Lecture notes
PPD 2021-2022

Thomas Zuber

Contents

Chapter 1

Growth?

Time is continuous.

The first question of growth theory may be stated as: "Are we richer today than our ancestors were a thousand years ago?" The answer is yes. The order of magnitude is: forty times richer. The second question of growth theory is: "Why?". This is the central question of growth *theory*.

What do we mean by "richer"? To keep things simple, and for the rest of this course, we will think of wealth as *average* wealth. In other words growth theory is mainly concerned neither with the movements of output *per se*, nor in changes in the *distribution* of output, but of output *per capita*, *per head*, *per person*, *per worker*... Not total output Y_t, but total output divided by the number of persons L_t who participate in production and, were the receipts of this production to be evenly distributed among them, should enjoy the consumption of an equal share of it. Let us denote this quantity by $y_t = Y_t/L_t$.

1.1 The one factor model of growth

How does growth theory answer this second question? The main ingredient of the answer lies in what is called an aggregate production function. An aggregate production is not the production function used by any particular firm but a stable relationship between aggregate output and aggregate inputs. In other words an aggregate production function is a model of Y_t which restricts the variable causes of total output to be aggregate quantities of inputs. For the sake of simplicity let us start with the assumption that aggregate output Y_t is produced using aggregate labor L_t as the only input:

$$Y_t = F(L_t) \tag{1.1}$$

If this relation holds true then at any time t output per capita can be written :

$$y_t = \frac{Y_t}{L_t} = \frac{F(L_t)}{L_t} \tag{1.2}$$

which is a function of population size only.[1] How does wealth vary with population? To answer to this question we can express the growth rate of output per capita $\dot{y}_t/y_t$ as a function of the growth rate of population $\dot{L}_t/L_t$.[2] [3] Before doing so notice that in general the growth rate of a ratio of variables

[1] Notice that we implicitly assume here that "aggregate labor" is nothing but total population. If this were not the case — for instance if child labor happens to be forbidden — then the expression for output per capita should be written:

$$y_t = \frac{F(L_t)}{N_t} = \frac{F(L_t)}{L_t}\frac{L_t}{N_t}$$

where N_t denotes total population. In this case output per capita is a function of *average labor productivity* or *average output* $F(L_t)/L_t$ and the "employment rate" L_t/N_t.

[2] By convention a dotted continuous time variable $\dot{x}_t$ will always denote the first order time derivative of the corresponding un-dotted variable x_t, i.e.:

$$\frac{dx_t}{dt} = \dot{x}_t$$

[3] Growth rates? Not exactly. In the rest of this course when you hear or read the words "growth rate" you should think "instantaneous growth rate". In continuous time the *instantaneous* growth rate of a variable x_t is defined as

$$\frac{\dot{x}_t}{x_t}$$

To understand what an *instantaneous* growth rate actually is consider that you are at time t and want to know the value that x will take at time $t+\Delta t$ where $\Delta t > 0$ is very small. From knowledge of the value of x_t and the *instantaneous* growth rate $\frac{\dot{x}_t}{x_t}$ you can compute the value of $x_{t+\Delta t}$ as:

$$x_{t+\Delta t} = [1 + \frac{\dot{x}_t}{x_t}\Delta t]x_t$$

Before showing why this is indeed true note that by analogy to the discrete time case where you would write something like :

$$x_{t+1} = [1+g]x_t$$

is the difference of its numerator's and denominator's respective growth rates. In our case

$$\frac{\dot{y}_t}{y_t} = \frac{\dot{Y}_t}{Y_t} - \frac{\dot{L}_t}{L_t} \tag{1.3}$$

To show that this is indeed the case take the time derivative of y_t:

$$\dot{y}_t = \frac{\dot{Y}_t L_t - Y_t \dot{L}_t}{L_t^2}$$

divide by y_t and rearrange to get:

---

where g is the "growth rate" of x, here the quantity case the quantity $\frac{\dot{x}_t}{x_t}\Delta t$ is the "growth rate" of x between time t and $t + \Delta t$. The *instantaneous* growth rate of x is the growth rate of x ***per unit of time*** — that is to say, the growth rate of x over a small interval of time divided by the length of this interval :

$$\frac{x_{t+\Delta t} - x_t}{x_t} \frac{1}{\Delta t} \sim \frac{\dot{x}_t}{x_t}$$

when Δt goes to zero. Re-arranging this expression gives you:

$$x_{t+\Delta t} \sim [1 + \frac{\dot{x}_t}{x_t} \Delta t] x_t$$

which is holds with equality in the limiting case where Δt goes to zero.

Another way to see what an instantaneous growth rate means is to consider a variable x_t which grows at the constant instantaneous growth rate g. In this case:

$$x_{t+\Delta t} = x_t e^{g\Delta t}$$

and using a first order approximation of the exponential when Δt goes to zero one falls back on:

$$x_{t+\Delta t} = [1 + g\Delta t] x_t$$

To conclude this long footnote, when hearing that g is the ***instantaneous*** growth rate of a variable x you should *not* think "This means that" :

$$x_{t+\Delta t} = [1 + g] x_t$$

but instead that:

$$x_{t+\Delta t} = [1 + g\Delta t] x_t$$

$$\frac{\dot{y}_t}{y_t} = \frac{\dot{Y}_t L_t - Y_t \dot{L}_t}{L_t^2}\frac{L_t}{Y_t}$$
$$= \frac{\dot{Y}_t}{Y_t} - \frac{\dot{L}_t}{L_t}$$

At this point we are already halfway from where we would like to be. Equation 1.3 shows that output per capita will grow if and only if the growth of total output outpaces the growth of the labor force:

$$\frac{\dot{y}_t}{y_t} > 0 \Longleftrightarrow \frac{\dot{Y}_t}{Y_t} > \frac{\dot{L}_t}{L_t}$$

This is true in general: average output grows if output grows faster than population — or declines slower, if both growth rates are negative. We are

Output growth and population growth

All what is left to do is to determine how *total* output varies with population. We known from 1.1, i.e. the "aggregate production function" that at all levels of population L_t:

$$Y_t = F(L_t)$$

so that applying the chain rule:

$$\dot{Y}_t = \dot{L}_t F'(L_t)$$

where F' is the first derivative of F. Dividing both sides by Y_t and using the production F to substitute for Y we get:

$$\frac{\dot{Y}_t}{Y_t} = \frac{\dot{L}_t F'(L_t)}{Y_t} = \frac{\dot{L}_t F'(L_t)}{F(L_t)}$$

Dividing and multiplying the right hand side by L_t this yields the following expression for the growth rate of total output:

$$\frac{\dot{Y}_t}{Y_t} = \frac{\dot{L}_t F'(L_t)}{F(L_t)}\frac{L_t}{L_t}$$
$$= \frac{L_t F'(L_t)}{F(L_t)}\frac{\dot{L}_t}{L_t}$$

so that:

$$\frac{\dot{Y}_t}{Y_t} = \beta(L_t)\frac{\dot{L}_t}{L_t} \tag{1.4}$$

where $\beta(L_t) = \frac{F'(L_t)L_t}{F(L_t)}$ denotes the elasticity of output with respect to labor.

The elasticity of output with respect to labor

Elasticities of output with respect to factors of production probably are some of the most important quantities in growth theory. In our case $\beta(L_t)$ tells us by how much Y_t changes in percentage when labor L_t increases by 1%.[4] Assume that $F' > 0$ so that an increase in labor always increases output, if only by a very small amount. As this assumption tells us $\beta(L_t)$ is strictly positive, we are left with three possible cases

1. when $\beta(L_t) < 1$ total output increases **less than one for one** with labor;
2. when $\beta(L_t) = 1$ total output increases **one for one** with labor;
3. when $\beta(L_t) > 1$ total output increases **more than one for one** with labor.

[4] To see this you can rewrite

$$\begin{aligned}\beta(L_t) &= \frac{F'(L_t)L_t}{F(L_t)} \\ &= \frac{\frac{dF(L_t)}{dL_t}L_t}{F(L_t)} \\ &= \frac{\frac{F(L_{t+\Delta t})-F(L_t)}{L_{t+\Delta t}-L_t}L_t}{F(L_t)} \\ &= \frac{\frac{F(L_{t+\Delta t})-F(L_t)}{F(L_t)}}{\frac{L_{t+\Delta t}-L_t}{L_t}}\end{aligned}$$

where for illustrative purposes we assume that a small change in L occurs between time t and time $t+\Delta t$. In this case you can see that the elasticity $\beta(L_t)$ can be exactly computed as the ratio of the growth rate of $F(L_t)$ to the growth rate of L_t, over a small interval of time Δt. This being said we could simply write in a more compact form :

$$\beta(L) = \frac{\frac{dF}{F}}{\frac{dL}{L}}$$

where the variations dF and dL can arise because of time or across countries, firms, etc.

The elasticity of output with respect to labor

Before moving on with our simple growth model and bringing it to the data we should pause to highlight some important features of this β parameter.

First notice that we write β as function of the level of population L_t because in general there is no reason for the relationship between changes in labor supply and changes in output to be similar at all levels of L_t. One could imagine for instance that when labor is low, a 1% increase in labor increases output by more than 1%, whereas at very high levels of L_t a 1% increase in labor increases output by less than 1%. Although this is indeed a possibility, we will mostly abstract from it in the rest of this course. In order to focus on the understanding of some important properties of this parameter I will from now on drop the explicit dependence on L_t and simply write β instead. A little be further down in these lectures notes, we will see that assuming a constant elasticity β has very important consequences for our production function F.

The elasticity β and Ricardian rents

The fact that β lies below or above one tells us something very important about the behavior of F. Take the $\beta < 1$ case for instance. By definition this means that:

$$\frac{F'(L_t)L_t}{F(L_t)} < 1$$

which you can rewrite as

$$F'(L_t) < \frac{F(L_t)}{L_t}$$

The $\beta < 1$ is thus equivalent to saying that at every level of L_t the marginal product of labor is less than the average product of labor. In this situation a competitive firm paying its input at its marginal product will be making strictly positive profits. To see this we ca again rewrite the $\beta < 1$ condition as:

$$F(L_t) - F'(L_t)L_t > 0$$

where $F(L_t)$ will be the amount of production sold at price normalized to 1, L_t the amount of labor used and $F'(L_t)$ the price of this labor input on a competitive labor market. In economic history the difference between the average product of a factor and its marginal product is known as the "Ricardian rent". In general the existence of a Ricardian rent, which we have just shown to be equivalent to to the $\beta < 1$ case, will allow competitive firms to make positive profits. In the $\beta = 1$ case competitive firms will make zero profits because the total receipts of production are redistributed to labor through wages ($F(L_t) = L_t F'(L_t)$). Finally the $\beta > 1$ case implies that competitive firms will make negative profits because all units of input will be paid above their marginal product.

The elasticity β and returns to scale

The value of β and the existence of a Ricardian rent is tightly linked to another property of the production function: the degree of its returns to scale. Returns to scale characterize the behavior of a production function when you scale up its inputs by a factor $\lambda > 1$. We can distinguish 3 cases:

1. Decreasing returns when multiplying inputs by λ multiplies output by less than λ :
$$F(\lambda L) < \lambda F(L)$$
2. Constant returns when multiplying inputs by λ multiplies output by exactly λ :
$$F(\lambda L) = \lambda F(L)$$
3. Increasing returns when multiplying inputs by λ multiplies output by more than λ :
$$F(\lambda L) > \lambda F(L)$$

The fact that β lies below or above 1 will determine in which of these three situations the production function F falls. Assume for instance that $\beta < 1$. In this case the average product of labor will be decreasing. Indeed, the derivative of $F(L)/L$ is:

$$\frac{F'(L)L - F(L)}{L^2}$$

which is negative when:

$$F'(L)L - F(L) < 0 \iff F'(L) < \frac{F(L)}{L}$$
$$\iff \beta < 1$$

Or in other words, the average product of labor decreases as labor rises as soon as the marginal product of labor lies below its average product: adding less and less productive units to the production process lowers the average productivity of all units.

A direct consequence of the fact that the average product of labor is decreasing in the level of labor used in production is that, for any $\lambda > 1$:

$$\frac{F(\lambda L)}{\lambda L} < \frac{F(L)}{L}$$

Hence that:

$$F(\lambda L) < \lambda F(L)$$

which is just the definition of decreasing returns to scale. The notion of returns to scale is crucial to understand some of the debates that have agitated growth theory since the 1980s. As we will see shortly the possibility of output per capita growth in our very simple model can be traced back to to the value of β and so in the end to the decreasing, constant or increasing nature of F's returns to scale.

Finally solving for $\dot{y}_t/y_t$

Substituting 1.4 in 1.3 we finally get that the growth rate of output per capita and growth rate of labor are related through:

$$\frac{\dot{y}_t}{y_t} = (\beta - 1)\frac{\dot{L}_t}{L_t} \tag{1.5}$$

Taking the growth rate of population $n = \dot{L}_t/L_t$ as exogenous this yields a simple expression for $\dot{y}_t/y_t$:

$$\frac{\dot{y}_t}{y_t} = (\beta - 1)n$$

When $\beta < 1$ so that total output increases less than one for one with the quantity of labor used in production, or equivalently when returns to scale are decreasing, $\dot{y}_t/y_t$ is negative. In this case output per capita decreases when population increases. Increasing population increases total output but the additional output is less than what would have been needed to keep output per capita constant. As we will see later in the course this is the quintessential mechanism at work in a Malthusian growth regime.

When $\beta = 1$ returns to scale are constant and so does output per capita. Increasing population increases total output and the additional output is just enough to keep output per capita at its original level. Can growth not depend on population growth.

Finally when $\beta > 1$, returns to scale are increasing and an increase in population delivers more output than what was needed to keep output per capita constant. Our simple model predicts that growth will arise if and only if returns to scale are increasing and population is growing.

Which of these three possible cases will be relevant in practice? It is hard to say in general. All the more so because, despite our simplifying hypothesis, it would be perfectly possible for $\beta(L_t)$ to vary with L_t. The elasticity of output with respect to labor could for instance be greater than one for low values of L_t, equal to one for a medium size labor force, and lower than one when population grows up to large levels. In such a situation and if population grows at a constant positive rate from a low initial level, then output per capita will start to grow ($\beta(L_t) > 1$), will then be constant for some period of time ($\beta(L_t) = 1$), and will end up decreasing as population grows to infinity ($\beta(L_t) < 1$).

Matching the data

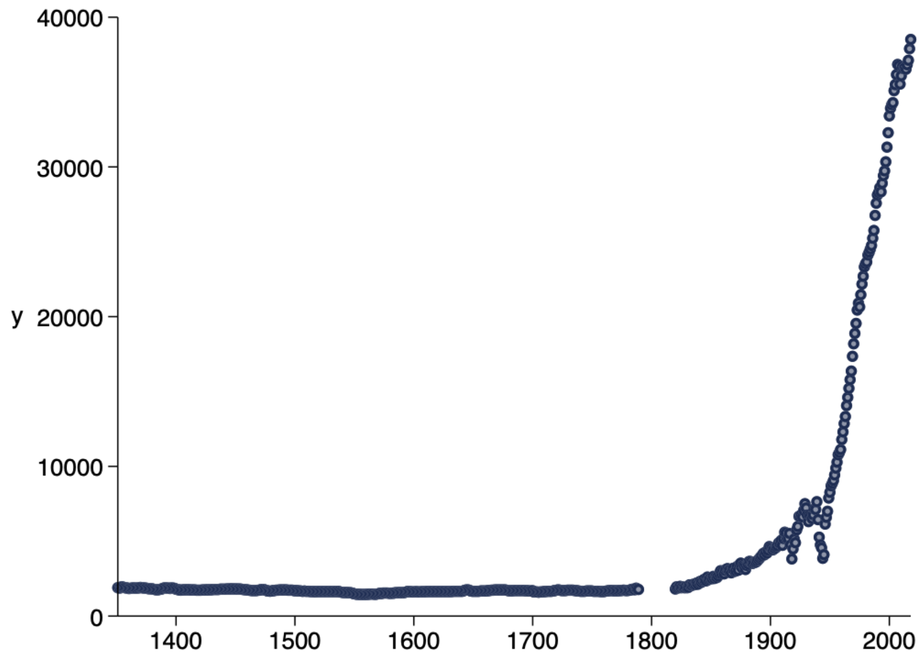
This of course is not what we observe in the data. At least since the industrial revolution population and output per capita have grown hand in hand. Figure 1.1 below plots the evolution of output per capita y_t in France from 1350 to 2019.

This simple scatter plot suggests that:

- output per capita was roughly constant until the 19th century;
- started to take of during the 19th century;
- accelerated exponentially after World War II.

How do these statements translate in growth rates terms? An easy way to assess changes in the growth rate of a variable over time is to plot not the *level* this variable but its natural logarithm. Indeed when looking at a plot of

Figure 1.1: Output per capita in France



Output per capita in France from 1350 to 2019 expressed in thousands of 2011 dollars. Maddison project estimates taken from Bolt and van Zanden (2020).

$\log(y_t)$ against time rather than at its level, the slope of the resulting curve is precisely the growth rate of y_t.[5] Figure 1.2 does precisely that with output per capita in France.

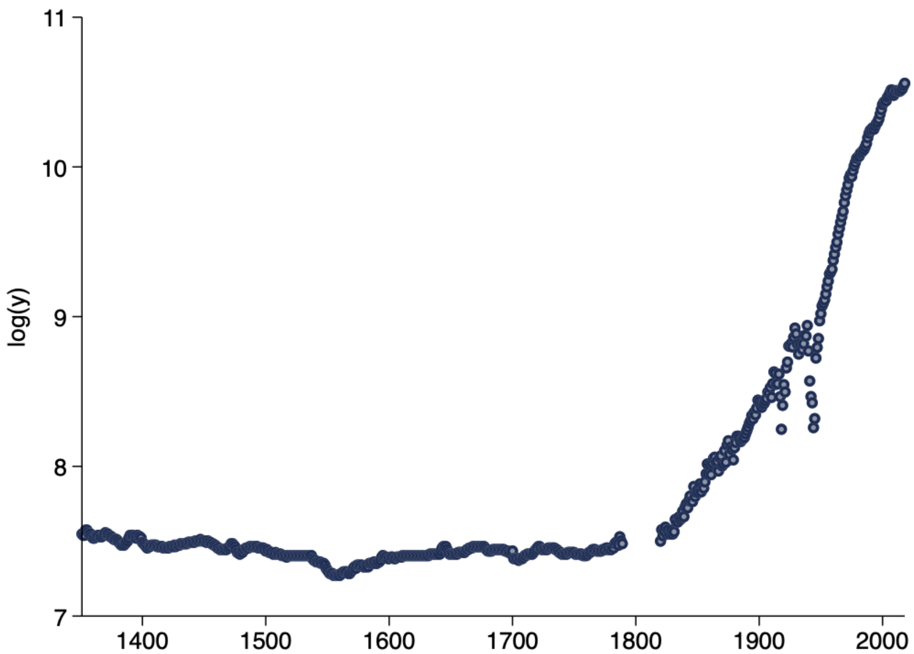
When looking at Figure 1.2 it would tempting to fit a straight line with zero slope going from the 14th to the end of the 18th century, a straight line with a moderately positive slope during the 19th century and a straight line with much steeper slope after WWII. In other words, Figure 1.2 shows that over long periods of time output per capita has grown at constant but different rates. To put precise numbers on this general statement Table 1.1 reports the results of a series of century by century split sample regressions of $\log(y_t)$ on time (i.e. years).

Confirming the visual inspection of Figure 1.2, Table 1.1 shows that

[5]To see this just take the time derivative of $\log(y_t)$:

$$\frac{d\log(y_t)}{dt} = \frac{\dot{y}_t}{y_t}$$

Figure 1.2: Log output per capita in France



Log output per capita in France from 1350 to 2019 expressed in log thousands of 2011 dollars. Maddison project estimates taken from Bolt and van Zanden (2020).

Table 1.1: Output per capita growth in France

	15th	16th	17th	18th	19th	20th	21st
year	-0.000250	-0.000963	0.000627	0.000615	0.0109	0.0222	0.00616
	(0.0000664)	(0.000141)	(0.0000521)	(0.0000946)	(0.000194)	(0.000754)	(0.000676)
N	100	100	100	90	80	100	19

This table reports the coefficient of split sample regressions of log output per capita in France on a continuous year variable for each century. Maddison project estimates of ouput per capita taken from Bolt and van Zanden (2020). Standard errors in parenthesis.

output per capita growth was essentially zero prior to 1800 (< 0.1% per year and actually negative for long stretches of time), increased to 1.1% after the onset of the industrial revolution in the 19th century, averaged 2.2% during the 20th century, and has since then fallen down to 0.6% per year.

Can our bare bones "theory" fit this pattern of output per capita growth? To know this we need to match these numbers with what our model would predict. Let us concentrate on the post Industrial Revolution period. Estimates of population growth for the 19th, 20th and 21th century in France are 0.3%, 0.5%, and 0.6% per year respectively. We know from equation 1.5 that:

$$\frac{\dot{y}_t}{y_t} = [\beta - 1]\frac{\dot{L}_t}{L_t}$$

If our model is true, the value of β consistent with the observed growth rates of output per capita and population can be backed out as:

$$\beta = \frac{\dot{y}_t/y_t}{\dot{L}_t/L_t} + 1$$

Applying this formula to the growth rate of output per capita and population observed in France gives us an average value of the elasticity of output with respect to labor of 4.7 in the 19th century, 5.4 in the 20th century and 2 in the 21st century. As could be expected from the theory these values are well above one and imply increasing returns to labor.

Another a more direct way to get to a similar result is to actually plot the production function that our theory assumes, i.e. Y_t against L_t. Figure 1.3 does precisely that using the same 19th, 20th and 21st century data/

Taken at face value Figure 1.3 suggests that there is indeed a stable relationship between Y_t and L_t across time. Taking logs on both sides and regressing $\log(Y_t)$ on $\log(L_t)$ we find a coefficient of 5.7. Notice that because:

$$\frac{d\log(Y_t)}{d\log(L_t)} = \frac{dY_t/Y_t}{dL_t/L_t} = \frac{F'(L_t)L_t}{F(L_t)} = \beta$$

regressing $\log(Y_t)$ on $\log(L_t)$ is just another way to recover the elasticity parameter β of our production function F.

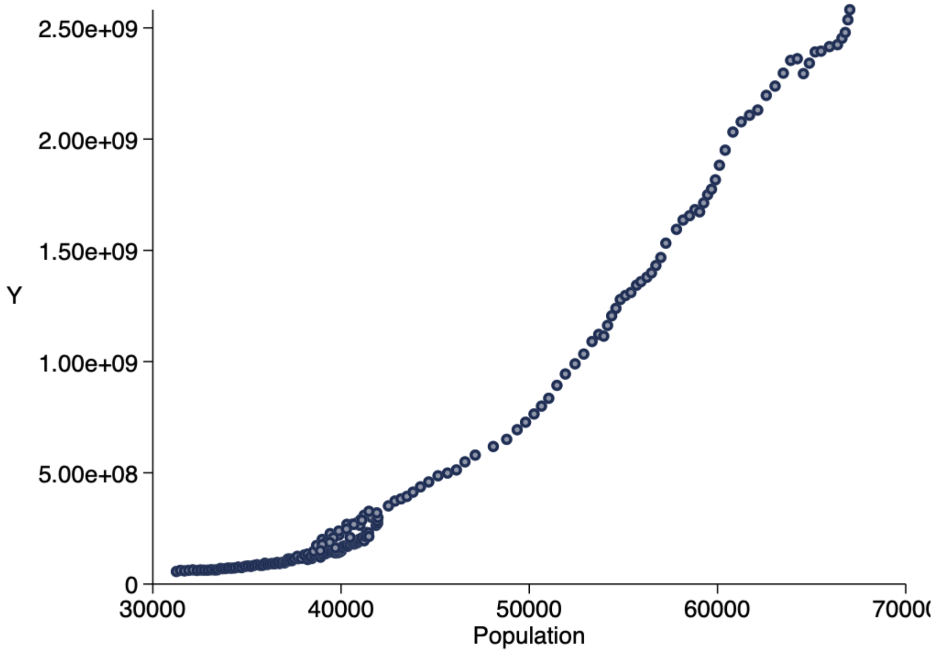
Now that we have some candidate values for β can we recover the full function F? If $\beta(L_t)$ varies with L the answer is "in general no". But if we are willing to assume that $\beta(L_t)$ is constant, as our simple econometric model regressing $\log(Y_t)$ on $\log(L_t)$ implicitly does, then knowledge of the constant value β of the elasticity $\beta(L_t)$ is enough to recover the whole production function!

We are looking for a function which verifies

$$\frac{F'(L)L}{F(L)} = \beta$$

for all possible levels of labor L. A simple integration of this expression shows that the only function that satisfies this constant elasticity property is the so-called "Cobb-Douglas" function:

Figure 1.3: Output against population in France



Total output against total population in France from 1800 to 2019. Output is expressed in thousands of 2011 dollars. Maddison project estimates taken from Bolt and van Zanden (2020).

$$F(L_t) = L_t^{\beta}$$

where we have arbitrarily normalized $F(1)$ to 1.[6] Taking our model at face value and matching the observed correlation between output and population would hence imply that:

$$Y_t = L_t^{5.7}$$

Sanity check

At this point we should stop to assess the relative success or failure of our theory. On the one hand its simplicity allowed us to explain growth through a single parameter whose value could be backed out from observed data using a simple ratio of growth rates or a simple regression involving the log of output and the log of population. If we take F to be of constant elasticity β and trust our average estimate yielding $\beta = 5.7$, then achieving higher standards of living for all should not be very complicated. One only needs to keep population growing!

While increasing returns to labor allow for unlimited growth of output per capita, a value of β above 1 is inherently problematic. To see why we are going to take our "production function" model of output seriously and embed it in a simple market economy.

This economy has two agents a firm and a worker. The firms produces output Y thanks to our production function F and sells this output at price p. In order to do so the firm needs to use labor L which it rents from the worker at rate w. Our last assumptions is that the markets for consumption

[6] To prove this claim start from

$$\beta = \frac{F'(L)L}{F(L)}$$

Rewrite this as:

$$\frac{\beta}{L} = \frac{F'(L)}{F(L)}$$

Integrate between 1 and L:

$$\int_1^L \frac{\beta}{l} dl = \int_1^L \frac{F'(l)}{F(l)} dl$$

This yields:

$$\beta \log(L) = \log(F(L)) - \log(F(1))$$

Normalize $F(1)$ to one and take the exponential on both sides to get:

$$F(L) = L^{\beta}$$

goods Y and labor L are competitive — i.e. agents are price takers — and that the firm seeks to maximize its profits. The firm's profit

$$\max_L pF(L) - wL \tag{1.6}$$

The associated necessary first order condition for a maximum to be attained is:

$$pF'(L) = w$$

in other words the marginal product of labor should be equal to the unit price of labor, the real wage:

$$F'(L) = \frac{w}{p}$$

Normalizing p to 1 equilibrium profits write:

$$\pi = F(L) - wL = F(L) - F'(L)L$$

which just the Ricardian rent we already discussed above in relation to β. We know from this discussion that a value of β above 1 implies a negative Ricardian rent and hence negative profits for a competitive firm. If you think that firms need some level of profits in order to survive a competitive economy with $\beta = 5.7$ appears as a theoretical conundrum.

But more than theoretically impossible, the $\beta > 1$ result appears to be inconsistent with a simple empirical test. Rearranging terms of the firm's first order condition, one can find an expression relating β and the labor share of value added:

$$\frac{wL}{pY} = \frac{F'(L)L}{F(L)} = \beta$$

This formula tells us that *in equilibrium* β pins down the labor share of production. In practice observed labor shares are quite stable and around 2/3 of output. This gives us another strong reason to reject high values of β as impossible.

A last but important technical remark. In the Cobb-Douglas case with a constant elasticity, $F(L) = L^\beta$ will be a concave function if and only if $\beta < 1$. From a mathematical point of view concavity is important because the first order condition associated to the maximization problem 1.6 will be sufficient to characterize an interior optimum only if F is concave. Indeed, if β were strictly greater than one, 1.6 would be a convex maximization problem so that solving for the associated first order condition would be equivalent to finding a minimum rather than a maximum level of profits... [7]

From an economic perspective concavity is linked to the idea of decreasing marginal productivity. A value of β above one would imply that each additional unit of labor employed is more productive will be more productive than previously employed units. In this one factor context, decreasing returns, the Ricardian rent, concavity and decreasing marginal productivity are all governed by the same parameter β. This will not be the case anymore when we extend our production function to include other factors of production — in which case a concave function with decreasing marginal productivity in each input may very well exhibit constant returns to scale to all inputs taken together.

At this point we have reached a good understanding of a useless model. This a good first step...

1.2 The two factors model of growth

In some ways the one factor growth model that we just examined asks to much from labor. The very high value of β needed to match the observed pattern of output per capita growth in this model is a sign that other important factors distinct from but correlated with population growth are loaded on this single parameter. One of these factors is capital. Out of deference to a famous German writer we shall denote this factor K_t. Still out of deference to this same German writer we should note that contrary to other factors of production such as land, capital can be produced and accumulated. This means that contrary to land which is more or less fixed, capital can grow and henceforth contribute the growth of output per worker.

Bringing capital into our model means that we need to modify our assumptions on the production process. To do so will extend the production function F to include this new factor of production:

$$Y_t = F(K_t, L_t)$$

Before studying the dynamic behavior of output per capita, as we did in the case of our one factor model of growth, let us first ask what kind assumptions may seem reasonable to impose on F.

[7] Under convexity of the production function profit maximizing quantities are corner solutions —i.e. 0 or $+\infty$ depending on the exact shape of f — and the elasticity β does not pin down anything.

Assumptions on F

In the last section one important hypothesis we discussed was returns to scale. We saw that growth in output per capita was only possible if we were willing to accept increasing returns to labor. When we add capital to our story this is no longer true. Indeed, with two factors of production instead of one it will be possible to have increasing or constant returns to all factors of production while keeping the assumption of decreasing returns to each factor taken separately.

Yet another digression on returns to scale

Before looking at concrete example of F satisfying this property it is useful to restrict our attention to the particular class of "homogeneous" production functions. This class of function will be useful to understand the consequences of assuming decreasing, constant or increasing returns to scale in production. A production function F is said to be homogeneous of degree θ if and only if, for all λ:

$$F(\lambda K, \lambda L) = \lambda^{\theta} F(K, L) \tag{1.7}$$

We immediately see from this definition that homogeneity of degree 1 is equivalent to the general definition of constant returns to scale function, while homogeneity of degree greater and lower than 1 imply increasing and decreasing returns to scale respectively.[8] Compared to the general notion of returns to scale, homogeneity of degree θ just puts more mathematical structure on the way production changes when input vary in a proportional way. Restricting our attention the class of homogeneous production functions is useful because it allow us to express F as a linear combination of its partial derivatives. To do so start to differentiate 1.7 with respect to λ

$$\frac{\partial F}{\partial K}(\lambda K, \lambda L)K + \frac{\partial F}{\partial L}(\lambda K, \lambda L)L = \theta \lambda^{\theta-1} F(K, L)$$

and setting λ to 1 gives us[9]

[8] The converse is not true, increasing returns to scale do not imply homogeneity of degree greater than 1 and decreasing returns to scale do not imply homogeneity of degree lower than 1. The equivalence only holds in the constant returns case.

[9] These two equations are respectively the proof and statement of the Euler theorem for a function of two variables.

$$\frac{\partial F}{\partial K}(K,L)K + \frac{\partial F}{\partial K}(K,L)L = \theta F(K,L) \tag{1.8}$$

This equation is just the statement of the Euler theorem for homogeneous functions in a two variables setting.[10] We see from expression that homogeneity is very useful tool to characterize the way F relates to its partial derivatives which, in economic terms, are just the marginal products of its inputs.

When an homogeneous production function has increasing returns to scale ($\theta > 1$) the sum of payments made to its factors of production is greater than total production, implying negative profits in a competitive economy:

$$F(K,L) < \frac{\partial F}{\partial K}(K,L)K + \frac{\partial F}{\partial K}(K,L)L$$

When a production function is homogeneous of degree one ($\theta = 1$) or, equivalently, when a production function exhibits constant returns to scale, the sum of payments made to its factors of production is just equal to total production, implying zero profits in a competitive economy:

$$F(K,L) = \frac{\partial F}{\partial K}(K,L)K + \frac{\partial F}{\partial L}(K,L)L$$

Finally, when an homogeneous production function has decreasing returns to scale ($\theta < 1$) the sum of payments made to its factors of production is lower than total production, implying positive profits in a competitive economy:

[10] The full Euler theorem also states that conversely if F verifies this partial differential equation 1.8 then it is homogeneous of degree θ. This converse statement is important because it implies that the partial derivatives of an homogeneous function of degree θ are themselves homogeneous of degree $\theta - 1$. To see why take the partial derivative of 1.8 with respect to K. This yields:

$$F_{KK}K + F_K + F_{LK}L = \theta F_K$$

where lower scripts K and L are a short hand for partial derivatives. Using the fact that twice differentiable functions verify $F_{KL} = F_{LK}$ we get:

$$F_{KK}K + F_{KL}L = (\theta - 1)F_K$$

Applying the converse of the Euler theorem to F_K, we conclude that F_K is homogeneous of degree $\theta - 1$. And similarly for F_L... In practice this means that when working with constant returns to scale production functions, the marginal products of K and L are homogeneous of degree 0: i.e. they only depend on the relative values of capital and labor, not on their magnitude.

$$F(K,L) > \frac{\partial F}{\partial K}(K,L)K + \frac{\partial F}{\partial L}(K,L)L$$

So far all these results are just a generalization to a two factors setting of our discussion of Ricardian rents in the one factor model. Conveniently the Euler theorem also gives us a very useful way to measure returns to scale in practice. Rewriting 1.8 but dividing both sides by total output gives:

$$\frac{\partial F}{\partial L}\frac{K}{F} + \frac{\partial F}{\partial L}\frac{L}{F} = \theta$$

where for convenience I suppressed the explicit dependence of F and F's partial derivatives on the level of inputs (K,L). This last equation shows that the degree of returns to scales θ associated to an homogeneous production function is just the sum of its factor elasticities which can easily be recovered in the data.[11]

Increasing, decreasing or constant returns ?

We are now in a position to impose some assumptions on our production function F. To clarify things let us consider that the elasticities of F with respect to capital and labor are constant. We will denote these elasticities α and β respectively. As in the one factor case this corresponds of very particular choice of F which will take a Cobb-Douglas form:

$$F(K,L) = K^{\alpha}L^{\beta}$$

It is straightforward to verify that at all levels of K and L.:

$$\alpha = \frac{\partial F}{\partial K}\frac{K}{F}$$

and

$$\beta = \frac{\partial F}{\partial L}\frac{L}{F}$$

and that the degree of returns to scale associated to F is $\theta = \alpha + \beta$.[12] Fol-

[11]Where out of simplicity we dropped the explicit dependence of the partial derivatives of F with respect to the level of its inputs K and L.

[12]Multiplying both K and L by $\lambda > 1$ gives:

$$F(\lambda K, \lambda L) = [\lambda K]^{\alpha}[\lambda L]^{\beta} = \lambda^{1-\alpha+\alpha}K^{\alpha}L^{\beta}$$

which is greater than $\lambda F(K,L)$ if and only if $\alpha + \beta \geq 1$.

lowing our previous discussion F admits increasing, constant or decreasing returns if $\alpha + \beta$ is above, equal to, or below 1.

Referring back to our discussion on the link between decreasing returns to scale and decreasing marginal productivity, we see in this two factor case that returns to each input taken separately can be decreasing — this will be the case if $\beta < 1$ and $\alpha < 1$ — while overall returns to scale $\alpha + \beta$ can be decreasing, constant or even increasing. Intuitively this is because in a world with more than one factor, decreasing marginal productivity in each input can be compensated by cross factors complementarities. In other words, even if increasing labor and capital simultaneously will decrease average output because of decreasing marginal product of each input taken separately, if capital and labor are complements (i.e. if $F_{KL} > 0$) a rising level capital increases the marginal productivity of labor and conversely. If complementarities are strong enough this exactly compensates for the losses arising from each input's decreasing marginal product. If complementarities are even stronger, you can end up with increasing returns to scale despite marginal products being decreasing for each input.

Now, are increasing returns an appealing assumption in this setting? As in the one factor model increasing returns are hard to justify a priori and imply negative equilibrium profits in a competitive setting. What's more, as in the one factor case, assuming increasing returns is somewhat akin to "assuming" growth instead of explaining it.[13]

Are decreasing returns a more appealing assumption? To some extent yes because they are compatible with a well defined competitive equilibrium. But at another level they imply that doubling the number of machines and workers of an economy will result in a lower output per capita level. This is not a particularly appealing consequence if we think that capital and labor are the only two important factors of production in an industrialized economy. Absent a limiting factor of production such as land or natural resources, multiplying by two capital and labor should allow us to multiply by two total output so as to leave output per capita unaffected. This "replication" argument has lead growth economists to prefer the constant returns case as their baseline assumption. For the rest of this lecture (and for many to come) we will henceforth assume constant returns in capital and labor. With the further simplification of constant elasticities with respect to both factors this narrows down our choice of F to:

[13] As we will see later in these lectures a good share of endogeneous growth theory is devoted to understanding how increasing returns can arise as an equilibrium outcome instead of bluntly assuming them in the production function.

$$F(K,L) = K^{\alpha} L^{1-\alpha} \tag{1.9}$$

a two factors Cobb-Douglas function where we have set the labor elasticity β to $1-\alpha$ in order for overall returns to scale to be constant.

The central role of capital accumulation

What does this choice of F imply for growth? Let us proceed as we did in the one factor case and solve for the growth rate of output per capita as a function of (i) the growth rate of population, (ii) the growth rate of capital and (iii) the elasticities of output with respect to capital and labor. To do so take the time derivative of total output:

$$\dot{Y}_t = \frac{\partial F}{\partial K}\dot{K}_t + \frac{\partial F}{\partial L}\dot{L}_t$$

and re-arrange to make growth rates and elasticities appear:

$$\frac{\dot{Y}_t}{Y_t} = \frac{\partial F}{\partial K}\frac{K_t}{F}\frac{\dot{K}_t}{K_t} + \frac{\partial F}{\partial L}\frac{L_t}{F}\frac{\dot{L}_t}{L_t}$$

Under our constant returns to scale the capital and labor elasticities of output sum to 1. So using the fact that:

$$\frac{\partial F}{\partial L}\frac{L_t}{F} = 1 - \frac{\partial F}{\partial K}\frac{K_t}{F}$$

we can rewrite:

$$\frac{\dot{Y}_t}{Y_t} = \frac{\partial F}{\partial K}\frac{K_t}{F}[\frac{\dot{K}_t}{K_t} - \frac{\dot{L}_t}{L_t}] + \frac{\dot{L}_t}{L_t}$$

Subtracting $\dot{L}_t/L_t$ on both sides and noticing that

$$\frac{\dot{K}_t}{K_t} - \frac{\dot{L}_t}{L_t}$$

is just the growth rate of the capital/labor ratio $k_t = K_t/L_t$ we finally arrive at an expression for $\dot{y}_t/y_t$:

$$\frac{\dot{y}_t}{y_t} = \frac{\partial F}{\partial K}\frac{K_t}{F}\frac{\dot{k}_t}{k_t}$$

The fact that output per capita growth can be expressed as a direct function of capital per capita growth and the elasticity of output with respect to capital is a direct consequence of the constant returns to scale assumption.[14] Adding the fact that the elasticity of output with respect to capital is α in our Cobb-Douglas case gives a simple expression for $\dot{y}_t/y_t$:

$$\frac{\dot{y}_t}{y_t} = \alpha \frac{\dot{k}_t}{k_t}$$

implying that output per capita growth will be exactly proportional to the growth of capital per capita. Hence the two factors model of growth tells a story where societies get richer because of capital accumulation — or to be more precise, because capital accumulates at a faster rate than population growth. Unlike in the one factor model, here the growth rate of population does not affect output per capita directly. It does so only indirectly through its negative influence on the measure of capital per person. Another interesting property of this simple two factors model is that output per capita growth does not depend on increasing returns to factors of production but can arise endogenously in a "realistic" setting with well defined factor shares. In the rest of these lectures we will devote a great deal of time to understanding the determinants of capital accumulation.

Solow (1957), at last...

Does our two factors model of growth give a better of the actual data than our one factor model? A way to answer this question is look at the value of the capital share α needed to match the observed pattern of y and k. Indeed our two factor model implies that:

$$\alpha = \frac{\dot{y}_t}{y_t} / \frac{\dot{k}_t}{k_t}$$

To recover an estimate of α we will need to gather data on the evolution of productive capital over time. A simple way to do so is the use the Penn World Tables. Over the 1950/2019 output per capita growth averaged 2% per year in the US. Over the same period the capital/labor ratio grew by

[14]To understand why assume that capital and labor grow at the same constant rate g so that the capital/labor ratio is constant. In this case constant returns to scale implies that total output will also grow at rate g, so that output per capita will grow at rate $g - g = 0$... This is the reason why under constant returns to scale, capital needs to grow faster than labor in order for output per capita to increase.

1.65% per year. This implies that if our model were true, α would have to be equal to $2/1.65 \sim 1.21$. While this estimate does seem far less outrageous than what was needed in the one factor model in order to match the data, it is still at odds with observed capital shares which averaged 30% in US over the last 70 years.

Another way to tackle this same problem is to turn around the question: "Given that capital accounts on average for 30% of production, to what extent can capital accumulation explain the observed variation of output per capita?" This is the question asked by Robert Solow in a very famous and important 1957 paper. In this paper Solow's approach is very close to our two factor model. The only difference is that Solow allows for an unobserved "technological" factor to affect the overall productivity of capital and labor. More concretely Solow (1957) assumes that output is produced according to:

$$Y_t = A_t F(K_t, L_t) \tag{1.10}$$

where A_t is a factor affecting the level production at a given level of capital and labor inputs. In other words a positive shift in A_t will increase production holding K_t and L_t constant. At this stage the interpretation of A_t as a the level of technology may seem arbitrary. Indeed A_t will include actual technological improvements but may also reflect other non technological factors which affect total output at a given level of inputs. As Solow himself points out in his 1957 paper, A_t is just a way to measure changes in total output that are linked to shifts in the production function rather than to changes in inputs. The idea is simple and powerful. Consider the augmented two factor model 1.10. When output per capita changes, this change can either be explained by increased capital per capita — as we saw in the two factors model — or by a shift in the whole production function. But given that the elasticity of output with respect to capital is close to 0.3, changes in capital per capita $\dot{k}_t/k_t$ can only explain $0.3\dot{k}_t/k_t$ of the observed changes in output $\dot{y}_t/y_t$. The unexplained part, if any, must be attributed to shifts in the production function. To see this mathematically we can use 1.10 to express the growth rate of output per capita in this augmented two factors model. Taking the time derivative and rearranging terms yields:

$$\frac{\dot{y}_t}{y_t} = \frac{\dot{A}_t}{A_t} + \frac{\partial F}{\partial K}\frac{K_t}{F}\frac{\dot{K}_t}{K_t} + \frac{\partial F}{\partial L}\frac{L_t}{F}\frac{\dot{L}_t}{L_t}$$

As in the previous section, and for the same reasons, we will continue to assume constant returns to scale in capital and labor. As before, this assumption allows us to express the growth rate of output per capita more compactly as:

$$\frac{\dot{y}_t}{y_t} = \frac{\dot{A}_t}{A_t} + \frac{\partial F}{\partial K}\frac{K_t}{F}\frac{\dot{k}_t}{k_t}$$

This is the central equation of Solow's 1957 paper. Using data on output per capita and capital intensity in the United-States. Notice that unlike the Cobb-Douglas case were we can replace the elasticity of output with respect to capital by its constant value α, Solow does not assume constant elasticities in his production function. Instead Solow (1957) uses time varying capital shares to identify the capital elasticity of output at different points in time.[15] The original data set used by Solow appears in Table I of the 1957 paper and covers the 1909/1949 period. Solow uses private non farm GDP per man hour to measure y and α_t the time varying share of property income in total income to measure the capital elasticity of output. For want of a direct measure of capital utilization, Solow adjusts the measured capital stock by the employment rate at each date. This adjusted measure of capital is then divided by the total number of hours worked to obtain a measure of capital per man hour. Computing an estimate of $\dot{A}_t/A_t$ as:

$$\frac{\dot{A}_t}{A_t} = \frac{\dot{y}_t}{y_t} - \alpha_t\frac{\dot{k}_t}{k_t}$$

This exercise yields an average growth rate of technology around 1.5% per year over the 1909/1949 period. Normalizing A_{1909} to 1, these estimates imply that by 1949 the US level of technology almost doubled to reach $A_{1949} = 1.809$.

Solow, however, does not stop there. His estimate of A_t allows him to recover the shape of aggregate production function F. Rewriting 1.10 as:

$$\frac{Y_t}{A_t} = F(K_t, L_t)$$

dividing by L_t on both sides and making use of the constant returns to scale assumption yields:

$$\frac{y_t}{A_t} = F(k_t, 1)$$

[15] Recall that this is valid only under the hypothesis of perfect competition, i.e. if factor prices equal marginal productivity in value.

The $F(k_t, 1)$ is usually designated as the "intensive" form of the production F, i.e. the output per capita as a function of the capital labor ratio. In general we will denote this function by $f(k)$. Recall that if F is approximately equal to a Cobb-Douglas function with constant input elasticities (which is the case in Solow's data) then $F(K,L) = K^{1-\alpha}L^{\alpha}$ and the intensive form of F will write k^{α}. Under such a Cobb-Douglas approximation Solow's data should satisfy:

$$\frac{y_t}{A_t} = F(k_t, 1) = f(k_t) = k_t^{\alpha}$$

In other words, plotting the estimated value of y_t/A_t against the value of k_t should "look like" the intensive form of a Cobb-Douglas aggregate production function. Figure 1.4 below reproduces the original scatter plot of normalized output per man hour against capital per man hour which you can find in Solow's 1957 paper. Not surprisingly, fitting a Cobb-Douglas through this scatter plots yields an estimated value of α of 0.35 (Solow (1957), Table 2), consistent with the observed share of capital in total output.

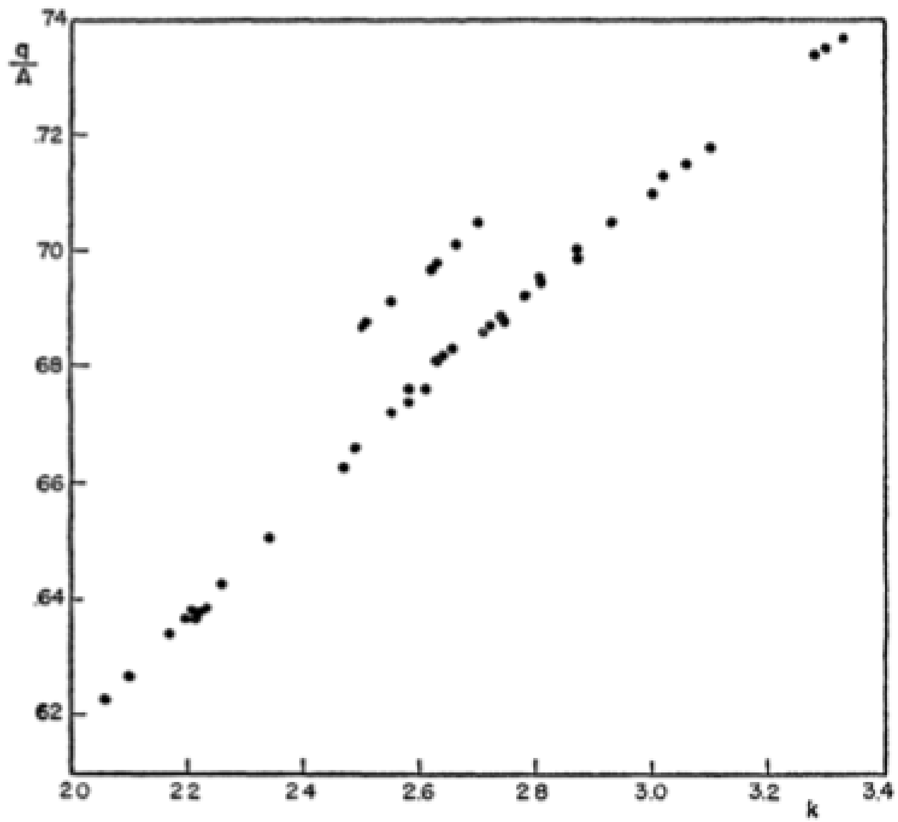
How much of observed growth in output per capita can be traced back to technology and capital respectively? Between 1909 and 1949, Solow's measure of y grew from 0.623 to 1.275 thousands of 1939 dollars. The normalized measure of output per man hour y/A, however, only grew to 0.705 thousands of dollars. Overall technical change explains

$$\frac{1.275 - 0.705}{1.275 - 0.623}$$

that is to say 87.5% of output per capita growth over the period, while capital accumulation accounts for the rest.

Solow (1957) is both fascinating and disappointing. Fascinating because from very simple and clear assumptions Solow is able to recover "by hand" the aggregate production function of the US economy. Disappointing because his calculations imply that capital accumulation contributes only modestly to the growth of output per capita, 87.5% of which remains to be explained.

Figure 1.4: The aggregate production function



This figure reproduces Chart 4 of Solow's 1957 original paper.

Chapter 2

Solow's model of economic growth

> (I have never been sure exactly what it is that is 'balanced' along such a path, but we need a term for solutions with this constant growth rate property and this is as good as any.)
>
> Lucas (1988)

So far our understanding of economic growth relied on a model of aggregate production. However, the theoretical relationship that we hypothesized between aggregate output and aggregate inputs — i.e. the aggregate production function — proved to be unable to match the observed pattern of economic growth on its own. Indeed in the last section we saw that more than 80% of actual growth remains unexplained by the growth in observable inputs. Following Solow's 1957 growth accounting exercise, we may label these unexplained 80% "technical change" or more simply "knowledge". Even if we were willing to accept this interpretation of Solow's residual, a satisfactory theory of growth would need to explain why technology grows and why capital accumulates. In other words, the aggregate production function gives us a theory of y given A and k, but neither a theory of A nor a theory of k. The rest of this course will be devoted to building and analyzing models which explicitly account for these two sources of economic growth. Models in which technological improvements arise endogenously from the equilibrium behavior of agents are usually called "endogenous growth models". We will study them in the second part of these lectures. This label, however, is somewhat misleading. In this lecture and the following we will study models in which growth in output per capita arises from the endogenous accumulation of factors of production. As we will see the reason why this second type of models doesn't deserve the same laudatory "endogenous growth" label is not because they lack an endogenous component, but because they fail to explain the possibility of long term growth.

2.1 Focusing on capital accumulation: the original Solow model.

As we saw in the last lecture, growth in output per capita can at least in part be traced back to capital accumulation. Capital accumulation in turn should depend on the rate at which economic agents decide to invest rather than consume. Following this logic the saving rate s of an economy should be an important determinant of growth. This is the hypothesis that Solow explores in his 1956 paper.[1]

Capital accumulation

Abstracting from technological progress Solow (1956) assumes that output is produced from capital and labor according to:

$$Y_t = F(K_t, L_t) \tag{2.1}$$

where F is an increasing, concave, constant returns to scale production function. As we explained in our discussion of the two factors model, concavity which implies decreasing marginal products of capital and labor taken separately, is perfectly compatible with constant returns to scale to capital and labor when taken jointly. More precisely this will be case when capital and labor are complements in production — i.e. when an increase in one of the

[1] If you read Solow (1956), you will see that Solow's main concern was to understand to understand the consequences of different saving rates on the possibility of balanced growth. At the time this concern arose from the separate models of Harrod and Domar showing that balanced growth, which neither creates inflation from labor shortages, nor deflation and unemployment from insufficient aggregate demand, was only possible under one single knife-edge saving rate. Harrod and Domar's conclusion, as Solow (1956) shows, only holds under the very strict assumption of perfect complementarity between capital and labor in the aggregate production function (i.e. Leontieff). By allowing for some degree of capital/labor substitutability, Solow showed that any saving rate is compatible with balanced growth.

inputs will raise the marginal productivity of the other. From our previous discussion of Ricardian rents and concavity we know that:

1. Concavity implies that the first order conditions associated to a firm's profit maximization problem are sufficient to characterize a interior maximum.

2. Constant returns to scale imply the Ricardian rent is zero, and so will be profits in a competitive equilibrium, with factor shares summing to one.

On the demographic side Solow assumes that population grows at constant instantaneous rate n.[2] Mathematically this just means that:

$$\frac{\dot{L}_t}{L_t} = n$$

or, equivalently, that starting from some initial value: L_0

$$L_t = L_0 e^{nt}$$

Finally capital accumulation is determined by two opposite forces. On the one hand the representative agent of this economy is assumed to save a fraction s of income at each date, thereby increasing the current capital stock by sY_t. On the the other hand the existing capital stock depreciates at a constant rate δ so that δK_t of the installed amount of capital disappears at every point in time.

In an discrete time context these assumptions would translate into a equation for period $t+1$ capital K_{t+1} as function of period t production $Y_t = F(K_t, L_t)$ and capital stock K_t:

$$K_{t+1} = sF(K_t, L_t) + (1-\delta)K_t$$

To derive Solow's continuous time equivalent of this discrete time capital accumulation equation we have to study what happens to the capital stock over a small interval of time Δt, i.e. between time t and time $t + \Delta t$. Start from time t at which capital is K_t and labor L_t. Output at time t will be equal to $Y_t = F(K_t, L_t)$. Production over the small time interval Δt will be equal to

$$\Delta t F(K_t, L_t)$$

[2] Note that we will maintain the assumption of exogenous population growth almost all of the rest of these lectures.

a fraction s of which will saved and added to the existing capital stock. Because capital depreciates at the *instantaneous* rate δ the loss in capital between t and $t+\Delta t$ will be:

$$\delta \Delta t K_t$$

Concretely this means that absent positive savings the capital stock would *decline* at the constant rate δ.[3] Pulling savings and depreciation together we get an equation for capital at time $t+\Delta t$

$$K_{t+\Delta t} = s\Delta t F(K_t, L_t) + (1-\delta\Delta t)K_t$$

which can be re-arranged as:

$$\frac{K_{t+\Delta t} - K_t}{\Delta t} = sF(K_t, L_t) - \delta K_t$$

Taking the limit as Δt goes to zero we get an expression for $\dot{K}_t$ the time derivative of the capital stock in the Solow model:

$$\dot{K}_t = sF(K_t, L_t) - \delta K_t \tag{2.2}$$

The quantity $\dot{K}_t$ represents investment in new capital net of the depreciation of the existing capital **per unit of time**.[4]

To fix ideas divide 2.2 by K_t on both sides in order to get an expression for the growth rate of capital:

[3] In other words, absent savings we would have:

$$K_{t+\Delta t} = (1-\delta\Delta t)K_t$$

You might recognize that formula is the "definition" of what would happen to a variable growing at the negative instantaneous growth rate $-\delta$ in continuous time. For a full treatment you can refer back to footnote 3.

[4] If you want to "recover" the value of $K_{t+\Delta t}$ from given values of K_t and $\dot{K}_t$ you would simply use:

$$K_{t+\Delta t} = K_t + \dot{K}_t \Delta t$$

which is just a first order approximation of $K_{t+\Delta t}$, i.e. when Δt is small, true up to an $o(\Delta)$. Again, going back to footnote 3 you can rewrite this as:

$$K_{t+\Delta t} = (1 + \frac{\dot{K}_t}{K_t}\Delta t)K_t$$

where $\dot{K}_t/K_t$ is the instantaneous growth rate of K_t.

$$\frac{\dot{K}_t}{K_t} = s\frac{Y_t}{K_t} - \delta$$

The average growth rate of capital between 1949 and 2019 in the US was about 2.7% per year. During the same period the average output/capital ratio and the share of investment in GDP were both close to 1/4. Hence the Solow capital accumulation equation implies a value of δ close to $1/16 - 0.027 \sim 0.035$ per year. Not a completely unrealistic value.

From aggregate to per capita quantities

As the goal of growth theory is to understand the behavior of output per capita rather than output per se, it is often convenient to transform 2.1 and 2.2 in equations describing output per capita $y_t = Y_t/L_t$ and the evolution of the capital/labor ratio $k_t = K_t/L_t$. This transformation is greatly facilitated by our constant returns to scales assumption. Indeed dividing 2.1 by L_t on both sides we get:

$$y_t = \frac{Y_t}{L_t} = \frac{F(K_t, L_t)}{L_t} = F(\frac{K_t}{L_t}, 1) = F(k_t, 1) := f(k_t) \tag{2.3}$$

where we define f as the per capita production function which takes the capital/labor ratio as its unique argument. In the rest of these lectures we will use this per capita production function repeatedly. Concretely compared to the original aggregate production F, small f tells the quantity of output that a single worker is able to produce using a given quantity k of capital, machines... The fact that we assumed constant returns to scale in capital and labor means that big F can be reconstructed from the knowledge of the amount of capital used per worker and of the actual number of workers taking part in production. In other words, if one worker with capital k produces $f(k)$, then L workers each using k units of capital produce

$$Lf(k) = LF(k, 1) = F(L \times k, L) = F(K, L)$$

where $L \times k = K$ measures the aggregate capital stock used in production in economy where each employed worker uses k units of capital. Constant returns imply that is completely equivalent to produce L times with one worker and k units of capital or to produce one time with L workers and $L \times k = K$ units of capital.

Finally note again that in this model without technological improvements, constant returns to scale imply that growth in output per capita can only arise out of a rising capital/labor ratio: in other words, there is a one-to-one relationship between y_t and k_t.[5] So much for the per capita production function f.

Transforming the expression for $\dot{K}_t$ into an expression for $\dot{k}_t$ takes a few bit more lines of calculus:

$$\begin{aligned}
\dot{k}_t &= \frac{\dot{K}_t L_t - K_t \dot{L}_t}{L_t^2} \\
&= \frac{\dot{K}_t}{K_t}\frac{K_t}{L_t} - \frac{\dot{L}_t}{L_t}\frac{K_t}{L_t} \\
&= [s\frac{F(K_t, L_t)}{K_t} - \delta]\frac{K_t}{L_t} - n\frac{K_t}{L_t} \\
&= sF(\frac{K_t}{L_t}, 1) - (\delta + n)\frac{K_t}{L_t}
\end{aligned}$$

In other words

$$\dot{k}_t = sf(k_t) - (\delta + n)k_t \tag{2.4}$$

The intuition for 2.4 is straightforward the capital/labor ratio, i.e. the amount of capital *per capita* grows in proportion to saving *per capita* $sf(k_t)$ and declines either because of capital depreciation which affects all units of capital K_t or because of population growth which affects L_t the denominator of k_t. This is the central equation of the original Solow (1956) model of economic growth.

How should we think of equation 2.4 in practice? Rearranging its terms in order to get an expression for the growth rate of capital per capita we get that:

$$\frac{\dot{k}_t}{k_t} = s\frac{y_t}{k_t} - \delta - n$$

A quick glance at Penn World Tables allows us gauge the empirical relevance of such a relation. In the US, between 1950 and 2019, the growth rate of capital per worker averaged 1.6%, the share of gross investment in output was close to 25% (i.e. a simple measure for the saving rate s), the output/capital ratio was close to 25%, and population growth averaged 1.6%. Putting back these values in the equation for $\dot{k}_t/k_t$ allows us to compute a

[5] On this point you refer back to our discussion of the "two factors model" and more particularly to footnote 14.

depreciation rate consistent with the observed pattern of growth in the US. In practice this gives us:

$$
\begin{aligned}
\delta &= s\frac{y_t}{k_t} - n - \frac{\dot{k}_t}{k_t} \\
&= 0.25 * 0.25 - 0.01 - 0.016 \\
&= 0.0365
\end{aligned}
$$

Not an implausible value.

Competitive equilibrium

So far we have treated our economy in a very mechanical way. Capital per worker is accumulated because of a fixed saving rate, there are no markets and, as a consequence, neither equilibrium nor equilibrium prices. If you read Solow's original 1956 paper, however, you will find a discussion of factor prices for capital and labor. How come? Implicitly Solow assumes that a competitive structure governs the allocation of factors as well as the production of a unique consumption/investment good. The only crucial variable which is *not* the outcome of a well defined economic equilibrium is the saving rate s.

We already know that constant returns to scale ensure that:

$$F(K_t, L_t) = F_K K_t + F_L L_t$$

But competitive equilibrium factor prices, the wage w_t and rental price of capital r_t are given by:

$$w_t = F_L(K_t, L_t)$$

and

$$r_t = F_K(K_t, L_t)$$

So that total payments to the owners of capital and labor exactly equal total production:

$$F(K_t, L_t) = r_t K_t + w_t L_t$$

This allows us to write total savings realized by households as a fraction of total production instead of as fraction of total household income.

Before moving on to solving Solow's model, notice that, very much as we translated the evolution of aggregate capital into the evolution of capital per capita, factor prices can be directly expressed as functions of the the capital labor ratio k_t. To do so start with the competitive rental rate of capital for which:

$$r_t = F_K(K_t, L_t) = F_K(k_t, 1) = f'(k_t)$$

And using Euler's relation applied to F, competitive wages can then be expressed as:

$$w_t = F_L(K_t, L_t) = \frac{F(K_t, L_t)}{L_t} - \frac{F_K K_t}{L_t} = f(k_t) - f'(k_t)k_t$$

The dynamic behavior of capital in the Solow model

Mathematically equation 2.4 implies that the path of k_t predicted by Solow's 1956 model will be the solution of a non linear first order differential equation.[6] While equations like 2.4 do not in general admit explicit solutions[7], the behavior of k_t can usually be inferred by carefully looking at the implied movements of the variable of interest, in this case k_t. Indeed if the capital/labor ratio of our economy truly followed 2.4 then:

- k_t would be growing over time ($\dot{k}_t > 0$) whenever savings are greater than capital per capita losses due to depreciation and population growth:
$$sf(k_t) > (\delta + n)k_t$$

- k_t would be constant over time ($\dot{k}_t = 0$) whenever savings taken from current production exactly compensate the capital per capita losses due to depreciation and population growth:

$$sf(k_t) = (\delta + n)k_t$$

[6] "Non linear" because f does not have any reason to be linear and "first order" because equation 2.4 only involves the first order time derivative of k_t.

[7] An explicit solution would be a formula for k_t as a function of time and the parameters of the model

$$k_t = 3(s+n) \times t + \frac{\delta}{t}$$

for instance...

- k_t would be declining over time ($\dot{k}_t < 0$) whenever savings are lower than capital per capita losses due to depreciation and population growth:

$$sf(k_t) < (\delta + n)k_t$$

Which of these three cases will arise at any point in time will depend on the value k_t and on the shape of the per capita production function f.

Yet more assumptions on the production function

The dynamic equation 2.4 for $\dot{k}_t$ implies that the shape of the per capita production f will determine the behavior of capital accumulation in the Solow model. Per capita production $f(k)$, however, is just a short hand for $F(k,1)$, and thereby inherits its properties from the original production function. In particular the first and second derivatives of f are given by:

$$f'(k) = F_K(k,1)$$

and

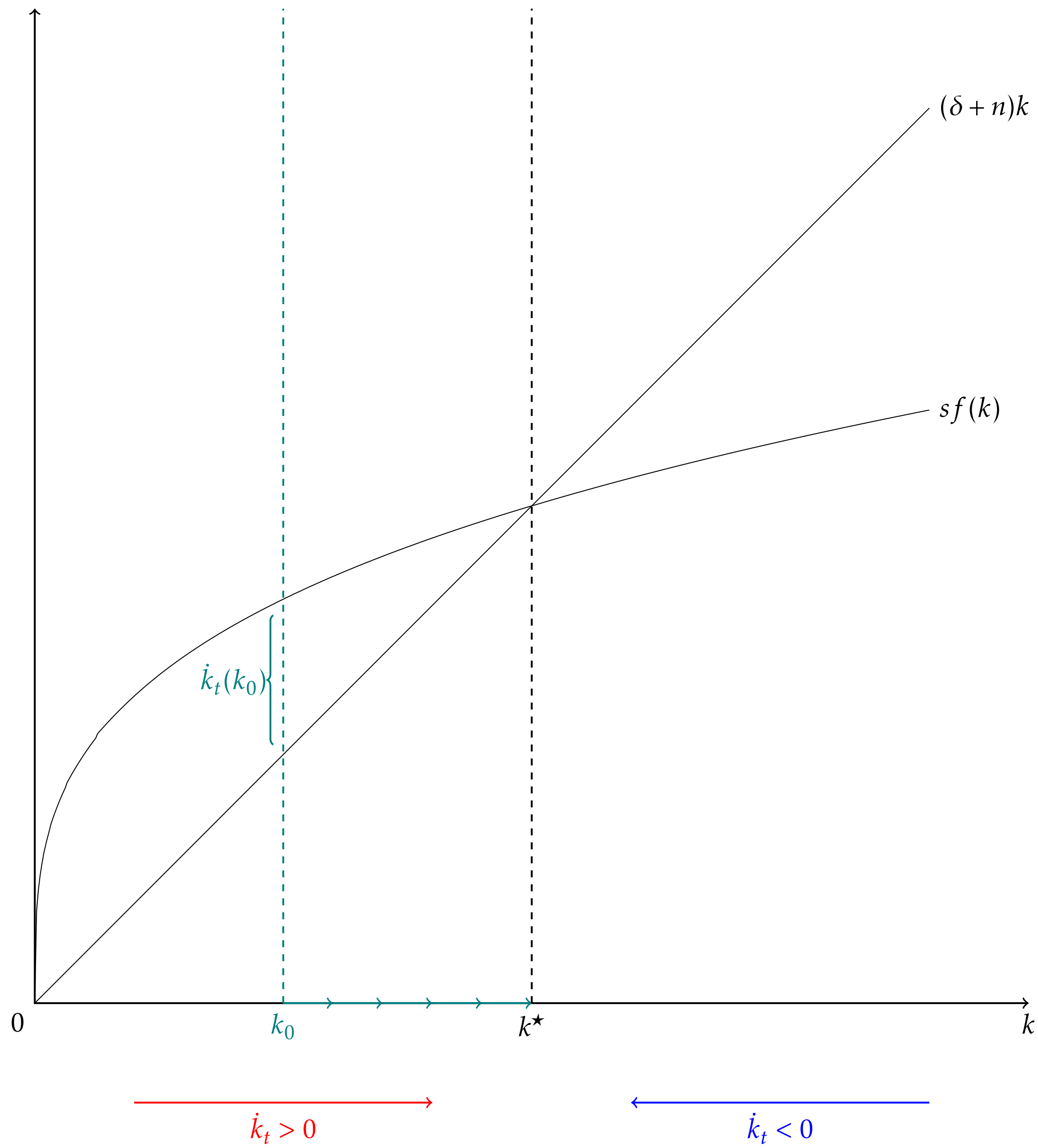
$$f''(k) = F_{KK}(k,1)$$

where F_K and F_{KK} stand for the first order and second order partial derivatives of F with respect to capital. But because F is a concave function, $F_{KK} < 0$ and so does f''. In other words decreasing marginal returns in capital translate into decreasing marginal returns in capital per capita.[8]

Figure 2.1 below graphs (i) the flow of savings $sf(k)$ arising out of such a concave per capita production function $f(k)$, as well as (ii) the $(\delta+n)k$ depreciation function. Referring back to 2.4, we see that at any level of capital per worker, the difference between the two curves gives us the value of $\dot{k}_t$. As you can see, the way we drew the per capita saving function sf on the diagram implies that the two curves cross at the level of per capita capital $k^\star$. Recall that at this level savings exactly compensate depreciation and population growth so that the capital labor ratio remains stable. As a consequence, if the capital/labor ratio k_t reaches this value at some point in time, equation 2.4 implies that it will then remain there for the rest of times. Points such as $k^\star$ where the time derivative of the capital/labor ratio vanishes (i.e. where $\dot{k}_t = 0$) are called "steady states".[9]

[8]This should come as a surprise: if the marginal product of capital is decreasing at any

Figure 2.1: THE SOLOW MODEL



given level of labor L, it should also be decreasing at $L = 1$.

[9]You may have noticed that $k = 0$ appears to be another steady state because of the way

Existence

Unfortunately, in order to ensure that a non trivial steady state exists in this economy our previous assumptions on f are not sufficient. If you go back to Figure 2.1 you will find that it is perfectly possible to draw an increasing and concave per capita savings function which does not cross the depreciation function $(\delta+n)k$. This is why growth theorists often impose some additional assumptions on the production function f. These assumptions are called Inada's conditions and require that:

1. The marginal product of capital be infinite at $K = 0$

$$\lim_{K \mapsto 0} F_K(K, L) = +\infty$$

   or in per capita terms:

$$\lim_{k \mapsto 0} f'(k) = +\infty$$

2. The marginal product of capital be zero when K goes to infinity:

$$\lim_{K \mapsto +\infty} F_K(K, L) = 0$$

   or in per capita terms:

$$\lim_{k \mapsto +\infty} f'(k) = 0$$

Inada's conditions are stronger than needed for Solow's central result (they are sufficient, not necessary) but will be useful later on when we will study the Ramsey model of economic growth. From an economic point of view Inada's conditions imply that, *holding labor fixed*, the marginal product of capital is very high for small amounts of capital and very low when the capital stock grows large. Mathematically this implies that at $k = 0$ the derivative of $sf(k)$ is very large and, in particular, larger than $\delta+n$ as depicted in Figure 2.1. If you add the common "free disposal" assumption that $f(0) = 0$, this implies that moving away from the origin the $sf(k)$ curve will lie above the $(\delta+n)k$ one. When capital per capita grows larger, decreasing returns imply that the derivative of savings per capita $sf(k)$ will decline.

we drew the function f (the free disposal assumption from your microeconomics class, $f(0) = 0$). Such trivial steady states will arise frequently in our the models we will study. We will discard them both because they are economically not very interesting and because as we will see, they generally are unstable.

The second Inada assumption implies that at some point this decline will bring the slope of $sf(k)$ below $\delta+n$. Up to this point the gap between the two curves was steadily increasing with k. After this point this gap will start to decrease. What's more, the fact that $sf'(k)$ will eventually go to zero implies that the gap between the two curves will continue to decrease as capital rises and will eventually turn negative. Because both curves are continuous this means that the two curves must cross at some point, so that at least one steady state $k^\star$ exists.

An easy way to visualize the influence of the level of capital per capita on its time derive $\dot{k}_t$ is to plot directly the difference between the two curves of Figure 2.1. This is done in Figure 2.2. As this second figure shows, $\dot{k}_t$ will be a non-monotonous function of the level of capital k. It will start to rise, reach a maximum when $sf'(k) = \delta + n$, start to decline, vanish at $k^\star$ and turn negative.

Uniqueness

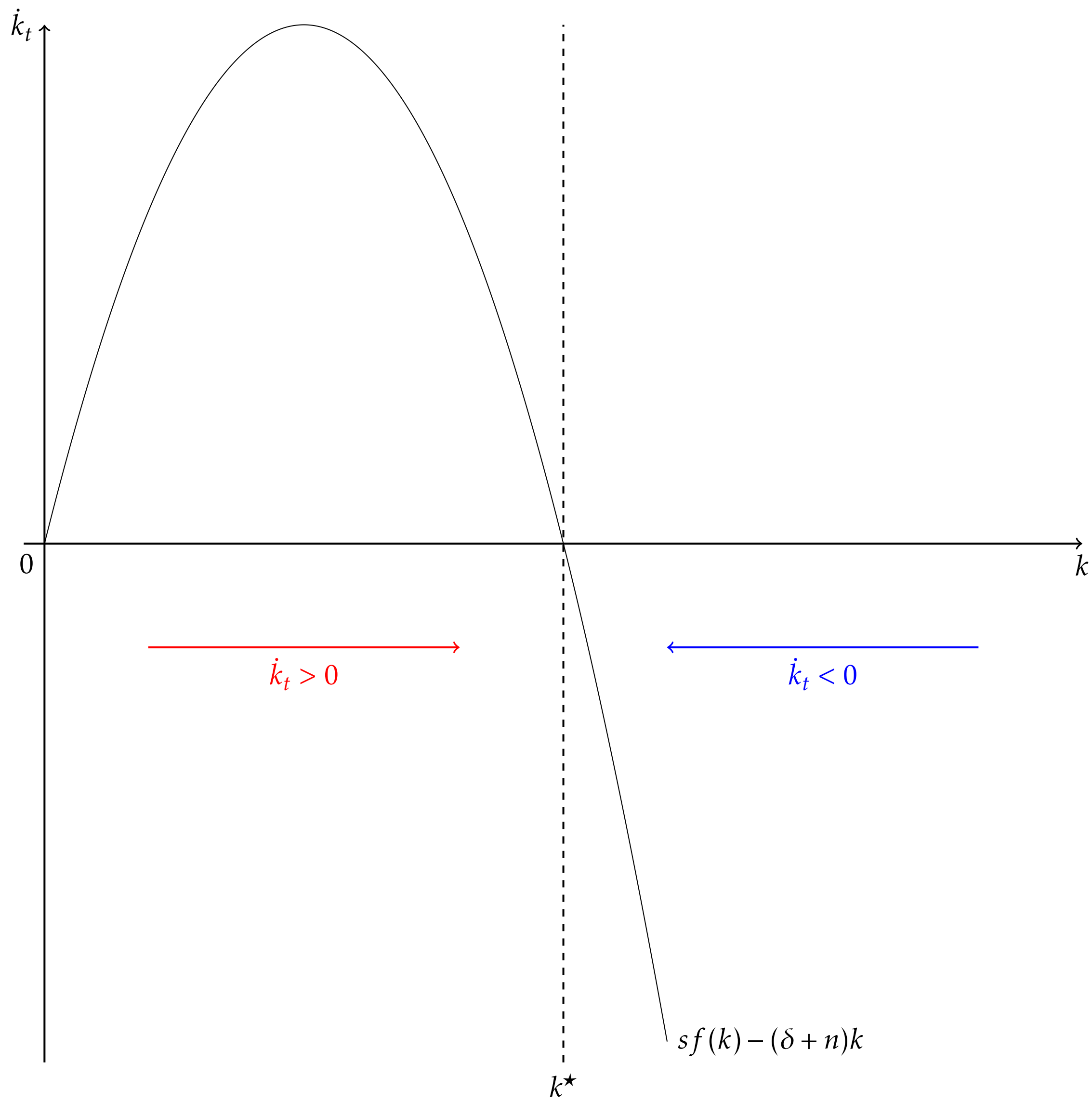
We just saw that Inada's condition imply that at least one steady state exists. Uniqueness however does not follow from these boundary conditions. In principle at least, a lot of things could happen between 0 and $+\infty$ on which Inada's conditions do not have much of say. Of course tha way we drew the $sf(k)$ curve in Figure 2.1 suggests that only one steady state can exist. This, however, is a direct consequence of our assumption that f is concave. More precisely, the concavity of f, implies that f' is monotonous. If you try to draw a continuous function f satisfying Inada's condition but giving rise to several steady states, you will see that it is only possible if you allow f' to be non monotonous — or in other words if the curvature of f is allowed to switch direction at least once. Importantly note that uniqueness of the steady state would also arise if f were convex.[10] What matters for uniqueness is that in both cases f' has to move in one direction, i.e. be monotonous.

Stability

So far our assumptions on f allowed us to show that a steady capital

[10]In which case you would have to flip around Inada's conditions in order to ensure existence...

Figure 2.2: Net savings in the Solow model



labor ratio exists and is unique. The last important question we have to solve is: does this steady state matter at all? Will capital per worker converge it if we let the economy run from an initial level of capital per worker k_0? This is were assuming concavity rather than convexity is crucial.

Indeed the concavity of f implies that the $sf(k)$ curve will cross the $(\delta+n)k$ one from above in Figure 2.1, or equivalently that $\dot{k}_t$ is negatively sloping[11] at $k^\star$ in Figure 2.2. A direct consequence is that $\dot{k}_t$ will be positive to the left of $k^\star$ and negative to its right. In other words, capital will be rising when taking values below the steady state and declining when taking values above the steady state. A steady state verifying these dynamic properties is said to be "stable". To see why, consider an economy starting with an initial level of capital per capita k_0 below $k^\star$ very much like the example depicted in Figure 2.1. Our assumptions on f imply that $\dot{k}_0 > 0$, and that this will be true as long as k_t remains below $k^\star$. For this reason capital per worker will grow over time until it reaches its steady state level $k^\star$. Symmetrically, an initial level of capital per worker above the steady state level $k^\star$ implies $\dot{k}_t < 0$ until k_t reaches $k^\star$. In this case capital will decline until it reaches its steady state level. Once this level is reached, whether from one side or the other, the very nature of $k^\star$ implies that capital will stay there forever.

All in all:

1. Inada's condition ensure that a steady state exists.
2. monotonicity of f' ensures that this steady state is unique,
3. and the concavity of f ensures that, at this steady state, the $\dot{k}_t$ curve is negatively sloping so that $k^\star$ is a stable steady state.

Solow's predictions

Solow's simple model of growth makes very strong prediction on what what should expect to see in the data:

1. Solow's model predicts that capital accumulation alone cannot lead no sustained growth. In other words Solow's growth model predicts that there should be "no growth" in the long run.

[11] A technical side note. Recall that equation 2.4 determines the time derivative of k_t as a function of the level of k_t. Concavity of f implies that:

$$\frac{d\dot{k}_t}{dk_t}(k^\star) < 0$$

as we will see later when we study more general models, the fact the derivative of $\dot{k}_t$ with respect k_t is negative when evaluated at the steady state is a necessary and sufficient condition for this steady state to be stable. For a general discussion of this topic see Appendix **??**.

2. As a consequence, important parameters like the saving rate or population growth rate do not have growth effects but only level ones.

3. Because countries should stop growing when reaching their steady state, we should expect poorer countries which are still far from their steady state to "catch up" with richer countries. In other words we should observe convergence in the levels of GDP per capita over time.

Let us look more closely at each of these predictions.

We have just shown that if our economy behaves according to Solow's assumption, capital per capita (or the capital labor ratio, or capital per worker...) will always converge to steady state level $k^\star$. What's more this steady state level of capital is defined as the unique solution to:

$$sf(k^\star) = (\delta + n)k^\star \tag{2.5}$$

A direct consequence is that the steady state, or long run, level of output per capita will be given by:

$$y^\star = f(k^\star)$$

Today, more than half a century after Solow's original contribution, this statement appears more as a textbook curiosity than a scientific discovery. But taking a step back and forgetting about what we know from textbooks, Solow's conclusion still remains a striking result. According to Solow's model, economies may grow for some time thanks to capital per worker accumulation, but this process of growth will eventually stop. As a consequence, the saving rate which was thought of as an essential determinant of an economy's growth rate does not actually influence growth in the long run — because there is no growth in the long run regardless of the level the saving rate. Fundamentally this is a direct consequence of decreasing returns to scale. Imagine that some astute reform — converting every catholic to protestantism, for instance — allows you to increase permanently the saving rate of country. The $sf(k)$ curve in Figure 2.1 will shift up, and so will the new steady state. Following this shock your economy will start to grow, but only for some time. The reason why is twofold. On the one hand, rising capital per worker implies that the absolute losses from depreciation and population growth will also rise. On the other hand, because of decreasing returns to capital, rising capital per worker will decrease the amount of new savings which can be added to the existing capital stock to compensate for depreciation. At some point the fall in capital's marginal product will be such that new savings exactly compensate depreciation and population growth, and the economy will stabilize at this new steady state. Hence a higher saving rate does not imply faster economic *growth* in the long run but only influences the *level* of output per capita that an economy will be able to sustain in the long run.

How do the parameters of the model influence the steady state capital labor ratio $k^\star$? A simple differentiation[12] of 2.5 with respect to s, δ and n shows that:

$$\frac{\partial k^\star}{\partial s} = \frac{f(k^\star)}{\delta + n - sf'(k^\star)} > 0$$

$$\frac{\partial k^\star}{\partial \delta} = \frac{k^\star}{sf'(k^\star) - \delta - n} < 0$$

$$\frac{\partial k^\star}{\partial n} = \frac{k^\star}{sf'(k^\star) - \delta - n} < 0$$

which are all deduced from the fact that at the steady state $sf'(k^\star) - \delta - n < 0$. We already commented on the fact that the saving rate has a positive impact on steady state capital. Capital depreciation and population growth have similar bu opposite effects. Higher depreciation and faster population growth will both make it harder to maintain a high level of steady state capital per worker and, as a consequence, steady state output per capita.

At a more general level, an important prediction of Solow's model is that poor countries should grow faster than richer ones which are closer to their steady state level. Recall that the growth rate of output per capita is given by:

$$\frac{\dot{y}_t}{y_t} = \frac{F_K K_t}{F} \frac{\dot{k}_t}{k_t}$$

and that Solow's capital accumulation equation implies that:

$$\frac{\dot{k}_t}{k_t} = s\frac{f(k_t)}{k_t} - \delta - n$$

[12] Mathematically speaking, this differentiation is not so "simple". Here we are using the implicit functions theorem which states under which conditions we are allowed to take the partial derivative of an equation with respect some of its parameters. The name "implicit functions" theorem comes from the fact that $k^\star$ is implicitly defined by 2.5 as a function of s δ and n.

We know from our discussion of the elasticity of output with respect to labor in the one factor model that an elasticity of output per capita with respect to capital per capita below one implies that average capital productivity $f(k_t)/k_t$ will decline as capital rises. Feeding this back into the expression for the growth rate of output per capita written above this implies the growth rate of a Solow economy slows down as it approaches its steady state.[13]

The fact that poorer economies will grow faster in the Solow model implies that countries should converge in per capita terms. How fast should economies converge? Approximating 2.4 around the steady state value $k^\star$ gives:

$$\dot{k}_t \sim -\lambda(k_t - k^\star)$$

where

$$\begin{aligned}
\lambda &= -\frac{\partial \dot{k}_t}{\partial k_t}(k^\star) \\
&= \delta + n - sf'(k^\star) \\
&= \delta + n - \frac{(\delta+n)k^\star}{f(k^\star)} f'(k^\star) \\
&= (1 - \frac{f'(k^\star)k^\star}{f(k^\star)})(\delta + n)
\end{aligned}$$

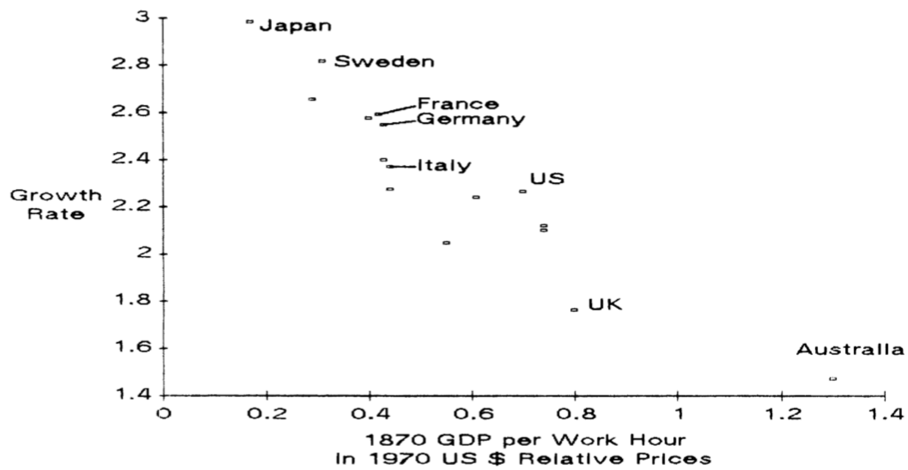
For a baseline value of the capital elasticity of 1/3 and if $\delta + n$ is close to 6% this implies a speed of convergence of 4% meaning that countries should close half the gap from their steady state in about 17 years.

Using Maddison's data published in 1982, Baumol (1986) sets out to test Solow's convergence prediction. To do so Baumol plots the 1870 level of GDP per capita of a set of sixteen countries against their growth rate over the next century. The result of this exercise is reported in Figure 2.3 below and strongly suggests that poorer countries tend to grow faster in subsequent years than their richer counterparts.

As pointed out by De Long (1988), however, this conclusion could well be completely driven by the fact that we are missing data on poor countries that did not grow fast over the period. To test this selection issue De Long was able to complete Maddison's original dataset with more poorer countries. The result of this exercise reported in Figure 2.4 casts some doubt on the

[13] As a matter of fact, Solow economies never actually their steady states... but only approach them asymptotically. We leave this to PS2.

Figure 2.3: BAUMOL'S TEST OF SOLOW'S CONVERGENCE PREDICTION



Notes: This figure reproduces Figure 2 of Baumol (1986). In this paper Baumol uses one of the first available versions of A. Maddison's famous database.

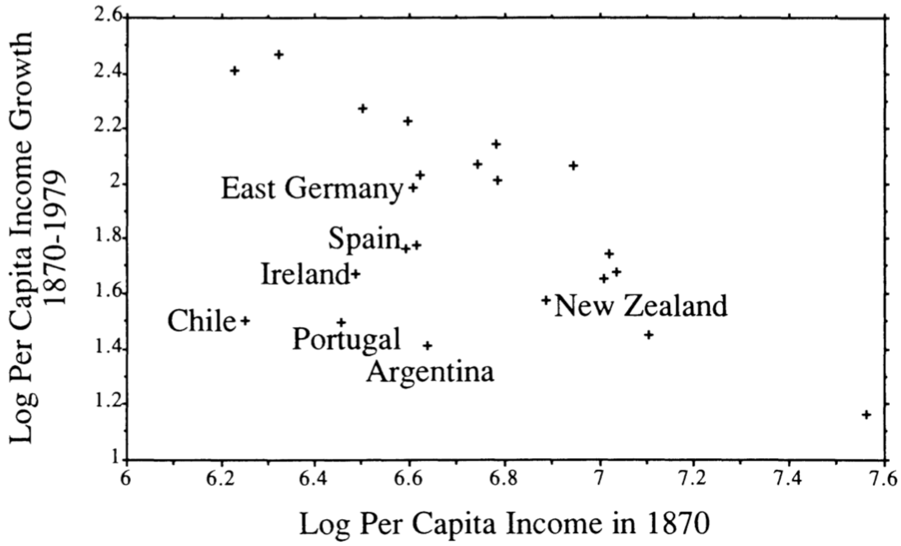
convergence claim. Whereas Baumol's original convergence coefficient was close to -1 and precisely estimated, De Long's correction cuts this negative correlation to -0.57. Another point made by De Long (1988) is that measurement error in 1870 could also be a source of bias in Baumol (1986)'s original converge result. If measurement error was greater in 1870 than in 1970, we will overstate the growth rate of countries with an unusually low measured level of output per capita in 1870, and understate the growth rate of countries with an unusually high measured level of output per capita in 1870. This phenomenon could give a rise to a negative correlation between the initial measured level of output per capita and the subsequent pattern of growth. De Long (1988) shows that assuming a standard measurement error of 0.15 log points in 1870 would be sufficient to completely kill Baumol's convergence result.

At this stage Solow's model empirical fit seems pretty poor.

The golden rule level of capital

Despite Solow's disappointing conclusion that savings do not affect long term growth, it is still the case that a higher saving rate will result in a higher level of capital and hence of production. Does it follow that countries should implement reforms aiming at such saving rate increases? While on the one hand savings increase the levels of capital and output per-worker, on the

Figure 2.4: De Long's test of Baumol's test of Solow's convergence prediction



Notes: This figure reproduces Figure 2 of De Long (1988). In this paper De Long adds Chile, Spain, Ireland, Portugal, Argentina, New Zealand and East Germany to Baumol's original set of 16 countries.

other hand, and by definition, savings decrease the amount of output which is actually consumed. Writing steady state per capita consumption $c^\star$ where:

$$c^\star = (1-s)f(k^\star)$$

we see that increases in the saving rate has offsetting effects on consumers' welfare. Indeed:

$$\frac{\partial c^\star}{\partial s} = (1-s)f'(k^\star)\frac{\partial k^\star}{\partial s} - f(k^\star)$$

where it is important to remember that $k^\star$ is a function of s. Because steady state consumption is a concave function of s (it take a bit of calculus to prove this claim) equating this equation to zero and solving for the saving rate should give us the "optimal" saving rate at which steady state consumption is maximized. Concretely the optimal saving rate of our Solow economy should solve

$$(1-s)f'(k^\star)\frac{\partial k^\star}{\partial s} = f(k^\star)$$

Substituting for $\frac{\partial k^\star}{\partial s}$ and re-arranging this expression we get:

$$f'(k^{\star\star}) = \delta + n \tag{2.6}$$

where $k^{\star\star}$ denotes the unique steady state level of capital which solves 2.6. If s is such that this equation is verified, then consumption is maximized at the steady state. In this situation increasing savings would increase output (as it always does) but decrease consumption and welfare. This is why the level of capital per worker $k^{\star\star}$ associated to condition 2.6 is usually called the "Golden rule" level of capital. When below this Golden rule level, an economy could do better in the long run by increasing savings and consuming forever more at the steady state. Symmetrically, when located above this Golden rule level an economy would do better by decreasing its saving rate, consuming its excess capital stock, stabilizing at the Golden rule level and consuming forever more at this new steady state. Economies with excess savings which have accumulated a capital stock above the Golden rule level are considered to be 'dynamically inefficient". As we will see in a future lecture there might be a host of good reasons why real life economies could end up in such an inefficient position. As a last general remark notice that in the case of an inefficient economy where $k^\star > k^{\star\star}$ and:

$$f'(k^\star) < \delta + n$$

the real rate of return on capital net of depreciation $f'(k^\star) - \delta$ will be lower than the population growth rate n. This provides an easy check for dynamic inefficiencies in practice.

2.2 Bringing growth back into the story

This simple version of Solow's 1956 model of growth makes a straightforward point: sustained growth cannot arise solely out capital accumulation. If you recall Solow's 1957 conclusion that 80% of observed US growth in the first half of the 20th century cannot be explained by capital accumulation, this should not come as a surprise.[14] Given the importance of technological change a natural next step would be incorporate knowledge into Solow's

[14] Let me stress again that the two papers make related but distinct points. While Solow (1957) shows that capital accumulation has been a relatively minor source of growth in practice, Solow (1956) shows that capital's direct contribution to growth should be zero in the long run. "Direct contribution" because as we will see shortly, if technological forces increase output per capita, capital will continue to accumulate. In this case capital accumulation is a consequence, rather than a cause of growth.

theoretical framework. One possible way do so could be to change our aggregate production function for the more general use by Solow (1957) and involving an exogenous technological term (see equation 1.10). This would give us total production Y_t as:

$$Y_t = A_t F(K_t, L_t) \tag{2.7}$$

This exactly the road taken by Solow in his 1956 theoretical paper. As we will see, however, this way of incorporating technology into the Solow model is inconsistent with some important empirical regularities that are observed in practice.

Kaldor's facts and balanced growth

In the previous lecture we have stressed that perfect competition imposes some empirical constraints on returns to scale and factor shares. In particular, under perfect competition in output and input markets, the labor share of production will pin down what a realistic elasticity of output with respect of capital should look like. Similarly constant returns to scale are important if one believes than in the long run factor shares should sum to one. But factor shares and profits are not the only facts that a good model of growth should be expected to match. In a very famous chapter Kaldor (1961) set out a list of six " stylized facts" which in his few were strong enough in the data to serve as "a starting point for the construction of theoretical models". Translated into our notations these facts were:

1. "No recorded tendency for falling rate of growth of productivity" (i.e. output Y_t and output per capita $y_t = Y_t/L_t$ both grow at a constant rate).

2. "A continued increase in the amount of capital per worker" (i.e capital per worker $k_t = K_t/L_t$ also grows at a constant rate).

3. "A steady rate of profit on capital" (i.e. steadiness of r_t the real rate of return on capital).

4. "Steady capital/output ratios over long periods" (i.e, K_t/Y_t is stable over time).

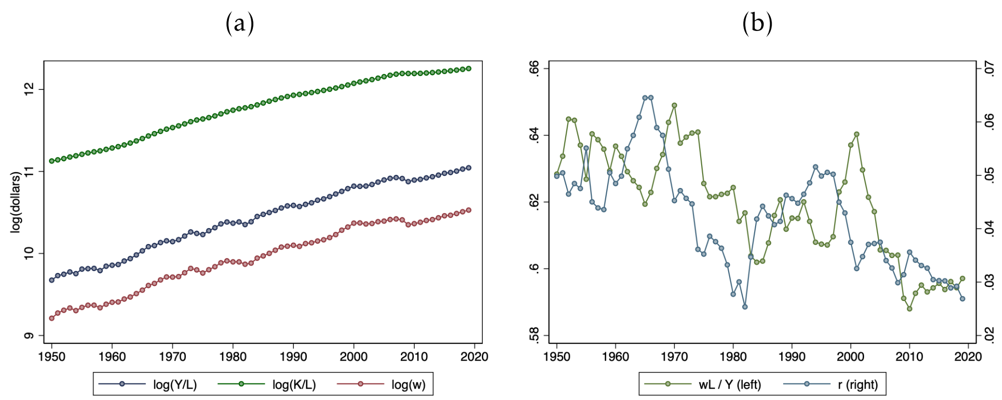
5. "Steadiness in the share of profits and the share of wages" (i.e. stable values of $r_t K_t/Y_t$ and $w_t L_t/Y_t$, where r_t is the rental rate of capital and

w are wages, of course this implies that wages are raising at the same speed as average labor productivity $y_t = Y_t/L_t$).

6. Growth rates of output per capita vary across countries and are positively correlated to investment rates.

These "facts" were laid down by Kaldor sixty years ago and have greatly influenced the way economists think about the process of economic growth. You can see in the two panels of Figure 2.5 below, that Kaldor's have to large extent been confirmed by post-1960s data (to a large extent only because as you can see in Figure 2.5(b) the labor share and real rate of return of capital both appear to be declining in the US in the long run).

Figure 2.5: Kaldor's stylized facts in the US



Notes: Panel (a) plots log output per worker, capital per worker and wages against time in the US (wages are computes as total labor compensation divided by population). Panel (b) plots the labor share of production (left scale) and the real rate of return to capital net of depreciation (right scale) against time in the US. Source: PWT.

What are the consequences of Kaldor's facts for growth theory? First the fact that output per worker grows at constant rate is incompatible with Solow's simple model without technological progress according to which the growth rate of output per capita should converge to zero. Second, constancy of the capital/output ratio implies that capital and output grow at the same rate. Third, if capital and output grow at the same rate, capital and output per capita also grow at the same rate. Fourth, as we already noted, the fact the labor share is constant implies that wages grow at the same rate as average labor productivity, i.e. output per capita. All in all, Kaldor's facts imply that a lot of variables, wages, output per capita, capital per capita... grow not only at constant rate, but at the same rate (empirically this means that the scatter plots in Figure 2.5(a) should not only be well approximated by straight lines but that these lines should be parallel to one another). In the language of growth theory, the fact capital, output and wages are predicted to grow at the same rate can be translated by saying that our economies should follow a “balanced growth path”.

Another way to come at the same conclusion is to look at Solow’s capital accumulation equation. In per capita terms this equation tells us that:

$$\dot{k}_t = sy_t - \delta k_t$$

or in other words that

$$\frac{\dot{k}_t}{k_t} = s\frac{y_t}{k_t} - \delta$$

This relation implies that for k_t to grow at a constant rate we need the output/capital ratio $y_t/k_t = Y_t/K_t$ to remain constant. A direct consequence is that if k_t grows at a constant rate, output per capita y_t needs to grow at the very same and constant rate. All in all balanced growth appears to be compatible with a Solow type capital accumulation equation if capital and output grow at the same rate.

Unfortunately this type of balanced growth turns out to impossible if technological change is embodied in the model as a term multiplying the original production function as in equation 2.7. In other words “neutral” technological change which affects the marginal productivities of labor and capital symmetrically turns out to be incompatible with balanced growth. In some ways, this should completely come as a surprise: the fact that the real rate of return on capital remains constant while the real wage is constantly increasing hints at the fact that technological improvements should not affect capital and labor symmetrically if Kaldor’s “stylized facts” are indeed verified.

Uzawa’s theorem

In a very short paper Uzawa (1961) showed that for balanced growth to be possible technological change had to take a “labor augmenting” or “Harrod enutral” form. Or in other words, that technology should enter the production function as a multiplicative increase in the effectiveness of the labor input:

$$Y_t = F(K_t, A_tL_t) \tag{2.8}$$

Why? Uzawa's reasoning can be summed in the following way. Assume that technological change affects the aggregate production function not in labor-augmenting way like in 2.7, but in a neutral multiplicative way (also called "Hicks neutral"), like in 2.7. In this case the output capital ratio is given by:

$$\frac{Y_t}{K_t} = A_t F(1, \frac{L_t}{K_t})$$

Because the output/capital ratio is assumed to be constant on a balanced growth path, this equation defines a locus of admissible points in the $(A_t, L_t/K_t)$ space. If the economy is on balanced growth paths, movements in A_t should move L_t/K_t along this admissible locus of points. But under the assumptions that the interest rate is constant and that capital is paid its marginal product then it should also be true that:

$$r = A_t F_K(K_t, L_t)$$

As F has constant returns to scale we know that F_K will be homogeneous of degree zero, or in other words, we know that F_K only depends on the ratio of its two arguments.[15] This allows us to rewrite the last equation as:

$$r = A_t F_K(1, \frac{L_t}{K_t})$$

Notice that this equation also gives a relation in the $(A_t, L_t/K_t)$ space which should be verified in along the balanced growth path. The difference is that now F_K appears rather than F/K. But there not reasons why the marginal and average product of capital should have similar isoquants, making balanced growth in this context impossible. Another way to see this is to divide the second condition by the first one. Doing this gives us an expression for the capital share of output — which should be constant — as a function of the capital labor ratio:

$$\frac{rK_t}{Y_t} = \frac{F_K(1, \frac{L_t}{K_t})}{F(1, \frac{L_t}{K_t})}$$

Again, in general there is now reason why this equation should continue to be verified as the capital labor rises at a constant rate.

Now switch to a labor augmenting production function like in 2.8. In this case our two relations become:

[15] See footnote 10.

$$\frac{Y_t}{K_t} = F(1, \frac{A_t L_t}{K_t})$$

and

$$r = F_K(1, \frac{A_t L_t}{K_t})$$

Under labor augmenting technological change, both the output/capital ratio and the rate of return to capital will be pinned down by the ratio of capital K_t to units of "effective labor" $A_t L_t$ (labor augmented by its technological efficiency). In this case a *constant* value of the effective capital labor ratio defined as

$$\frac{K_t}{A_t L_t}$$

is compatible both with a *constant* value of the output/capital ratio and *constant* value of the rate of return to capital. Of course, for capital per effective worker k_t to be constant, the capital labor ratio k_t needs to grow at the same rate as technology. But as soon as this is the case technological growth appears to be compatible both with a steady real rate of return and a fixed capital/output ratio, or in other words, technological growth can give rise to a balanced growth path.

Another way to come at the same conclusion is to reason directly from the labor augmenting technological change production function 2.8. Because F is assumed to have constant returns to scale, the fact that Y_t and K_t grow at the same rate suggests that whatever comes under the second argument of the function should also grow at this same rate. Without labor augmenting technological change, the second argument would be labor. But labor grows at the constant rate n, which has no reason to be equal to the growth rate of output and capital. For this reason, balanced growth where capital and output grow hand in hand is only possible if technology somewhat compensates for the fact that labor does not grow at the desired rate. If we call g_Y and g_K the (identical) rates of growth of output and capital on the balanced growth path, the rate of technological growth g compatible with balanced growth would have to verify:

$$g_Y = g_K = g + n$$

Another way to sum up Uzawa's claim is to say that the only possible way in which exogenous technological change may give rise to a balanced growth path with observed growth rate g_Y is if the underlying technical change is labor augmenting at rate:

$$g = g_Y - n$$

In other words, for a balanced growth path to exist under constant returns to scale, labor augmenting technological change must "bridge the gap" between the observed growth rates of output and capital on the one hand, and the growth rate of labor on the other hand.

Solving Solow's model with technological growth

We know from Solow's simple model that durable growth cannot arise without technological change. What is more, Uzawa's theorem showed us that if we observe capital and output growing at the same rate with a constant rate of return to capital, the only way that this can be caused by technological improvements is if technology takes a labor augmenting form. More than this, Uzawa's theorem show us that if the balanced growth rate of output is g_Y and if the exogenous growth rate of labor is n, then the rate of labor augmenting technological change generating this pattern of growth g_Y must be $g = g_Y - n$. In this sense Uzawa's theorem goes from the observed patterns of growth to its hypothetical underlying causes. Let us now walk back the opposite way and show that a Solow model augmented with exogenous labor augmenting technological change g will give rise to a balanced growth path economy.

Under labor augmenting technological change Solow's original production function will be replaced by 2.8. Compared to the original model, however, the capital accumulation equation remains unchanged. Very much like in Solow's original model we will continue to assume capital depreciation at rate δ, a fixed saving rate s and constant population growth at rate n. But unlike in Solow's original setting we will now assume that labor augmenting technological change grows at the constant exogenous rate g.

As we already pointed out in our discussion of Uzawa's result, under ongoing labor augmenting technological change, the stable quantity of interest will not be the capital/labor ratio K_t/L_t but the ratio of capital to effective units of labor so that we will redefine the k_t as

$$k_t = K_t/(A_t L_t)$$

Following a series of steps very similar to those that allowed us to express the time derivative of the capital/labor ratio as a function of its level in

Solow's original model, we can show that under exogenous technological change the effective capital labor ratio follows:

$$\dot{k}_t = sf(k_t) - (\delta + n + g)k_t \tag{2.9}$$

Except for the fact that g now appears on the right hand side equation 2.9 is in every respect similar to 2.4. As a consequence the dynamics of the effective capital/labor ratio in the Solow model augmented with labor augmenting technological change will be very similar to those followed by the simple capital/labor ratio in Solow's original model. Inada's conditions and concavity will imply that a stable steady state level of the effective capital labor ratio $k^\star$ exists and is solution to:

$$sf(k^\star) = (\delta + n + g)k^\star$$

Once the economy reaches this stable level $k^\star$, technology A_t and the capital/labor ratio will grow hand in hand at the same rate g. How will this affect output per capita? Notice that on the balanced growth path output per capita will be determined by

$$\frac{Y_t}{L_t} = F(A_t\frac{K_t}{A_tL_t}, A_t)F(A_tk^\star, A_t) = A_tF(k^\star, 1) = A_tf(k^\star)$$

As a consequence, output per capita will also grow at the constant rate g. And so will competitive wages which will be determined in equilibrium as:

$$w_t = A_tF_L(1, 1/k^\star)$$

Finally, as expected from Uzawa's theorem, the steady state level of the effective capital/labor ratio $k^\star$ will pin down a constant rate of return on capital as well as a constant capital/output ratio. All in all, Solow's model enriched with labor augmenting technological change predicts that starting from some initial level of capital K_0, labor L_0 and technology A_0, a country should converge to a balanced growth path verifying Kaldor's main stylized facts. Importantly, once countries have reached this balanced growth path, their growth rates and levels of outcome per capita will be identical.

"Taking Robert Solow seriously..."

How useful is Solow's model in practice? We have just seen that in order to replicate observed patterns of balanced growth over long stretches of time, we need to add to Solow's original model some form of exogenous technological change. This a simple albeit arbitrary fix for Solow's first counterfactual prediction that we should not observe any growth in the long run. What about Solow's model's other empirical predicaments? Under exogenous technological change Solow's model predicts that all countries should after some time end up on the balanced growth path. Once on this balanced growth path, all countries should grow at the same rate g and exhibit the same capital per effective worker ratio $k^\star$. In other words Solow's augmented model also predicts that countries should converge in the long run, both in levels and in growth rates. Very much like in Solow's model without technological change, poorer countries farther away from the balanced growth path will tend to grow faster than their richer counterparts and so to "catch up". All in all, adding technology to Solow's original model is a good and easy fix for the "no growth" prediction, but in itself does not resolve other important failures like the apparent absence of convergence. This failure is one of the reasons which lead to the take off of the endogenous growth literature in the 1980s. Indeed, if the growth rates of different countries is endogenously determined in equilibrium rather than an exogenously given constant, we could expect small fundamental differences in economic behavior across countries (initial capital stock, saving rates, etc) to translate into permanent differences in growth rates. In this case, nothing would warranty convergence over time and across space, whether in levels or in growth rates. When carried along for many years, even a seemingly small difference in long run growth rates can give rise to significant income differences.

Before completely throwing away Solow's simple and elegant model, however, one should consider one last possibility. When taken rigorously, Solow's model predicts convergence *conditional* on countries fundamental parameters like the saving rate and population growth rate. Until now in these lecture notes and, more importantly, until the end of the 1980s in the literature, the possibility that differences in the saving and population growth rates may account for cross country income differences and the lack of absolute convergence was implicitly dismissed, without any formal test (see for instance Lucas (1988) on p.14). The reason why probably had to do with a lack of proper data to test the hypothesis combined with the simple observation that, under a capital share of 1/3, 10-fold differences in steady state output per capita would have to translate into 1000-fold differences in steady state capital labor ratios — an order of magnitude simply to great to be seriously considered. In 1992, however, going against this trend and making use of Summers and Heston (1988)'s newly available data set, Mankiw, Romer and Weil set out to formally "test" Solow's stark predictions. To do so Mankiw et al. (1992) start off with a Solow type aggregate production function compatible with balanced growth.

$$Y_t = K_t^{\alpha}[A_t L_t]^{1-\alpha}$$

The growth rate of technology g is assumed to be exogenous and constant and so is the saving rate s. With capital depreciation δ and population's growth rate n, the augmented Solow model predicts:

$$k^{\star} = (\frac{s}{n+\delta+g})^{\frac{1}{1-\alpha}}$$

As we already know this equation leads to a steady state level of capital per effective worker:

$$k^{\star} = (\frac{s}{n+\delta+g})^{\frac{1}{1-\alpha}}$$

so that on the balanced growth path output per capita will be given by:

$$\frac{Y_t}{L_t} = A_t(\frac{s}{n+\delta+g})^{\frac{\alpha}{1-\alpha}}$$

and taking the log of this expression yields

$$\log(\frac{Y_t}{L_t}) = \log(A_0) + gt + \frac{\alpha}{1-\alpha}\log(s) - \frac{\alpha}{1-\alpha}\log(n+\delta+g) \tag{2.10}$$

In order to bring this equation to the data Mankiw et al. (1992) assume that the log of each country's initial technology level can be decomposed as

$$\log(A_0) = a + \epsilon$$

where ϵ is orthogonal to population growth and savings. If this is indeed the case, then unbiased estimates of equation 2.10's parameters can be recovered by OLS. What's more, the estimated coefficients on population growth and savings should be of opposite in sign and equal in magnitude. Finally, taking a baseline capital share close to 1/3, both estimates value of $\alpha/(1-\alpha)$ should not be far from 0.5.

Measuring the saving rate s as each country's share of investment in GDP and assume δ and g are constant across countries, Mankiw et al. (1992) estimate this equation in 1985. While the coefficients they find for the log saving rate and log population growth rate are significant, of similar magnitude and opposite in sign their actual values implies a capital share α close to 60%. This discrepancy suggests that a factor correlated with countries' saving and demographic behavior has been omitted from the regression. To fix this Mankiw et al. (1992) do not look for an credible instrument but turn to more a general model in order to get an explicit hand on the omitted factor. This factor is "human capital".

To quote Becker (1962) the idea of human capital encompasses all factors which increase real income through the "embedding of resources in people". Even if A_t increases the productivity every single unit of labor, it is not "human capital" in the sense that A_t need not be accumulated in order to produce its effects. Inherent in the idea of human capital is that "embedding" abstract knowledge, techniques, know-how, production processes "into" workers takes, if not tears and blood, at least some amount sweat and some amount time. In other words, making knowledge useful for production purposes is itself a costly process which can be reversed if one stopped investing into it. Concretely human capital relates to schooling, all forms of training, should it be on-the-job or off of it. What's more its omission from 2.10 would likely be problematic. Indeed, one can expect that physical capital and technical knowledge being complements in production, the accumulation of one should correlated to the accumulation of the other. In order to test the importance of human capital in explaining the partial failure of Solow's baseline model to explain cross country income differences at a point in time Mankiw et al. (1992) propose a simple extension of the model which explicitly includes the accumulated stock of human capital as a third factor of production. To do so they use a new production function

$$Y_t = K_t^{\alpha} H_t^{\beta} [A_t L_t]^{1-\alpha-\beta}$$

where H_t denotes human capital and β the human capital elasticity of output. The labor elasticity $1-\alpha-\beta$ is set in order to keep overall returns to scale constant. If we denote by y_t, k_t and h_t the corresponding quantities of output, physical and human capital per units of efficient labor, then this production function simply writes

$$y_t = k_t^{\alpha} h_t^{\beta}$$

Mankiw et al. (1992) assume that the accumulation of physical and human capital is governed by two potentially different saving rates s_k and s_h, but that both capital stocks depreciate at the same speed δ. Under this assumption one can write in per effective units of labor terms:

$$\dot{k}_t = s_k y_t - (n+\delta+g)k_t$$

and

$$\dot{h}_t = s_h y_t - (n+\delta+g)h_t$$

These two equations define steady values $k^\star$ and $h^\star$ given by:

$$k^\star = [\frac{s_h^\beta s_k^{1-\beta}}{n+\delta+g}]^{\frac{1}{1-\alpha-\beta}}$$

and

$$h^\star = [\frac{s_k^\alpha s_h^{1-\alpha}}{n+\delta+g}]^{\frac{1}{1-\alpha-\beta}}$$

So that human capital accumulation is indeed correlated to physical capital accumulation. This steady state is globally stable in the (k,h) space, meaning that countries will converge to it regardless of their initial levels of physical and human capital (k_0, h_0).[16] Plugging these steady state levels of physical and human capital into the effective per capita production function yields an expression for its steady state level $y^\star$

$$y^\star = s_k^{\frac{\alpha}{1-\alpha-\beta}} s_h^{\frac{\beta}{1-\alpha-\beta}} [\frac{1}{n+\delta+g}]^{\frac{\alpha+\beta}{1-\alpha-\beta}}$$

As in Solow's baseline model this equation can be used to express log output per capita as a function of time and of the parameters of the model:

$$\log(\frac{Y_t}{L_t}) = \log(A_0)+gt+\frac{\alpha}{1-\alpha-\beta}\log(s_k)+\frac{\beta}{1-\alpha+\beta}\log(s_h)-\frac{\alpha+\beta}{1-\alpha-\beta}\log(n+\delta+g) \tag{2.11}$$

Mankiw et al. (1992) use skill wage differentials to measure β which they find to be in the range of 1/3. With this estimate of β, and keeping a capital share close to 1/3, the OLS coefficients on log human and physical capital investment should both be close to 1, while the coefficient on log population growth should be close to -2. Going back to the data and using schooling rates to proxy the investment rate in human capital s_h, Mankiw et

[16] The stability curves of physical and human capital are given by

$$h = [\frac{n+\delta+g}{s_k}]^{\frac{1}{\beta}} k^{\frac{1-\alpha}{\beta}}$$

and

$$k = [\frac{n+\delta+g}{s_h}]^{\frac{1}{\alpha}} h^{\frac{1-\beta}{\alpha}}$$

The associated phase diagram reveals global stability of the unique non trivial steady state and has been seen in class.

al. (1992) find that their augmented empirical model fits the data in a way that appears to be coherent with the underlying theory. What's more this simple empirical exercise is able to explain close to 80% of observed cross country income differences in 1985.

After showing that an augmented Solow model can fit cross country income differences at a point in time, Mankiw et al. (1992) re-open the Baumol/De Long convergence debate which was thought of by Lucas (1988) as definitely settled in favor of De Long's no convergence result. In the third part of the paper, Mankiw, Romer and Weil re-estimate Baumol's regression on De Long's enriched sample of countries but this time control for differences in investment, schooling and population growth, as their theory predicts one should do. These controls turn out to have a huge impact on the final result. Mankiw et al. (1992) find that countries do do converge conditional on their "structural differences". Of course, the claim is not entirely causal, and knowing how "structural" are these differences is a crucial line of research for growth theory. Investment, schooling and population growth rates are all likely to be determined in equilibrium. Whether or not this relegates Mankiw et al. (1992)'s conditional convergence result to the long list of foregone OLS fantasies is a question we cannot answer without a more comprehensive theory of growth.

Bibliography

Baumol, William J, "Productivity growth, convergence, and welfare: what the long-run data show," *The american economic review*, 1986, pp. 1072–1085.

Becker, Gary S, "Investment in human capital: A theoretical analysis," *Journal of political economy*, 1962, *70* (5, Part 2), 9–49.

Bolt, Jutta and Jan Luiten van Zanden, "Maddison style estimates of the evolution of the world economy. A new 2020 update," *Maddison-Project Working Paper WP-15, University of Groningen, Groningen, The Netherlands*, 2020.

Kaldor, Nicholas, "Capital accumulation and economic growth," in "The theory of capital," Springer, 1961, pp. 177–222.

Long, J Bradford De, "Productivity growth, convergence, and welfare: comment," *The American Economic Review*, 1988, *78* (5), 1138–1154.

Lucas, Robert E, "On the mechanics of economic development," *Journal of monetary economics*, 1988, *22* (1), 3–42.

Mankiw, N Gregory, David Romer, and David N Weil, "A contribution to the empirics of economic growth," *The quarterly journal of economics*, 1992, *107* (2), 407–437.

Solow, Robert M, "A contribution to the theory of economic growth," *The quarterly journal of economics*, 1956, *70* (1), 65–94.

— , "Technical change and the aggregate production function," *The review of Economics and Statistics*, 1957, pp. 312–320.

Summers, Robert and Alan Heston, "A new set of international comparisons of real product and price levels estimates for 130 countries, 1950–1985," *Review of income and wealth*, 1988, *34* (1), 1–25.

Uzawa, Hirofumi, “Neutral inventions and the stability of growth equilibrium,” *The Review of Economic Studies*, 1961, *28* (2), 117–124.