(t) = a(t+1) - a(t) \approx s(t)$.

As time unit choose one year. Going back to discrete time we have that if wealth grows at a constant rate $g > 0$ per year, then after i periods of length one year, with annual compounding, we have

$$a_i = a_0(1+g)^i, \quad i = 0, 1, 2, \ldots . \tag{3.2}$$

If instead compounding (adding saving to the principal) occurs n times a year, then after i periods of length $1/n$ year and a growth rate of g/n per such period,

$$a_i = a_0(1 + \frac{g}{n})^i. \tag{3.3}$$

With t still denoting time measured in years passed since date 0, we have $i = nt$ periods. Substituting into (3.3) gives

$$a(t) = a_{nt} = a_0(1+\frac{g}{n})^{nt} = a_0\left[(1+\frac{1}{m})^m\right]^{gt}, \qquad \text{where } m \equiv \frac{n}{g}.$$

We keep g and t fixed, but let $n \to \infty$ and thus $m \to \infty$. Then, in the limit there is continuous compounding and it can be shown that

$$a(t) = a_0 e^{gt}, \tag{3.4}$$

where e is a mathematical constant called the base of the natural logarithm and defined as $e \equiv \lim_{m\to\infty}(1 + 1/m)^m \simeq 2.7182818285....$

The formula (3.4) is the continuous-time analogue to the discrete time formula (3.2) with annual compounding. A geometric growth factor is replaced by an exponential growth factor.

We can also view the formulas (3.2) and (3.4) as the solutions to a difference equation and a differential equation, respectively. Thus, (3.2) is the solution to the linear difference equation $a_{i+1} = (1+g)a_i$, given the initial value a_0. And (3.4) is the solution to the linear differential equation $\dot{a}(t) = ga(t)$, given the initial condition $a(0) = a_0$. Now consider a time-dependent growth rate, $g(t)$. The corresponding differential equation is $\dot{a}(t) = g(t)a(t)$ and it has the solution

$$a(t) = a(0)e^{\int_0^t g(\tau)d\tau}, \tag{3.5}$$

where the exponent, $\int_0^t g(\tau)d\tau$, is the definite integral of the function $g(\tau)$ from 0 to t. The result (3.5) is called the *basic accumulation formula* in continuous time and the factor $e^{\int_0^t g(\tau)d\tau}$ is called the *growth factor* or the *accumulation factor.*

3.1.2 Compound interest and discounting

Let $r(t)$ denote the *short-term real interest rate in continuous time* at time t. To clarify what is meant by this, consider a deposit of $V(t)$ euro on a drawing account in a bank at time t. If the general price level in the economy at time t is $P(t)$ euro, the *real* value of the deposit is $a(t) = V(t)/P(t)$ at time t. By definition the *real rate of return* on the deposit in continuous time (with continuous compounding) at time t is the (proportionate) instantaneous rate at which the real value of the deposit expands per time unit when there is no withdrawal from the account. Thus, if the instantaneous nominal interest rate is $i(t)$, we have $\dot{V}(t)/V(t) = i(t)$ and so, by the fraction rule in continuous time (cf. Appendix A),

$$r(t) = \frac{\dot{a}(t)}{a(t)} = \frac{\dot{V}(t)}{V(t)} - \frac{\dot{P}(t)}{P(t)} = i(t) - \pi(t), \tag{3.6}$$

where $\pi(t) \equiv \dot{P}(t)/P(t)$ is the instantaneous inflation rate. In contrast to the corresponding formula in discrete time, this formula is exact. Sometimes $i(t)$ and $r(t)$ are referred to as the nominal and real *interest intensity*, respectively, or the nominal and real *force of interest.*

Calculating the terminal value of the deposit at time $t_1 > t_0$, given its value at time t_0 and assuming no withdrawal in the time interval $[t_0, t_1]$, the accumulation formula (3.5) immediately yields

$$a(t_1) = a(t_0)e^{\int_{t_0}^{t_1} r(t)dt}.$$

When calculating *present values* in continuous time analysis, we use compound discounting. We simply reverse the accumulation formula and go from the compounded or terminal value to the present value $a(t_0)$. Similarly, given a consumption plan, $(c(t))_{t=t_0}^{t_1}$, the present value of this plan as seen from time t_0 is

$$PV = \int_{t_0}^{t_1} c(t)\ e^{-rt}dt, \tag{3.7}$$

presupposing a constant interest rate. Instead of the geometric discount factor, $1/(1+r)^t$, from discrete time analysis, we have here an exponential discount factor, $1/(e^{rt}) = e^{-rt}$, and instead of a sum, an integral. When the interest rate varies over time, (3.7) is replaced by

$$PV = \int_{t_0}^{t_1} c(t)\ e^{-\int_{t_0}^{t} r(\tau)d\tau} dt.$$

In (3.7) $c(t)$ is discounted by $e^{-rt} \approx (1+r)^{-t}$ for r "small". This might not seem analogue to the discrete-time discounting in (**??**) where it is c_{t-1} that is discounted by $(1+r)^{-t}$, assuming a constant interest rate. When taking into account the timing convention that payment for c_{t-1} in period $t-1$ occurs at the end of the period (= time t), there is no discrepancy, however, since the continuous-time analogue to this payment is $c(t)$.

3.2 The allowed range for parameter values

The allowed range for parameters may change when we go from discrete time to continuous time with continuous compounding. For example, the usual equation for aggregate capital accumulation in continuous time is

$$\dot{K}(t) = I(t) - \delta K(t), \qquad K(0) = K_0 \text{ given,} \tag{3.8}$$

where $K(t)$ is the capital stock, $I(t)$ is the gross investment at time t and $\delta \geq 0$ is the (physical) capital depreciation rate. Unlike in discrete time, here $\delta > 1$ is conceptually allowed. Indeed, suppose for simplicity that $I(t) = 0$ for all $t \geq 0$; then (3.8) gives $K(t) = K_0 e^{-\delta t}$. This formula is meaningful for any $\delta \geq 0$. Usually, the time unit used in continuous time macro models is one year (or, in business cycle theory, rather a quarter of a year) and then a realistic value of δ is of course < 1 (say, between 0.05 and 0.10). However, if the time unit applied to the model is large (think of a Diamond-style OLG model), say 30 years, then $\delta > 1$ may fit better, empirically, if the model is converted into continuous time with the same time unit. Suppose, for example, that physical capital has a half-life of 10 years. With 30 years as our time unit, inserting into the formula $1/2 = e^{-\delta/3}$ gives $\delta = (\ln 2) \cdot 3 \simeq 2$.

In many simple macromodels, where the level of aggregation is high, the relative price of a unit of physical capital in terms of the consumption good is 1 and thus constant. More generally, if we let the relative price of the capital good in terms of the consumption good at time t be $p(t)$ and allow $\dot{p}(t) \neq 0$, then we have to distinguish between the physical depreciation of capital, δ, and the *economic depreciation*, that is, the loss in economic value of a machine per time unit. The economic depreciation will be $d(t) = p(t)\delta - \dot{p}(t)$, namely the economic value of the physical wear and tear (and technological obsolescence, say) minus the capital gain (positive or negative) on the machine.

Other variables and parameters that by definition are bounded from below in discrete time analysis, but not so in continuous time analysis, include rates of return and discount rates in general.

3.3 Stocks and flows

An advantage of continuous time analysis is that it forces the analyst to make a clear distinction between *stocks* (say wealth) and *flows* (say consumption or saving). Recall, a *stock* variable is a variable measured as a quantity at a given point in time. The variables $a(t)$ and $K(t)$ considered above are stock variables. A *flow* variable is a variable measured as quantity *per time unit* at a given point in time. The variables $s(t)$, $\dot{K}(t)$ and $I(t)$ are flow variables.

One can not add a stock and a flow, because they have *different denominations*. What exactly is meant by this? The elementary measurement units in economics are *quantity units* (so many machines of a certain kind or so many liters of oil or so many units of payment, for instance) and *time units* (months, quarters, years). On the basis of these we can form *composite measurement units*. Thus, the capital stock, K, has the denomination "quantity of machines", whereas investment, I, has the denomination "quantity of machines per time unit" or, shorter, "quantity/time". A growth rate or interest rate has the denomination "(quantity/time)/quantity" = "time^{-1}". If we change our time unit, say from quarters to years, the value of a flow variable as well as a growth rate is changed, in this case quadrupled (presupposing annual compounding).

In continuous time analysis expressions like $K(t)+I(t)$ or $K(t)+\dot{K}(t)$ are thus illegitimate. But one can write $K(t+\Delta t) \approx K(t)+(I(t)-\delta K(t))\Delta t$, or $\dot{K}(t)\Delta t \approx (I(t) - \delta K(t))\Delta t$. In the same way, suppose a bath tub at time t contains 50 liters of water and that the tap pours $\frac{1}{2}$ liter per second into the tub for some time. Then a sum like $50\ \ell + \frac{1}{2}\ (\ell/\text{sec})$ does not make sense. But the *amount* of water in the tub after one minute is meaningful. This amount would be $50\ \ell + \frac{1}{2} \cdot 60\ ((\ell/\text{sec})\times\text{sec}) = 80\ \ell$. In analogy, economic flow variables in continuous time should be seen as *intensities* defined for every t in the time interval considered, say the time interval $[0,\ T)$ or perhaps $[0,\ \infty)$. For example, when we say that $I(t)$ is "investment" at time t, this is really a short-hand for "investment intensity" at time t. The actual investment in a time interval $[t_0, t_0 + \Delta t)$, i.e., the invested amount *during*

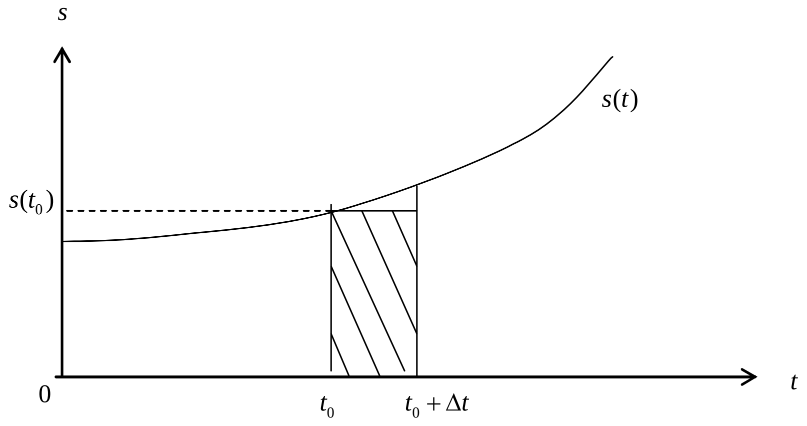


Figure 3.1: With Δt "small" the integral of $s(t)$ from t_0 to $t_0+\Delta t$ is $\approx$ the hatched area.

this time interval, is the integral, $\int_{t_0}^{t_0+\Delta t} I(t)dt \approx I(t_0)\Delta t$. Similarly, the flow of individual saving, $s(t)$, should be interpreted as the saving *intensity* at time t. The actual saving in a time interval $[t_0, t_0 + \Delta t)$, i.e., the saved (or accumulated) amount during this time interval, is the integral, $\int_{t_0}^{t_0+\Delta t} s(t)dt$. If Δt is "small", this integral is approximately equal to the product $s(t_0)\cdot\Delta t$, cf. the hatched area in Figure 3.1.

The notation commonly used in discrete time analysis blurs the distinction between stocks and flows. Expressions like $a_{i+1} = a_i + s_i$, without further comment, are usual. Seemingly, here a stock, wealth, and a flow, saving, are added. In fact, however, it is wealth at the beginning of period i and the saved *amount during* period i that are added: $a_{i+1} = a_i + s_i \cdot \Delta t$. The tacit condition is that the period length, Δt, is the time unit, so that $\Delta t = 1$. But suppose that, for example in a business cycle model, the period length is one quarter, but the time unit is one year. Then saving in quarter i is s_i $= (a_{i+1} - a_i) \cdot 4$ per year.

3.4 The choice between discrete and continuous time analysis

In empirical economics, data typically come in discrete time form and data for flow variables typically refer to periods of constant length. One could argue that this discrete form of the data speaks for discrete time rather than continuous time modelling. And the fact that economic actors often think and plan in period terms, may seem a good reason for putting at least microeconomic analysis in period terms. Nonetheless real time is continuous. And it can hardly be said that the *mass* of economic actors think and plan with one and the same period. In macroeconomics we consider the *sum* of the actions. In this perspective the continuous time approach has the advantage of allowing variation *within* the usually artificial periods in which the data are chopped up. And for example centralized asset markets equilibrate almost instantaneously and respond immediately to new information. For such markets a formulation in continuous time seems preferable.

There is also a risk that a discrete time model may generate *artificial* oscillations over time. Suppose the "true" model of some mechanism is given by the differential equation

$$\dot{x} = \alpha x, \qquad \alpha < -1. \tag{3.9}$$

The solution is $x(t) = x(0)e^{\alpha t}$ which converges in a monotonic way toward 0 for $t \to \infty$. However, the analyst takes a discrete time approach and sets up the seemingly "corresponding" discrete time model

$$x_{t+1} - x_t = \alpha x_t.$$

This yields the difference equation $x_{t+1} = (1 + \alpha)x_t$, where $1 + \alpha < 0$. The solution is $x_t = (1 + \alpha)^t x_0$, $t = 0, 1, 2, \ldots$. As $(1 + \alpha)^t$ is positive when t is even and negative when t is odd, oscillations arise in spite of the "true" model generating monotonous convergence towards the steady state $x^* = 0$.

It should be added, however, that this potential problem *can* always be avoided within discrete time models by choosing a sufficiently *short* period length. Indeed, the solution to a differential equation can always be obtained as the limit of the solution to a corresponding difference equation for the period length approaching zero. In the case of (3.9) the approximating difference equation is $x_{i+1} = (1 + \alpha\Delta t)x_i$, where Δt is the period length, $i = t/\Delta t$, and $x_i = x(i\Delta t)$. By choosing Δt small enough, the solution comes arbitrarily close to the solution of (3.9). It is generally more difficult to go in the opposite direction and find a differential equation that approximates a given difference equation. But the problem is solved as soon as a differential equation has been found that has the initial difference equation as an approximating difference equation.

From the point of view of the economic contents, the choice between discrete time and continuous time may be a matter of taste. From the point of view of mathematical convenience, the continuous time formulation, which has worked so well in the natural sciences, seems preferable. At least this is so in the absence of uncertainty. For problems where uncertainty is important, discrete time formulations are easier to work with unless one is familiar with stochastic calculus.

3.5 Appendix

A. Growth arithmetic in continuous time

Let the variables z, x, and y be differentiable functions of time t. Suppose $z(t)$, $x(t)$, and $y(t)$ are positive for all t. Then:

PRODUCT RULE $z(t) = x(t)y(t) \Rightarrow \frac{\dot{z}(t)}{z(t)} = \frac{\dot{x}(t)}{x(t)} + \frac{\dot{y}(t)}{y(t)}$.

Proof. Taking logs on both sides of the equation $z(t) = x(t)y(t)$ gives $\ln z(t) = \ln x(t) + \ln y(t)$. Differentiation w.r.t. t, using the chain rule, gives the conclusion. □

The procedure applied in this proof is called *logarithmic differentiation* w.r.t. t.

FRACTION RULE $z(t) = \frac{x(t)}{y(t)} \Rightarrow \frac{\dot{z}(t)}{z(t)} = \frac{\dot{x}(t)}{x(t)} - \frac{\dot{y}(t)}{y(t)}$.

The proof is similar.

POWER FUNCTION RULE $z(t) = x(t)^\alpha \Rightarrow \frac{\dot{z}(t)}{z(t)} = \alpha\frac{\dot{x}(t)}{x(t)}$.

The proof is similar.

In continuous time these simple formulas are exactly true. In discrete time the analogue formulas are only approximately true and the approximation can be quite bad unless the growth rates of x and y are small.

B. Solution formulas for linear differential equations of first order

For a general differential equation of first order, $\dot{x}(t) = \varphi(x(t), t)$, with $x(t_0) = x_{t_0}$ and where φ is a continuous function, we have, at least for t in an interval $(-\varepsilon, +\varepsilon)$ for some $\varepsilon > 0$,

$$x(t) = x_{t_0} + \int_{t_0}^{t} \varphi(x(\tau), \tau) d\tau. \qquad (*)$$

To get a confirmation, calculate $\dot{x}(t)$ from (*).

For the special case of a linear differential equation of first order, $\dot{x}(t) + a(t)x(t) = b(t)$, we can specify the solution. Three sub-cases of rising complexity are:

1. $\dot{x}(t) + ax(t) = b$, with $a \neq 0$ and initial condition $x(t_0) = x_{t_0}$. Solution:

$$x(t) = (x_{t_0} - x^*)e^{-a(t-t_0)} + x^*, \text{ where } x^* = \frac{b}{a}.$$

If $a = 0$, we get, directly from (*), the solution $x(t) = x_{t_0} + bt$.[1]

2. $\dot{x}(t) + ax(t) = b(t)$, with initial condition $x(t_0) = x_{t_0}$. Solution:

$$x(t) = x_{t_0}e^{-a(t-t_0)} + e^{-a(t-t_0)} \int_{t_0}^{t} b(s)e^{a(s-t_0)}ds.$$

Special case: $b(t) = ce^{ht}$, with $h \neq -a$ and initial condition $x(t_0) = x_{t_0}$. Solution:

$$x(t) = x_{t_0}e^{-a(t-t_0)} + e^{-a(t-t_0)}c \int_{t_0}^{t} e^{(a+h)(s-t_0)}ds = (x_{t_0} - \frac{c}{a+h})e^{-a(t-t_0)} + \frac{c}{a+h}e^{h(t-t_0)}.$$

3. $\dot{x}(t) + a(t)x(t) = b(t)$, with initial condition $x(t_0) = x_{t_0}$. Solution:

$$x(t) = x_{t_0}e^{-\int_{t_0}^{t} a(\tau)d\tau} + e^{-\int_{t_0}^{t} a(\tau)d\tau} \int_{t_0}^{t} b(s)e^{\int_{t_0}^{s} a(\tau)d\tau}ds.$$

Special case: $b(t) = 0$. Solution:

$$x(t) = x_{t_0}e^{-\int_{t_0}^{t} a(\tau)d\tau}.$$

Even more special case: $b(t) = 0$ and $a(t) = a$, a constant. Solution:

$$x(t) = x_{t_0}e^{-a(t-t_0)}.$$

Remark 1 For $t_0 = 0$, most of the formulas will look simpler.

Remark 2 To check whether a suggested solution *is* a solution, calculate the time derivative of the suggested solution and add an arbitrary constant. By appropriate adjustment of the constant, the final result should be a replication of the original differential equation together with its initial condition.

[1] Some non-linear differential equations can be transformed into this simple case. For simplicity let $t_0 = 0$. Consider the equation $\dot{y}(t) = \alpha y(t)^{\beta}$, $y_0 > 0$ given, $\alpha \neq 0, \beta \neq 1$ (a Bernoulli equation). To find the solution for $y(t)$, let $x(t) \equiv y(t)^{1-\beta}$. Then, $\dot{x}(t) = (1-\beta)y(t)^{-\beta}\dot{y}(t) = (1-\beta)y(t)^{-\beta}\alpha y(t)^{\beta} = (1-\beta)\alpha$. The solution for this is $x(t) = x_0 + (1-\beta)\alpha t$, where $x_0 = y_0^{1-\beta}$. Thereby the solution for $y(t)$ is $y(t) = x(t)^{1/(1-\beta)} = \left(y_0^{1-\beta} + (1-\beta)\alpha t\right)^{1/(1-\beta)}$, which is defined for $t > -y_0^{1-\beta}/((1-\beta)\alpha$.

Chapter 4

Balanced growth theorems

In this chapter we shall discuss three fundamental propositions about balanced growth. In view of the generality of the propositions, they have a broad field of application.

The chapter covers the stuff in Acemoglu's §2.7.3. Our propositions 1 and 2 are slight extensions of part 1 and 2, respectively, of what Acemoglu calls Uzawa's Theorem I (Acemoglu, 2009, p. 60). Proposition 3 essentially corresponds to what Acemoglu calls Uzawa's Theorem II (Acemoglu, 2009, p. 63).

4.1 Balanced growth and constancy of key ratios

First we shall define the terms "steady state" and "balanced growth" as they are usually defined in growth theory. With respect to "balanced growth" this implies a minor deviation from the way Acemoglu briefly defines it informally on his page 57. The main purpose of the present chapter is to lay bare the connections between these two concepts as well as their relation to the hypothesis of Harrod-neutral technical progress and Kaldor's stylized facts.

4.1.1 The concepts of steady state and balanced growth

A basic equation in many one-sector growth models for a closed economy in continuous time is

$$\dot{K} = I - \delta K = Y - C - \delta K \equiv S - \delta K, \tag{4.1}$$

where K is aggregate capital, I aggregate gross investment, Y aggregate output, C aggregate consumption, S aggregate gross saving ($\equiv Y - C$), and

$\delta \geq 0$ is a constant physical capital depreciation rate.

Usually, in the theoretical literature on dynamic models, a *steady state* is defined in the following way:

Definition 3 *A steady state of a dynamic model is a stationary solution to the fundamental differential equation(s) of the model.*

Or briefly: a steady state is a stationary point of a dynamic process.

Let us take the Solow growth model as an example. Here gross saving equals sY, where s is a constant, $0 < s < 1$. Aggregate output is given by a neoclassical production function, F, with CRS and Harrod-neutral technical progress: $Y = F(K, AL) = ALF(\tilde{k}, 1) \equiv TLf(\tilde{k})$, where L is the labor force, A is the level of technology, and $\tilde{k} \equiv K/(AL)$ is the (effective) capital intensity. Moreover, $f' > 0$ and $f'' < 0$. Solow assumes $L(t) = L(0)e^{nt}$ and $A(t) = A(0)e^{gt}$, where $n \geq 0$ and $g \geq 0$ are the constant growth rates of the labor force and technology, respectively. By log-differentiating $\tilde{k}$ w.r.t. t,[1] we end up with the *fundamental differential equation* ("law of motion") of the Solow model:

$$\dot{\tilde{k}} = sf(\tilde{k}) - (\delta + g + n)\tilde{k}. \tag{4.2}$$

Thus, in the Solow model, a (non-trivial) steady state is a $\tilde{k}^* > 0$ such that, if $\tilde{k} = \tilde{k}^*$, then $\dot{\tilde{k}} = 0$.

The most common definition in the literature of balanced growth for an aggregate economy is the following:

Definition 4 *A balanced growth path is a path $(Y, K, C)_{t=0}^{\infty}$ along which the quantities Y, K, and C are positive and grow at constant rates (not necessarily positive and not necessarily the same).*

Acemoglu, however, defines (p. 57) balanced growth in the following way: "balanced growth refers to an allocation where output grows at a constant rate and capital-output ratio, the interest rate, and factor shares remain constant". My problem with this definition is that it mixes growth of quantities with distribution aspects (interest rate and factor income shares). And it is not made clear what is meant by the output-capital ratio if the relative price of capital goods is changing over time. So I stick to the standard definition above which is known to function well in many different contexts.

[1] Or by directly using the fraction rule, see Appendix A to Chapter 3.

4.1.2 A general result about balanced growth

We now leave the specific Solow model. The interesting fact is that, given the dynamic resource constraint (4.1), we have *always* that if there is balanced growth with positive gross saving, then the ratios Y/K and C/Y are constant (by "*always*" is meant: independently of how saving is determined and of how the labor force and technology change). And also the other way round: as long as gross saving is positive, constancy of the Y/K and C/Y ratios is enough to ensure balanced growth. So balanced growth and constancy of key ratios are essentially equivalent.

This is a very practical general observation. And since Acemoglu does not state any balanced growth theorem at this general level, we shall do it, in a precise way, here, together with a proof. Letting g_x denote the growth rate of the (positively valued) variable x, i.e., $g_x \equiv \dot{x}/x$, we claim:

Proposition 1 *(the balanced growth equivalence theorem). Let* $(Y, K, C)_{t=0}^{\infty}$ *be a path along which* Y, K, C, *and* $S \equiv Y - C$ *are positive for all* $t \geq 0$. *Then, given the accumulation equation (4.1), the following holds:*

(i) if there is balanced growth, then $g_Y = g_K = g_C$, *and the ratios* Y/K *and* C/Y *are constant;*

(ii) if Y/K *and* C/Y *are constant, then* Y, K, *and* C *grow at the same constant rate, i.e., not only is there balanced growth, but the growth rates of* Y, K, *and* C *are the same.*

Proof Consider a path $(Y, K, C)_{t=0}^{\infty}$ along which Y, K, C, and $S \equiv Y - C$ are positive for all $t \geq 0$. (i) Assume there is balanced growth. Then, by definition, g_Y, g_K, and g_C are constant. Hence, by (4.1), we have that $S/K = g_K + \delta$ is constant, implying

$$g_S = g_K. \tag{4.3}$$

Further, since $Y = C + S$,

$$\begin{aligned} g_Y &= \frac{\dot{Y}}{Y} = \frac{\dot{C}}{Y} + \frac{\dot{S}}{Y} = g_C\frac{C}{Y} + g_S\frac{S}{Y} = g_C\frac{C}{Y} + g_K\frac{S}{Y} \qquad \text{(by (4.3))} \\ &= g_C\frac{C}{Y} + g_K\frac{Y-C}{Y} = \frac{C}{Y}(g_C - g_K) + g_K. \end{aligned} \tag{4.4}$$

Now, let us provisionally assume that $g_K \neq g_C$. Then (4.4) gives

$$\frac{C}{Y} = \frac{g_Y - g_K}{g_C - g_K}, \tag{4.5}$$

a constant, so that $g_Y = g_C$. But this result implies, by (4.5), that $C/Y = 1$, i.e., $C = Y$. In view of (4.1), however, this outcome contradicts the given condition that $S > 0$. Hence, our provisional assumption is wrong, and we have $g_K = g_C$. By (4.4), this implies $g_Y = g_K = g_C$, but now without the condition $C/Y = 1$ being implied. It follows that Y/K and C/K are constant. Then, also $C/Y = (Y/K)/(C/K)$ is constant.

(ii) Suppose Y/K and C/Y are constant. Then $g_Y = g_K = g_C$, so that C/K is a constant. We now show that this implies that g_K is constant. Indeed, from (4.1), $S/Y = 1 - C/Y$, so that also S/Y is constant. It follows that $g_S = g_Y = g_K$, so that S/K is constant. By (4.1),

$$\frac{S}{K} = \frac{\dot{K} + \delta K}{K} = g_K + \delta,$$

so that g_K is constant. This, together with constancy of Y/K and C/Y, implies that also g_Y and g_C are constant. □

Remark. It is part (i) of the proposition which requires the assumption $S > 0$ for all $t \geq 0$. If $S = 0$, we would have $g_K = -\delta$ and $C \equiv Y - S = Y$, hence $g_C = g_Y$ for all $t \geq 0$. Then there would be balanced growth if the common value of g_C and g_Y had a constant growth rate. This growth rate, however, could easily differ from that of K. Suppose $Y = AK^{\alpha}L^{1-\alpha}$, $g_A = \gamma$ and $g_L = n$ (γ and n constants). Then we would have $g_C = g_Y = \gamma - \alpha\delta + (1-\alpha)n$, which could easily be strictly positive and thereby different from $g_K = -\delta \leq 0$ so that (i) no longer holds. □

The nice feature is that this proposition holds for *any* model for which the simple dynamic resource constraint (4.1) is valid. No assumptions about for example CRS and other technology aspects or about market form are involved. Further, the proposition suggests that if one accepts Kaldor's stylized facts as a description of the past century's growth experience, and if one wants a model consistent with them, one should construct the model such that it can generate balanced growth. For a model to be capable of generating balanced growth, however, technological progress must be of the Harrod-neutral type (i.e., be labor-augmenting), at least in a neighborhood of the balanced growth path. For a fairly general context (but of course not as general as that of Proposition 1), this was shown already by Uzawa (1961). The next section presents a modernized version of Uzawa's contribution.

4.2 The crucial role of Harrod-neutrality

Let the aggregate production function be

$$Y(t) = \tilde{F}(K(t), L(t); t). \tag{4.6}$$

The only technology assumption needed is that $\tilde{F}$ has CRS w.r.t. the first two arguments ($\tilde{F}$ need not be neoclassical for example). As a representation of technical progress, we assume $\partial\tilde{F}/\partial t > 0$ for all $t \geq 0$ (i.e., as time proceeds, unchanged inputs result in more and more output). We also assume that the labor force evolves according to

$$L(t) = L(0)e^{nt}, \tag{4.7}$$

where n is a constant. Further, non-consumed output is invested and so (4.1) is the dynamic resource constraint of the economy.

Proposition 2 *(Uzawa's balanced growth theorem) Let* $(Y(t), K(t), C(t))_{t=0}^{\infty}$*, where* $0 < C(t) < Y(t)$ *for all* $t \geq 0$*, be a path satisfying the capital accumulation equation (4.1), given the CRS-production function (5.2) and the labor force path in (4.7). Then:*

(i) a necessary condition for this path to be a balanced growth path is that along the path it holds that

$$Y(t) = \tilde{F}(K(t), L(t); t) = \tilde{F}(K(t), A(t)L(t); 0), \tag{4.8}$$

where $A(t) = e^{gt}$ *with* $g \equiv g_Y - n$*;*

(ii) for any $g > 0$ *such that there is a* $q > g + n + \delta$ *with the property that* $\tilde{F}(1, k^{-1}; 0) = q$ *for some* $k > 0$ *(i.e., at any* t*, hence also* $t = 0$*, the production function* $\tilde{F}$ *in (5.2) allows an output-capital ratio equal to* q*), a sufficient condition for the existence of a balanced growth path with output-capital ratio* q*, is that the technology can be written as in (4.8) with* $A(t) = e^{gt}$*.*

Proof (i)[2] Suppose the path $(Y(t), K(t), C(t))_{t=0}^{\infty}$ is a balanced growth path. By definition, g_K and g_Y are then constant, so that $K(t) = K(0)e^{g_K t}$ and $Y(t) = Y(0)e^{g_Y t}$. We then have

$$Y(t)e^{-g_Y t} = Y(0) = \tilde{F}(K(0), L(0); 0) = \tilde{F}(K(t)e^{-g_K t}, L(t)e^{-nt}; 0), \tag{4.9}$$

[2] This part draws upon Schlicht (2006), who generalized a proof in Wan (1971, p. 59) for the special case of a constant saving rate.

where we have used (5.2) with $t = 0$. In view of the precondition that $S(t) \equiv Y(t) - C(t) > 0$, we know from (i) of Proposition 1, that Y/K is constant so that $g_Y = g_K$. By CRS, (4.9) then implies

$$Y(t) = \tilde{F}(K(t)e^{g_Y t}e^{-g_K t}, L(t)e^{g_Y t}e^{-nt}; 0) = \tilde{F}(K(t), e^{(g_Y - n)t}L(t); 0).$$

We see that (4.8) holds for $A(t) = e^{gt}$ with $g \equiv g_Y - n$.

(ii) Suppose (4.8) holds with $A(t) = e^{gt}$. Let $g > 0$ be given such that there is a $q > g + n + \delta$ with the property that $\tilde{F}(1, k^{-1}; 0) = q$ for some $k > 0$. Then our first claim is that with $K(0) = kL(0)$, $s \equiv (g + n + \delta)/q$, and $S(t) = sY(t)$, (4.1), (4.7), and (4.8) imply $Y(t)/K(t) = q$ for all $t \geq 0$. Indeed, by construction

$$\frac{Y(0)}{K(0)} = \frac{\tilde{F}(K(0), L(0); 0)}{K(0)} = \tilde{F}(1, k^{-1}; 0) = q = \frac{\delta + g + n}{s}. \tag{4.10}$$

It follows that $sY(0)/K(0) - \delta = g + n$. So, by (4.1), we have $\dot{K}(0)/K(0) = sY(0)/K(0) - \delta = g + n$, implying that K initially grows at the same rate as effective labor input, $A(t)L(t)$. Then, in view of $\tilde{F}$ being homogeneous of degree one w.r.t. its first two arguments, also Y grows initially at this rate. As an implication, the ratio Y/K does not change, but remains equal to the right-hand side of (4.10) for all $t \geq 0$. Consequently, K and Y continue to grow at the same constant rate, $g + n$. As $C = (1 - s)Y$, C grows forever also at this constant rate. Hence, the path $(Y(t), K(t), C(t))_{t=0}^{\infty}$ is a balanced growth path, as was to be proved. □

The form (4.8) indicates that along a balanced growth path, technical progress must be purely "labor augmenting", that is, Harrod-neutral. It is in this case convenient to define a new CRS function, F, by $F(K(t), A(t)L(t)) \equiv \tilde{F}(K(t), A(t)L(t); 0)$. Then (i) of the proposition implies that at least along the balanced growth path, we can rewrite the production function this way:

$$Y(t) = \tilde{F}(K(t), L(t); t) = F(K(t), A(t)L(t)), \tag{4.11}$$

where $A(t) = e^{gt}$ with $g \equiv g_Y - n$.

It is important to recognize that the occurrence of Harrod-neutrality says nothing about what the *source* of technological progress is. Harrod-neutrality should not be interpreted as indicating that the technological progress emanates specifically from the labor input. Harrod-neutrality only means that technical innovations predominantly are such that not only do labor and capital in combination become more productive, but this happens to *manifest* *itself* at the aggregate level in the form (4.11).[3]

What is the intuition behind the Uzawa result that for balanced growth to be possible, technical progress must have the purely labor-augmenting form? First, notice that there is an asymmetry between capital and labor. Capital is an accumulated amount of non-consumed output. In contrast, in simple macro models labor is a non-produced production factor which (at least in this context) grows in an exogenous way. Second, because of CRS, the original formulation, (5.2), of the production function implies that

$$1 = \tilde{F}(\frac{K(t)}{Y(t)}, \frac{L(t)}{Y(t)}; t). \tag{4.12}$$

Now, since capital is accumulated non-consumed output, it inherits the trend in output such that $K(t)/Y(t)$ must be constant along a balanced growth path (this is what Proposition 1 is about). Labor does not inherit the trend in output; indeed, the ratio $L(t)/Y(t)$ is free to adjust as time proceeds. When there is technical progress ($\partial\tilde{F}/\partial t > 0$) along a balanced growth path, this progress must manifest itself in the form of a changing $L(t)/Y(t)$ in (4.12) as t proceeds, precisely because $K(t)/Y(t)$ *must* be constant along the path. In the "normal" case where $\partial\tilde{F}/\partial L > 0$, the needed change in $L(t)/Y(t)$ is a *fall* (i.e., a rise in $Y(t)/L(t)$). This is what (4.12) shows. Indeed, the fall in $L(t)/Y(t)$ must exactly offset the effect on $\tilde{F}$ of the rising t, when there is a fixed capital-output ratio.[4] It follows that along the balanced growth path, $Y(t)/L(t)$ is an increasing implicit function of t. If we denote this function $A(t)$, we end up with (4.11).

The generality of Uzawa's theorem is noteworthy. The theorem assumes CRS, but does not presuppose that the technology is neoclassical, not to speak of satisfying the Inada conditions.[5] And the theorem holds for exogenous as well as endogenous technological progress. It is also worth mentioning that the proof of the sufficiency part of the theorem is *constructive*. It provides a method to construct a hypothetical balanced growth path (BGP from now).[6]

A simple implication of the Uzawa theorem is the following. Interpreting the $A(t)$ in (4.8) as the "level of technology", we have:

[3] For a CRS Cobb-Douglas production function with technological progress, Harrod-neutrality is present whenever the output elasticity w.r.t capital (often denoted α) is constant over time.

[4] This way of presenting the intuition behind the Uzawa result draws upon Jones and Scrimgeour (2008).

[5] Many accounts of the Uzawa theorem, including Jones and Scrimgeour (2008), presume a neoclassical production function, but the theorem is much more general.

[6] Part (ii) of Proposition 2 is ignored in Acemoglu's book.

COROLLARY Along a BGP with positive gross saving and the technology level, $A(t)$, growing at the rate g, output grows at the rate $g + n$ while labor productivity, $y \equiv Y/L$, and consumption per unit of labor, $c \equiv C/L$, grow at the rate g.

Proof That $g_Y = g + n$ follows from (i) of Proposition 2. As to the growth rate of labor productivity we have

$$y_t = \frac{Y(0)e^{g_Y t}}{L(0)e^{nt}} = y(0)e^{(g_Y - n)t} = y(0)e^{gt}.$$

Finally, by Proposition 1, along a BGP with $S > 0$, c must grow at the same rate as y. □

We shall now consider the implication of Harrod-neutrality for the income shares of capital and labor when the technology is neoclassical and markets are perfectly competitive.

4.3 Harrod-neutrality and the functional income distribution

There is one facet of Kaldor's stylized facts we have so far not related to Harrod-neutral technical progress, namely the long-run "approximate" constancy of both the income share of labor, wL/Y, and the rate of return to capital. At least with neoclassical technology, profit maximizing firms, and perfect competition in the output and factor markets, these properties are inherent in the combination of constant returns to scale, balanced growth, and the assumption that the relative price of capital goods (relative to consumption goods) equals one. The latter condition holds in models where the capital good is nothing but non-consumed output, cf. (4.1).[7]

To see this, we start out from a neoclassical CRS production function with Harrod-neutral technological progress,

$$Y(t) = F(K(t), A(t)L(t)). \tag{4.13}$$

With $w(t)$ denoting the real wage at time t, in equilibrium under perfect competition the labor income share will be

$$\frac{w(t)L(t)}{Y(t)} = \frac{\frac{\partial Y(t)}{\partial L(t)}L(t)}{Y(t)} = \frac{F_2(K(t), A(t)L(t))A(t)L(t)}{Y(t)}. \tag{4.14}$$

[7] The reader may think of the "corn economy" example in Acemoglu, p. 28.

In this simple model, without natural resources, capital (gross) income equals non-labor income, $Y(t) - w(t)L(t)$. Hence, if $r(t)$ denotes the (net) rate of return to capital at time t, then

$$r(t) = \frac{Y(t) - w(t)L(t) - \delta K(t)}{K(t)}. \tag{4.15}$$

Denoting the capital (gross) income share by $\alpha(t)$, we can write this $\alpha(t)$ (in equilibrium) in three ways:

$$\begin{aligned}
\alpha(t) &\equiv \frac{Y(t) - w(t)L(t)}{Y(t)} = \frac{(r(t)+\delta)K(t)}{Y(t)}, \\
\alpha(t) &= \frac{F(K(t), A(t)L(t)) - F_2(K(t), A(t)L(t))A(t)L(t)}{Y(t)} = \frac{F_1(K(t), A(t)L(t))K(t)}{Y(t)}, \\
\alpha(t) &= \frac{\frac{\partial Y(t)}{\partial K(t)}K(t)}{Y(t)},
\end{aligned} \tag{4.16}$$

where the first row comes from (4.15), the second from (4.13) and (4.14), the third from the second together with Euler's theorem.[8] Comparing the first and the last row, we see that in equilibrium

$$\frac{\partial Y(t)}{\partial K(t)} = r(t) + \delta.$$

In this condition we recognize one of the first-order conditions in the representative firm's profit maximization problem under perfect competition, since $r(t) + \delta$ can be seen as the firm's required gross rate of return.[9]

In the absence of uncertainty, the equilibrium real interest rate in the bond market must equal the rate of return on capital, $r(t)$. And $r(t) + \delta$ can then be seen as the firm's cost of disposal over capital per unit of capital per time unit, consisting of interest cost plus capital depreciation.

Proposition 3 *(factor income shares and rate of return under balanced growth) Let the path* $(K(t), Y(t), C(t))_{t=0}^{\infty}$ *be a BGP in a competitive economy with the production function (4.13) and with positive saving. Then, along the BGP, the* $\alpha(t)$ *in (4.16) is a constant,* $\alpha \in (0, 1)$. *The labor income share will be* $1 - \alpha$ *and the (net) rate of return on capital will be* $r = \alpha q - \delta$, *where* q *is the constant output-capital ratio along the BGP.*

[8] From Euler's theorem, $F_1K + F_2AL = F(K, AL)$, when F is homogeneous of degree one.

[9] With natural resources, say land, entering the set of production factors, the formula, (4.15), for the rate of return to capital should be modified by subtracting rents from the numerator.

Proof By CRS we have $Y(t) = F(K(t), A(t)L(t)) = A(t)L(t)F(\tilde{k}(t), 1) \equiv A(t)L(t)f(\tilde{k}(t))$. In view of part (i) of Proposition 2, by balanced growth, $Y(t)/K(t)$ is some constant, q. Since $Y(t)/K(t) = f(\tilde{k}(t))/\tilde{k}(t)$ and $f'' < 0$, this implies $\tilde{k}(t)$ constant, say equal to $\tilde{k}^*$. But $\partial Y(t)/\partial K(t) = f'(\tilde{k}(t))$, which then equals the constant $f'(\tilde{k}^*)$ along the BGP. It then follows from (4.16) that $\alpha(t) = f'(\tilde{k}^*)/q \equiv \alpha$. Moreover, $0 < \alpha < 1$, where $0 < \alpha$ follows from $f' > 0$ and $\alpha < 1$ from the fact that $q = Y/K = f(\tilde{k}^*)/\tilde{k}^* > f'(\tilde{k}^*)$, in view of $f'' < 0$ and $f(0) \geq 0$. Then, by the first equality in (4.16), $w(t)L(t)/Y(t) = 1 - \alpha(t) = 1 - \alpha$. Finally, by (4.15), the (net) rate of return on capital is $r = (1 - w(t)L(t)/Y(t))Y(t)/K(t) - \delta = \alpha q - \delta$. □

This proposition is of interest by displaying a link from balanced growth to constancy of factor income shares and the rate of return, that is, some of the "stylized facts" claimed by Kaldor. Note, however, that although the proposition implies constancy of the income shares and the rate of return, it does not *determine* them, except in terms of α and q. But both q and, generally, α are endogenous and depend on $\tilde{k}^*$,[10] which will generally be unknown as long as we have not specified a theory of saving. This takes us to theories of aggregate saving, for example the simple Ramsey model, cf. Chapter 8 in Acemoglu's book.

4.4 What if technological change is embodied?

In our presentation of technological progress above we have implicitly assumed that all technological change is *disembodied.* And the way the propositions 1, 2, and 3, are formulated assume this.

As noted in Chapter 2, *disembodied technological change* occurs when new technical knowledge advances the combined productivity of capital and labor independently of whether the workers operate old or new machines. Consider again the aggregate dynamic resource constraint (4.1) and the production function (5.2):

$$\begin{aligned} \dot{K}(t) &= Y(t) - C(t) - \delta K(t), && (*) \\ Y(t) &= \tilde{F}(K(t), L(t); t), \qquad \partial \tilde{F}/\partial t > 0. && (**) \end{aligned}$$

Here $Y(t) - C(t)$ is aggregate gross investment, $I(t)$. For a given level of $I(t)$, the resulting amount of new capital goods per time unit $(\dot{K}(t) + \delta K(t))$, measured in efficiency units, is independent of *when* this investment occurs. It is

[10] As to α, there is of course a trivial exception, namely the case where the production function is Cobb-Douglas and α therefore is a given parameter.

thereby not affected by technological progress. Similarly, the interpretation of $\partial\tilde{F}/\partial t > 0$ in (**) is that the higher technology level obtained as time proceeds results in higher productivity of *all* capital and labor. Thus also firms that have only old capital equipment benefit from recent advances in technical knowledge. No new investment is needed to take advantage of the recent technological and organizational developments.[11]

In contrast, we say that technological change is *embodied*, if taking advantage of new technical knowledge requires construction of new investment goods. The newest technology is incorporated in the design of newly produced equipment; and this equipment will not participate in subsequent technological progress. Whatever the source of new technical knowledge, investment becomes an important bearer of the productivity increases which this new knowledge makes possible. Without new investment, the potential productivity increases remain potential instead of being realized.

As also noted in Chapter 2, we may represent embodied technological progress (also called investment-specific technological change) by writing capital accumulation in the following way,

$$\dot{K}(t) = q(t)I(t) - \delta K(t), \tag{4.17}$$

where $I(t)$ is gross investment at time t and $q(t)$ measures the "quality" (productivity) of newly produced investment goods. The increasing level of technology implies increasing $q(t)$ so that a given level of investment gives rise to a greater and greater additions to the capital stock, K, measured in efficiency units. As in our aggregate framework, q capital goods can be produced at the same minimum cost as one consumption good, we have $p \cdot q = 1$, where p is the equilibrium price of capital goods in terms of consumption goods. So embodied technological progress is likely to result in a steady decline in the relative price of capital equipment, a prediction confirmed by the data (see, e.g., Greenwood et al., 1997).

This raises the question how the propositions 1, 2, and 3 fare in the case of embodied technological progress. The answer is that a generalized version of Proposition 1 goes through. Essentially, we only need to replace (4.1) by (4.17) and interpret K in Proposition 1 as the *value* of the capital stock, i.e., we have to replace K by $\tilde{K} = pK$.

But the concept of Harrod-neutrality no longer fits the situation without further elaboration. Hence to obtain analogies to Proposition 2 and Proposition 3 is a more complicated matter. Suffice it to say that with em-

[11] In the standard versions of the Solow model and the Ramsey model it is assumed that all technological progress has this form - for no other reason than that this is by far the simplest case to analyze.

bodied technological progress, the class of production functions that are consistent with balanced growth is smaller than with disembodied technological progress.

4.5 Concluding remarks

In the Solow model as well as in many other models with disembodied technological progress, a steady state and a balanced growth path imply each other. Indeed, they are two sides of the same process. There *exist* cases, however, where this equivalence does not hold (some open economy models and some models with *embodied* technical change). Therefore, it is recommendable always to maintain a terminological distinction between the two concepts, steady state and balanced growth.[12]

Note that the definition of balanced growth refers to *aggregate* variables. At the same time as there is balanced growth at the aggregate level, *structural change* may occur. That is, a changing sectorial composition of the economy is under certain conditions compatible with balanced growth (in a generalized sense) at the aggregate level, cf. the "Kuznets facts" (see Kongsamut et al., 2001, and Acemoglu, 2009, Chapter 20).

In view of the key importance of Harrod-neutrality, a natural question is: has growth theory uncovered any *endogenous* tendency for technical progress to converge to Harrod-neutrality? Fortunately, in his Chapter 15 Acemoglu outlines a theory about a mechanism entailing such a tendency, the theory of "directed technical change". Jones (2005) suggests an alternative mechanism.

4.6 References

Acemoglu, D., 2009, *Introduction to Modern Economic Growth*, Princeton University Press: Oxford.

Barro, R., and X. Sala-i-Martin, 2004, *Economic Growth*, second edition, MIT Press: Cambridge (Mass.)

Gordon, R. J., 1990. *The Measurement of Durable goods Prices.* Chicago University Press: Chicago.

[12] Here we depart from Acemoglu, p. 65, where he says that he will use the two terms "interchangingly". We also depart from Barro and Sala-i-Martin (2004, pp. 33-34) who *define* a steady state as synonymous with a balanced growth path as the latter was defined above.

Greenwood, J., Z. Hercowitz, and P. Krusell, 1997. Long-Run Implications of Investment-Specific Technological Change. *American Economic Review* 87 (3), 342-362.

Groth, C., and R. Wendner, 2014. Embodied Learning by Investing and Speed of Convergence, *J. of Macroeconomics* (forthcoming).

Jones, C. I., 2005, The shape of production functions and the direction of technical change. *Quarterly Journal of Economics*, no. 2, 517-549.

Jones, C. I., and D. Scrimgeour, 2008, The steady-state growth theorem: Understanding Uzawa (1961), *Review of Economics and Statistics 90* (1), 180-182.

Kongsamut, P., S. Rebelo, and D. Xie, 2001, Beyond balanced growth. *Review of Economic Studies 48,* 869-882.

Schlicht, E., 2006, A variant of Uzawa's theorem, *Economics Bulletin 6,* 1-5.

Uzawa, H., 1961, Neutral inventions and the stability of growth equilibrium, *Review of Economic Studies 28,* No. 2, 117-124.

Wan, H. Y. Jr., 1971, *Economic Growth*, Harcourt Brace: New York.

Chapter 5

The concepts of TFP and growth accounting: Some warnings

5.1 Introduction

This chapter discusses the concepts of Total Factor Productivity, TFP, and TFP growth, and ends up with three warnings regarding uncritical use of them.

First, however, we should provide a precise definition of the TFP *level* which is in fact a tricky concept. Unfortunately, Acemoglu (p. 78) does not make a clear distinction between TFP *level* and TFP *growth.* Moreover, Acemoglu's point of departure (p. 77) assumes *a priori* that the way the production function is time-dependent can be represented by a one-dimensional index, $A(t)$. The TFP concept and the applicability of growth accounting are, however, not limited to this case.

For convenience, in this chapter we treat time as continuous (although the timing of the variables is indicated merely by a subscript).[1]

5.2 TFP level and TFP growth

Let Y_t denote aggregate output (value added in fixed prices) at time t in a sector or the economy as a whole. Suppose Y_t is determined by the function

$$Y_t = \tilde{F}(K_t, H_t; t), \tag{5.1}$$

[1] I thank Niklas Brønager for useful discussions related to this chapter.

where K_t is an aggregate input of physical capital and H_t an index of quality-adjusted labor input.[2] The "quality-adjustment" of the input of labor (man-hours per year) aims at taking educational level and work experience into account. In fact, both output and the two inputs are aggregates of heterogeneous elements. The involved conceptual and measurement difficulties are huge and there are different opinions in the growth accounting literature about how to best deal with them. Here we ignore these problems. The third argument in (5.1) is time, t, indicating that the production function $\tilde{F}(\cdot\,,\cdot\,;t)$ is time-dependent. Thus "shifts in the production function", due to changes in efficiency and technology ("technical change" for short), can be taken into account. We treat time as continuous and assume that $\tilde{F}$ is a neoclassical production function. When the partial derivative of $\tilde{F}$ w.r.t. the third argument is positive, i.e., $\partial\tilde{F}/\partial t > 0$, technical change amounts to technical *progress*. We consider the economy from a purely supply-side perspective.[3]

We shall here concentrate on the fundamentals of TFP and TFP growth. These can in principle be described without taking the heterogeneity and changing quality of the labor input into account. Hence we shall from now on ignore this aspect and simplifying *assume* that labor is homogeneous and labor quality is constant. So (5.1) is reduced to the simpler case,

$$Y_t = \tilde{F}(K_t, L_t; t), \tag{5.2}$$

where L_t is the number of man-hours per year. As to measurement of K_t, some adaptation of the *perpetual inventory method*[4] is typically used, with some correction for under-estimated quality improvements of investment goods in national income accounting. The output measure is (or at least should be) corrected correspondingly, also for under-estimated quality improvements of consumption goods.

[2] Natural resources (land, oil wells, coal in the ground, etc.) constitute a third primary production factor. The role of this factor is in growth accounting often subsumed under K.

[3] Sometimes in growth accounting the left-hand side variable, Y, in (5.2) is the gross product rather than value added. Then non-durable intermediate inputs should be taken into account as a third production factor and enter as an additional argument of $\tilde{F}$ in (5.2). Since non-market production is difficult to measure, the government sector is usually excluded from Y in (5.2). Total Factor Productivity is by some authors called *Multifactor Productivity* and abbreviated MFP.

[4] Cf. Chapter 2.

5.2.1 TFP growth

The notion of Total Factor Productivity at time t, TFP_t, is intended to indicate a *level* of productivity. Nevertheless there is a tendency in the literature to evade a direct definition of this level and instead go straight away to a decomposition of output *growth*. Let us start the same way here but not forget to come back to the issue about what can be meant by the level of TFP.

The growth rate of a variable Z at time t will be denoted $g_{Z,t}$. Taking logs and differentiating w.r.t. t in (5.2) we get

$$\begin{aligned} g_{Y,t} &\equiv \frac{\dot{Y}_t}{Y_t} = \frac{1}{Y_t}\left[\tilde{F}_K(K_t, L_t; t)\dot{K}_t + \tilde{F}_L(K_t, L_t; t)\dot{L}_t + \tilde{F}_t(K_t, L_t; t)\cdot 1\right] \\ &= \frac{K_t\tilde{F}_K(K_t, L_t; t)}{Y_t}g_{K,t} + \frac{L_t\tilde{F}_L(K_t, L_t; t)}{Y_t}g_{L,t} + \frac{\tilde{F}_t(K_t, L_t; t)}{Y_t} \\ &\equiv \varepsilon_{K,t}g_{K,t} + \varepsilon_{L,t}g_{L,t} + \frac{\tilde{F}_t(K_t, L_t; t)}{Y_t}, \end{aligned} \tag{5.3}$$

where $\varepsilon_{K,t}$ and $\varepsilon_{L,t}$ are shorthands for $\varepsilon_K(K_t, L_t; t) \equiv \frac{K_t\tilde{F}_K(K_t,L_t;t)}{\tilde{F}(K_t,L_t;t)}$ and $\varepsilon_L(K_t, L_t; t)$ $\equiv \frac{L_t\tilde{F}_L(K_t,L_t;t)}{\tilde{F}(K_t,L_t;t)}$, respectively, that is, the partial output elasticities w.r.t. the two production factors, evaluated at the factor combination (K_t, L_t) at time t. Finally, $\tilde{F}_t(K_t, L_t; t) \equiv \partial\tilde{F}/\partial t$, that is, the partial derivative w.r.t. the third argument of the function $\tilde{F}$, evaluated at the point (K_t, L_t, t).

The equation (5.3) is the *basic growth-accounting relation*, showing how the output growth rate can be decomposed into the "contribution" from growth in each of the inputs and a residual. The TFP *growth rate* is defined as the residual

$$g_{\text{TFP},t} \equiv g_{Y,t} - (\varepsilon_{K,t}g_{K,t} + \varepsilon_{L,t}g_{L,t}) = \frac{\tilde{F}_t(K_t, L_t; t)}{Y_t}, \tag{5.4}$$

So the TFP growth rate is what is left when from the output growth rate is subtracted the "contribution" from growth in the factor inputs weighted by the output elasticities w.r.t. these inputs. This is sometimes interpreted as reflecting that part of the output growth rate which is *explained* by technical progress. One should be careful, however, not to identify a descriptive accounting relationship with deeper causality. Without a complete model, at most one can say that the TFP growth rate measures that fraction of output growth that is *not directly attributable* to growth in the capital and labor inputs. So:

> The TFP growth rate can be interpreted as reflecting the "*direct* contribution" to current output growth from current technical change (in a broad sense including learning by doing and organizational improvement).

Let us consider how the actual estimation of $g_{\mathrm{TFP},t}$ can be carried out. The output elasticities w.r.t. capital and labor, $\varepsilon_{K,t}$ and $\varepsilon_{L,t}$, will, under perfect competition and absence of externalities, equal the income shares of capital and labor, respectively. Time series for these income shares and for Y, K, and L, hence also for $g_{Y,t}$, $g_{K,t}$, and $g_{L,t}$, can be obtained (directly or with some adaptation) from national income accounts. This allows straightforward measurement of the residual, $g_{\mathrm{TFP},t}$.[5]

The decomposition in (5.4) was introduced already by Solow (1957). Since the TFP growth rate appears as a residual, it is sometimes called the *Solow residual*. As a residual it may reflect the contribution of many things, some wanted (current technical innovation in a broad sense including organizational improvement), others unwanted (such as varying capacity utilization, omitted inputs, measurement errors, and aggregation bias).

5.2.2 The TFP level

Now let us consider the *level* of TFP, that "something" for which we have calculated its growth rate without yet having defined what it really is. But knowing the growth rate of TFP for all t in a certain time interval, we in fact have a differential equation in the TFP level of the form $dx(t)/dt = g(t)x(t)$, namely:

$$d(\mathrm{TFP}_t)/dt = g_{\mathrm{TFP},t} \cdot \mathrm{TFP}_t.$$

The solution of this simple linear differential equation is[6]

$$\mathrm{TFP}_t = \mathrm{TFP}_0 e^{\int_0^t g_{\mathrm{TFP},\tau} d\tau}. \tag{5.5}$$

For a given initial value $\mathrm{TFP}_0 > 0$ (which may be normalized to 1 if desired), the time path of TFP is determined by the right-hand side of (5.5). Consequently:

> The TFP level at time t can interpreted as reflecting the cumulative "*direct* contribution" to output since time 0 from cumulative technical change since time 0.

[5] Of course, data are in discrete time. So to make actual calculations we have to translate (5.4) into discrete time. The weights $\varepsilon_{K,t}$ and $\varepsilon_{L,t}$ can then be estimated by two-years moving averages of the factor income shares as shown in Acemoglu (2009, p. 79).

[6] See Appendix B of Chapter 3 in these lecture notes or Appendix B to Acemoglu.

Why do we say "*direct* contribution"? The reason is that the cumulative technical change since time 0 may also have an *indirect* effect on output, namely via affecting the output elasticities w.r.t. capital and labor, $\varepsilon_{K,t}$ and $\varepsilon_{L,t}$. Through this channel cumulative technical change affects the role of input growth for output growth. This possible indirect effect over time of technical change is not included in the TFP concept.

To clarify the matter we will compare the TFP calculation under Hicks-neutral technical change with that under other forms of technical change.

5.3 The case of Hicks-neutrality*

In the case of Hicks neutrality, by definition, technical change can be represented by the evolution of a one-dimensional variable, B_t, and the production function in (5.2) can be specified as

$$Y_t = \tilde{F}(K_t, L_t; t) = B_t F(K_t, L_t). \tag{5.6}$$

Here the TFP level is at any time, t, identical to the level of B_t if we normalize the initial values of both B and TFP to be the same, i.e., $\text{TFP}_0 = B_0 > 0$. Indeed, calculating the TFP growth rate, (5.4), on the basis of (5.6) gives

$$g_{\text{TFP},t} = \frac{\tilde{F}_t(K_t, L_t; t)}{Y_t} = \frac{\dot{B}_t F(K_t, L_t)}{B_t F(K_t, L_t)} = \frac{\dot{B}_t}{B_t} \equiv g_{B,t}, \tag{5.7}$$

where the second equality comes from the fact that K_t and L_t are kept fixed when the *partial* derivative of $\tilde{F}$ w.r.t. t is calculated. The formula (5.5) now gives

$$\text{TFP}_t = B_0 \cdot e^{\int_0^t g_{B,\tau} d\tau} = B_t.$$

The nice feature of Hicks neutrality is thus that we can write

$$\text{TFP}_t = \frac{\tilde{F}(K_t, L_t; t)}{\tilde{F}(K_t, L_t; 0)} = \frac{B_t F(K_t, L_t)}{B_0 F(K_t, L_t)} = B_t, \tag{5.8}$$

using the normalization $B_0 = 1$. That is:

> *Under Hicks neutrality, current TFP appears as the ratio between the current output level and the hypothetical output level that would have resulted from the current inputs of capital and labor in case of no technical change since time 0.*

So in the case of Hicks neutrality the economic meaning of the TFP level is straightforward. The reason is that under Hicks neutrality the output elasticities w.r.t. capital and labor, $\varepsilon_{K,t}$ and $\varepsilon_{L,t}$, are *independent* of technical change.

5.4 Absence of Hicks-neutrality*

The above very intuitive interpretation of TFP is only valid under Hicks-neutral technical change. Neither under general technical change nor even under Harrod- or Solow-neutral technical change (unless the production function is Cobb-Douglas so that both Harrod and Solow neutrality imply Hicks-neutrality), will current TFP appear as the ratio between the current output level and the hypothetical output level that would have resulted from the current inputs of capital and labor in case of no technical change since time 0.

To see this, let us return to the general time-dependent production function in (5.2). Let X_t denote the ratio between the current output level at time t and the hypothetical output level, $\tilde{F}(K_t, L_t; 0)$, that would have obtained with the current inputs of capital and labor in case of no change in the technology since time 0, i.e.,

$$X_t \equiv \frac{\tilde{F}(K_t, L_t; t)}{\tilde{F}(K_t, L_t; 0)}. \tag{5.9}$$

So X_t can be seen as a factor of joint-productivity growth from time 0 to time t evaluated at the time-t input combination.

If this X_t should always indicate the level of TFP at time t, the growth rate of X_t should equal the growth rate of TFP. Generally, it does not, however. Indeed, defining $G(K_t, L_t) \equiv \tilde{F}(K_t, L_t; 0)$, by the rule for the time derivative of fractions[7], we have

$$\begin{aligned}
g_{X,t} &\equiv \frac{d\tilde{F}(K_t, L_t; t)/dt}{\tilde{F}(K_t, L_t; t)} - \frac{dG(K_t, L_t)/dt}{G(K_t, L_t)} \\
&= \frac{1}{Y_t}\left[\tilde{F}_K(K_t, L_t; t)\dot{K}_t + \tilde{F}_L(K_t, L_t; t)\dot{L}_t + \tilde{F}_t(K_t, L_t; t)\cdot 1\right] \\
&\quad -\frac{1}{G(K_t, L_t)}\left[G_K(K_t, L_t)\dot{K}_t + G_L(K_t, L_t)\dot{L}_t\right] \\
&= \varepsilon_K(K_t, L_t; t)g_{K,t} + \varepsilon_L(K_t, L_t; t)g_{L,t} + \frac{\tilde{F}_t(K_t, L_t; t)}{Y_t} \\
&\quad -(\varepsilon_K(K_t, L_t; 0)g_{K,t} + \varepsilon_L(K_t, L_t; 0)g_{L,t}) \qquad (5.10) \\
&= \left(\varepsilon_K(K_t, L_t; t) - \varepsilon_K(K_t, L_t; 0)\right)g_{K,t} + (\varepsilon_L(K_t, L_t; t) - \varepsilon_L(K_t, L_t; 0))g_{L,t} + g_{\text{TFP},t} \\
&\neq g_{\text{TFP},t} \qquad \text{generally,}
\end{aligned}$$

where $g_{\text{TFP},t}$ is given in (5.4). Unless the partial output elasticities w.r.t. capital and labor, respectively, are unaffected by technical change, the conclusion is that TFP_t will differ from our X_t defined in (5.9). So:

[7] See Appendix A to Chapter 3 of these lecture notes.

> *In the absence of Hicks neutrality, current TFP does not generally appear as the ratio between the current output level and the hypothetical output level that would have resulted from the current inputs of capital and labor in case of no technical change since time 0.*

A closer look at X_t vs. TFP$_t$

As X_t in (5.9) is the time-t output arising from the time-t inputs relative to the fictional time-0 output from the same inputs, we consider X_t along with TFP as two alternative joint-productivity indices. From (5.10) we see that

$$g_{\text{TFP},t} = g_{X,t} - (\varepsilon_K(K_t, L_t; t) - \varepsilon_K(K_t, L_t; 0))\, g_{K,t} - (\varepsilon_L(K_t, L_t; t) - \varepsilon_L(K_t, L_t; 0)) g_{L,t}.$$

So the growth rate of TFP equals the growth rate of the joint-productivity index X corrected for the cumulative impact of technical change since time 0 on the direct contribution to time-t output growth from time-t input growth. This impact comes about when the output elasticities w.r.t. capital and labor, respectively, are affected by technical change, that is, when $\varepsilon_K(K_t, L_t; t) \neq \varepsilon_K(K_t, L_t; 0)$ and/or $\varepsilon_L(K_t, L_t; t) \neq \varepsilon_L(K_t, L_t; 0)$.

Under Hicks-neutral technical change there will be no correction because the output elasticities are *independent* of technical change. In this case TFP coincides with the index X. In the absence of Hicks-neutrality the two indices differ. This is why we in Section 2.2 characterized the TFP level as the cumulative "*direct* contribution" to output since time 0 from cumulative technical change, thus excluding the possible indirect contribution coming about via the potential effect of technical change on the output elasticities w.r.t. capital and labor and thereby on the contribution to output from input growth.

Given that the joint-productivity index X is the more intuitive joint-productivity measure, why is TFP the more popular measure? There are at least two reasons for this. First, it can be shown that the TFP measure has more convenient balanced growth properties. Second, X is more difficult to measure. To see this we substitute (5.3) into (5.10) to get

$$g_{X,t} = g_{Y,t} - (\varepsilon_K(K_t, L_t; 0) g_{K,t} + \varepsilon_L(K_t, L_t; 0) g_{L,t}). \tag{5.11}$$

The relevant output elasticities, $\varepsilon_K(K_t, L_t; 0) \equiv \frac{K_t \tilde{F}_K(K_t, L_t; 0)}{\tilde{F}(K_t, L_t; 0)}$ and $\varepsilon_L(K_t, L_t; 0) \equiv \frac{L_t \tilde{F}_L(K_t, L_t; 0)}{\tilde{F}(K_t, L_t; 0)}$, are hypothetical constructs, referring to the technology as it was at time 0, but with the factor combination observed at time t, not at time 0. The nice thing about the Solow residual is that under the assumptions of perfect competition and absence of externalities, it allows measurement by using data on prices and quantities alone, that is, without knowledge of the production function. To evaluate g_X, however, we need estimates of the hypothetical output elasticities, $\varepsilon_K(K_t, L_t; 0)$ and $\varepsilon_L(K_t, L_t; 0)$. This requires knowledge about how the output elasticities depend on the factor combination and time, respectively, that is, knowledge about the production function.

Now to the warnings concerning application of the TFP measure.

5.5 Three warnings

Balanced growth at the aggregate level, hence Harrod neutrality, seems to characterize the growth experience of the UK and US over at least a century (Kongsamut et al., 2001; Attfield and Temple, 2010). At the same time the aggregate elasticity of factor substitution is generally estimated to be significantly less than one (see, e.g., Antras, 2004). This amounts to rejection of the Cobb-Douglas specification of the aggregate production function and so, at the aggregate level, Harrod neutrality rules out Hicks neutrality.

Warning 1 Since Hicks-neutrality is empirically doubtful at the aggregate level, TFP_t can often *not* be identified with the simple intuitive joint-productivity measure X_t, defined in (5.9) above.

Warning 2 When Harrod neutrality obtains, relative TFP growth rates across sectors or countries can be quite deceptive.

Suppose there are n countries and that country i has the aggregate production function

$$Y_{it} = F^{(i)}(K_{it}, A_t L_{it}) \qquad i = 1, 2, ..., n,$$

where $F^{(i)}$ is a neoclassical production function with CRS and A_t is the level of labor-augmenting technology which, for simplicity, we assume shared by all the countries (these are open and "close" to each other). So technical progress is Harrod-neutral. Let the growth rate of A be a constant $g > 0$. Many models imply that $\tilde{k}_i \equiv K_{it}/(A_t L_{it})$ tends to a constant, $\tilde{k}_i^*$, in the long run, which we assume is also the case here. Then, for $t \to \infty$, $k_{it} \equiv K_{it}/L_{it}$ $\equiv \tilde{k}_{it} A_t$ where $\tilde{k}_{it} \to \tilde{k}_i^*$ and $y_{it} \equiv Y_{it}/L_{it} \equiv \tilde{y}_{it} A_t$ where $\tilde{y}_{it} \to \tilde{y}_i^* = f^{(i)}(\tilde{k}_i^*)$; here $f^{(i)}$ is the production function on intensive form. So in the long run g_{k_i} and g_{y_i} tend to $g_A = g$.

Formula (5.4) then gives the TFP growth rate of country i in the long run as

$$\begin{aligned} g_{\mathrm{TFP}_i} &\equiv g_{Y_i} - (\alpha_i^* g_{K_i} + (1-\alpha_i^*) g_{L_i}) = g_{Y_i} - g_{L_i} - \alpha_i^*(g_{K_i} - g_{L_i}) \\ &= g_{y_i} - \alpha_i^* g_{k_i} = (1-\alpha_i^*) g, \qquad (5.12) \end{aligned}$$

where α_i^* is the output elasticity w.r.t. capital, $f^{(i)\prime}(\tilde{k}_i)\tilde{k}_i/f^{(i)}(\tilde{k}_i)$, evaluated at $\tilde{k}_i = \tilde{k}_i^*$. Under labor-augmenting technical progress, the TFP growth rate thus varies negatively with the output elasticity w.r.t. capital (the capital income share under perfect competition). Owing to differences in product and industry composition, the countries have different α_i^*'s. In view of (5.12), for two different countries, i and j, we get

$$\frac{TFP_i}{TFP_j} \rightarrow \begin{cases} \infty \text{ if } \alpha_i^* < \alpha_j^*, \\ 1 \text{ if } \alpha_i^* = \alpha_j^*, \\ 0 \text{ if } \alpha_i^* > \alpha_j^*, \end{cases} \qquad (5.13)$$

for $t \rightarrow \infty$.[8] Thus, in spite of long-run growth in the essential variable, y, being the same across the countries, their TFP growth rates are very different. Countries with low α_i^*'s appear to be technologically very dynamic and countries with high α_i^*'s appear to be lagging behind. It is all due to the difference in α across countries; a higher α just means that a larger fraction of $g_{y_i} = g_{k_i} = g$ becomes "explained" by g_{k_i} in the growth accounting (5.12), leaving a smaller residual. And the level of α has nothing to do with technical progress.

We conclude that comparison of TFP levels across countries or time may misrepresent the intuitive meaning of productivity and technical progress when output elasticities w.r.t. capital differ and technical progress is Harrod-neutral (even if technical progress were at the same time Hicks-neutral as is the case with a Cobb-Douglas specification). It may be more reasonable to just compare levels of Y/L across countries and time.

Warning 3 Growth accounting is - as the name says - just about accounting and measurement. So do not confuse growth *accounting* with *causality* in growth analysis. To talk about causality we need a theoretical model supported by the data. On the basis of such a model we can say that this or that set of exogenous factors through the propagation mechanisms of the model cause this or that phenomenon, including economic growth. In contrast, considering the growth accounting identity (5.3) in itself, none of the terms have

[8] If F is Cobb-Douglas with output elasticity w.r.t. capital equal to α_i, the result in (5.12) can be derived more directly by first defining $B_t = A_t^{1-\alpha_i}$, then writing the production function in the Hicks-neutral form (5.6), and finally use (5.7).

priority over the others w.r.t. a causal role. And there are important omitted variables. There are simple illustrations in Exercises III.1 and III.2.

In a complete model with exogenous technical progress, part of $g_{K,t}$ will be *induced* by this technical progress. If technical progress is endogenous through learning by investing, as in Arrow (1962), there is mutual causation between $g_{K,t}$ and technical progress. Yet another kind of model might explain both technical progress and capital accumulation through R&D, cf. the survey by Barro (1999).

5.6 References

Antràs, P., 2004, Is the U.S. aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution, *Contributions to Macroeconomics*, vol. 4, no. 1, 1-34.

Attfield, C., and J.R.W. temple, 2010, Balanced growth and the great ratios: New evidence for the US and UK, *J. of Macroeconomics*, vol. 32, 937-956.

Barro, R.J., 1999, Notes on growth accounting, *J. of Economic Growth*, vol. 4 (2), 119-137.

Bernard, A. B., and C. I. Jones, 1996a, Technology and Convergence, *Economic Journal*, vol. 106, 1037-1044.

Bernard, A. B., and C. I. Jones, 1996b, Comparing Apples to Oranges: productivity convergence and measurement across industries and countries, *American Economic Review,* vol. 86, no. 5, 1216-1238.

Greenwood, J., and P. Krusell, 2006, Growth accounting with investment-specific technological progress: A discussion of two approaches, *J. of Monetary Economics.*

Hercowitz, Z., 1998, The 'embodiment' controversy: A review essay, *J. of Monetary Economics,* vol. 41, 217-224.

Hulten, C.R., 2001, Total factor productivity. A short biography. In: Hulten, C.R., E.R. Dean, and M. Harper (eds.), *New Developments in Productivity Analysis,* Chicago: University of Chicago Press, 2001, 1-47.

Kongsamut, P., S. Rebelo, and D. Xie, 2001, Beyond Balanced Growth, *Review of Economic Studies*, vol. 68, 869-882.

Sakellaris, P., and D.J. Wilson, 2004, Quantifying embodied technological progress, *Review of Economic Dynamics*, vol. 7, 1-26.

Solow, R.M., 1957, Technical change and the aggregate production function, *Review of Economics and Statistics,* vol. 39, 312-20.

Chapter 6

Transitional dynamics. Barro-style growth regressions

In this chapter we discuss three issues, all of which are related to the transitional dynamics of a growth model:

- Do poor countries necessarily tend to approach their steady state from below?
- How fast (or rather how slow) are the transitional dynamics in a growth model?
- What exactly is the theoretical foundation for a Barro-style growth regression analysis?

The Solow growth model may serve as the analytical point of departure for the first two issues and to some extent also for the third.

6.1 Point of departure: the Solow model

As is well-known, the fundamental differential equation for the Solow model is

$$\dot{\tilde{k}}(t) = sf(\tilde{k}(t)) - (\delta + g + n)\tilde{k}(t), \qquad \tilde{k}(0) = \tilde{k}_0 > 0, \tag{6.1}$$

where $\tilde{k}(t) \equiv K(t)/(A(t)L(t))$, $f(\tilde{k}(t)) \equiv F(\tilde{k}(t), 1)$, $A(t) = A_0 e^{gt}$, and $L(t) = L_0 e^{nt}$ (standard notation). The production function F is neoclassical with CRS and the parameters satisfy $0 < s < 1$ and $\delta + g + n > 0$. The production function on intensive form, f, therefore satisfies $f(0) \geq 0, f' > 0, f'' < 0$, and

$$\lim_{\tilde{k} \to 0} f'(\tilde{k}) > \frac{\delta + g + n}{s} > \lim_{\tilde{k} \to \infty} f'(\tilde{k}). \tag{A1}$$

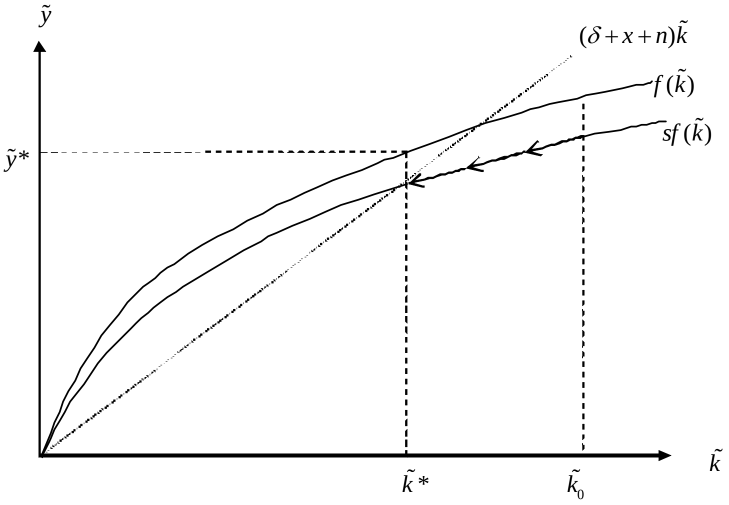


Figure 6.1: Note: x means g.

Then there exists a unique non-trivial steady state, $\tilde{k}^* > 0$, that is, a unique positive solution to the equation

$$sf(\tilde{k}^*) = (\delta + g + n)\tilde{k}^*. \tag{6.2}$$

Furthermore, given an arbitrary $\tilde{k}_0 > 0$, we have for all $t \geq 0$,

$$\dot{\tilde{k}}(t) \gtreqless 0 \text{ for } \tilde{k}(t) \lesseqgtr \tilde{k}^*, \tag{6.3}$$

respectively. The steady state, $\tilde{k}^*$, is thus *globally asymptotically stable* in the sense that for all $\tilde{k}_0 > 0$, $\lim_{t\to\infty} \tilde{k}(t) = k^*$ and this convergence is *monotonic* (in the sense that $\tilde{k}(t) - \tilde{k}^*$ does not change sign during the adjustment process).

From now on the dating of $\tilde{k}$ is suppressed unless needed for clarity. Figure 6.1 illustrates the dynamics as seen from the perspective of (6.1) (in this and the two next figures, x should read g. Figure 6.2 illustrates the dynamics emerging when we rewrite (6.1) this way:

$$\dot{\tilde{k}} = s\left(f(\tilde{k}) - \frac{\delta + g + n}{s}\tilde{k}\right) \gtreqless 0 \text{ for } \tilde{k} \lesseqgtr \tilde{k}^*.$$

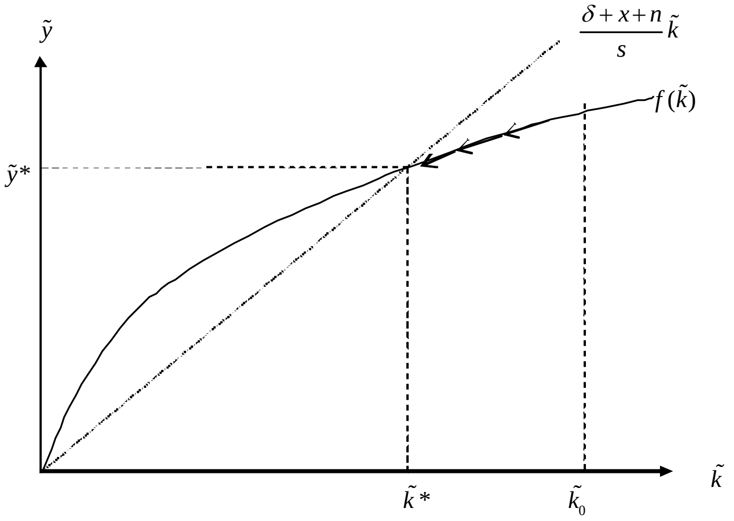


Figure 6.2: Note: x means g.

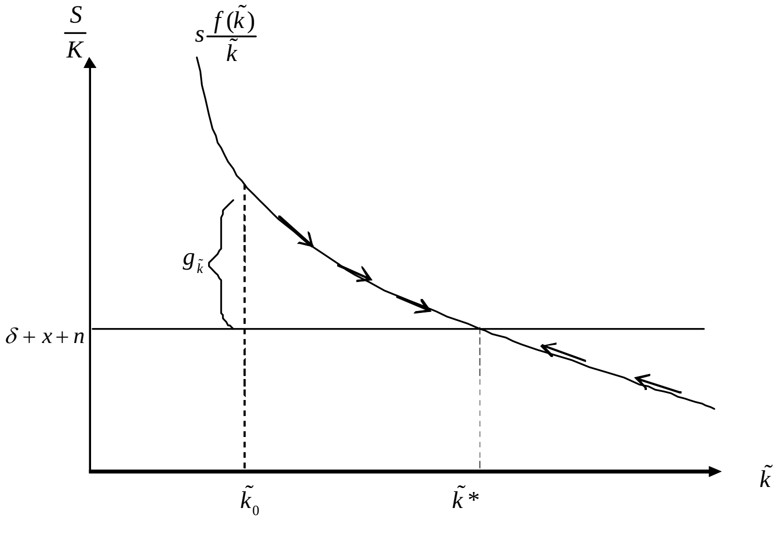


Figure 6.3: Note: x means g.

In Figure 6.3 yet another illustration is exhibited, based on rewriting (6.1) this way:

$$\frac{\dot{\tilde{k}}}{\tilde{k}} = s\frac{f(\tilde{k})}{\tilde{k}} - (\delta + g + n),$$

where $sf(\tilde{k})/\tilde{k}$ is gross saving per unit of capital, $S/K \equiv (Y - C)/K$.

An important variable in the analysis of the adjustment process towards steady state is the output elasticity w.r.t. capital:

$$\frac{K}{Y}\frac{\partial Y}{\partial K} = \frac{\tilde{k}}{f(\tilde{k})}f'(\tilde{k}) \equiv \varepsilon(\tilde{k}), \tag{6.4}$$

where $0 < \varepsilon(\tilde{k}) < 1$ for all $\tilde{k} > 0$.

6.2 Do poor countries tend to approach their steady state from below?

From some textbooks (for instance Barro and Sala-i-Martin, 2004) one gets the impression that poor countries tend to approach their steady state *from below*. But this is *not* what the Penn World Table data seems to indicate. And from a theoretical point of view the size of $\tilde{k}_0$ relative to $\tilde{k}^*$ is certainly ambiguous, whether the country is rich or poor. To see this, consider a poor country with initial effective capital intensity

$$\tilde{k}_0 \equiv \frac{K_0}{A_0 L_0}.$$

Here K_0/L_0 will typically be small for a poor country (the country has not yet accumulated much capital relative to its fast-growing population). The technology level, A_0, however, *also* tends to be small for a poor country. Hence, whether we should expect $\tilde{k}_0 < \tilde{k}^*$ or $\tilde{k}_0 > \tilde{k}^*$ is not obvious *apriori*. Or equivalently: whether we should expect that a poor country's GDP at an arbitrary point in time grows at a rate higher or lower than the country's steady-state growth rate, $g + n$, is not obvious *apriori*.

While Figure 6.3 illustrates the case where the inequality $\tilde{k}_0 < \tilde{k}^*$ holds, Figure 6.1 and 6.2 illustrate the opposite case. There *exists* some empirical evidence indicating that poor countries tend to approach their steady state *from above*. Indeed, Cho and Graham (1996) find that "on average, countries with a lower income per adult are above their steady-state positions, while countries with a higher income are below their steady-state positions".

The prejudice that poor countries *apriori* should tend to approach their steady state from below seems to come from a confusion of conditional and unconditional β convergence. The Solow model predicts - and data supports - that within a group of countries with similar structural characteristics (approximately the same f, A_0, g, s, n, and δ), the initially poorer countries will grow faster than the richer countries. This is because the poorer countries (small $y(0) = f(\tilde{k}_0)A_0$) will be the countries with relatively small initial capital-labor ratio, k_0. As all the countries in the group have approximately the same A_0, the poorer countries thus have $\tilde{k}_0 \equiv k_0/A_0$ relatively small, i.e., $\tilde{k}_0 < \tilde{k}^*$. From $y \equiv Y/L \equiv \tilde{y}A = f(\tilde{k})A$ follows that the growth rate in output per worker of these poor countries tends to exceed g. Indeed, we have generally

$$\frac{\dot{y}}{y} = \frac{\dot{\tilde{y}}}{\tilde{y}} + g = \frac{f'(\tilde{k})\dot{\tilde{k}}}{f(\tilde{k})} + g \gtreqless g \text{ for } \dot{\tilde{k}} \gtreqless 0, \text{ i.e., for } \tilde{k} \lesseqgtr \tilde{k}^*.$$

So, *within* the group, the poor countries tend to approach the steady state, $\tilde{k}^*$, *from below*.

The countries in the world as a whole, however, differ a lot w.r.t. their structural characteristics, including their A_0. Unconditional β convergence is definitely rejected by the data. Then there is no reason to expect the poorer countries to have $\tilde{k}_0 < \tilde{k}^*$ rather than $\tilde{k}_0 > \tilde{k}^*$. Indeed, according to the mentioned study by Cho and Graham (1996), it turns out that the data for the relatively poor countries favors the latter inequality rather than the first.

6.3 Convergence speed and adjustment time

Our next issue is: How fast (or rather how slow) are the transitional dynamics in a growth model? To put it another way: according to a given growth model with convergence, how fast does the economy approach its steady state? The answer turns out to be: not very fast - to say the least. This is a rather general conclusion and is confirmed by the empirics: adjustment processes in a growth context are quite time consuming.

In Acemoglu's textbook we meet the concept of speed of convergence at p. 54 (under an alternative name, rate of adjustment) and p. 81 (in connection with Barro-style growth regressions). Here we shall go more into detail with the issue of speed of convergence.

Again the Solow model is our frame of reference. We search for a formula for the *speed of convergence* of $\tilde{k}(t)$ and $y(t)/y^*(t)$ in a closed economy described by the Solow model. So our analysis is concerned with *within-country* *convergence*: how fast do variables such as $\tilde{k}$ and y approach their steady state paths in a closed economy? The key adjustment mechanism is linked to diminishing returns to capital (falling marginal productivity of capital) in the process of capital accumulation. The problem of *cross-country convergence* (which is what "β convergence" and "σ convergence" are about) is in principle more complex because also such mechanisms as technological catching-up and cross-country factor movements are involved.

6.3.1 Convergence speed for $\tilde{k}(t)$

The ratio of $\dot{\tilde{k}}(t)$ to $(\tilde{k}(t) - \tilde{k}^*) \neq 0$ can be written

$$\frac{\dot{\tilde{k}}(t)}{\tilde{k}(t) - \tilde{k}^*} = \frac{d(\tilde{k}(t) - \tilde{k}^*)/dt}{\tilde{k}(t) - \tilde{k}^*}, \tag{6.5}$$

since $d\tilde{k}^*/dt = 0$. We define the *instantaneous speed of convergence* at time t as the (proportionate) rate of *decline* of the distance $\left|\tilde{k}(t) - \tilde{k}^*\right|$ at time t and we denote it $\mathrm{SOC}_t(\tilde{k})$.[1] Thus,

$$\mathrm{SOC}_t(\tilde{k}) \equiv -\frac{d\left(\left|\tilde{k}(t) - \tilde{k}^*\right|\right)/dt}{\left|\tilde{k}(t) - \tilde{k}^*\right|} = -\frac{d(\tilde{k}(t) - \tilde{k}^*)/dt}{\tilde{k}(t) - \tilde{k}^*}, \tag{6.6}$$

where the equality sign is valid for monotonic convergence.

Generally, $\mathrm{SOC}_t(\tilde{k})$ depends on both the absolute size of the difference $\tilde{k} - \tilde{k}^*$ at time t and its sign. But if the difference is already "small", $\mathrm{SOC}_t(\tilde{k})$ will be "almost" constant for increasing t and we can find an approximate measure for it. Let the function $\varphi(\tilde{k})$ be defined by $\varphi(\tilde{k}) \equiv sf(\tilde{k}) - m\tilde{k}$, where $m \equiv \delta + g + n$. A first-order Taylor approximation of $\varphi(\tilde{k})$ around $\tilde{k} = \tilde{k}^*$ gives

$$\varphi(\tilde{k}) \approx \varphi(\tilde{k}^*) + \varphi'(\tilde{k}^*)(\tilde{k} - \tilde{k}^*) = 0 + (sf'(\tilde{k}^*) - m)(\tilde{k} - \tilde{k}^*).$$

[1] Synonyms for speed of convergence are *rate of convergence, rate of adjustment* or *adjustment speed.*

For $\tilde{k}$ in a small neighborhood of the steady state, $\tilde{k}^*$, we thus have

$$\begin{aligned}\dot{\tilde{k}} &= \varphi(\tilde{k}) \approx (sf'(\tilde{k}^*) - m)(\tilde{k} - \tilde{k}^*) \\ &= (\frac{sf'(\tilde{k}^*)}{m} - 1)m(\tilde{k} - \tilde{k}^*) \\ &= (\frac{\tilde{k}^* f'(\tilde{k}^*)}{f(\tilde{k}^*)} - 1)m(\tilde{k} - \tilde{k}^*) \qquad \text{(from (6.2))} \\ &\equiv (\varepsilon(\tilde{k}^*) - 1)m(\tilde{k} - \tilde{k}^*) \qquad \text{(from (6.4))}.\end{aligned}$$

Applying the definition (6.6) and the identity $m \equiv \delta + g + n$, we now get

$$\mathrm{SOC}_t(\tilde{k}) = -\frac{d(\tilde{k}(t) - \tilde{k}^*)/dt}{\tilde{k}(t) - \tilde{k}^*} \approx (1 - \varepsilon(\tilde{k}^*))(\delta + g + n) \equiv \beta(\tilde{k}^*) > 0. \quad (6.7)$$

This result tells us how fast, approximately, the economy approaches its steady state if it starts "close" to it. If, for example, $\beta(\tilde{k}^*) = 0.02$ per year, then 2 percent of the gap between $\tilde{k}(t)$ and $\tilde{k}^*$ vanishes per year. We also see that everything else equal, a higher output elasticity w.r.t. capital implies a lower speed of convergence.

In the limit, for $\left|\tilde{k} - \tilde{k}^*\right| \rightarrow 0$, the instantaneous speed of convergence coincides with what is called the *asymptotic speed of convergence*, defined as

$$\mathrm{SOC}^*(\tilde{k}) \equiv \lim_{\left|\tilde{k}-\tilde{k}^*\right| \rightarrow 0} \mathrm{SOC}_t(\tilde{k}) = \beta(\tilde{k}^*). \quad (6.8)$$

Multiplying through by $-(\tilde{k}(t) - \tilde{k}^*)$, the equation (6.7) takes the form of a homogeneous linear differential equation (with constant coefficient), $\dot{x}(t) = \beta x(t)$, the solution of which is $x(t) = x(0)e^{\beta t}$. With $x(t) = \tilde{k}(t) - \tilde{k}^*$ and "=" replaced by "≈", we get in the present case

$$\tilde{k}(t) - \tilde{k}^* \approx (\tilde{k}(0) - \tilde{k}^*)e^{-\beta(\tilde{k}^*)t}. \quad (6.9)$$

This is the approximative time path for the gap between $\tilde{k}(t)$ and $\tilde{k}^*$ and shows how the gap becomes smaller and smaller at the rate $\beta(\tilde{k}^*)$.

One of the reasons that the speed of convergence is important is that it indicates what weight should be placed on transitional dynamics of a growth model relative to the steady-state behavior. The speed of convergence matters for instance for the evaluation of growth-promoting policies. In growth models with diminishing marginal productivity of production factors, successful growth-promoting policies have transitory growth effects and permanent level effects. Slower convergence implies that the full benefits are slower to arrive.

6.3.2 Convergence speed for $\log \tilde{k}(t)$

We have found an approximate expression for the convergence speed of $\tilde{k}$. Since models in empirical analysis and applied theory are often based on log-linearization, we might ask what the speed of convergence of $\log \tilde{k}$ is. The answer is: approximately the same and asymptotically exactly the same as that of $\tilde{k}$ itself! Let us see why.

A first-order Taylor approximation of $\log \tilde{k}(t)$ around $\tilde{k} = \tilde{k}^*$ gives

$$\log \tilde{k}(t) \approx \log \tilde{k}^* + \frac{1}{\tilde{k}^*}(\tilde{k}(t) - \tilde{k}^*). \tag{6.10}$$

By definition

$$\begin{aligned} \mathrm{SOC}_t(\log \tilde{k}) &= -\frac{d(\log \tilde{k}(t) - \log \tilde{k}^*)/dt}{\log \tilde{k}(t) - \log \tilde{k}^*} = -\frac{d\tilde{k}(t)/dt}{\tilde{k}(t)(\log \tilde{k}(t) - \log \tilde{k}^*)} \\ &\approx -\frac{d\tilde{k}(t)/dt}{\tilde{k}(t)\frac{\tilde{k}(t)-\tilde{k}^*}{\tilde{k}^*}} = \frac{\tilde{k}^*}{\tilde{k}(t)}\mathrm{SOC}_t(\tilde{k}) \to \mathrm{SOC}^*(\tilde{k}) = \beta(\tilde{k}^*) \text{ for } \tilde{k}(t) \to \tilde{k}^*, \end{aligned} \tag{6.11}$$

where in the second line we have used, first, the approximation (6.10), second, the definition in (6.7), and third, the definition in (6.8).

So, at least in a neighborhood of the steady state, the instantaneous rate of decline of the logarithmic distance of $\tilde{k}$ to the steady-state value of $\tilde{k}$ approximates the instantaneous rate of decline of the distance of $\tilde{k}$ itself to its steady-state value. The asymptotic speed of convergence of $\log \tilde{k}$ coincides with that of $\tilde{k}$ itself and is exactly $\beta(\tilde{k}^*)$.

In the Cobb-Douglas case (where $\varepsilon(\tilde{k}^*)$ is a constant, say α) it is possible to find an explicit solution to the Solow model, see Acemoglu p. 53 and Exercise II.2. It turns out that the instantaneous speed of convergence in a finite distance from the steady state is a constant and equals the asymptotic speed of convergence, $(1 - \alpha)(\delta + g + n)$.

6.3.3 Convergence speed for $y(t)/y^*(t)$

The variable which we are interested in is usually not so much $\tilde{k}$ in itself, but rather labor productivity, $y(t) \equiv \tilde{y}(t)A(t)$. In the interesting case where $g > 0$, labor productivity does not converge towards a constant. We therefore focus on the ratio $y(t)/y^*(t)$, where $y^*(t)$ denotes the hypothetical value of labor productivity at time t, conditional on the economy being on its steady-state path, i.e.,

$$y^*(t) \equiv \tilde{y}^* A(t). \tag{6.12}$$

We have

$$\frac{y(t)}{y^*(t)} \equiv \frac{\tilde{y}(t)A(t)}{\tilde{y}^*A(t)} = \frac{\tilde{y}(t)}{\tilde{y}^*}. \tag{6.13}$$

As $\tilde{y}(t) \rightarrow \tilde{y}^*$ for $t \rightarrow \infty$, the ratio $y(t)/y^*(t)$ converges towards 1 for $t \rightarrow \infty$.

Taking logs on both sides of (6.13), we get

$$\begin{aligned}
\log\frac{y(t)}{y^*(t)} &= \log\frac{\tilde{y}(t)}{\tilde{y}^*} = \log\tilde{y}(t) - \log\tilde{y}^* \\
&\approx \log\tilde{y}^* + \frac{1}{\tilde{y}^*}(\tilde{y}(t) - y^*) - \log\tilde{y}^* \quad \text{(first-order Taylor approx. of } \log\tilde{y}\text{)} \\
&= \frac{1}{f(\tilde{k}^*)}(f(\tilde{k}(t)) - f(\tilde{k}^*)) \\
&\approx \frac{1}{f(\tilde{k}^*)}(f(\tilde{k}^*) + f'(\tilde{k}^*)(\tilde{k}(t) - \tilde{k}^*) - f(\tilde{k}^*)) \quad \text{(first-order approx. of } f(\tilde{k})) \\
&= \frac{\tilde{k}^*f'(\tilde{k}^*)}{f(\tilde{k}^*)}\frac{\tilde{k}(t) - \tilde{k}^*}{\tilde{k}^*} \equiv \varepsilon(\tilde{k}^*)\frac{\tilde{k}(t) - \tilde{k}^*}{\tilde{k}^*} \\
&\approx \varepsilon(\tilde{k}^*)(\log\tilde{k}(t) - \log\tilde{k}^*) \quad \text{(by (6.10))}. \qquad (6.14)
\end{aligned}$$

Multiplying through by $-(\log\tilde{k}(t) - \log\tilde{k}^*)$ in (6.11) and carrying out the differentiation w.r.t. time, we find an approximate expression for the growth rate of $\tilde{k}$,

$$\begin{aligned}
\frac{d\tilde{k}(t)/dt}{\tilde{k}(t)} &\equiv g_{\tilde{k}}(t) \approx -\frac{\tilde{k}^*}{\tilde{k}(t)}\mathrm{SOC}_t(\tilde{k})(\log\tilde{k}(t) - \log\tilde{k}^*) \\
&\rightarrow -\beta(\tilde{k}^*)(\log\tilde{k}(t) - \log\tilde{k}^*) \quad \text{for } \tilde{k}(t) \rightarrow \tilde{k}^*, \qquad (6.15)
\end{aligned}$$

where the convergence follows from the last part of (6.11). We now calculate the time derivative on both sides of (6.14) to get

$$\begin{aligned}
d(\log\frac{y(t)}{y^*(t)})/dt &= d(\log\frac{\tilde{y}(t)}{\tilde{y}^*})/dt = \frac{d\tilde{y}(t)/dt}{\tilde{y}(t)} \equiv g_{\tilde{y}}(t) \\
&\approx \varepsilon(\tilde{k}^*)g_{\tilde{k}}(t) \approx -\varepsilon(\tilde{k}^*)\beta(\tilde{k}^*)(\log\tilde{k}(t) - \log\tilde{k}^*). \quad (6.16)
\end{aligned}$$

from (6.15). Dividing through by $-\log(y(t)/y^*(t))$ in this expression, taking (6.14) into account, gives

$$-\frac{d(\log\frac{y(t)}{y^*(t)})/dt}{\log\frac{y(t)}{y^*(t)}} = -\frac{d(\log\frac{y(t)}{y^*(t)} - \log 1)/dt}{\log\frac{y(t)}{y^*(t)} - \log 1} \equiv \mathrm{SOC}_t(\log\frac{y}{y^*}) \approx \beta(\tilde{k}^*), \tag{6.17}$$

in view of $\log 1 = 0$. So the logarithmic distance of y from its value on the steady-state path at time t has approximately the same rate of decline as the logarithmic distance of $\tilde{k}$ from $\tilde{k}$'s value on the steady-state path at time t. The asymptotic speed of convergence for $\log y(t)/y^*(t)$ is exactly the same as that for $\tilde{k}$, namely $\beta(\tilde{k}^*)$.

What about the speed of convergence of $y(t)/y^*(t)$ itself? Here the same principle as in (6.11) applies. The asymptotic speed of convergence for $\log(y(t)/y^*(t))$ is the same as that for $y(t)/y^*(t)$ (and vice versa), namely $\beta(\tilde{k}^*)$.

With one year as our time unit, standard parameter values are: $g = 0.02$, $n = 0.01$, $\delta = 0.05$, and $\varepsilon(\tilde{k}^*) = 1/3$. We then get $\beta(\tilde{k}^*) = (1-\varepsilon(\tilde{k}^*))(\delta+g+n)$ $= 0.053$ per year. In the empirical Chapter 11 of Barro and Sala-i-Martin (2004), it is argued that a lower value of $\beta(\tilde{k}^*)$, say 0.02 per year, fits the data better. This requires $\varepsilon(\tilde{k}^*) = 0.75$. Such a high value of $\varepsilon(\tilde{k}^*)$ ($\approx$ the income share of capital) may seem difficult to defend. But if we reinterpret K in the Solow model so as to include *human* capital (skills embodied in human beings and acquired through education and learning by doing), a value of $\varepsilon(\tilde{k}^*)$ at that level may not be far out.

6.3.4 Adjustment time

Let τ_ω be the time that it takes for the fraction $\omega \in (0,1)$ of the initial gap between $\tilde{k}$ and $\tilde{k}^*$ to be eliminated, i.e., τ_ω satisfies the equation

$$\frac{\left|\tilde{k}(\tau_\omega) - \tilde{k}^*\right|}{\left|\tilde{k}(0) - \tilde{k}^*\right|} = \frac{\tilde{k}(\tau_\omega) - \tilde{k}^*}{\tilde{k}(0) - \tilde{k}^*} = 1 - \omega, \tag{6.18}$$

where $1 - \omega$ is the fraction of the initial gap still remaining at time τ_ω. In (6.18) we have applied that $sign(\tilde{k}(t) - \tilde{k}^*) = sign(\tilde{k}(0) - \tilde{k}^*)$ in view of monotonic convergence.

By (6.9), we have

$$\tilde{k}(\tau_\omega) - \tilde{k}^* \approx (\tilde{k}(0) - \tilde{k}^*)e^{-\beta(\tilde{k}^*)\tau_\omega}.$$

In view of (6.18), this implies

$$1 - \omega \approx e^{-\beta(\tilde{k}^*)\tau_\omega}.$$

Taking logs on both sides and solving for τ_ω gives

$$\tau_\omega \approx -\frac{\log(1-\omega)}{\beta(\tilde{k}^*)}. \tag{6.19}$$

This is the approximate *adjustment time* required for $\tilde{k}$ to eliminate the fraction ω of the initial distance of $\tilde{k}$ to its steady-state value, $\tilde{k}^*$, when the adjustment speed (speed of convergence) is $\beta(\tilde{k}^*)$.

Often we consider the *half-life* of the adjustment, that is, the time it takes for half of the initial gap to be eliminated. To find the half-life of the adjustment of $\tilde{k}$, we put $\omega = \frac{1}{2}$ in (6.19). Again we use one year as our time unit. With the previous parameter values, we have $\beta(\tilde{k}^*) = 0.053$ per year and thus

$$\tau_{\frac{1}{2}} \approx -\frac{\log \frac{1}{2}}{0.053} \approx \frac{0.69}{0.053} = 13,1 \text{ years.}$$

As noted above, Barro and Sala-i-Martin (2004) estimate the asymptotic speed of convergence to be $\beta(\tilde{k}^*) = 0.02$ per year. With this value, the half-life is approximately

$$\tau_{\frac{1}{2}} \approx -\frac{\log \frac{1}{2}}{0.02} \approx \frac{0.69}{0.02} = 34.7 \text{ years.}$$

And the time needed to eliminate three quarters of the initial distance to steady state, $\tau_{3/4}$, will then be about 70 years ($= 2 \cdot 35$ years, since $1 - 3/4 = \frac{1}{2} \cdot \frac{1}{2}$).

Among empirical analysts there is not general agreement about the size of $\beta(\tilde{k}^*)$. Some authors, for example Islam (1995), using a panel data approach, find speeds of convergence considerably larger, between 0.05 and 0.09. McQuinne and Whelan (2007) get similar results. There is a growing realization that the speed of convergence differs across periods and groups of countries. Perhaps an empirically reasonable range is $0.02 < \beta(\tilde{k}^*) < 0.09$. Correspondingly, a reasonable range for the half-life of the adjustment will be 7.6 years $< \tau_{\frac{1}{2}} < 34.7$ years.

Most of the empirical studies of convergence use a variety of cross-country regression analysis of the kind described in the next section. Yet the theoretical frame of reference is often the Solow model - or its extension with human capital (Mankiw et al., 1992). These models are closed economy models with exogenous technical progress and deal with "within-country" convergence. It is not obvious that they constitute an appropriate framework for studying cross-country convergence in a globalized world where capital mobility and to some extent also labor mobility are important and where some countries are pushing the technological frontier further out, while others try to imitate and catch up. At least one should be aware that the empirical estimates obtained may reflect mechanisms in addition to the falling marginal productivity of capital in the process of capital accumulation.

6.4 Barro-style growth regressions

Barro-style growth regression analysis, which became very popular in the 1990s, draws upon transitional dynamics aspects (including the speed of convergence) as well as steady state aspects of neoclassical growth theory (for instance the Solow model or the Ramsey model).

In his Section 3.2 of Chapter 3 Acemoglu presents Barro's growth regression equations in an unconventional form, see Acemoglu's equations (3.12), (3.13), and (3.14). The left-hand side appears as if it is just the growth rate of y (output per unit of labor) from one year to the next. But the true left-hand side of a Barro equation is the average compound annual growth rate of y over many years. Moreover, since Acemoglu's text is very brief about the formal links to the underlying neoclassical theory of transitional dynamics, we will spell the details out here.

Most of the preparatory work has already been done above. The point of departure is a neoclassical one-sector growth model for a closed economy:

$$\dot{\tilde{k}}(t) = s(\tilde{k}(t))f(\tilde{k}(t)) - (\delta + g + n)k(t), \qquad \tilde{k}(0) = \tilde{k}_0 > 0, \text{ given}, \tag{6.20}$$

where $\tilde{k}(t) \equiv K(t)/(A(t)L(t))$, $A(t) = A_0 e^{gt}$, and $L(t) = L_0 e^{nt}$ as above. The Solow model is the special case where the saving-income ratio, $s(\tilde{k}(t))$, is a constant $s \in (0, 1)$.

It is assumed that the model, (6.20), generates monotonic convergence, i.e., $\tilde{k}(t) \to \tilde{k}^* > 0$ for $t \to \infty$. Applying again a first-order Taylor approximation, as in Section 3.1, and taking into account that $s(\tilde{k})$ now may depend on $\tilde{k}$, as for instance it generally does in the Ramsey model, we find the asymptotic speed of convergence for $\tilde{k}$ to be

$$\mathrm{SOC}^*(\tilde{k}) = (1 - \varepsilon(\tilde{k}^*) - \eta(\tilde{k}^*))(\delta + g + n) \equiv \beta(\tilde{k}^*) > 0, \tag{*}$$

where $\eta(\tilde{k}^*) \equiv \tilde{k}^* s'(\tilde{k}^*)/s(\tilde{k}^*)$ is the elasticity of the saving-income ratio w.r.t. the effective capital intensity, evaluated at $\tilde{k} = \tilde{k}^*$. (In case of the Ramsey model, one can alternatively use the fact that $\mathrm{SOC}^*(\tilde{k})$ equals the absolute value of the negative eigenvalue of the Jacobian matrix associated with the dynamic system of the model, evaluated in the steady state. For a fully specified Ramsey model this eigenvalue can be numerically calculated by an appropriate computer algorithm; in the Cobb-Douglas case there exists even an explicit algebraic formula for the eigenvalue, see Barro and Sala-i-Martin, 2004). In a neighborhood of the steady state, the previous formulas remain valid with $\beta(\tilde{k}^*)$ defined as in (*). The asymptotic speed of convergence of for example $y(t)/y^*(t)$ is thus $\beta(\tilde{k}^*)$ as given in (*). For notational convenience, we will just denote it β, interpreted as a derived parameter, i.e.,

$$\beta = (1 - \varepsilon(\tilde{k}^*) - \eta(\tilde{k}^*))(\delta + g + n) \equiv \beta(\tilde{k}^*). \tag{6.21}$$

In case of the Solow model, $\eta(\tilde{k}^*) = 0$ and we are back in Section 3.

In view of $y(t) \equiv \tilde{y}(t)A(t)$, we have $g_y(t) = g_{\tilde{y}}(t) + g$. By (6.16) and the definition of β,

$$g_y(t) \approx g - \varepsilon(\tilde{k}^*)\beta(\log \tilde{k}(t) - \log \tilde{k}^*) \approx g - \beta(\log y(t) - \log y^*(t)), \tag{6.22}$$

where the last approximation comes from (6.14). This generalizes Acemoglu's Equation (3.10) (recall that Acemoglu concentrates on the Solow model and that his k^* is the same as our $\tilde{k}^*$).

With the horizontal axis representing time, Figure 6.4 gives an illustration of these transitional dynamics. As $g_y(t) = d\log y(t)/dt$ and $g = d\log y^*(t)/dt$, (6.22) is equivalent with

$$\frac{d(\log y(t) - \log y^*(t))}{dt} \approx -\beta(\log y(t) - \log y^*(t)). \tag{6.23}$$

So again we have a simple differential equation of the form $\dot{x}(t) = \beta x(t)$, the solution of which is $x(t) = x(0)e^{\beta t}$. The solution of (6.23) is thus

$$\log y(t) - \log y^*(t) \approx (\log y(0) - \log y^*(0))e^{-\beta t}.$$

As $y^*(t) = y^*(0)e^{gt}$, this can written

$$\log y(t) \approx \log y^*(0) + gt + (\log y(0) - \log y^*(0))e^{-\beta t}. \tag{6.24}$$

The solid curve in Figure 6.4 depicts the evolution of $\log y(t)$ in the case where $\tilde{k}_0 < \tilde{k}^*$ (note that $\log y^*(0) = \log f(\tilde{k}^*) + \log A_0$). The dotted curve exemplifies the case where $\tilde{k}_0 > \tilde{k}^*$. The figure illustrates per capita income convergence: low initial income is associated with a high subsequent growth rate which, however, diminishes along with the diminishing logarithmic distance of per capita income to its level on the steady state path.

For convenience, we will from now on treat (6.24) as an equality. Subtracting $\log y(0)$ on both sides, we get

$$\begin{aligned} \log y(t) - \log y(0) &= \log y^*(0) - \log y(0) + gt + (\log y(0) - \log y^*(0))e^{-\beta t} \\ &= gt - (1 - e^{-\beta t})(\log y(0) - \log y^*(0)). \end{aligned}$$

Dividing through by $t > 0$ gives

$$\frac{\log y(t) - \log y(0)}{t} = g - \frac{1 - e^{-\beta t}}{t}(\log y(0) - \log y^*(0)). \tag{6.25}$$

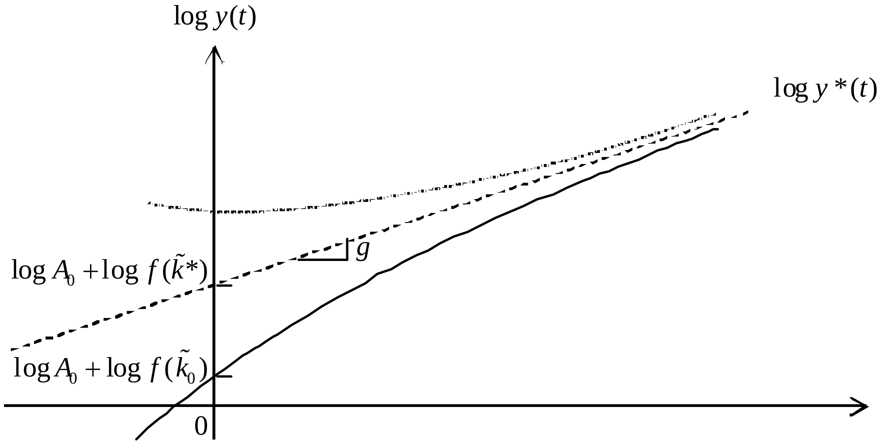

$\log y(t)$
$\log y^*(t)$
$\log A_0 + \log f(\tilde{k}^*)$
g
$\log A_0 + \log f(\tilde{k}_0)$
0
t

Figure 6.4

On the left-hand side appears the average compound annual growth rate of y from period 0 to period t, which we will denote $\bar{g}_y(0,t)$. On the right-hand side appears the initial distance of $\log y$ to its hypothetical level along the steady state path. The coefficient, $-(1-e^{-\beta t})/t$, to this distance is negative and approaches zero for $t \to \infty$. Thus (6.25) is a translation into growth form of the convergence of $\log y_t$ towards the steady-state path, $\log y_t^*$, in the theoretical model without shocks. Rearranging the right-hand side, we get

$$\bar{g}_y(0,t) = g + \frac{1-e^{-\beta t}}{t}\log y^*(0) - \frac{1-e^{-\beta t}}{t}\log y(0) \equiv b^0 + b^1 \log y(0),$$

where both the constant $b^0 \equiv g + \left[(1-e^{-\beta t})/t\right] \log y^*(0)$ and the coefficient $b^1 \equiv -(1-e^{-\beta t})/t$ are determined by "structural characteristics". Indeed, β is determined by δ, g, n, $\varepsilon(\tilde{k}^*)$, and $\eta(\tilde{k}^*)$ through (6.21), and $y^*(0)$ is determined by A_0 and $f(\tilde{k}^*)$ through (6.12), where, in turn, $\tilde{k}^*$ is determined by the steady state condition $s(\tilde{k}^*)f(\tilde{k}^*) = (\delta + g + n)\tilde{k}^*$, $s(\tilde{k}^*)$ being the saving-income ratio in the steady state.

With data for N countries, $i = 1, 2,\ldots, N$, a test of the *unconditional convergence hypothesis* may be based on the regression equation

$$\bar{g}_{y_i}(0,t) = b^0 + b^1 \log y_i(0) + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2), \tag{6.26}$$

where ϵ_i is the error term. This can be seen as a Barro growth regression equation in its simplest form. For countries in the entire world, the theoretical hypothesis $b^1 < 0$ is clearly not supported (or, to use the language of statistics, the null hypothesis, $b^1 = 0$, is not rejected).[2]

Allowing for the considered countries having different structural characteristics, the Barro growth regression equation takes the form

$$\bar{g}_{y_i}(0,t) = b_i^0 + b^1 \log y_i(0) + \epsilon_i, \qquad b^1 < 0, \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2). \tag{6.27}$$

In this "fixed effects" form, the equation has been applied for a test of the *conditional convergence hypothesis,* $b^1 < 0$, often supporting this hypothesis.

From the estimate of b^1 the implied estimate of the asymptotic speed of convergence, β, is readily obtained through the formula $b^1 \equiv (1 - e^{-\beta t})/t$. Even β, and therefore also the slope, b^1, does depend, theoretically, on country-specific structural characteristics. But the sensitivity on these do not generally seem large enough to blur the analysis based on (6.27) which abstracts from this dependency.

With the aim of testing hypotheses about growth determinants, Barro (1991) and Barro and Sala-i-Martin (1992, 2004) decompose b_i^0 so as to reflect the role of a set of measurable potentially causal variables,

$$b_i^0 = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_m x_{im},$$

where the α's are the coefficients and the x's are the potentially causal variables.[3] These variables could be measurable Solow-type parameters among those appearing in (6.20) or a broader set of determinants, including for instance the educational level in the labor force, and institutional variables like rule of law and democracy. Some studies include the initial within-country inequality in income or wealth among the x's and extend the theoretical framework correspondingly.[4]

From an econometric point of view there are several problematic features in regressions of Barro's form (also called the β convergence approach). These problems are discussed in Acemoglu pp. 82-85.

6.5 References

Alesina, A., and D. Rodrik, 1994, Distributive politics and economic growth, *Quarterly Journal of Economics*, vol. 109, 465-490.

[2] Cf. Acemoglu, p. 16. For the OECD countries, however, b^1 is definitely estimated to be negative (cf. Acemoglu, p. 17).

[3] Note that our α vector is called β in Acemoglu, pp. 83-84. So Acemoglu's β is to be distinquished from our β which denotes the asymptotic speed of convergence.

[4] See, e.g., Alesina and Rodrik (1994) and Perotti (1996), who argue for a negative relationship between inequality and growth. Forbes (2000), however, rejects that there should be a robust negative correlation between the two.

Barro, R. J., 1991, Economic growth in a cross section of countries, *Quarterly Journal of Economics,* vol. 106, 407-443.

Barro, R. J., X. Sala-i-Martin, 1992, Convergence, *Journal of Political Economy,* vol. 100, 223-251.

Barro, R., and X. Sala-i-Martin, 2004, *Economic Growth.* Second edition, MIT Press: Cambridge (Mass.).

Cho, D., and S. Graham, 1996, The other side of conditional convergence, *Economics Letters,* vol. 50, 285-290.

Forbes, K.J., 2000, A reassessment of the relationship between inequality and growth, *American Economic Review*, vol. 90, no. 4, 869-87.

Groth, C., and R. Wendner, 2012, Embodied learning by investing and speed of convergence, Working Paper.

Islam, N., 1995, Growth Empirics. A Panel Data Approach, *Quarterly Journal of Economics*, vol. 110, 1127-1170.

McQinn, K., K. Whelan, 2007, Conditional Convergence and the Dynamics of the Capital-Output Ratio, *Journal of Economic Growth,* vol. 12, 159-184.

Perotti, R., 1996, Growth, income distribution, and democracy: What the data say, *Journal of Economic Growth,* vol. 1, 149-188.

Chapter 7

Michael Kremer's population-breeds-ideas model

This chapter relates to Section 2 of Acemoglu's Chapter 4 and explains the details of what may also be called the *Simon-Kremer version* of the population-breeds-ideas model (cf. Acemoglu, p. 114).

7.1 The model

Suppose a pre-industrial economy can be described by:

$$\begin{aligned} Y_t &= A_t^\sigma L_t^\alpha Z^{1-\alpha}, \qquad \sigma > 0, 0 < \alpha < 1, & (7.1)\\ \dot{A}_t &= \lambda A_t^\varepsilon L_t, \qquad \lambda > 0, 0 < \varepsilon \leq 1, \qquad A_0 > 0 \text{ given}, & (7.2)\\ L_t &= \frac{Y_t}{\bar{y}} \equiv \varphi Y_t, \qquad \bar{y} > 0, & (7.3) \end{aligned}$$

where Y is aggregate output, A the level of technical knowledge, L the labor force (= population), Z the amount of land (fixed), and $\bar{y}$ subsistence minimum (so the φ in Acemoglu's equation (4.2) is simply the inverse of the subsistence minimum). Both Z and $\bar{y}$ are considered as constant parameters. Time is continuous and it is understood that a kind of Malthusian population mechanism (see below) is operative behind the scene.

The exclusion of capital from the aggregate production function, (7.1), reflects the presumption that capital (tools etc.) is quantitatively of minor importance in a pre-industrial economy. In accordance with the replication argument, the production function has CRS w.r.t. the rival inputs, labor and land. The factor A_t^σ measures total factor productivity. In view of (7.2), the technology level, A_t, is rising over time. The increase in A_t per time unit is seen to be an increasing function of the size of the population. This reflects the hypothesis that population breeds ideas; these are non-rival and enter the pool of technical knowledge available for society as a whole. The rate per capita, λA^{ε}, by which population breeds ideas is an increasing function of the already existing level of technical knowledge. This reflects the hypothesis that the larger is the stock of ideas the easier do new ideas arise (perhaps by combination of existing ideas).

Equation (7.3) is a shortcut description of a Malthusian population mechanism. Suppose the true mechanism is

$$\dot{L}_t = \beta(y_t - \bar{y})L_t \gtreqless 0 \quad \text{for} \quad y_t \gtreqless \bar{y}, \tag{7.4}$$

where $\beta > 0$ is the speed of adjustment, $y_t \equiv Y_t/L_t$ is per capita income, and $\bar{y} > 0$ is subsistence minimum. A rise in y_t above $\bar{y}$ will lead to increases in L_t, thereby generating downward pressure on Y_t/L_t and perhaps end up pushing y_t below $\bar{y}$. When this happens, population will be decreasing for a while and so return towards its sustainable level, $Y_t/\bar{y}$. Equation (7.3) treats this mechanism as if the population instantaneously adjusts to its sustainable level (as if $\beta \to \infty$). The model hereby gives a long-run picture, ignoring the Malthusian ups and downs in population and per capita income about the subsistence minimum. The important feature is that the technology level and thereby Y_t as well as the sustainable population will be rising over time. This speeds up the arrival of new ideas and so raises Y_t even faster although per-capita income remains at its long-run level, $\bar{y}$.[1]

For simplicity, we now normalize the constant Z to be 1.

7.2 Law of motion

The dynamics of the model can be reduced to one differential equation, the law of motion of technical knowledge. By (7.3), $L_t = \varphi Y_t = \varphi A_t^{\sigma} L_t^{\alpha}$. Consequently $L_t^{1-\alpha} = \varphi A_t^{\sigma}$ so that

$$L_t = \varphi^{\frac{1}{1-\alpha}} A_t^{\frac{\sigma}{1-\alpha}}. \tag{7.5}$$

Substituting this into (7.2) gives the law of motion of technical knowledge:

$$\dot{A}_t = \lambda \varphi^{\frac{1}{1-\alpha}} A_t^{\varepsilon + \frac{\sigma}{1-\alpha}} \equiv \hat{\lambda} A_t^{\varepsilon + \frac{\sigma}{1-\alpha}}. \tag{7.6}$$

[1] Extending the model with the institution of private ownership and competitive markets, the absence of a growing standard of living corresponds to the doctrine from classical economics called the *iron law of wages.* This is the theory (from Malthus and Ricardo) that scarce natural resources and the pressure from population growth causes real wages to remain at subsistende level.

Define $\mu \equiv \varepsilon + \frac{\sigma}{1-\alpha}$ and assume $\mu > 1$. Then (7.6) can be written

$$\dot{A}_t = \hat{\lambda} A_t^{\mu}, \tag{7.7}$$

which is a nonlinear differential equation in A.[2] Let $x \equiv A^{1-\mu}$. Then

$$\dot{x}_t = (1-\mu)A_t^{-\mu}\hat{\lambda}A_t^{\mu} = (1-\mu)\hat{\lambda}, \tag{7.8}$$

a constant. To find x_t from this, we only need simple integration:

$$x_t = x_0 + \int_0^t \dot{x}_\tau d\tau = x_0 + (1-\mu)\hat{\lambda}t.$$

As $A = x^{\frac{1}{1-\mu}}$ and $x_0 = A_0^{1-\mu}$, this implies

$$A_t = x_t^{\frac{1}{1-\mu}} = \left[A_0^{1-\mu} + (1-\mu)\hat{\lambda}t\right]^{\frac{1}{1-\mu}} = \frac{1}{\left[A_0^{1-\mu} - (\mu-1)\hat{\lambda}t\right]^{\frac{1}{\mu-1}}}. \tag{7.9}$$

7.3 The inevitable ending of the Malthusian regime

The result (7.9) helps us in understanding why the Malthusian regime must come to an end (at least if the model is an acceptable description of the Malthusian regime).

Although to begin with, A_t may grow extremely slowly, the growth in A_t will be *accelerating* because of the *positive feedback* (visible in (7.2)) from both rising population and rising A_t. Indeed, since $\mu > 1$, the denominator in (7.9) will be decreasing over time and approach zero in finite time, namely as t approaches the finite value $t^* = A_0^{1-\mu}/((\mu-1)\hat{\lambda})$. Figure 7.1 illustrates. The evolution of technical knowledge becomes explosive as t approaches t^*.

It follows from (7.5) and (7.1) that explosive growth in A implies explosive growth in L and Y, respectively. The acceleration in the evolution of Y will sooner or later make Y move fast enough so that the Malthusian population mechanism (which for biological reasons has to be slow) can not catch up. Then, what was in the Malthusian population mechanism, equation (7.4), earlier only a transitory excess of y_t over $\bar{y}$, will sooner or later become a permanent excess and take the form of sustained growth in y_t. This is known as the *take-off.*

[2] The differential equation, (7.7), is a special case of what is known as the *Bernoulli equation.* In spite of being a non-linear differential equation, the Bernoulli equation always has an explicit solution.

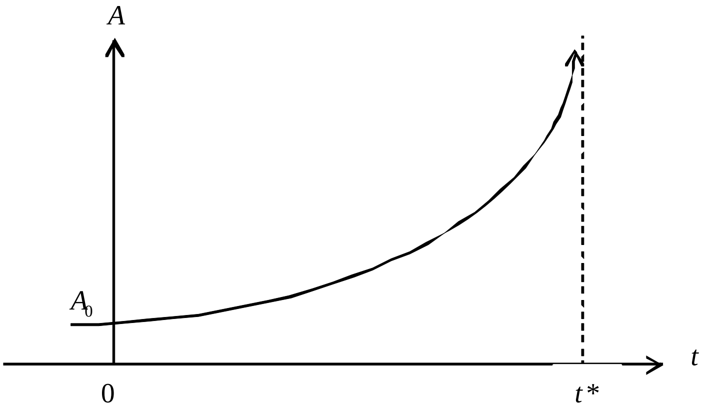


Figure 7.1

According to equation (7.4) the take-off should lead to a permanently rising population growth rate. As economic history has testified, however, along with the rising standard of living the demographics changed The *demographic transition* took place with fertility declining faster than mortality. This results in completely different dynamics about which the present model has nothing to say.[3] As to the demographic transition as such, explanations suggested by economists include: higher opportunity costs of raising children, the trade-off between "quality" (educational level) of the offspring and their "quantity" (Becker, Galor), skill-biased technical change, and improved contraception technology.

7.4 Closing remarks

The present model is about dynamics in the Malthusian regime of the pre-industrial epoch. The story told by the model is the following. When the feedback parameter, μ, is above one, the Malthusian regime has to come to an end because the battle between scarcity of land (or natural resources more generally) and technological progress will inevitably be won by the latter.[4]

The cases $\mu < 1$ and $\mu = 1$ are considered in Exercise III.3. The case $\mu = 1$ corresponds to Acemoglu's first version (p. 113) of the population-breeds-ideas model. In that version, σ has the value $1 - \alpha$ and $\varepsilon = 0$ (two

[3] Kremer (1993), however, also includes an extended model taking some of these changed dynamics into account.

[4] The mathematical background for the explosion result is explained in the appendix.

arbitrary knife-edge conditions). Then a constant growth rate in A, L, and Y is the result and y remains at $\bar{y}$ forever. Take-off never takes place.

On the basis of demographers' estimates of the growth in global population over most of human history, Kremer (1993) finds empirical support for $\mu > 1$. Indeed, in the opposite case, $\mu \leq 1$, there would *not* have been a rising world population growth rate since one million years B.C. to the industrial revolution. The data in Kremer (1993, p. 682) indicates that the population growth rate has been more or less proportional to the size of population until about the1960s.

7.5 Appendix

Mathematically, the background for the explosion result is that the solution to a first-order differential equation of the form $\dot{x}(t) = \alpha + bx(t)^c$, $c > 1$, $b \neq 0$, $x(0) = x_0$ given, is always explosive. Indeed, the solution, $x = x(t)$, will have the property that $x(t) \rightarrow \pm\infty$ for $t \rightarrow t^*$ for some fixed $t^* > 0$; and thereby the solution is defined only on a bounded time interval.

Take the differential equation $\dot{x}(t) = 1 + x(t)^2$ as an example. As is well-known, the solution is $x(t) = \tan t = \sin t / \cos t$, defined on the interval $(-\pi/2, \pi/2)$.

7.6 References

Becker, G. S., ...

Galor, O., 2011, *Unified Growth Theory*, Princeton University Press.

Kremer, M., 1993, Population Growth and Technological Change: One Million B.C. to 1990, *Quarterly Journal of Economics 108*, No. 3.

List of contents to be continued.