Economic Growth Lecture Note 1.

04.02.2013. Christian Groth

Introduction to Economic Growth[1]

This introductory lecture is a refresher on basic concepts.

Section 1 defines Economic Growth as a field of economics. In Section 2 formulas for calculation of compound average growth rates in discrete and continuous time are presented. Section 3 briefly presents two sets of stylized facts. Finally, Section 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 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 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

[1] I thank Niklas Brønager for useful comments on this lecture note.

the widening of the opportunities of human beings to lead freer and more worthwhile lives.

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 some 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 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.

1.2 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

	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: 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.

1870 and let $g_{Y/N}$ denote the average (compound) growth rate of Y/N per year since 1870. Table 1 gives these growth rates for four countries.

Figure 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 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, 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 while continuous compounding is used in the two figures. These two alternative methods of calculation are explained in the next section.

2 Calculation of the average growth rate

2.1 Discrete compounding

Let $y \equiv Y/N$. The average growth rate of y from period 0 to period t, with discrete compounding, is that G which satisfies

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

"Compounding" means adding the one-period "net return" (like with "compound interest"). Obviously, G will generally be quite different from the arithmetic average of the period-by-period growth rates. To underline this, G is sometimes called the "average compound growth rate" or the "geometric average growth rate".

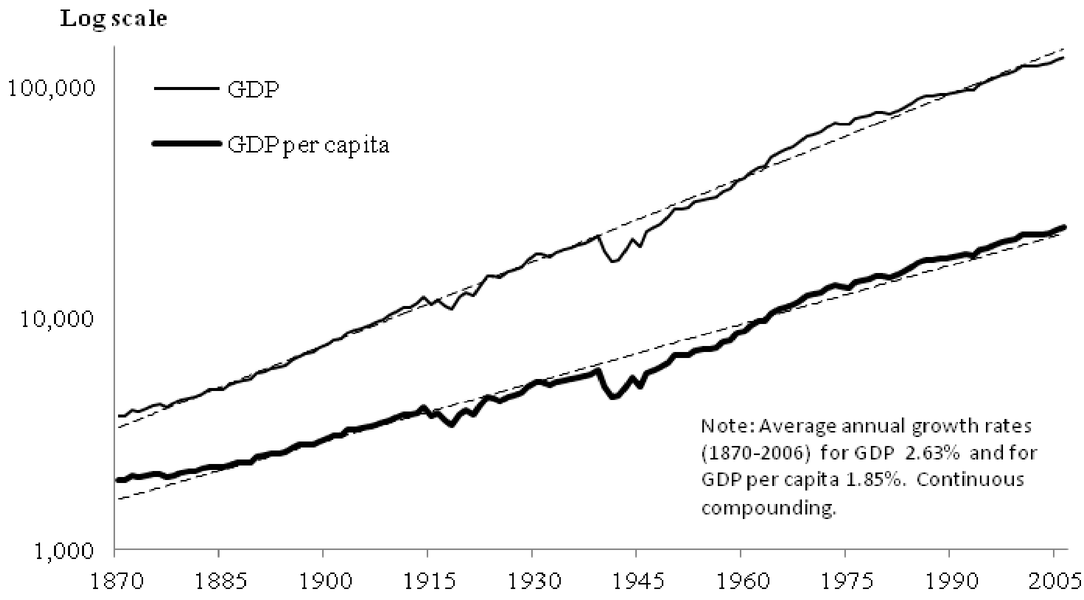


Figure 1: GDP and GDP per capita (1990 International Geary-Khamis dollars) in Denmark, 1870-2006. Source: Maddison, A. (2009). Statistics on World Population, GDP and Per Capita GDP, 1-2006 AD, www.ggdc.net/maddison.

Using a pocket calculator, the following steps in the calculation of G may be convenient. Take logs on both sides of (1) to get

$$\begin{aligned} \ln \frac{y_t}{y_0} &= t \ln(1+G) \;\Rightarrow \\ \ln(1+G) &= \frac{\ln \frac{y_t}{y_0}}{t} \;\Rightarrow \qquad (3) \\ G &= \text{antilog}\left(\frac{\ln \frac{y_t}{y_0}}{t}\right) - 1. \qquad (4) \end{aligned}$$

Note that t in the formulas (2) and (4) equals the number of periods *minus 1*.

2.2 Continuous compounding

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

$$y_t = y_0 e^{gt}, \qquad (5)$$

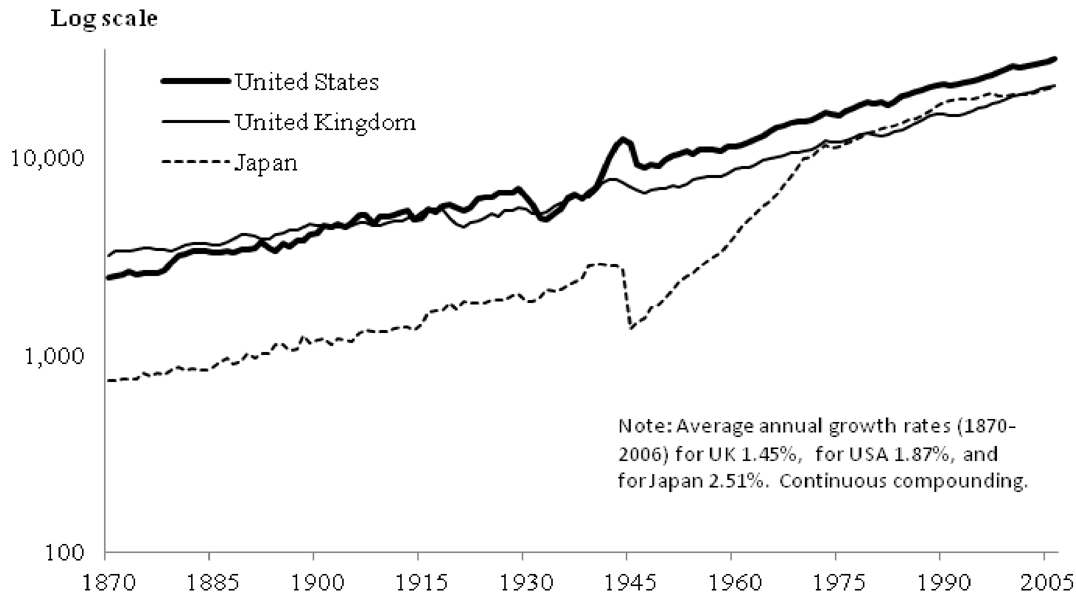


Figure 2: GDP per capita (1990 International Geary-Khamis dollars) in UK, USA and Japan, 1870-2006. Source: Maddison, A. (2009). Statistics on World Population, GDP and Per Capita GDP, 1-2006 AD, www.ggdc.net/maddison.

where e denotes the Euler number, i.e., the base of the natural logarithm.[2] Solving for g gives

$$g = \frac{\ln \frac{y_t}{y_0}}{t} = \frac{\ln y_t - \ln y_0}{t}. \tag{6}$$

The first formula in (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 (3) we see that $g = \ln(1 + G) < G$ for $G > 0$. Yet, by a first-order Taylor approximation about $G = 0$ we have

$$g = \ln(1 + G) \approx G \quad \text{for } G \text{ "small"}. \tag{7}$$

For a given data set the G calculated from (2) will be slightly above the g calculated from (6), cf. the mentioned difference between the growth rates in Table 1 and those in Figure 1 and Figure 2. The reason is that a given growth force is more powerful when

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

compounding is continuous rather than discrete. Anyway, the difference between G and g is usually immaterial. 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 $G = 0.10$ (think of China in recent decades), the corresponding g is 0.0953. But if G stands for the inflation rate and there is high inflation, the difference is substantial. During hyperinflation the monthly inflation rate may be, say, $G = 100\%$, but the corresponding g is only 69%.

For calculation with a pocket calculator the continuous compounding formula, (6), is slightly easier to use than the discrete compounding formulas, whether (2) or (4).

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 (3):

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

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 (6):

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

With $g = 0.0187$, cf. Table 1, we find

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

Again, with a pocket calculator the continuous compounding formula is slightly easier to use.

3 Some stylized facts of economic growth

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

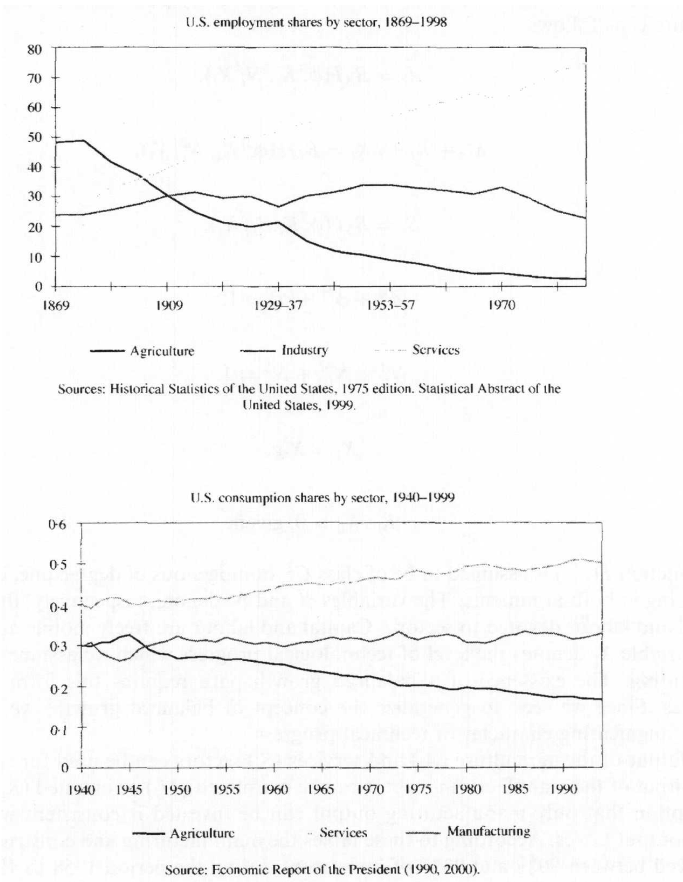


FIGURE 2
The Kuznets facts

Figure 3: The Kuznets facts. Source: Kongsamut et al., Beyond Balanced Growth, Review of Economic Studies, vol. 68, Oct. 2001, 869-82.

and communication). The shares of total consumption expenditure going to these three sectors have moved similarly. Differences in the demand elasticities are 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 3 illustrate the Kuznets facts.

3.2 Kaldor's stylized facts

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

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 grows at a more or less constant rate exceeding the growth rate of the labor input.

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).

4 Concepts of income convergence

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

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).

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.

[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.)

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).

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 = labor force 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*

[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.

of y across n countries in a given year is

$$\sigma_y \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{8}$$

where

$$\bar{y} \equiv \frac{\sum_i y_i}{n}, \tag{9}$$

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:

$$\sigma_{y/\bar{y}} \equiv \sqrt{\frac{1}{n}\sum_i(\frac{y_i}{\bar{y}} - 1)^2}. \tag{10}$$

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

$$CV_y \equiv \frac{\sigma_y}{\bar{y}}, \tag{11}$$

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

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

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.,

$$\sigma_{\ln y} \equiv \sqrt{\frac{1}{n}\sum_i(\ln y_i - \ln y^*)^2}, \tag{12}$$

where

$$\ln y^* \equiv \frac{\sum_i \ln y_i}{n}. \tag{13}$$

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).

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 (12) is modified to

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

where

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

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.

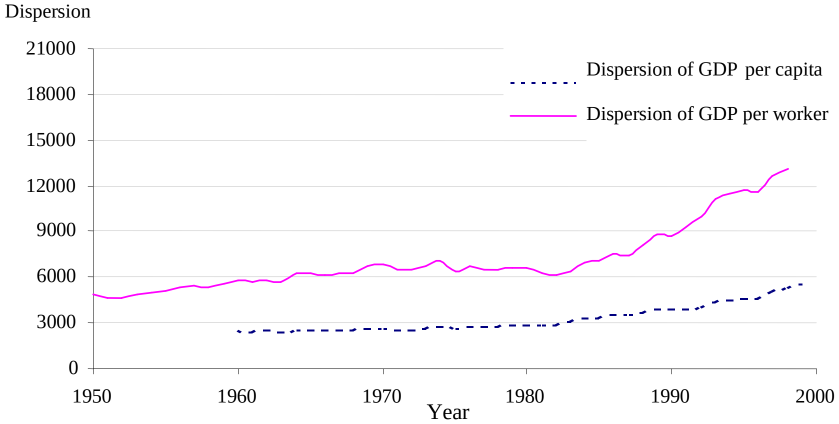
What the precise meaning of "structural characteristics" is, will depend on what model of the countries the researcher has in mind. According to the Solow model, a set of relevant "structural characteristics" are: the aggregate production function, the initial level of technology, the rate of technical 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 across-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.

4.5 A bird's-eye view of the data on convergence/divergence

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

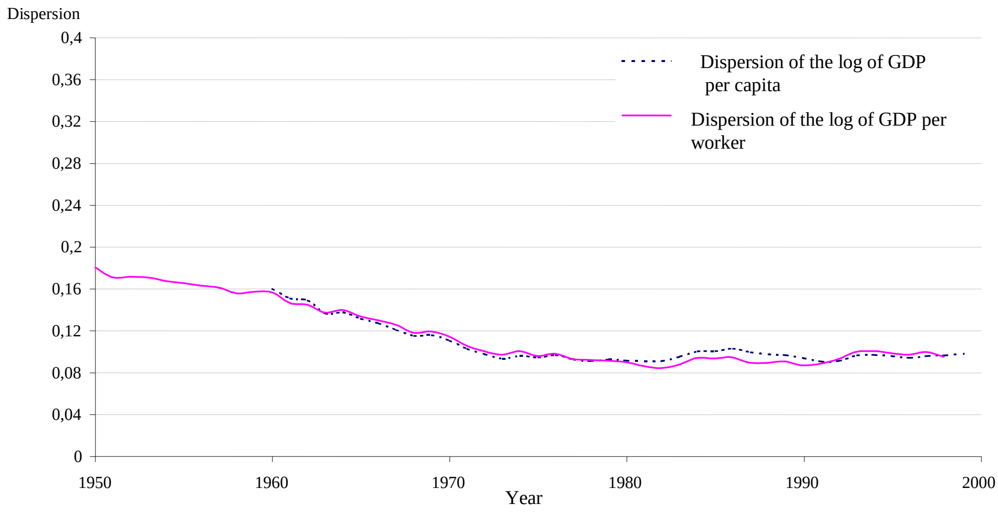
Figure 4 shows the time profile for the standard deviation of y itself for 12 EU countries, whereas Figure 5 and Figure 6 show the time profile of the standard deviation of $\log y$ and the time profile of the coefficient of variation, respectively. Comparing the upward trend in Figure 4 with the downward 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 income convergence. To put it another way: although there seems to be conditional income convergence with respect to the two scale-free measures, Figure 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]

[5]Unfortunately, sometimes misleading graphs or texts to graphs about across-country income conver-



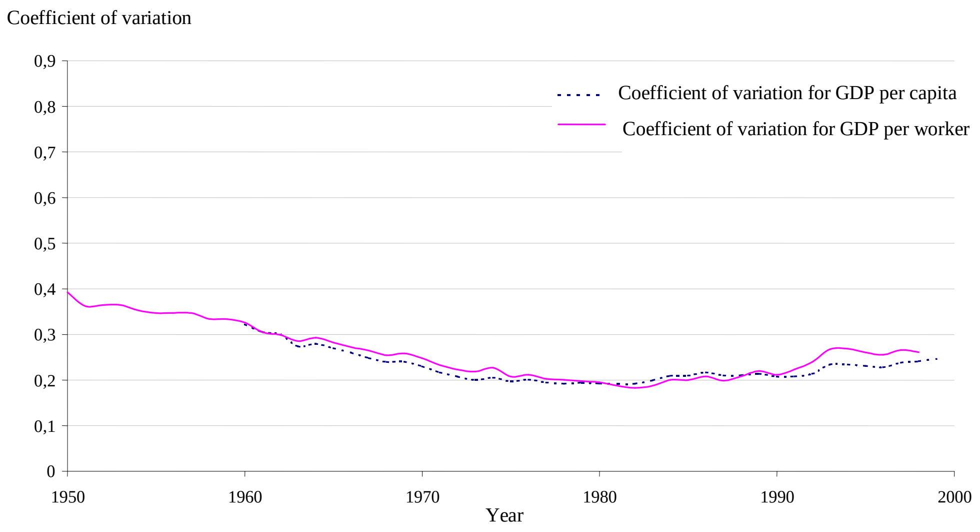
Remarks: Germany is not included in GDP per worker. GDP per worker is missing for Sweden and Greece in 1950, and for Portugal in 1998. The EU comprises Belgium, Denmark, Finland, France, Greece, Holland, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden, Germany, the UK and Austria.
Source: Pwt6, OECD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

Figure 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 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.

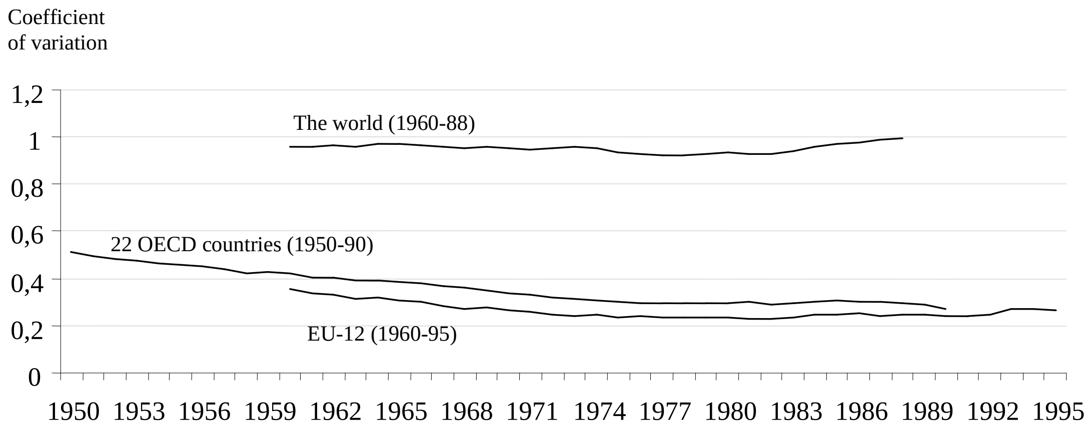
Figure 6: Coefficient of variation of GDP per capita and GDP per worker across 12 EU countries, 1950-1998.

Figure 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 4.1. A country that was poor in 1950 will typically have become a member of OECD only if it grew relatively fast afterwards.

4.6 Other 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.*

gence are published. In Problem Set I 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 7: Coefficient of variation of income per capita across different sets of countries.

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 convergence* at aggregate or industry level. 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).

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.

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.

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.

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,

Quah, D., 1996b,

Quah, D., 1996c,

Quah, D., 1997,

Romer, D., 2001, *Advanced Macroeconomics*, 2. ed., 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

Economic Growth   Lecture Note 2

04.02.2013.   Christian Groth

Terminology concerning technology. Continuous time modeling

Section 1 of this lecture note presents the terminology around technology which I find useful and will use in the lectures. At a few points I deviate somewhat from definitions in Acemoglu's book.

Section 2 provides a refresher of stuff that should be well-known from earlier courses in micro- and macroeconomics. Section 3 can be used as a formula manual for the case of CRS. Finally, in Section 4 we describe the transition from discrete time to continuous time analysis.

1 A two-factor production function

Consider a two-factor production function given by

$$Y = F(K, L), \tag{1}$$

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 (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 (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 (cf. Exercise Problem I.3).

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 continuity (and the "no input, no output" property).

1.1 A neoclassical production function

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:

$$F_K > 0, \quad F_L > 0, \tag{2}$$

$$F_{KK} < 0, \quad F_{LL} < 0. \tag{3}$$

(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 form qualitatively similar to that shown in Fig. 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 $\Delta Y \approx (\partial Y/\partial K)\Delta K = \partial Y/\partial K$.

1.2 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

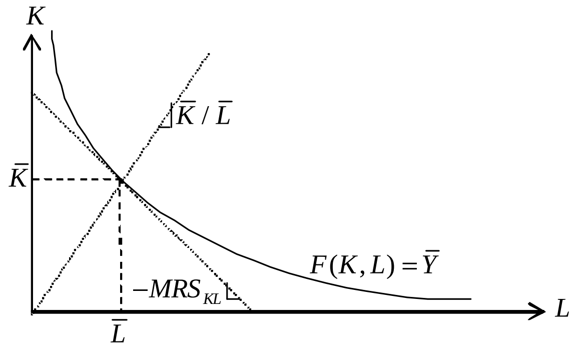


Figure 1: MRS_{KL} as the absolute slope of the isoquant.

L at the point (K, L) is defined as the absolute slope of the isoquant at that point, cf. Fig. 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

$$MRS_{KL} = -\frac{dK}{dL}_{|Y=\bar{Y}} = \frac{F_L(K, L)}{F_K(K, L)} > 0. \tag{4}$$

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.[1] 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).

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 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 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.

[1] The subscript $|Y = \bar{Y}$ in (4) indicates that we are moving along a given isoquant, $F(K, L) = \bar{Y}$.

1.3 The Inada conditions

A continuously differentiable production function is said to satisfy the *Inada conditions*[2] if

$$\lim_{K\to 0} F_K(K,L) = \infty,\ \lim_{K\to\infty} F_K(K,L) = 0, \tag{5}$$

$$\lim_{L\to 0} F_L(K,L) = \infty,\ \lim_{L\to\infty} F_L(K,L) = 0. \tag{6}$$

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 vanishing when the input of the factor increases without bound. Actually, (5) and (6) express *four* conditions, which it is preferable to consider separately and label one by one. In (5) 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 in (6) 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. Some economic growth textbooks, *including Acemoglu's*, define a neoclassical production function more narrowly by including the Inada conditions as a requirement for calling the production function neoclassical. In contrast, when in a given context we need one or another Inada condition, we state it explicitly as an additional assumption.

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 $\lambda > 0$,

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

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

As all inputs are scaled up or down by some factor, output is scaled up or down by the same factor. In their definition of a neoclassical production function some textbooks, *including Acemoglu's*, add constant returns to scale as a requirement. In contrast, when in a given context we need an assumption of constant returns to scale, this is stated as an additional assumption.

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, 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.

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

$$Y = AK^{\alpha}L^{\beta}, \qquad A > 0, 0 < \alpha < 1, 0 < \beta < 1, \tag{7}$$

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 the "total factor productivity". As an exercise, the reader should verify that (7) satisfies (a) and (b) above and is therefore a neoclassical production function. The function in (7) 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

$$Y = \min(AK, BL), \qquad A > 0, B > 0, \tag{8}$$

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.* 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. As (8) 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. □

2.1 The replication argument

The assumption of CRS is widely used in macroeconomics. The model builder may appeal to the *replication argument* saying that by, conceptually, doubling all the inputs, we should always be able to double the output, since we just "replicate" what we are already doing. One should be aware that in principle the CRS assumption is about *technology* − limits to the availability of resources is another question. The CRS assumption and the replication argument presuppose that *all* the relevant inputs are explicit as arguments in the production function and that these are changed equiproportionately. Concerning our present production function $F(\cdot)$, one could easily argue that besides capital and labor, also land is a necessary input and should appear as a separate argument.[3] Then,

[3]We think of "capital" as producible means of production, whereas "land" refers to non-producible natural resources, including for example building sites. If an industrial firm decides to duplicate what it

on the basis of the replication argument we should in fact expect DRS wrt. capital and labor alone. In manufacturing and services, empirically, this and other possible sources for departure from CRS may be minor and so many macroeconomists feel comfortable enough with assuming CRS wrt. 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 and an important part of national wealth.

Another problem with the replication argument is the following. 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 precise form. This gives occasion for introducing a useful local measure of returns to scale.

2.2 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

$$\eta(K,L) = \frac{\theta}{F(K,L)}\frac{dF(\theta K, \theta L)}{d\theta} \approx \frac{\Delta F(\theta K, \theta L)/F(K,L)}{\Delta\theta/\theta}, \text{ evaluated at } \theta = 1. \tag{9}$$

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

$$\text{if } \eta(K,L) \begin{cases} > 1, & \text{then there are locally } \textit{IRS}, \\ = 1, & \text{then there are locally } \textit{CRS}, \\ < 1, & \text{then there are locally } \textit{DRS}. \end{cases} \tag{10}$$

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

has been doing, it needs a piece of land to build another plant like the first.

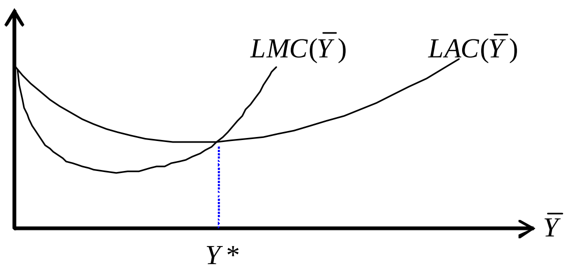


Figure 2: Locally CRS at optimal plant size.

partial output elasticities at that point:

$$\eta(K,L)=\frac{F_K(K,L)K}{F(K,L)}+\frac{F_L(K,L)L}{F(K,L)}. \tag{11}$$

This follows from the definition in (9) by taking into account that

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

Fig. 2 illustrates a popular case from microeconomics, a U-shaped average cost curve from the perspective of the individual firm (or plant): at low levels of output there are falling average costs (thus IRS), at higher levels rising average costs (thus DRS). 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 $F_L(\bar{K},\bar{L})/F_K(\bar{K},\bar{L})=w_L/w_K$. From microeconomics we know that the elasticity of scale at $(\bar{K},\bar{L})$ will satisfy:

$$\eta(\bar{K},\bar{L})=\frac{LAC(\bar{Y})}{LMC(\bar{Y})}, \tag{12}$$

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.

This provides a more subtle replication argument for CRS at the aggregate level. Even though technologies may differ across firms, the surviving firms in a competitive market will have the same average costs at the optimal plant size. In the medium and long run, changes in 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.

The empirical evidence concerning returns to scale is mixed (see the literature notes at the end of the chapter). Notwithstanding the theoretical and empirical ambiguities, the assumption of CRS wrt. 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.

3 Properties of the production function under CRS

Expedient inferences of the CRS assumption include:

(i) marginal costs are constant and equal to average costs (so the right-hand side of (12) 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 (11) 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 "gross-complementarity" between K and L;

(v) a two-factor production function known to have CRS and 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.[4]

[4]Proofs of these claims are 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 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:

$$F_K(K, L) = \frac{\partial Y}{\partial K} = \frac{\partial \left[Lf(k) \right]}{\partial K} = Lf'(k)\frac{\partial k}{\partial K} = f'(k). \tag{13}$$

And the marginal productivity of labor can be written

$$\begin{aligned} F_L(K, L) &= \frac{\partial Y}{\partial L} = \frac{\partial \left[Lf(k) \right]}{\partial L} = f(k) + Lf'(k)\frac{\partial k}{\partial L} \\ &= f(k) + Lf'(k)K(-L^{-2}) = f(k) - f'(k)k. \end{aligned} \tag{14}$$

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 (13) and (2). And the property $f'' < 0$ follows from (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

$$f(k) - f'(k)k > 0 \text{ for all } k > 0, \tag{15}$$

as well as

$$\lim_{k \to 0} \left[f(k) - f'(k)k \right] = f(0). \tag{16}$$

Indeed, from the mean value theorem[5] 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 $f'' < 0$. In view of $f(0) \geq 0$, this establishes (15). And from $f(k) > f(k) - f'(k)k > f(0)$ and continuity of f follows (16).

Under CRS the Inada conditions for MPK can be written

$$\lim_{k \to 0} f'(k) = \infty, \qquad \lim_{k \to \infty} f'(k) = 0. \tag{17}$$

[5] 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 $f'(\gamma) = (f(\beta) - f(\alpha))/(\beta - \alpha)$.

An input which must be positive for positive output to arise is called an *essential input.* The second part of (17), 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.[6] Combining these results, when *both* the upper Inada conditions hold and CRS obtains, then both capital and labor are essential inputs.[7]

Fig. 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 Fig. 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 Fig. 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 Fig. 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 wrt. the B/A ratios, 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 $k > B_n/A_n$, $MPK = 0$,[8] whereas for $k \to 0$, MPK stays equal to $A_1 < \infty$, thus questioning the lower Inada condition.

[6] Proofs of these claims are in the Appendix to Chapter 2 of my Lecture Notes in Macroeconomics.

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

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

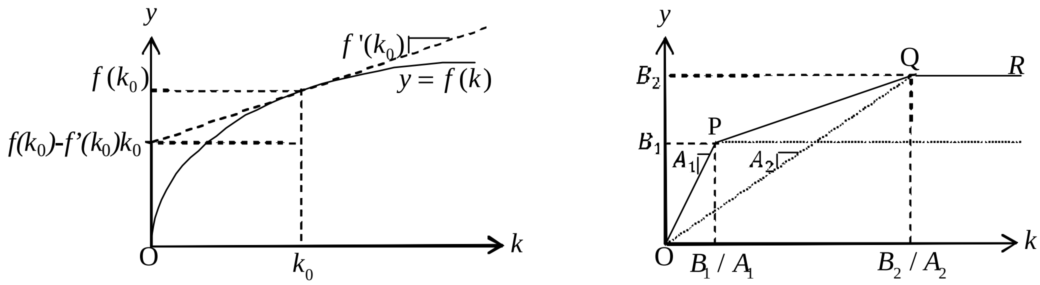


Figure 3: Two labor productivity curves based on CRS technologies. Left: neoclassical technology with Inada conditions for MPK satisfied. Right: a combination of two efficient Leontief techniques.

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* can be shown to imply, under CRS, that output has no upper bound when either input goes to infinity for fixed amount of the other input.

4 Long-run vs. short-run production functions

The production functions considered up to now are based on the questionable assumption that the substitutability between capital and labor are 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,

$$Y = \min(Au\bar{K}, BL), \qquad A > 0, B > 0, \tag{18}$$

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.

In the short run the choice variable is essentially u. 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 $L = A\bar{K}/B$; the marginal productivity 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 (18). These pairs satisfy

$$f(k) = Ak = B. \tag{19}$$

So, for an increasing sequence of k's, $k_1, k_2, \ldots, k_i, \ldots$, the corresponding pairs are $(A_i, B_i) = (f(k_i)/k_i, f(k_i))$, $i = 1, 2, \ldots$[9] We say that ex ante, 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 (18). 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 1 gives

[9] The points P and Q in the right-hand panel of Fig. 3 can be interpreted as constructed this way from the neoclassical production function in the left-hand panel of the figure.

an overview of the alternative cases.

Table 1. Technologies classified according to
factor substitutability ex ante and ex post

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

To take technological change as time proceeds into account, we may replace (19) and (18) by $f(k,t) = A_t k = B_t$ and $Y = \min(A_t u \bar{K}_t, B_t L)$, 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 least one strict equality), it may pay the firm to invest in the new technology at the same time as some old plant and machinery are dismantled.

The ex post clay aspect of many technologies seems clearly relevant for short-run analysis. For long-run analysis it is less important since within a long time horizon it is the accumulated new investments that matter.

5 On continuous time analysis

Because dynamic analysis is often easier in continuous time, most growth models are stated in continuous time. In this section we describe some of the conceptual aspects of continuous time analysis.

Let us start from a discrete time framework: the run of time is divided into successive periods of constant length, taken as the time-unit. Let financial wealth at the beginning of period i be denoted a_i, $i = 0, 1, 2, ...$ Then wealth accumulation in discrete time can be written

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

where s_i is (net) saving in period i.

5.1 Transition to continuous time

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 be equal to one time unit. Then $a(i\Delta t) = a_i$. 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 $\Delta t \to 0$. Assuming $a(\cdot)$ is a differentiable function, then $\lim_{\Delta t \to 0} \Delta a(t)/\Delta t$ exists and is denoted the *derivative of* $a(\cdot)$ at t, usually written $da(t)/dt$ or just $\dot{a}(t)$. That is,

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

By implication, wealth accumulation in continuous time is written

$$\dot{a}(t) = s(t), \qquad a(0) = a_0 \text{ given}, \tag{20}$$

where $s(t)$ is the saving 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 let us choose one year. Going back to discrete time, if wealth grows at the constant rate $g > 0$ per year, then after i periods of length one year (with annual compounding)

$$a_i = a_0(1 + g)^i, \qquad i = 0, 1, 2, \ldots . \tag{21}$$

When compounding is n times a year, corresponding to a period length of $1/n$ year, then after i *such* periods:

$$a_i = a_0(1 + \frac{g}{n})^i. \tag{22}$$

With t still denoting time (measured in years) that has passed since the initial date (here date 0), we have $i = nt$ periods. Substituting into (22) 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 (and so m) $\to \infty$. Then, in the limit there is continuous compounding and

$$a(t) = a_0 e^{gt}, \tag{23}$$

where e is the "exponential" defined as $e \equiv \lim_{m \to \infty}(1 + 1/m)^m \simeq 2.718281828....$

The formula (23) is the analogue in continuous time (with continuous compounding) to the discrete time formula (21) with annual compounding. Thus, a geometric growth factor is replaced by an exponential growth factor.

We can also view the formulas (21) and (23) as the solutions to a difference equation and a differential equation, respectively. Thus, (21) is the solution to the simple linear difference equation $a_{i+1} = (1 + g)a_i$, given the initial value a_0. And (23) is the solution to the simple linear differential equation $\dot{a}(t) = ga(t)$, given the initial condition $a(0) = a_0$. With a time-dependent growth rate, $g(t)$, the corresponding differential equation is $\dot{a}(t) = g(t)a(t)$ with solution

$$a(t) = a_0 e^{\int_0^t g(\tau)d\tau}, \tag{24}$$

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 (24) is called the *basic growth formula* in continuous time and the factor $e^{\int_0^t g(\tau)d\tau}$ is called the *growth factor* or the *accumulation factor*.

Notice that 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

$$\dot{K}(t) = I(t) - \delta K(t), \qquad K(0) = K_0 \text{ given}, \tag{25}$$

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, in (25) $\delta > 1$ is conceptually allowed. Indeed, suppose for simplicity that $I(t) = 0$ for all $t \geq 0$; then (25) gives $K(t) = K_0 e^{-\delta t}$ (exponential decay). 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, 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 overlapping generations 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. Then with 30 years as our time unit, inserting into the formula $1/2 = e^{-\delta/3}$ gives $\delta = (\ln 2) \cdot 3 \simeq 2$.

5.2 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 and saving). A *stock* variable is a variable measured as just 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 $I(t)$ above are flow variables.

One cannot add a stock and a flow, because they have *different denomination.* What exactly is meant by this? The elementary measurement units in economics are *quantity units* (so and so many machines of a certain kind or so and so many liters of oil or so and so many units of payment) 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". In contrast, investment I has the denomination "quantity of machines per time unit" or, shorter, "quantity/time". If we change our time unit, say from quarters to years, the value of a flow variable is quadrupled (presupposing annual compounding). A growth rate or interest rate has the denomination "(quantity/time)/quantity" = "time^{-1}".

Thus, in continuous time analysis expressions like $K(t) + I(t)$ or $K(t) + \dot{K}(t)$ are 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 contains 50 liters of water and the tap pours $\frac{1}{2}$ liter per second into the tub. Then a sum like $50\ \ell + \frac{1}{2}$ (ℓ/sec.) does not make sense. But the *amount* of water in the tub after one minute is meaningful. This amount would be $50\ \ell + \frac{1}{2} \cdot 60$ ((ℓ/sec.)×sec.) $= 90\ \ell$. 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* this time interval, is the integral, $\int_{t_0}^{t_0+\Delta t} I(t)dt \approx I(t)\Delta t$. Similarly, $s(t)$, that is, the flow of individual saving, 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 Fig. 4.

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. But, it is really 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.

In empirical economics data typically come in discrete time form and data for flow

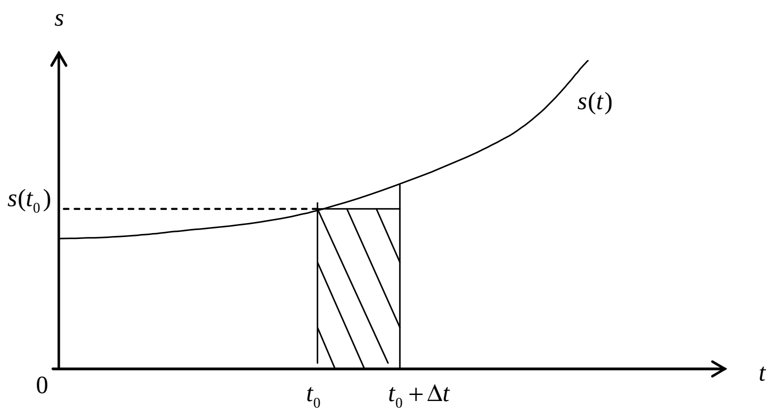


Figure 4: With Δt "small" the integral of $s(t)$ from t_0 to $t_0 + \Delta t$ is $\approx$ the hatched area.

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 be a good reason for putting at least microeconomic analysis in period terms. Yet, 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 and then a formulation in continuous time may be preferable. This also allows variation *within* the usually artificial periods in which the data are chopped up.[10] For example, stock markets (markets for bonds and shares) are more naturally modelled in continuous time because such markets equilibrate almost instantaneously; they respond immediately to new information.

In his discussion of this modelling issue, Allen (1967) concluded that from the point of view of the economic contents, the choice between discrete time or continuous time analysis may be a matter of taste. But from the point of view of mathematical convenience, the continuous time formulation, which has worked so well in the natural sciences, is preferable.[11]

6 Appendix: Growth formulas 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:

[10] Allowing for such variations may be necessary to avoid the *artificial* oscillations which sometimes arise in a discrete time model due to a too large period length.

[11] At least this is so in the absence of uncertainty. For problems where uncertainty is important, discrete time formulations are easier if one is not familiar with stochastic calculus.

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, cf. Appendix A to Chapter 4.

7 References

Allen, R.G.D., 1967, *Macro-economic Theory. A mathematical Treatment*, Macmillan, London.