Chapter 13

Growth

Timo Boppart and Peter J. Klenow

13.1 Motivation

In this chapter we will study long-run economic growth. Figure 13.1, which replicates Figure 2.1 introduced in Chapter 2, illustrates the striking phenomenon of sustained growth in U.S. GDP per capita over the past two centuries at a steady rate of about 1.5 percent a year. What is behind this pattern of steady growth?

Standards of living have improved not only in advanced economies like the U.S., but also in developing economies in the last half-century. Countries did not, however, all grow at the same rate or exhibit the same income levels. The chapter also deals with the determinants of differences in GDP per capita levels around the globe.

In this chapter we first take a look at some stylized facts about growth and development in the U.S. and elsewhere. We then lay out a version of the neoclassical growth model with investment-specific technical change and use this framework to analyze the data. Finally, we move on to endogenous growth theory, discussing canonical models and how they can be brought to the data.

Motivating the study of economic growth is straightforward. The process of economic growth transformed the U.S. far beyond what is visible in the economic indicators. Understanding the forces behind this phenomenon is of first-order importance to human welfare. Moreover, understanding why some countries are poor and some countries are rich and what policy could do to influence this is a many-trillion dollar question for the entire social sciences. A nice quote that makes this point is from Lucas (1988): "Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, what, exactly? If not, what is it about the 'nature of India' that makes it so? The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else."

13.2 Empirical patterns

As mentioned, Figure 13.1 depicts the steady growth of U.S. GDP per capita over the last two centuries. Appendix Figure 13.A.1 makes it clear that one should not take such growth for granted, as it globally began slowly around a thousand years ago before accelerating over the last 200 years (after the so called Industrial Revolution).

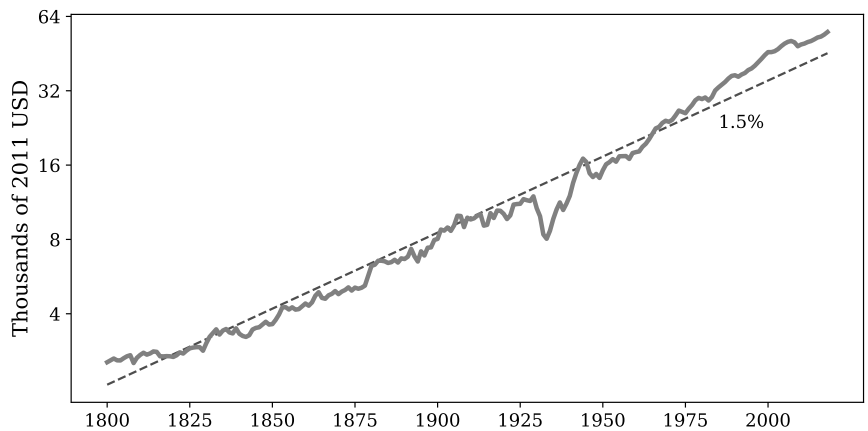

Figure 13.1: U.S. Real GDP per Capita

Source: Maddison Project Database (2020). The vertical axis is in ratio scale.

We can get a sense of whether growth was faster or slower in other countries compared to the U.S. using data from the Penn World Table. Figure 13.2 plots GDP per worker for regions of the world *relative to the U.S.* from 1960 through 2019. The U.S. is normalized to 1 in each year. Each region has a balanced panel of countries, with the countries weighted by their employment. One can see that Europe converged toward the U.S. from 1960 to 1990 or so, but since has grown parallel to the U.S. Latin America as a whole mostly moved sideways over the sample. East Asia, led by China, Japan, and South Korea, rose from 1/16th of U.S. GDP per worker in 1960 to almost 1/4th in 2019. Sub-Saharan Africa grew at a similar rate to the U.S. from 1960 to 1980, then fell behind from 1980 to 2000 (falling from 1/8th to 1/16th) before stabilizing. Finally, South Asia tracked at about 1/16th of the U.S. until the mid-1970s, lost ground through the mid-1980s, then soared to over 1/8th led by India's surging growth. The broad picture that emerges is that most regions generally grew along with the U.S., though at differing rates.

Figure 13.3 provides a more detailed country-by-country look. The vertical axis is GDP per capita in 2019 and the horizontal axis is GDP per capita in 1960, with the U.S. normalized to 1 in both years. Countries along the 45 degree line maintained the same growth rate as the U.S. This was the "typical" pattern in that a regression of log 2019 GDP per capita on the 1960 version yields a coefficient of about 1. That said, many countries grew considerably faster (e.g., South Korea or Singapore) or slower (e.g., the Congo or Burundi). Rich countries generally hewed close to the U.S. growth rate, whereas developing countries exhibited more dispersion.

Figure 13.3 displays neither strong divergence nor strong convergence of per capita incomes over time. The early empirical cross-country growth literature emphasized a distinction between unconditional and conditional convergence. As we already saw in Figure 3.6 earlier, this figure suggests no unconditional convergence.[1] The absence of divergence is

[1] The literature further distinguished "beta" and "sigma" convergence. Beta convergence refers to a

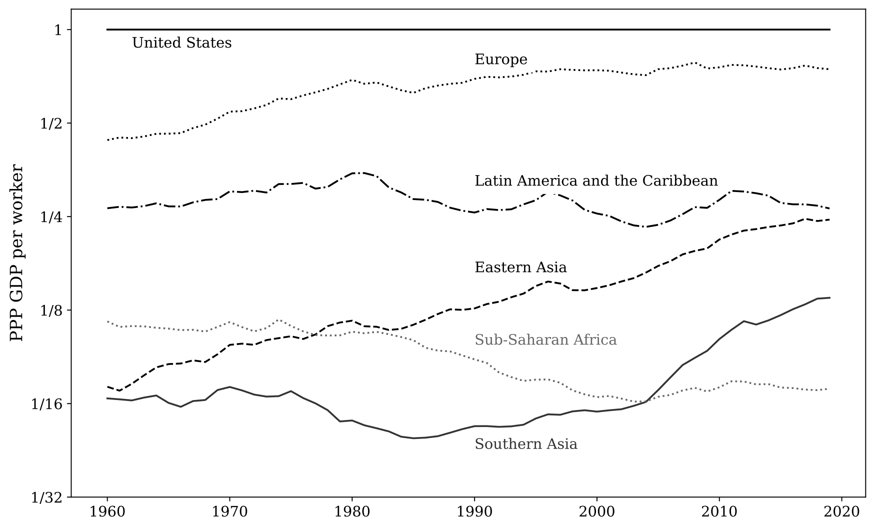

Figure 13.2: PPP GDP per Worker (U.S. = 1)

Source: Penn World Table 10.0. PPP GDP per worker is calculated as the ratio of the "rgdpo" and "emp" variables relative to the U.S. in all years.

consistent with a common trend, say due to technology that gradually diffuses across countries. See Nath, Ramey, and Klenow (2023) for evidence that country income differences are persistent, whereas country growth differences are largely transitory.

Rather than looking at each country, it is useful to plot the distribution of income across the world population. Appendix Figure 13.A.2 does so under the simplifying assumption that income is the same within a country. The U.S. is normalized to 1 in each year. One can see the distribution shifting to the right because of the rapid growth in China and India. The distribution is left-skewed but decreasingly so over time.[2]

The plots with individual countries highlight the tremendous dispersion in living standards across countries. In Appendix Figure 13.A.4 we focus on the PPP GDP per capita in the 20 largest countries by population in 2019. With the U.S. normalized to 1, one can see that income differs by a factor of 64 between the Congo and the U.S. Even with the rapid growth in China and India in recent decades, they have only attained about 1/4th and 1/8th the level of U.S. per capita income in 2019.

These facts on growth rates and income differences call out for explanation. In a deeper sense they may stem from policies such as taxation; government investments in education, health, and infrastructure; regulation and government ownership; and openness to trade and foreign direct investment. But short of that, we can do growth and development accounting to determine the more proximate sources of growth and development differences.

To be concrete, imagine a simple Cobb-Douglas aggregate production function that is

tendency for countries with initially higher per capita incomes to grow more slowly. Sigma convergence involves a narrowing of the dispersion of per capita incomes over time. One could have beta convergence without sigma convergence because of ongoing shocks. But our Figure 13.3 and the literature find a lack of both beta and sigma convergence.

[2]For contrast, Appendix Figure 13.A.3 plots the histogram in the case when each country is weighted equally. The distribution still shifts to the right, but is much less skewed in each year.

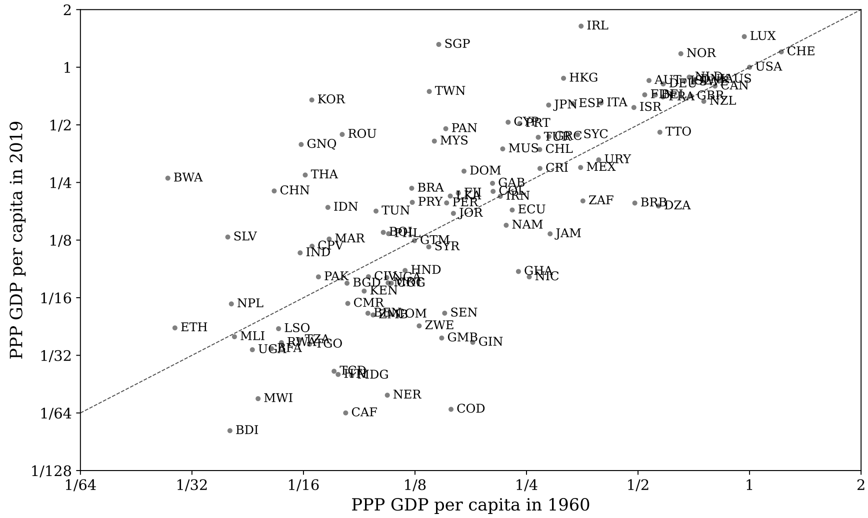

Figure 13.3: PPP GDP per Capita in 1960 and 2019 (U.S. = 1)

Note: The data comes from the Penn World Table 10.0. PPP GDP per capita is calculated as the ratio of the “rgdpo” and “pop” variables relative to the U.S. in both years. The dotted line is the 45 degree line. The simple correlation across the two years is 0.75 and the elasticity is 0.97 (0.091).

common to all countries i and years t:

$$Y_{it} = K_{it}^{\alpha} \left(A_{it} H_{it}\right)^{1-\alpha}, \tag{13.1}$$

where Y is real output, K is real physical capital, H is total human capital (efficiency units summed across all workers), and A is residual labor-augmenting TFP. Here α is the production elasticity of output with respect to physical capital.[3]

Dividing equation (13.1) by L and rearranging implies

$$\frac{Y_{it}}{L_{it}} = \left(\frac{K_{it}}{Y_{it}}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{H_{it}}{L_{it}}\right) A_{it}. \tag{13.2}$$

Here L is total employment so that H/L is (average) human capital per worker. As described in Section 4.3.3 above, the balanced growth path of a neoclassical growth model features a stable K/Y that does not depend on the levels of A or H/L. Meanwhile, models such as Mankiw, Romer, and Weil (1992) yield a stationary level of human capital per worker that does not depend on A or K/Y. Hence, the decomposition in (13.2) is compatible with steady states under standard forms of endogenous investments in physical and human capital. In the data, there is substantial persistence in capital-output ratios over time within countries (see Appendix Figure 13.A.5). This is consistent with countries being close to their respective balanced growth paths in terms of capital accumulation. Countries do not seem to transition

[3] Assuming a Cobb-Douglas functional form is convenient but not strictly necessary. As in Solow (1957), one can calculate residual TFP for any neoclassical (i.e., constant returns) production function. Moreover, even with a Cobb-Douglas specification the production elasticity could vary over time or across countries.

to a common capital-output ratio. For this reason we will do the accounting below based on (13.2) rather than (13.1).[4]

Taking logs and differentiating with respect to time we obtain a form convenient for growth accounting:

$$\Delta \log\left(\frac{Y_{i,t}}{L_{i,t}}\right) = \frac{\alpha}{1-\alpha}\Delta \log\left(\frac{K_{i,t}}{Y_{i,t}}\right) + \Delta \log\left(\frac{H_{i,t}}{L_{i,t}}\right) + \Delta \log\left(A_{i,t}\right). \tag{13.3}$$

One can then average the right side components over time, back out TFP growth $\Delta \log(A)$, and compare it to the left hand side $\Delta \log(Y/L)$. Table 13.1 does this for the U.S. in recent decades. The table applies to total private businesses for the years 1948–2020 and α is approximated by physical capital's cost share.[5] Increases in the capital-output ratio contributed modestly—less than one-tenth of average growth over the entire sample. Human capital per worker contributed twice as much, primarily through rising years of education. Still, this leaves about three-fourths of growth coming from residual TFP. This echoes what Solow (1957) famously found in earlier U.S. data.[6]

Table 13.1 further shows that TFP growth accounts for most of the medium-run shifts in the rate of growth in output per hour over time in the U.S. When growth is high (such as from 1948–1973 or 1995–2007) or low (1973–1995 or 2007–2020), it is primarily due to the pace of TFP growth. Researchers such as Jorgenson, Ho, and Stiroh (2008) have attributed TFP growth from 1995–2007 to the contribution of ICT (Information and Communication Technology), both directly in the ICT-producing sector and downstream in ICT-using sectors such as retail trade.[7]

By taking logs of equation (13.2) we can decompose differences in the level of development across countries at a point in time. In doing so the Penn World Table follows Klenow and Rodriguez-Clare (1997), Hall and Jones (1999), and Bils and Klenow (2000) in using an efficiency units formulation (all levels of human capital are perfectly substitutable) for simplicity. Their method, unlike that of Mankiw et al. (1992), explicitly ties human capital to evidence on the Mincerian wage return to years of schooling.

Table 13.2 looks across 117 countries in 2019. This exercise is dubbed income or development accounting as it backs out TFPs as in Solow's growth accounting method but is applied to levels as opposed to growth rates. The columns pertain to log GDP per worker, the log capital-output ratio multiplied by (1−labor's share)/(labor's share), the log of human capital per worker, and residual labor-augmenting TFP.[8] The second row reports the elasticity of each log variable with respect to GDP per worker. These elasticities add up

[4]Doing accounting based on (13.2) attributes a larger role to differences in TFP and human capital relative to physical capital. Equation (13.2) takes into account that higher levels of TFP and human capital will induce more physical capital accumulation in steady state.

[5]Using a cost share as opposed to an income share prevents any price-cost markups that affect revenue relative to costs from biasing our estimate of the elasticity.

[6]Because of its mysterious nature, Abramovitz (1956) dubbed TFP a "measure of our ignorance." Below we try to unpack some of what we have learned about determinants of TFP.

[7]This resolved the Solow Paradox that "you can see the computer age everywhere but in the productivity statistics." See Solow (1987).

[8]Under competitive markets we can approximate the production elasticity of output with respect to labor using payments to labor relative to GDP. We average this across countries to arrive at a single estimate.

Table 13.1: Growth Accounting for the U.S.

Period	Y/L	Contributions from K/Y	H/L	A
1948–2020	2.37	0.21	0.38	1.79
1948–1973	3.28	-0.18	0.27	3.19
1973–1995	1.54	0.46	0.36	0.72
1995–2007	2.80	0.32	0.40	2.08
2007–2020	1.64	0.43	0.59	0.63

Note: The data comes from the U.S. Bureau of Labor Statistics. Y/L denotes real output per hour, K/Y the real capital-output ratio, H/L human capital per worker (which grows predominantly from rising years of schooling), and A is residual labor-augmenting TFP inclusive of contributions from R&D and intellectual property. The contribution of physical capital is scaled by $\alpha/(1-\alpha)$, where α is the average cost share for physical capital over the sample, equal to 0.34. See equation (13.2).

to 1 by construction. The coefficients are the same as the contribution of each term in a variance decomposition in which we split the covariance terms between all three components equally. By this metric, capital intensity generates only 14% of differences in income across countries. Schooling is more important, being responsible for 22% of differences. Like the U.S. over time, however, the largest contribution is from residual TFP at 64%.

The last row of Table 13.2 gives the 90/10 ratios of each variable, where 90 and 10 refer to the 90th and 10th percentile countries in terms of 2019 PPP GDP per worker. An advantage of looking at 90/10 ratios is that they are less sensitive to outliers. Remarkably, if we put the ratios in log space, they imply almost the same contributions. For example, residual TFP contributes around 64% ($\log(4.92)/\log(12) \approx 0.64$).

Studies incorporating differences in the quality of schooling (Schoellman, 2012) and human capital accumulated on the job (Lagakos, Moll, Porzio, Qian, and Schoellman, 2018) arrive at larger contributions from human capital—on the order of 50%—thereby winnowing the role of residual TFP down to about 40%. Appendix Figure 13.A.7 provides the elasticities in Table 13.2 by year. They are broadly similar in 1960 and 2019, though the importance of human capital has diminished whereas the role of capital intensity fell from 1990 to 2010 before rising.

13.3 Neoclassical growth with investment-specific technical change

Chapter 2 discussed long-run stylized facts observed in advanced economies (sometimes referred to as the Kaldor facts) and how they discipline the typically imposed structure in growth models. The Solow model in Chapter 3, for example, settles down to a stable capital-output ratio. The neoclassical growth model, as developed by David Cass and Tjalling Koopmans in the 1960s, microfounded the consumption-saving decision and estab-

Table 13.2: Development Accounting in 2019

Statistic	Y/L	Contributions from K/Y	H/L	A
Variance of log	1.00	0.14	0.08	0.57
Elasticity wrt Y/L		0.14	0.22	0.64
90/10 ratio	12.00	1.40	1.74	4.92

Note: The data comes from the Penn World Table 10.0. The sample is 117 countries in 2019. Output per worker is constructed using the "rgdpo" and "emp" variables. The capital to output ratio is constructed using the "cn", "rgdpo" and "labsh" variables. The human capital index corresponds to the "hc" variable. We use the population-weighted average labor share of 0.53 across countries in 2019, which implies $\alpha = 0.47$.

lished conditions for a steady state optimal saving rate.[9] As the aggregate saving rate is not equal to a universal constant but rather determined by the behavior of many agents in an economy, micro-founding the corresponding intertemporal decision is important in the light of a classic Lucas critique (see Chapter 1). Furthermore, specifying intertemporal preferences allows us to use the theory to make normative statements about welfare.

In the following we specify and solve a specific version of the neoclassical growth model in continuous time that allows for investment-specific technical change (as in Greenwood, Hercowitz, and Krusell, 1997). This framework is then used to address the time series and cross-country data. We choose a version of the model with investment-specific technical change because it speaks to a striking pattern in the data on relative prices between the two distinct final uses: consumption and investment.[10] In the U.S. the relative price of investment fell systematically since 1970 at an average annual rate of about 1.3% (see Figure 13.4). Investment can be further split up into equipment and structures. The decrease in the relative price of investment is driven by equipment which saw a decrease of its relative price to consumption at almost 2.5% a year. In contrast, the relative price of structures (the left out category in Figure 13.4) increased over time. Figure 13.5 shows that the relative prices also systematically differ across countries. In the poorest countries the price of investment relative to consumption is roughly twice as high as in the U.S. This suggests that the poorest countries are particularly inefficient in turning primary production factors into investment goods (Hsieh and Klenow, 2007). This has consequences for development accounting as we expect richer countries with the same nominal saving rate to arrive on average at a higher real capital-output ratio. The model we lay out in this section allows us to speak to these facts.

Preferences As in Chapter 9, we denote a variable X that changes over time as $X(t)$ and its time derivative $\dot{X}(t) \equiv dX(t)/dt$. We consider a representative household, and abstract

[9] The intertemporal problem of optimal savings had actually been studied much earlier by Ramsey (1928), far before Solow wrote his seminal paper. The neoclassical growth model is therefore often called the Ramsey-Cass-Koopmans model, or just the Ramsey growth model.

[10] The plain vanilla version of the neoclassical growth model is discussed in Chapter 4.

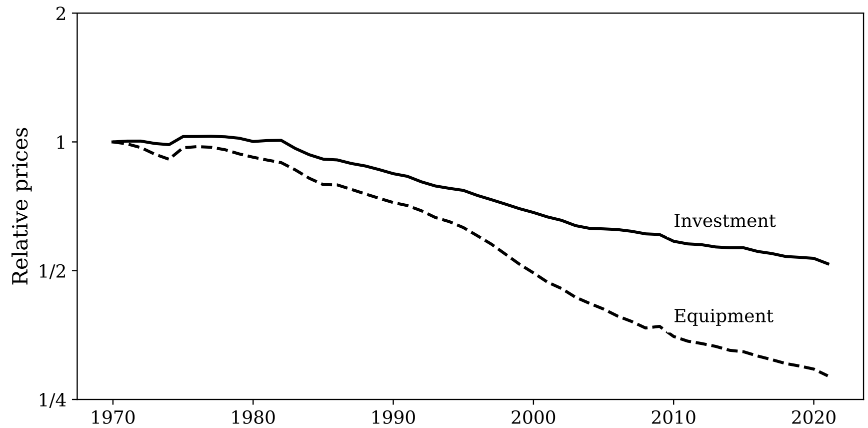

Figure 13.4: Relative price of equipment and investment in the U.S.

Notes: The data comes from the U.S. Bureau of Economic Analysis (NIPA Table 1.1.4). The series are relative to the price of consumption, and are each normalized to 1 in 1970. The price level for structures nearly doubled over 50 years by growing 1.2% per year relative to the consumption deflator. In contrast, the annual growth rate of the relative price of equipment and total investment were about −2.5% and −1.3%, respectively.

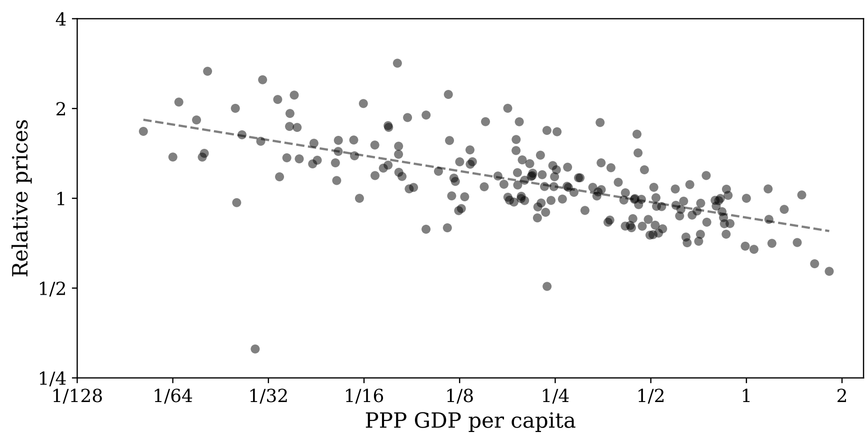

Figure 13.5: Price of investment relative to consumption in 2019 across countries

Notes: The data comes from the Penn World Table 10.0. The U.S. is normalized to 1 for both variables. The price of investment relative to consumption is calculated from the "pl_i" and "pl_c" variables. PPP GDP per capita is calculated as the ratio of the "rgdpo" and "pop" variables. The estimated elasticity is equal to −0.17 with a standard error of 0.02.

from endogenous labor supply. The household problem is given by

$$\max_{\{a(t),c(t)\}_{t=0}^{\infty}} \int_0^{\infty} e^{-(\rho-n)t} \frac{c(t)^{1-\sigma}-1}{1-\sigma} dt, \tag{13.4}$$

subject to

$$\dot{a}(t) = (r(t) - n)a(t) + w(t)h - c(t), \forall t, \text{ with } a(0) \text{ given},$$

and the no-Ponzi game condition

$$\lim_{T\to\infty} \left\{ e^{-\int_0^T (r(\tau)-n)d\tau} a(T) \right\} \geq 0.$$

Here ρ is the discount rate and n is the population growth rate. The initial population size is normalized to unity. The period budget constraints are stated in per-capita terms so a denotes per-capita wealth and h the level of human capital per worker. The household's problem takes the form that is familiar from Chapter 9.

In the household budget the price of the consumption good is normalized to one. The household converts its savings in investment goods at price P_x, and rent its capital stock out to firms. The household receives from the firm the rental rate $R(t)$ but has to incur depreciation δ on its capital plus faces a change in the relative price of $\dot{P}_x(t)$. Hence, the condition that links the return r to the rental rate, the depreciation rate, and the investment price reads

$$P_x(t)r(t) = R(t) - \delta P_x(t) + \dot{P}_x(t). \tag{13.5}$$

Technology Final output Y is competitively produced by a representative firm with the Cobb-Douglas technology

$$Y(t) = K(t)^\alpha \left(A_y e^{\gamma_y t} L(t)\right)^{1-\alpha}. \tag{13.6}$$

Here K denotes capital input and L labor input in production. The parameter $\alpha \in (0,1)$ controls the output elasticity of capital. The term $A_y e^{\gamma_y t}$ captures the technology in Harrod-neutral form (consisting of a level A_y and a rate of technological change $\gamma_y > 0$).

Output Y can be transformed one for one into consumption goods and one for $A_x e^{\gamma_x t}$ into investment goods. As both technologies are linear in Y goods, the price of Y will equalize to the price of C under perfect competition, which we choose as a numéraire, i.e., $P_y(t) = P_c(t) = 1$, $\forall t$ and the price of the investment good will be given by $P_x(t) = e^{-\gamma_x t}/A_x$. The object Y is then a measure of GDP expressed in terms of consumption units. Here $\gamma_x \geq 0$ represents investment-specific technical change; with $\gamma_x > 0$ the relative price of investment declines over time. The production side can then be characterized as a static profit maximization problem of a representative firm:

$$\pi(t) = \max_{K(t),L(t)} \left\{ K(t)^\alpha \left(A_y e^{\gamma_y t} L(t)\right)^{1-\alpha} - R(t)K(t) - w(t)L(t) \right\}. \tag{13.7}$$

Note that, given our choice of numéraire, (per-capita) wealth a and all the prices w, r, R, and P_x are stated in relative terms and measured in units of consumption goods.

Equilibrium definition A competitive equilibrium is then defined as a path of prices and quantities that:

1. solves the household problem (13.4) and the firm problem (13.7).

2. fulfills asset and labor market clearing conditions, i.e., $K(t)e^{-\gamma_x t}/A_x = a(t)e^{nt}$ and $L(t) = e^{nt}h$.

3. fulfills the return condition $r(t) = R(t)A_x e^{\gamma_x t} - \delta - \gamma_x$.

Here $a(t)e^{nt}$ is aggregate wealth (in units of the consumption good). This amount is invested in capital goods at price $e^{-\gamma_x t}/A_x$. Finally, to arrive at the return condition in (3.) we replaced $P_x(t)$ in (13.5) by $e^{-\gamma_x t}/A_x$.

As the model features growth we need to impose the following restriction to ensure that the household problem is well-defined[11]

$$n - \rho + \frac{\alpha(1-\sigma)}{1-\alpha}\gamma_x + (1-\sigma)\gamma_y < 0. \tag{13.8}$$

Model solution The household's problem can be solved using the technique we learned in Chapter 9. In Appendix Section 13.A.1 we solve for the equilibrium dynamics and show that they boil down the following differential equations in $K(t)$ and $c(t)$:

$$\frac{\dot{c}(t)}{c(t)} = \frac{A_x e^{\gamma_x t}\alpha\left(\frac{A_y e^{(\gamma_y+n)t}h}{K(t)}\right)^{1-\alpha} - \delta - \gamma_x - \rho}{\sigma} \tag{13.9}$$

and

$$\frac{\dot{K}(t)}{K(t)} = A_x\left(\left(\frac{e^{(\gamma_x/(1-\alpha)+\gamma_y+n)t}A_y h}{K(t)}\right)^{1-\alpha} - \frac{e^{(\gamma_x+n)t}c(t)}{K(t)}\right) - \delta. \tag{13.10}$$

The initial capital stock $K(0)$ is exogenously given and the additional terminal condition is given by

$$\lim_{T\to\infty}\left\{\frac{K(T)}{A_x}e^{-(\gamma_x+\rho)T}c(T)^{-\sigma}\right\} = 0. \tag{13.11}$$

Equation (13.9) is the consumption Euler equation whereas (13.10) states the law of motion of the capital stock. As the welfare theorems apply in this framework the same system of differential equations in $K(t)$ and $c(t)$ could have been found more directly by solving the planner's problem.[12] However, as we plan to confront the theory's prediction with data on prices and real quantities it is essential that we have solved for the decentralized equilibrium.

Balanced growth path We define a balanced growth path the standard way as a path along which all prices and quantities grow at constant rates. Does this economy admit a balanced growth path and if so at what rates do physical capital K, output Y and per-capita consumption c grow? Given that the model features investment-specific technical change these questions are non-trivial. Can the right-hand side of (13.10) be constant for all t? One notes that a constant ratio $e^{(\gamma_y+n)t}/K(t)$ is not consistent with a constant growth rate in K. Hence, the standard candidate in which capital grows at the combined rate of population and Harrod-neutral technological change is inconsistent with a balanced growth path. However, a valid candidate is with

$$\frac{\dot{K}(t)}{K(t)} = \frac{\gamma_x}{1-\alpha} + \gamma_y + n \equiv g_K$$

[11] We will see below that the long-run growth rate in per-capita consumption is $\frac{\alpha}{1-\alpha}\gamma_x + \gamma_y$. Then, condition (13.8) ensures that utility is bounded.

[12] Appendix Section 13.A.2 states and solves the planner problem for this economy.

and

$$\frac{\dot{c}(t)}{c(t)} = \frac{\alpha\gamma_x}{1-\alpha} + \gamma_y \equiv g_c.$$

As a consequence, we can write the system of differential equations in detrended variables $\tilde{c}(t) \equiv c(t)/e^{g_c t}$ and $\tilde{k}(t) \equiv K(t)/e^{g_K t}$ as:

$$\frac{\dot{\tilde{c}}(t)}{\tilde{c}(t)} = \frac{\alpha A_x \left(\frac{A_y h}{\tilde{k}(t)}\right)^{1-\alpha} - \delta - \gamma_x - \rho}{\sigma} - \frac{\alpha\gamma_x}{1-\alpha} - \gamma_y, \tag{13.12}$$

$$\frac{\dot{\tilde{k}}(t)}{\tilde{k}(t)} = A_x\left(\left(\frac{A_y h}{\tilde{k}(t)}\right)^{1-\alpha} - \frac{\tilde{c}(t)}{\tilde{k}(t)}\right) - \delta - \frac{\gamma_x}{1-\alpha} - \gamma_y - n, \tag{13.13}$$

with $\tilde{k}(0)$ given and the terminal condition[13]

$$\lim_{T\to\infty}\left\{\tilde{k}(T)\tilde{c}(T)^{-\sigma}e^{(n+(1-\sigma)\alpha\gamma_x/(1-\alpha)+(1-\sigma)\gamma_y-\rho)T}\right\} = 0.$$

This system of differential equations is indeed consistent with a stationary point that we can denote by $\tilde{c}^\star$ and $\tilde{k}^\star$. The terminal condition is fulfilled along this balanced growth path as we imposed condition (13.8). By setting $\dot{\tilde{k}}(t)/\tilde{k}(t)$ and $\dot{\tilde{c}}(t)/\tilde{c}(t)$ equal to zero we get

$$\tilde{k}^\star = A_x^{\frac{1}{1-\alpha}} A_y h \left(\frac{\alpha}{\frac{\alpha\sigma\gamma_x}{1-\alpha} + \gamma_x + \sigma\gamma_y + \delta + \rho}\right)^{\frac{1}{1-\alpha}}, \tag{13.14}$$

and $\tilde{c}^\star$ follows directly from

$$\tilde{c}^\star = \left(\tilde{k}^\star\right)^\alpha (A_y h)^{1-\alpha} - (g_K + \delta)\frac{\tilde{k}^\star}{A_x}. \tag{13.15}$$

In general, in a neoclassical growth framework with CRRA preferences any production function which fulfills standard assumptions can support a balanced growth path. However, as the model here features investment-specific technical change (with $\gamma_x > 0$) this is no longer the case and the assumption of a Cobb-Douglas technology in (13.6) is key in order to generate a balanced growth path. This relates back to Uzawa's theorem (see Appendix 3.A) that requires all the technological change to be of the labor augmenting type. In the special case of a Cobb-Douglas production function factor augmenting technical change is not distinctly defined and capital-augmenting technical change can as well be expressed in labor-augmenting terms by raising it to the power $\alpha/(1-\alpha)$.

[13] Adding the constant $A_x^{\frac{1}{1-\alpha}} A_y h$ to the denominator of the definition of the detrended capital stock would allow us to write the expression in a more compact way. However, further below we entertain a thought experiment where we allow countries to differ in these constants and it is then helpful to explicitly see how these factors affect the detrended capital stock along the balanced growth path.

BGP predictions Along the balanced growth path a constant fraction $s^\star = \alpha(g_K + \delta)/\left(\sigma g_c + \gamma_x + \delta + \rho\right)$ is saved out of income before depreciation.[14] The Solow model simply imposes such a constant saving rate. Here, in contrast, we derive this $s^\star$ endogenously and it will in general move along the transition.

We can compare the steady state capital stock $\tilde{k}^\star$ with the Golden Rule capital stock. Maximizing (13.15) with respect to $\tilde{k}^\star$ gives the Golden Rule capital stock

$$\tilde{k}^{gold} = A_x^{\frac{1}{1-\alpha}} A_y h \left(\frac{\alpha}{g_K + \delta}\right)^{\frac{1}{1-\alpha}}. \tag{13.16}$$

This would be achieved if the entire capital income was saved, i.e., $s^{gold} = \alpha$. As $\frac{\alpha\sigma\gamma_x}{1-\alpha} + \gamma_x + \sigma\gamma_y + \delta + \rho > g_K + \delta$ is ensured by condition (13.8) the steady state capital stock is strictly below the golden rule capital stock, which is reminiscent of the discussion in Appendix 4.A.4.

What are the theory's predictions for key aggregates and prices along the balanced growth path? As the model features two final uses—consumption and investment—and their relative price is changing over time, one has to be careful when evaluating the model prediction for real quantities. Total GDP when measured in units of consumption goods is captured by Y. Similarly, w and r denote the wage and interest rate measured in consumption units.[15] The wage rate grows at the same rate as per-capita consumption g_c, whereas the interest rate, $r^\star = \sigma g_c + \rho$ is constant along the balanced growth path (expressing how much consumption is obtained tomorrow by forgiving one unit of consumption today). In contrast, R denotes the rental rate of one unit of physical capital (also measured in terms of tomorrow's consumption units) and is falling along the balanced growth path at constant rate γ_x. The difference is precisely explained by the relative price of investment which falls at rate γ_x too, i.e., investing one unit of physical capital is getting in terms of consumption units less and less costly over time in terms of consumption units.

Also the price of the capital stock relative to the deflator of the consumption good, $e^{-\gamma_x t}A_x^{-1}$ decreases over time at the rate of investment-specific technical change γ_x. The *nominal* capital-output ratio $e^{-\gamma_x t}A_x^{-1}K(t)/Y(t)$ is constant along the balanced growth path and equal to $s^\star/(g_K + \delta)$.[16] "Total real output" is not a traded commodity in the model so a price deflator for total GDP has not been defined in the theory. We can still mimic in the model what statistical offices do in practice: As the nominal shares of investment and consumption are constant along the balanced growth path a reasonable definition of the growth rate of the GDP deflator is $(\dot{P}_x/P_x)^{s^\star}(\dot{P}_c/P_c)^{1-s^\star} = -(1-s^\star)\gamma_x$.[17] With such a definition, it is straightforward to see that the theory predicts that the ratio of real capital to real output grows at rate $(1-s^\star)\gamma_x$.

Transitional dynamics The transitional dynamics are similar as in a version of the neoclassical growth model without investment-specific technical change. If the economy starts

[14] To see this equate the right-hand side of (13.15) to $(1-s^\star)\left(\tilde{k}^\star\right)^\alpha (A_y h)^{1-\alpha}$ and solve for $s^\star$.

[15] The expressions for the equilibrium wage and interest rate can be found in Appendix 13.A.1.

[16] A similar formula was obtained in the Solow model (see Section 3.2). Importantly, here the formula applies to the nominal capital-output ratio and not to a measure of the real capital-output ratio.

[17] The second expression follows as we chose P_c as numéraire.

out with a detrended capital stock below $\tilde{k}^\star$ capital accumulates faster and $\tilde{k}$ increases over time until it asymptotically reaches $\tilde{k}^\star$.

For the special case with $\sigma = \alpha$ the model allows for a closed-form solution along the transition. This be shown by guessing and verifying the system (13.12) and (13.13) is globally consistent with the consumption rule $\tilde{c}(t) = \tilde{k}(t)\left(\delta(1-\alpha) + \gamma_x(1-\alpha) - \alpha n + \rho\right)/(\alpha A_x)$. For more general cases where no closed-form solution exists one can still analytically analyze the local dynamics by linearizing the system around the steady state. This is illustrated in Appendix 13.A.3. The linearized system can also be used to assess the (local) speed of convergence. A simple calibration done in Appendix 13.A.3 suggests that the speed of convergence is fast, implying a half-life of about 5.5 years.

Discussion of the framework The model we laid out assumes a closed economy. One obvious justification of this assumption is that the world as a whole is indeed a closed economy. For large economies like the U.S. international trade flows are quantitatively not that large relative to GDP and trade may be considered not to be of first-order importance for investment and capital accumulation. It is a standard approach to address the cross-country data through the lens of closed economy theory and this is also what we will pursue below.[18]

The framework we laid out assumes a representative household but as the CRRA utility function implies overall homothetic utility, preferences allow for aggregation.[19] The CRRA period utility function is key in order to microfound a constant long-run saving rate (with growth). We abstracted from endogenous labor supply. However, as discussed in Section 2.1.8, an endogenous labor supply decision could be allowed for and can be reconciled with a balanced growth path (see Boppart and Krusell, 2020).

The level of human capital h is assumed to be constant but the model allows to study transitional movements in this level. Allowing instead for sustained growth in h would potentially lead to very different dynamics which we comment on in the section below on endogenous growth.

The model consists of multiple sectors (an investment sector and a consumption sector). However, along the balanced growth path the nominal investment rate is constant which precludes a systematic long-run reallocation of production factors across sectors—a phenomenon called structural change. It is indeed a salient empirical observation that development is accompanied by shifts in employment shares, first from agriculture to manufacturing and at a later stage into services. How to square this unbalanced feature of growth at the sectoral level with the balanced nature of growth in the aggregate? There indeed exist restrictions on technologies and preferences such that there exists a balanced growth path at the aggregate level despite structural change at the sectoral level (see Kongsamut, Rebelo, and Xie, 2001, Ngai and Pissarides, 2007, and Boppart, 2014).[20]

[18] The interaction between countries is arguably more important to influence "technology". We will comment on technological spill-overs and adoption further below when we discuss theories of endogenous growth.

[19] I.e., we could as well assume that there is a unit interval of heterogeneous households endowed with h_i units of human capital and $a_i(0)$ units of initial wealth. As long as $\int_0^1 h_i\, di = h$ and $\int_0^1 a_i(0)\, di = a(0)$ the aggregate dynamics would be identical to the ones in the economy specified above.

[20] The standard definition of an exact balanced growth path as in Section 3.2 needs to be relaxed when studying structural change as imposing that all aggregate variables have to grow at constant rates restricts

Finally, a relevant extension is to explicitly modeling energy and natural resources as production factors. Chapter 25 discusses such models.

13.4 Confronting the investment-specific technical change model with the data

How does the model in the previous section allow us to think about the data? What explains the big income differences across countries? What is behind the observed growth miracles and growth disasters? By addressing these questions we also comment on the extensive empirical literature on growth.

One way to think about income differences through the lens of the theory is as transitional dynamics due to differences in the *initial* capital stock. Suppose all countries are identical except for their initial capital stock. As all countries then converge to the same balanced growth path, we expect to see a strong pattern of convergence of income levels across countries. As emphasized earlier, however, we do not find strong support for absolute (i.e., unconditional) convergence in the cross-country data (Figure 13.3). And even in samples where there is evidence of unconditional convergence (such as across U.S. states from 1880 to 1980, among OECD countries from 1950 to 1980, or in the full cross-country sample since the 2000s), the observed speed of convergence is not nearly as fast as the one implied by the theory.[21]

At the same time the world log income distribution does not systematically fan out over time (see Appendix Figure 13.A.2). Hence, there are clear forces that hold the income distribution together over time (and convergence due to the diminishing marginal product of capital is one such force). The lack of *divergence* in Figure 13.3 suggests the rate of technical change does not differ systematically between rich and poor countries. It is therefore reasonable to think of all countries to share the same parameters γ_y and γ_x in the model above. The labor income share is across countries not systematically related to the level of development suggesting that the parameter α can also be treated as common across countries. Technology differences could still matter, but would rather be reflected in the model by *level* differences in A_y or A_x.

Why do countries persistently differ in their income levels? One possibility is capital-output ratios, say, e.g., due to country differences in population growth n.[22] Moreover, given common values of γ_y, γ_x, and α, different levels of human capital h, A_y, or A_x lead to different

the sectoral shares to be constant. One such modification of the definition is to define an exact balanced growth path as a path along which the physical capital stock grows at a constant rate (see Alder, Boppart, and Müller, 2022). Alternatively, one could simply not insist on exact balanced growth but rather rely on such a path to exist asymptotically.

[21] Remember, a simple calibration in the model above suggests a half-life of about 5.5 years. Barro and Sala-i-Martin (1992) provide evidence of convergence across U.S. states, Baumol (1986) across OECD countries, and Kremer, Willis, and You (2022) in the full cross-country sample over the past 25 years.

[22] See the expression for the capital stock along the balanced growth path in (13.14) to see how the rate of population growth affects the capital-output ratio.

income levels. One can express output per worker (measured in consumption units) as

$$\frac{Y(t)}{e^{nt}} = \left(\frac{P_k(t)K(t)}{Y(t)}\right)^{\frac{\alpha}{1-\alpha}} \cdot h \cdot \underbrace{A_y A_x^{\frac{\alpha}{1-\alpha}}}_{=A} \cdot e^{\left(\gamma_y + \gamma_x \frac{\alpha}{1-\alpha}\right)t}, \tag{13.17}$$

where $P_k(t) = A_x^{-1} e^{-\gamma_x t}$. Along a balanced growth path, the nominal capital-output ratio $\frac{P_k(t)K(t)}{Y(t)}$ is constant and equal to $s^\star/(g_K+\delta) = \alpha/\left(\frac{\alpha\sigma\gamma_x}{1-\alpha} + \gamma_x + \sigma\gamma_y + \delta + \rho\right)$ and independent of the levels of human capital or TFP. For a given α, equation (13.17) then decomposes per-worker income differences in terms of consumption units in a theory-consistent way into contributions from physical capital, human capital, and a residual technology term. The results are shown in Appendix Table 13.A.1. The message that a majority of income difference is unaccounted for by physical and human capital is reinforced compared to the results in Table 13.2. The share explained by residual TFP is 72% compared to 64% before. This is because the relative price of capital (compared to the GDP deflator) is higher in poorer countries. The contribution of physical capital falls to 7% from 14% previously (whereas the contribution of human capital is unchanged at 22%).

An advantage of this two-sector development accounting is that one can decompose the residual A into general TFP differences A_y and differences specific to the investment sector A_x. One way to discipline the differences in A_x is to bring in the relative prices P_x/P_c from Figure 13.5. The data suggests that the elasticity of $A_x^{\alpha/(1-\alpha)}$ with respect to $P_Y Y/(P_C L)$ is 0.10. This implies that on average 0.10 out of 0.72 or about 14% of the A differences are due to rich countries being particularly efficient at producing investment goods.

Transition dynamics due to capital accumulation might be useful for thinking about the rapid catch-up growth of certain countries. For instance, Japan's GDP per capita relative to the U.S. rose from about 20% in 1950 to 80% in 1990. Rapid capital deepening should, however, be accompanied by a sharp decline in the return to capital, which was not observed in Japan.[23] Thus, even though convergence due to the diminishing marginal product of capital is an important feature of the neoclassical growth model, transitional dynamics due to different initial physical capital levels cannot account for growth miracles such as Japan's. The more realistic way to think about episodes of fast (catch-up) growth is evidently to see them as transitional changes in A_y or A_x. The previous section provided some analytics for such a theory.

Similar to development accounting, (13.17) can be used to decompose growth in per-capita output (in consumption units) in the U.S. over time into contributions from capital deepening, improvements in labor "quality," and residual TFP growth. Appendix Table 13.A.2 presents the results. Over the post-war period the U.S. nominal capital-output ratio was remarkably stable. In contrast, we saw in Table 13.1 that there was modest capital deepening in terms of the U.S. real capital-output ratio (in particular since the 1970s). The contribution of human capital is, again, largely unaffected compared to Table 13.1. This leaves the clear majority of growth accounted for by the TFP residual A. Growth

[23] In order to explain a per-capita income difference of a factor of 5, the relative capital stock per worker needs to fulfill $0.2 = (\tilde{k}_{1950}^{JAP}/\tilde{k}_{1950}^{USA})^\alpha$. Everything else equal, based on (13.A.1) the relative rental rate is $(R_{1950}^{JAP}/R_{1950}^{USA}) = (0.2)^{-\frac{1-\alpha}{\alpha}}$. With $\alpha = 1/3$ this implies in 1950 a rental rate in Japan that is 25 (!) times larger than in the U.S. By 1990 this ratio should have declined to less than 2 (see King and Rebelo, 1993).

in per-capita output measured in consumption prices of 2.07% per year can be used to discipline $\gamma_y + \gamma_x \frac{\alpha}{1-\alpha}$ in the model of the previous section. Relative prices allow us to further decompose this growth rate into contributions from γ_y and investment-specific technical change, respectively. Figure 13.4 suggest that $\gamma_x = 0.013$. With an α of about 1/3 this suggests $\gamma_y = 0.014$ and that about 31% of per-capita consumption growth is driven by investment-specific technical change.

To recap, residual TFP accounts for the bulk of differences in income levels and growth rates across countries. What is behind this mysterious residual? For the time series, a natural candidate is technological change (either neutral or investment-specific). In the next section we present theories of endogenous technical change. Technology could also play an important role in income levels across countries. Parente and Prescott (1994), Eaton and Kortum (1996, 1999), and Howitt (2000) are classic models of international technology diffusion. Barro and Sala-i-Martin (1995) chapter 8 and Acemoglu (2008) chapter 18 provide textbook treatments. More recent modeling efforts include Akcigit, Ates, and Impullitti (2018), Buera and Oberfield (2020), Hsieh, Klenow, and Shimizu (2022), and Hsieh, Klenow, and Nath (2023).

On the empirical side, Comin and Hobijn (2010) and Comin and Mestieri (2014) provide direct evidence on the use of specific technologies across countries and time. Keller (2004) and Comin and Mestieri (2018) survey the evidence. The Eaton and Kortum papers referenced above document cross-country patenting patterns and how they relate to country size and income. Evenson and Gollin (2003) and Gollin, Hansen, and Wingender (2021) provide evidence for the diffusion of hybrid seeds across many countries, and their impact on agricultural yields. Bloom and Van Reenen (2007) document differences in management practices across countries.

Differences in allocative efficiency could also contribute to income differences across economies. For example, countries may differ in the efficiency with which production factors (capital, labor, intermediates) are allocated across firms due to various market distortions and government policies. Hsieh and Klenow (2009), Alfaro, Charlton, and Kanczuk (2009), and Bartelsman, Haltiwanger, and Scarpetta (2013) provide evidence across countries. See Restuccia and Rogerson (2017) for a survey.

One can ask deeper questions about where these differences in physical capital, human capital, technology, and allocative efficiency come from. A vibrant literature connects these proximate determinants of income to underlying policies and institutions. Acemoglu et al. (2001) is a prominent empirical example, and Chapters 22 and 23 in Acemoglu (2008) survey some of the ways in which institutions and policies can be modeled. Geography is often an underlying factor affecting income directly or indirectly through institutions, such as in Gallup, Sachs, and Mellinger (1999).

13.5 Endogenous growth

AK model The simplest endogenous growth theory is obtained in the economy of the previous section by setting $\alpha = 1$, considering a single output good by setting $A_x = 1$, and shutting down exogenous technological change, i.e., setting $\gamma_y = \gamma_x = 0$. If we premultiply the production function by an additional constant A, it becomes $Y(t) = AK(t)$, which is why the literature refers to this as the AK model. The planner's problem of this economy then reads

$$\max_{\{K(t),c(t)\}_{t=0}^{\infty}} \int_0^{\infty} e^{-(\rho-n)t} \frac{c(t)^{1-\sigma}-1}{1-\sigma} dt, \tag{13.18}$$

subject to $\dot{K}(t) \leq AK(t) - e^{nt}c(t) - \delta K(t)$, given $K(0)$, and some non-negativity constraints $K(t) \geq 0, \forall t$. For the problem to be well defined and utility to be bounded we have to assume $\sigma(A-\delta-n) > A-\delta-\rho$. It is then straightforward to solve the planner's Hamiltonian to arrive at the Euler equation

$$\frac{\dot{c}(t)}{c(t)} = \frac{A-\delta-\rho}{\sigma}, \tag{13.19}$$

the resource constraint $\dot{K}(t) = AK(t) - e^{nt}c(t) - \delta K(t)$, and the transversality condition

$$\lim_{T\to\infty} \left\{ c(T)^{-\sigma} e^{-\rho T} K(T) \right\} = \lim_{T\to\infty} \left\{ c(0)^{-\sigma} e^{-(A-\delta)T} K(T) \right\} = 0. \tag{13.20}$$

These expressions characterize the planner's solution in the AK model.

Interestingly, consumption grows at a constant rate irrespective of the level of the initial capital stock (see (13.19)). Hence, the AK model features no transition dynamics and the optimal consumption rule is $c(t) = (A-\delta-n-(A-\delta-\rho)/\sigma)e^{-nt}K(t)$. Capital likewise grows right away at the constant rate $\dot{K}(t)/K(t) = n + (A-\delta-\rho)/\sigma$.[24] The planner's solution could also be decentralized with competitive markets.

The AK model is an endogenous growth model as the rate of long-run growth is no longer an exogenous constant but a function of preferences and technology parameters and does respond to policies.[25] The reason why endogenous sustained growth is feasible in this economy is because there is no diminishing marginal product of capital. The chapter on the Solow model showed that a constant saving rate then leads to an (endogenous) balanced growth path (see Figure 3.2).

The production function that is linear in capital violates the standard diminishing returns assumption, and implies no role for labor to play in production. One would need massive increasing returns to scale (internally or externally), as in Romer (1986), to reconcile linear production in capital with the empirical capital share on the order of 1/3. Related, labor is very much an important factor in production that commands about 2/3 of all income. So the AK model is knife-edge and unrealistic in several dimensions. Still, it is helpful to discuss it as some more sophisticated theories of endogenous growth (to be presented shortly) have a reduced form AK structure.

[24] As always, it is easy to show that this solution fulfills the system of equations plus the transversality condition (given we imposed $\sigma(A-\delta-n) > A-\delta-\rho$). One can also show that this is the only solution. As consumption grows at a constant rate we can plug this into the constraint to get $\dot{K}(t)/K(t) = A - \delta - e^{(n+(A-\delta-\rho)/\sigma)t}c(0)$. This differential equation has a closed form solution given by $K(t) = \Xi e^{(A-\delta)t} + (A-\delta-n-(A-\delta-\rho)/\sigma)^{-1} e^{(n+(A-\delta-\rho)/\sigma)t}c(0)$, where Ξ is a constant of integration. The transversality condition is then only fulfilled if the constant of integration Ξ is equal to zero (which coincides with the solution above where capital grows at a constant rate).

[25] For instance, consider a decentralized version of this economy, where capital income is taxed at rate τ and the tax revenues are rebated lump-sum to the household. It is straightforward to show that the resulting growth rate is monotonically decreasing in τ.

Perhaps more realistically, the AK model can be viewed as a reduced form of a model with constant returns to physical and human capital combined.[26] That is, $Y(t) = F(K(t), H(t))$ where $H = hL$ is total human capital which can be accumulated at a steady rate without a bound. Such an economy may support a balanced growth path along which output, physical capital, and human capital all grow at a constant endogenous rate. But can human capital grow at a steady rate? The answer appears to be yes if there is increasing investment in human capital, such as from rising years of schooling. Much less clear, however, is whether human capital grows if the level of human capital investment per person is fixed over time. Each generation arguably starts at the same level of human capital, so they would need to obtain more human capital with each year of education or experience. Mincerian returns to education and experience exhibit no such secular trends.

Bils and Klenow (2000) argued that higher schooling has a level effect rather than a growth effect in a panel of counties. Klenow and Rodriguez-Clare (2005) also made a case for level effects of schooling. For example, they pointed out that education differences are quite persistent across countries, while income growth rate differences are more transitory. And schooling levels are generally trending up over time, while growth rates are not.[27] In contrast, schooling levels *are* highly correlated with income levels across countries.[28]

A property of all such AK-type endogenous growth theories is that they are entirely factor accumulation based. This means that, if inputs are measured properly, a growth accounting exercise should reveal zero TFP growth. These class of models also predicts a rather elastic response of the long-run growth rate to changes in taxes on profits or R&D subsidies that are not seen in the data (see Stokey and Rebelo, 1995). There is no role for patents, R&D expenditures, ideas, or purposeful innovation. We next study models which do provide a such role.

13.5.1 Endogenous growth through expanding varieties

Romer (1990) put the non-rival nature of ideas at center stage and modeled purposeful R&D investment of profit maximizing firms in general equilibrium. We discuss here a version of Romer (1990) that abstracts from physical capital.

Suppose the aggregate production function is given by

$$Y(t) = \frac{A}{1-\phi} L_y(t)^{\phi} \int_0^{N(t)} x(\nu, t)^{1-\phi}\, d\nu. \tag{13.21}$$

This aggregate production function is defined over the inputs labor L_y and an interval of different "machines" of measure N with quantity $x(\nu)$ for machine ν. We use the terminology

[26]See Appendix 13.A.4 where we comment on other purely accumulation based endogenous growth theories that look in a reduced form like the simple AK model.

[27]This is akin to Jones (1995) noting that R&D investments trend up over time but the long-run growth rate does not. He inferred that this likewise suggested a level effect of R&D investment on productivity.

[28]Imagine the law of motion of human capital $\dot{h} = Bh^{\phi}s^{\omega}$. The key question is whether $\phi = 1$ or $\phi < 1$. If $\phi = 1$ then higher levels of human capital investment s lead to perpetually faster growth. If, instead, $\phi < 1$ then higher levels of s yield higher levels of h but no faster growth in the long run. Again, this is analogous to semi-endogenous growth due to R&D investment that we will discuss further below.

machine even though, for simplicity, the inputs are like material in that they depreciate fully after use. We assume $\phi \in (0,1)$.

The production function (13.21) merits some discussion: First, the factor $1/(1-\phi)$ is added for notational simplicity and is not essential (one can always just redefine the constant A). Second, the production function has constant returns to scale when considering labor L_y and all the machines $\{x(\nu)\}_{\nu=0}^{N}$—in line with a simple replication argument. However, here the number of machine varieties N will be endogenous and will grow over time. Including this endogenous N, equation (13.21) features overall *increasing returns* to scale.[29] Third, one can view the production function as Cobb-Douglas over labor and a machine composite X, i.e., $Y = \frac{A}{1-\phi} L_y^{\phi} X^{1-\phi}$. The bundle of machines is a CES (Dixit-Stiglitz) aggregator

$$X = \left(\int_0^N x(\nu)^{\frac{\epsilon-1}{\epsilon}} d\nu \right)^{\frac{\epsilon}{\epsilon-1}}.$$

One arrives at (13.21) if one directly connects the elasticity of substitution across machines, ϵ, to the output elasticity of labor, ϕ, by setting $\epsilon = 1/\phi$. One can relax this assumption and still obtain a balanced growth path in the environment with research labor we study below.

Below we will refer to N as the knowledge stock in the economy. In this model knowledge is embodied in the number of available intermediate inputs. An economy with a higher knowledge stock will exhibit higher TFP.

Market structure The final output good is competitively produced by a representative firm. At each point in time this firm solves (so we suppress the time index)

$$\max_{L_y, \{x(\nu)\}_{\nu=0}^{N}} \left\{ \frac{A}{1-\phi} L_y^{\phi} \int_0^N x(\nu)^{1-\phi} \, d\nu - w L_y - \int_0^N p(\nu) x(\nu) \, d\nu \right\}, \tag{13.22}$$

where we chose the price of the final output as the numéraire. The first-order conditions of this problem are

$$\phi Y / L_y = w \tag{13.23}$$

and

$$A L_y^{\phi} x(\nu)^{-\phi} = p(\nu), \ \forall \nu. \tag{13.24}$$

In contrast to the final good, each machine variety is produced by only one firm, which therefore acts as a monopolistic competitor. The producer of a given variety ν maximizes current period profits

$$\pi(\nu) = \max_{p(\nu), x(\nu)} \{ p(\nu) x(\nu) - \psi x(\nu) \}, \text{ subject to (13.24).} \tag{13.25}$$

Here the assumption is that machines can be produced at constant marginal cost ψ in units of the numéraire (final output good).

[29]To see this, suppose all machines cost p_x and a budget of E is equally split on all machines. We then have $x(\nu) = E/(p_x N)$, $\forall \nu$ and aggregate output becomes $Y = \frac{A}{1-\phi} L_y^{\phi} N^{\phi} (E/p_x)^{1-\phi}$. This highlights the love-of-variety inherent in specification (13.21): output is boosted if more types of machines are available (holding prices and the total budget constant).

Inventing varieties By hiring $1/(\eta N(t))$ labor units for R&D an entrant can invent a new machine. This implies for the law of motion of available varieties

$$\dot{N}(t) = \eta N(t) L_r(t), \text{ given } N(0) > 0, \tag{13.26}$$

where L_r is total labor used in R&D and $\eta > 0$. The N term in (13.26) represents a knowledge spillover: more available varieties make researchers more productive in terms of generating new varieties.[30]

After invention the entrant receives a perpetual patent to exclusively produce the invented variety. The value of such a patent is given at date t by the present discounted value of future profits

$$V(t) = \int_t^\infty e^{-\int_t^s r(\tau)d\tau} \pi(s) ds, \tag{13.27}$$

where r is the real interest rate and we suppressed the variety index ν due to symmetry. By time differentiating (13.27) one obtains the Hamilton-Jacobi-Bellman (HJB) equation: $V(t)r(t) = \pi(t) + \dot{V}(t)$. The return on V must be equal to the flow of profits $\pi(t)$ plus capital gains $\dot{V}(t)$. The stream of future profits in (13.27) incentivize the R&D investments in the first place. Under free entry in R&D we can assume that there is a representative R&D firm solving in each point in time

$$\max_{L_r} \{V(t)\eta N(t) L_r(t) - w(t) L_r(t)\} \tag{13.28}$$

subject to

$$L_r(t) \geq 0.$$

Household problem The model is closed with a standard household side. The representative household supplies inelastically L units of labor earning a wage rate w. The household owns all the firms, whose total value represents household wealth

$$a(t) = \int_0^{N(t)} V(\nu, t)\, d\nu, \tag{13.29}$$

that earns a combined real return of r. The household then solves

$$\max_{\{a(t),c(t)\}_{t=0}^\infty} \int_0^\infty e^{-\rho t} \frac{c(t)^{1-\sigma} - 1}{1-\sigma} \tag{13.30}$$

subject to

$$\dot{a}(t) = r(t)a(t) + w(t)L - c(t),$$

and a standard no-Ponzi game condition.[31] We deliberately abstract from population growth for now; more on that below.

A decentralized equilibrium is defined as a path of prices and quantities that jointly solve the household and the firm problem, and is consistent with factor market clearing. Appendix

[30] We will relax the one-for-one proportionality of this spillover in N below.

[31] The no-Ponzi game condition can be expressed as $\lim_{T\to\infty} \left\{ e^{-\int_0^T r(\tau)d\tau} a(T) \right\} \geq 0$.

13.A.5 gives a formal equilibrium definition. To ensure that the household problem is well defined (i.e., utility is bounded) we need to impose the parameter restriction

$$\rho > (1-\sigma)\frac{(1-\phi)\eta L - \rho}{1-\phi+\sigma}. \tag{13.31}$$

This condition ensures that the discount rate is larger than the endogenous consumption growth rate times $1-\sigma$. Furthermore, we focus in the following on the interesting case with strictly positive growth and $L_r > 0$, which will be ensured as long as

$$\rho < (1-\phi)\eta L. \tag{13.32}$$

This condition is fulfilled and innovations are profitable as long as the R&D efficiency η and the market size L are large enough compared to the discount rate.

Note that the decentralized equilibrium we characterize involves monopolistically competitive firms and knowledge spillovers. As a consequence, as we will see below, the decentralized equilibrium will not coincide with the planner's solution.

Solving the model We already solved the final output producer problem above, yielding the first-order conditions (13.23) and (13.24). We did so in order to specify the firm problem of the machine producers who take the (inverse) demand (13.24) into account.

Solving the problem of a machine producer gives the first-order conditions for any point in time

$$x(\nu) = \left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1}{\phi}} L_y, \ \forall \nu, \tag{13.33}$$

and the resulting optimal price is given by $p(\nu) = \psi/(1-\phi)$, i.e., a constant markup over marginal cost.[32] Profits are given by

$$\pi(\nu) = \phi A \left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1}{\phi}-1} L_y. \tag{13.34}$$

This model features a market size effect, i.e., profits from having developed a new machine variety are increasing in the number of workers, L_y, operating this machine. This force plays a role for the direction of technical change in the presence of high and low skilled labor in Acemoglu (1998). See Chapter 21 for further discussion.

Because of symmetry—$x(\nu)$ and $p(\nu)$ are the same for all ν—we can suppress the variety index ν and write final output (13.21) as

$$Y(t) = \frac{A}{1-\phi}\left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1-\phi}{\phi}} L_y(t)N(t), \tag{13.35}$$

and the wage rate as

$$w(t) = \phi\frac{A}{1-\phi}\left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1-\phi}{\phi}} N(t). \tag{13.36}$$

[32] The optimal price fulfills the Lerner condition that price is equal to $\frac{\partial x}{\partial p}\frac{p}{x}/\left(\frac{\partial x}{\partial p}\frac{p}{x}+1\right)$ times marginal cost. Under the imposed CES structure the price elasticity of demand $\frac{\partial x}{\partial p}\frac{p}{x}$ is equal to the constant $-1/\phi$. The markup is then $(1/\phi)/(1/\phi-1)$. This markup rule is familiar from Chapter 6.

Plugging (13.33) and (13.35) into the final good market clearing condition gives consumption as

$$c(t) = \left(\frac{1}{(1-\phi)^2} - 1 \right) \psi \left(\frac{A(1-\phi)}{\psi} \right)^{\frac{1}{\phi}} L_y(t) N(t). \tag{13.37}$$

The household problem gives rise to the same Euler equation as before:

$$\frac{\dot{c}(t)}{c(t)} = \frac{r(t) - \rho}{\sigma}, \tag{13.38}$$

as well as the transversality condition

$$\lim_{T \to \infty} \left\{ e^{-\int_0^T r(\tau) d\tau} a(T) \right\} = 0.$$

The first-order condition for the R&D firm's problem is given by $V(t) \leq \frac{w(t)}{\eta N(t)}$, with $L_r(t)(V(t) - w(t)/(\eta N(t))) = 0$ (taking care of the Kuhn-Tucker condition). In the case with strictly positive growth and R&D investment—which we focus on in the following—a firm value then needs to equalize the fixed cost of inventing a variety:

$$V(t) = \frac{w(t)}{\eta N(t)}. \tag{13.39}$$

Substituting in the wage from (13.36), this implies

$$V(t) = \frac{\phi A}{\eta(1-\phi)} \left(\frac{A(1-\phi)}{\psi} \right)^{\frac{1-\phi}{\phi}}, \ \forall t, \tag{13.40}$$

i.e., a *constant* equilibrium firm value over time. Substituting this constant firm value and the equilibrium profits into the HJB equation then gives

$$r(t) = \frac{\pi(t)}{V(t)} = \eta(1-\phi) L_y(t). \tag{13.41}$$

Finally, we obtain from substituting the labor market clearing condition in the law of motion of N

$$\frac{\dot{N}(t)}{N(t)} = \eta(L - L_y(t)). \tag{13.42}$$

This system of equations can be simplified. Using (13.37), (13.41), and (13.42) in the Euler equation, we obtain a single first-order differential equation in L_y given by

$$\frac{\dot{L}_y(t)}{L_y(t)} = \frac{\eta(1-\phi+\sigma)L_y(t) - \rho}{\sigma} - \eta L. \tag{13.43}$$

Hence, there is indeed a balanced growth path along which L_y is constant with

$$L_y^\star = \frac{\sigma L + \rho/\eta}{1 - \phi + \sigma}. \tag{13.44}$$

There are no transitional dynamics and the economy will grow right away along the balanced growth path (similar to the AK model above).[33] R&D labor along the balanced growth path is given by[34]

$$L_r^\star = \frac{(1-\phi)L - \rho/\eta}{1-\phi+\sigma}.$$

The interest rate along the balanced growth path is then

$$r^\star = (1-\phi)\frac{\eta\sigma L + \rho}{1-\phi+\sigma}. \tag{13.45}$$

The number of varieties, output, consumption, and wealth grow at rate

$$\frac{\dot{N}(t)}{N(t)} = \frac{\dot{Y}(t)}{Y(t)} = \frac{\dot{c}(t)}{c(t)} = \frac{\dot{a}(t)}{a(t)} = \frac{(1-\phi)\eta L - \rho}{1-\phi+\sigma}. \tag{13.46}$$

The transversality condition is ensured by (13.31).

We see that the decentralized growth rate is decreasing in the preference parameters ρ and σ.[35] The endogenous growth rate (13.46) is increasing in the labor force L, a so-called strong scale effect. Given an R&D intensity, $L_r^\star/L$, profits are increasing in scale, thereby attracting more R&D investment. Furthermore, the share of labor allocated to R&D $L_r^\star/L$ is increasing in L. When applied to closed economies, the model predicts bigger countries will grow faster and exhibit higher R&D intensity. These predictions are clearly not borne out in the data (neither in the cross-section of countries not in the time series within a country). The section below which limits the knowledge spillovers will address this critique.

Due to the increasing overall returns to scale in the production function (13.21) and the monopolistically competitive firms the decentralized equilibrium is not efficient. The planner solves:

$$\max_{\{L_y(t),c(t),\{x(\nu,t)\}_{\nu=0}^{N(t)},N(t)\}_{t=0}^\infty} \int_0^\infty e^{-\rho t}\frac{c(t)^{1-\sigma}-1}{1-\sigma}dt, \tag{13.47}$$

subject to the resource constraint

$$\frac{A}{1-\phi}L_y(t)^\phi \int_0^{N(t)} x(\nu,t)^{1-\phi}\, d\nu = c(t) + \int_0^{N(t)} \psi x(\nu,t)\, d\nu, \tag{13.48}$$

[33]Transition dynamics do not arise because the knowledge spillovers in (13.26) are proportional to N and because there is no population growth. We give up both assumptions below. One can directly see from (13.43) that this balanced growth path is the only equilibrium path. For any initial $L_y < L_y^\star$, L_y will shrink over time and go to zero and so does r and the transversality condition is violated. For any initial $L_y > L_y^\star$, L_y will systematically grow over time and go to L implying a "high" interest rate in (13.41), and yet no further growth because there is zero research labor, and in turn an inconsistency with the Euler equation.

[34]In the analysis above we considered an interior solution with $L_r > 0$. Can there also be a corner solution with $L_r = 0$? In such a case there is no growth and therefore $r^\star = \rho$, implying firm value

$$V = \pi/r^\star = \phi A \left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1}{\phi}-1} L,$$

which under (13.32) is strictly larger than $w(t)/\eta N(t)$. Hence, the condition $V(t) \leq w(t)/\eta N(t)$ is violated and (13.32) ensures an equilibrium with strictly positive growth.

[35]Intuitively, with more impatience or more curvature in the momentary utility function the consumption path is less steep.

and the ideas production function (13.42) and a given $N(0) > 0$. (Note that there we already used the labor constraint to substitute out $L_r(t)$.) For the solution to this problem to be well defined and leading to positive growth, we need to impose the parameter restrictions

$$(1-\sigma)\eta L < \rho < \eta L. \tag{13.49}$$

In Appendix 13.A.6 we solve the planner problem and show that it implies a balanced growth path with

$$g^{SP} = \frac{\eta L - \rho}{\sigma}, \tag{13.50}$$

and that the quantity of each machine is given by

$$x(\nu, t) = \left(\frac{A}{\psi}\right)^{\frac{1}{\phi}} L_y^{SP} = \left(\frac{A}{\psi}\right)^{\frac{1}{\phi}} \frac{(\sigma-1)L + \rho/\eta}{\sigma}. \tag{13.51}$$

Again, one can also show that there are no transitional dynamics and the economy grows right away along the balanced growth path (irrespective of the initial $N(0)$).

In the decentralized equilibrium above we saw that $x(\nu, t) = (A(1-\phi)/\psi)^{\frac{1}{\phi}} L_y^{\star}$. Hence, for given production labor L_y the decentralized machine quantity is too low by a factor of $(1-\phi)^{\phi} < 1$. This is due to the monopolistically competitive firms charging a markup factor of $1/(1-\phi)$ over marginal cost. This problem of under-usage of machines could be fixed by subsidizing machines by a rate τ that fulfills $1-\tau = (1-\phi)^{-1}$ and finance this subsidy by lump-sum taxes. However, even such a subsidy would not restore overall efficiency in this economy. This is because on top of the distortion due to markups there is the knowledge spillover in (13.42) that the planner takes into account, whereas in the decentralized equilibrium the R&D effort is only determined by the (discounted stream of) profits. Contrasting the decentralized equilibrium with the planner solution we indeed see that the decentralized growth rate is generally too low. As a consequence, restoring efficiency requires an additional R&D subsidy on top of the machine subsidy.

It is straightforward to add physical capital to the endogenous growth model above where additional transitional dynamics can arise due to the convergence of the capital stock to its steady state level.

The endogenous growth theory above is micro-founding the forward looking R&D decision of monopolistically competitive firms and characterizes how it determines the long-run growth rate of the economy. Markups serve a double role; they lead to efficiency losses due to too high machine prices, but they also lead to profits and incentivize R&D expenditures. The model however remains rather stylized in many dimensions and the resulting policy conclusions are rather stark. Along a balanced growth path the HJB equation, (13.39) and (13.36) imply

$$\frac{\pi^{\star}}{r^{\star}} = \frac{\phi}{\eta}\frac{A}{1-\phi}\left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1-\phi}{\phi}}. \tag{13.52}$$

Hence, the ratio of profits and the interest rate is equal to a constant. Taxing profits would lower the interest rate to therefore lower the growth rate (see (13.38)). Similarly, lowering the markup firms can charge to $\mu < (1-\phi)^{-1}$ through say antitrust policy would monotonically lower the growth rate. Or making patents of the innovators expiring would also systematically decrease the growth rate. The quality ladder model we discuss below—which features business stealing—can imply more nuanced and non-monotonic policy trade-offs.

Less than proportional knowledge spillovers in ideas production

Here we introduce two modifications to the framework above introduced by Jones (1995). First, suppose instead of (13.26) that the "ideas" production function is given by

$$\dot{N}(t) = \eta N(t)^{\epsilon} L_r(t), \text{ given } N(0) > 0, \tag{13.53}$$

with $\epsilon < 1$ capturing that it requires—as the knowledge stock advances—more and more R&D labor to sustain a constant rate of knowledge growth. With $0 < \epsilon < 1$ there are still positive (but limited) knowledge spillovers, i.e., with a higher knowledge stock, N, each R&D worker generates more ideas. In contrast, with $\epsilon < 0$ the specification can also capture a "fishing out" effect (negative technology spillovers) where it gets harder for a given R&D worker to find a new idea the higher is the knowledge stock. The special case with $\epsilon = 0$ captures a situation without any knowledge spillovers and with $\epsilon = 1$ we are back in the framework above with proportional knowledge spillovers. Second, suppose further that there is population growth at rate $n > 0$ such that the labor market clearing condition becomes $L(t) = Le^{nt} = L_y(t) + L_r(t)$. To ensure that the household problem is bounded we assume $\rho - n > n/(1-\epsilon)$.

With these two modifications to the set-up, how do the equilibrium dynamics look and is there still a balanced growth path along which all variables grow at constant rates? It is important to note that—as there is population growth—L_y and L_r both have to grow at rate n along a balanced growth path (if not at least one of L_y or L_r cannot change at a constant rate). The flow profits are still given by (13.34) which are now however time dependent as L_y is growing. Given a constant interest rate the value of a patent at time t along the balance growth path is then given by (see (13.27))

$$V(t) = \frac{\pi(t)}{r^{\star} - n} = \phi A \left(\frac{A(1-\phi)}{\psi} \right)^{\frac{1}{\phi}-1} \frac{L_y(t)}{r^{\star} - n}. \tag{13.54}$$

Again this value is now time dependent as the market size is expanding due to population growth. In an equilibrium with positive growth $V(t)$ has to equalize the marginal cost of innovating which is given by $w(t)/\eta N(t)^{\epsilon}$ under the new spillover specification. Substituting in the wage rate (which is still given by (13.36)) the optimality condition of the representative R&D firm reduces to

$$\frac{L_y(t)}{r^{\star} - n} = \frac{1}{\eta(1-\phi)} N(t)^{1-\epsilon}. \tag{13.55}$$

Can this equation be fulfilled for all t? By differentiating with respect to time we see the answer is yes if and only if

$$\frac{\dot{N}(t)}{N(t)} = \frac{n}{1-\epsilon}. \tag{13.56}$$

Then per-capita consumption, wages and per-capita output all grow at this rate $n/(1-\epsilon)$ and the interest rate $r^{\star} = \rho + \sigma n/(1-\epsilon)$ is pinned down by the Euler equation. It is straightforward to check that the transversality condition is indeed fulfilled along this path.

This model version is sometimes called a semi-endogenous growth model because the long-run growth rate is just determined by two parameters: the population growth rate n and a technology parameter ϵ. Intuitively, larger knowledge spillovers (a higher ϵ) increases the growth rate. With $\epsilon < 1$ the ideas production function runs into decreasing returns and the only way to sustain long-run growth is by pushing with an increasing number of researchers over time. Hence, without population growth, i.e., with $n = 0$, sustained long-run growth is not feasible. The long-run growth rate is efficient and there is no role for policy to influence it.[36] Hence, the model does not have rich predictions about how policy and the market structure interact with the long-run growth rate. But the model's prediction is in line with the time series data from the U.S. or the OECD countries that shows a rather stable productivity growth while population (and the number of researchers) is growing at a steady rate. Another strength of the model is that it adds demographics to the potential determinants of long-run growth (see Peters and Walsh (2023) for an application that links population growth to firm entry rates and growth).

In contrast to the model with proportional knowledge spillovers we studied before, it is important to note that this model now features transitional dynamics. Substituting the growth rate (13.56) into (13.53) and using the labor market clearing gives

$$\frac{n}{\eta(1-\epsilon)} = N(t)^{\epsilon-1}(Le^{nt} - L_y(t)), \tag{13.59}$$

which together with (13.55) specifies for given population Le^{nt} a unique $N(t)$ that supports the balanced growth. Whenever the initial N deviates from this level there will be transitional dynamics. Furthermore, as ϵ may be relatively close to one the transitional dynamics may be slow (unlike the convergence of the physical capital stock in the neoclassical growth model).[37] Policy will affect the long-run output level in this framework. Equation (13.59) implies along the balanced growth path $N(t) = (\eta(1-\epsilon)L_r(t)/n)^{\frac{1}{1-\epsilon}}$. Substituting this $N(t)$ into (13.35) allows us to express total output per worker in final goods production along the balanced growth path as

$$\frac{Y(t)}{L_y(t)} = \mathcal{A}\left(\frac{L_r(t)}{L(t)}\right)^{\frac{1}{1-\epsilon}} L(t)^{\frac{1}{1-\epsilon}}, \tag{13.60}$$

where $\mathcal{A}$ is some constant. Hence, in this model there is no scale effect on the long-run growth rate but for a given research intensity $L_r(t)/L(t)$ the model predicts higher per-capita income *levels* in larger economies. Similarly, as policy can affect research intensity it can affect the long-run *level* of income.

In the endogenous growth framework studied in this section, all long-run growth is generated by an expanding (input) variety. Though this is stark for exposition, there is indeed

[36] The planner problem is given by

$$\max_{\{L_y(t),c(t),\{x(\nu,t)\}_{\nu=0}^{N}(t),N(t)\}_{t=0}^{\infty}} \int_0^\infty e^{-(\rho-n)t}\frac{c(t)^{1-\sigma}-1}{1-\sigma}dt, \tag{13.57}$$

subject to

$$\frac{A}{1-\phi}L_y(t)^{\phi}\int_0^{N(t)} x(\nu,t)^{1-\phi}d\nu = c(t) + \int_0^{N(t)}\psi x(\nu,t)d\nu, \tag{13.58}$$

$\dot{N}(t) = \eta N(t)^{\epsilon}(Le^{nt} - L_y(t))$ and a given $N(0) > 0$. In the Appendix 13.A.7 we solve this planner problem and show that its solution also supports a balanced growth path with $\frac{\dot{N}(t)}{N(t)} = \frac{n}{1-\epsilon}$.

[37] Bloom, Jones, and Van Reenen (2020) estimate for example $\epsilon = 0.8$ in the semiconductor industry.

evidence that variety is growing and contributing to growth. Broda and Weinstein (2006) find that the U.S. import growth comes from a rising number of country-product pairs over time, across final consumer goods, intermediate goods, and business equipment. Such evidence breathes empirical life into influential models of trade in varieties such as Krugman (1980), Rivera-Batiz and Romer (1991), and Melitz (2003).

Broda and Weinstein (2010) document an expanding set of products at U.S. grocery stores. Related, Hottman, Redding, and Weinstein (2016) show that about one-third of variation in firm size and growth can be traced to the number of distinct consumer packaged goods in their portfolio. Hsieh and Rossi-Hansberg (2023) document that major retailers and service producers have opened more establishments in the U.S., especially since 2000. Given that such products and services are often nontradable, new locations offer local consumers growing choice. Hsieh and Rossi-Hansberg emphasize that such rival investments in local establishments are distinct from the nonrival corporate investments in products, services, and store layout. Garcia-Macia, Hsieh, and Klenow (2019), Klenow and Li (2021), and Peters and Walsh (2023) all estimate that between one-fifth and one-third of U.S. growth comes from rising variety.

13.5.2 Growth through quality ladders and creative destruction

The Romer model sketched out in the previous subsection endogenizes innovation, but entirely through new varieties. It does not feature any quality improvements in products after they are invented. Sales and profits of existing varieties are unaffected by the arrival of "horizontal" new varieties.

In contrast, it seems routine to observe companies improving their products over time. Think of successive generations of car models and microprocessors. The Romer model also abstracts from product and firm exit due to "creative destruction." Schumpeter's vision, formalized by Aghion and Howitt (1992), features innovation by new producers that displaces closely substitutable products of existing producers. That is, "vertical" innovation that destroys the sales and profits of existing producers. Figure 13.6 shows annual firm exit rates from the U.S. Census Business Dynamics Statistics (BDS) database for 1980–2022. Over 10% of small firms (1–4 employees) exit in a typical year, and around 2% of firms with 20 or more employees.

There is also substantial job reallocation among surviving establishments, a fact first documented by Davis, Haltiwanger, and Schuh (1998). Here, the job reallocation rate is defined as the sum of the job creation rate and the job destruction rate minus the absolute difference between the two. The job creation rate is the sum of employment gains at surviving and opening establishments from year $t-1$ to year t divided by average aggregate employment across the two years. The job destruction rate is defined similarly, only with a numerator equal to the sum of employment losses at surviving and exiting establishments from $t-1$ to t.[38] Figure 13.7 plots the annual rate of job reallocation for 5-year periods from 1980–2019 in the BDS. It averages 20–30%, but is on a downward trajectory. Job reallocation occurs mostly within narrow industries, so it reflects competition among close competitors rather than broad sectoral shifts such as from goods to services. Garcia-Macia et al. (2019) argue

[38] See Chapter 22 for the formal definitions.

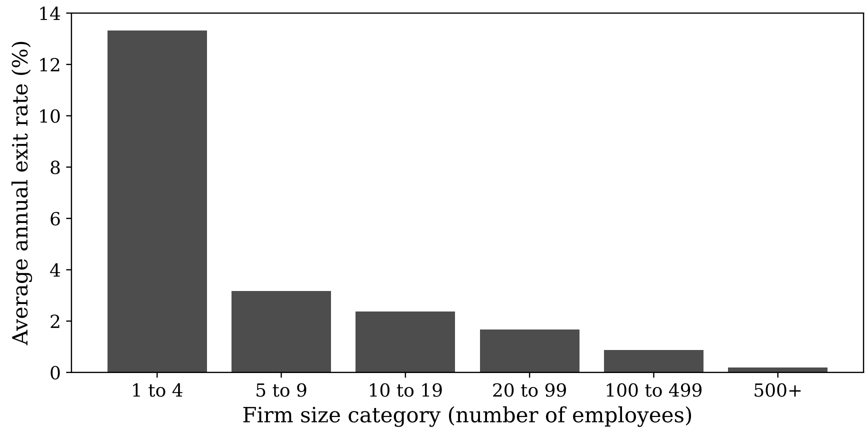

Figure 13.6: Firm exit rates in the U.S., 1980–2022

Source: U.S. Business Dynamics Statistics (BDS).

that fitting such high exit and job destruction rates requires a prominent role for creative destruction in overall growth.[39]

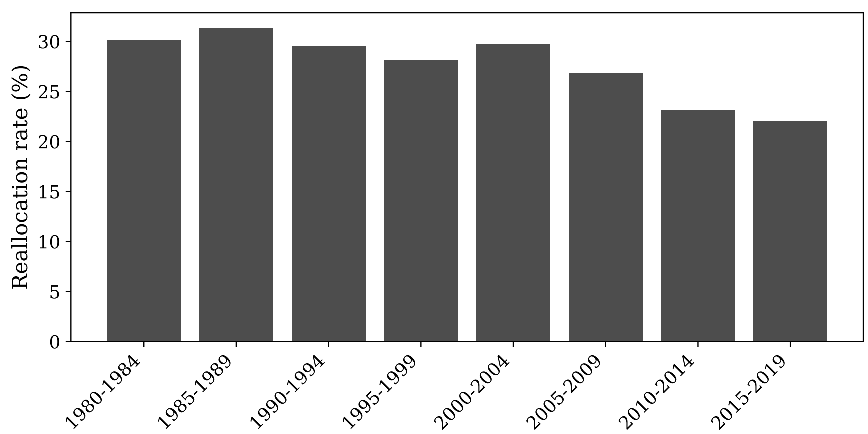

Figure 13.7: Job reallocation in the U.S., 1980–2019

Source: U.S. Business Dynamics Statistics (BDS).

Figure 13.8 compares the average size of new firms (age 0), firms age 1–5 years, 6–10 years, and 11+ years in the U.S. Young firms have 5–10 workers on average, whereas the oldest firms average 35–40 workers. Hsieh and Klenow (2014) report that this pattern stems equally from selection (smaller firms are more likely to exit) and survivor growth. Figure 13.9 plots the employment share of entrants (defined as firms aged 0 years) in the U.S. over 5-year periods from 1980–2019. Figure 13.10 plots the employment share of exiting firms in the U.S. over 5-year periods from 1980–2019. Their share fell from a bit over 3% to about 2%. The combination of falling job reallocation, entry, and exit rates has generated concern that declining "business dynamism" has contributed to slower productivity growth in the

[39]Chapter 22 presents the time-series patterns of establishment entry and exit, as well as the job creation and job destruction rates.

U.S. in recent decades. Decker, Haltiwanger, Jarmin, and Miranda (2016) and Akcigit and Ates (2023) present related evidence and a model.

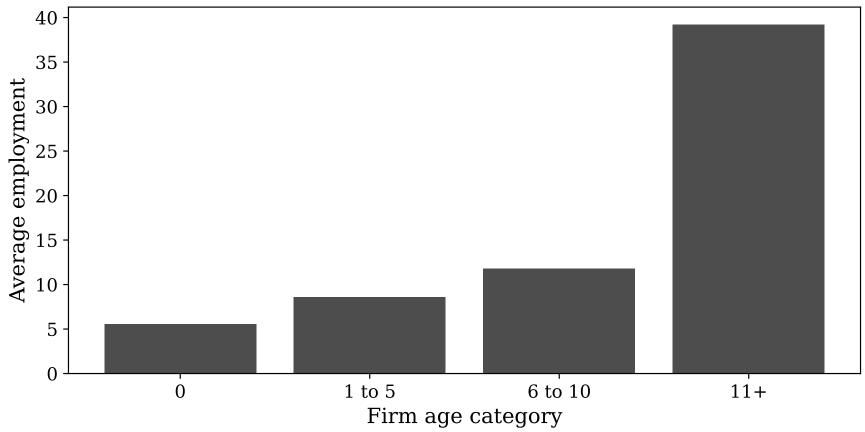

Figure 13.8: Firm size and firm age, 1988–2022

Source: U.S. Business Dynamics Statistics (BDS).

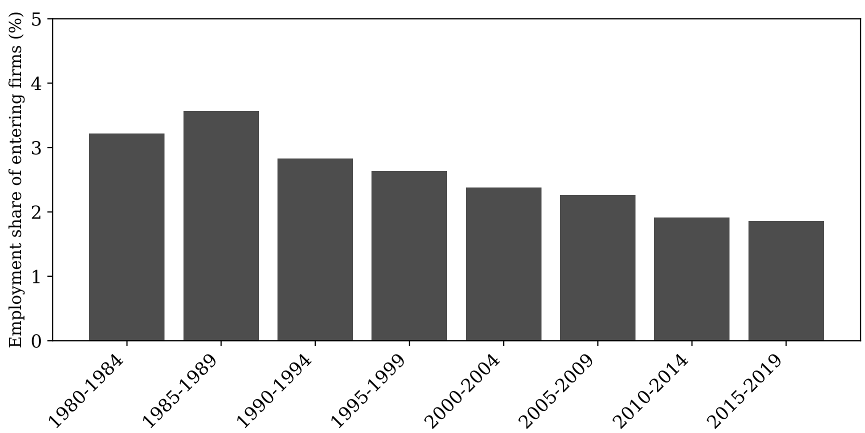

Figure 13.9: Entrant employment share in the U.S., 1980–2019

Source: U.S. Business Dynamics Statistics (BDS).

Finally, Figure 13.11 plots industry productivity growth against industry exit rates, both averaged over 5-year periods within 1988–2022. This involves 160 sectors and seven 5-year periods for a total of 1,104 sector-periods. The plot displays 20 bin-scatter points. Consistent with theories of creative destruction, industries with high exit rates tend to exhibit faster productivity growth. See Adhami (2025) for documentation of these facts and the use of entry and exit rates to infer knowledge spillovers from incumbents to entrants.

To speak to the substantial churn of market shares, even within narrow industries, we now sketch a model of growth through creative destruction.

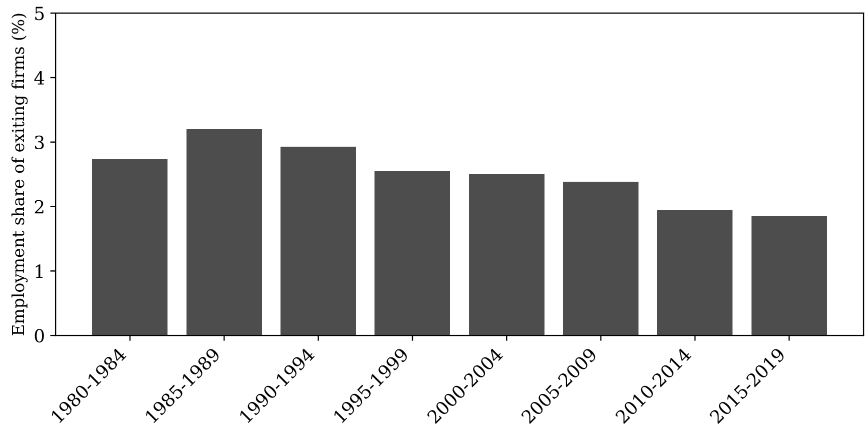

Figure 13.10: Exiting firm employment share in the U.S., 1980–2019

Source: U.S. Business Dynamics Statistics (BDS).

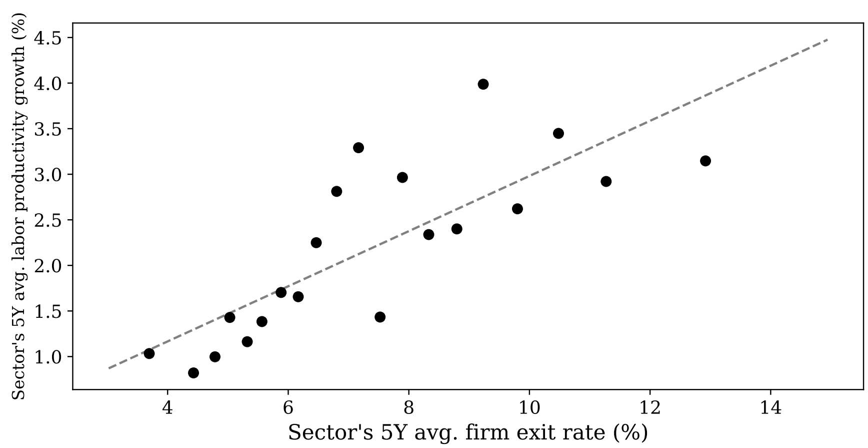

Figure 13.11: Growth and exit rates in the U.S., 1988–2022

Source: Labor productivity growth rates from the U.S. Bureau of Labor Statistics and exit rates from the U.S. Business Dynamics Statistics (BDS).

Quality Ladders model Final good producers use a fixed measure one of intermediate good varieties with evolving quality levels $q(\nu, t)$:

$$Y(t) = \frac{A}{1-\phi} L(t)^{\phi} \int_0^1 q(\nu, t) x(\nu, t)^{1-\phi} d\nu, \tag{13.61}$$

where $0 < \phi < 1$ so that there is diminishing returns to any one intermediate line ν. A fixed amount of labor $L(t) = L$ can be used to produce final goods.

Output can be used for consumption, intermediate goods, and research:

$$Y(t) = C(t) + X(t) + Z(t).$$

This is sometimes called a "lab equipment" model because intermediate goods (which for simplicity depreciate fully here rather than being durable equipment) are used to make research goods. Unlike in the Romer model, here there is no knowledge spillover from past innovation to current research. Intermediate goods, in turn, are produced according to

$$X(t) = \int_0^1 \psi q(\nu, t) x(\nu, t) d\nu,$$

so that one unit of intermediate good ν requires $\psi \cdot q(\nu, t)$ units of the final good.

Aggregate research effort is merely the sum of effort directed at improving each intermediate good's quality

$$Z(t) = \int_0^1 Z(\nu, t)\, d\nu.$$

Quality improves on a product line in discrete increments in proportion $\lambda > 1$:

$$q(\nu, t) = \lambda^{m(\nu,t)} q(\nu, 0) \quad \forall \nu, t.$$

See Figure 13.12. Here $m(\nu, t)$ is the cumulative number of quality steps taken on line ν between time 0 and t. The flow rate of innovations on product line ν at moment t is proportional to research effort on that line:

$$z(\nu, t) \equiv \eta \frac{Z(\nu, t)}{q(\nu, t)},$$

where $\eta > 0$. Note that it is more difficult to innovate on a line the higher the quality attained on that line.

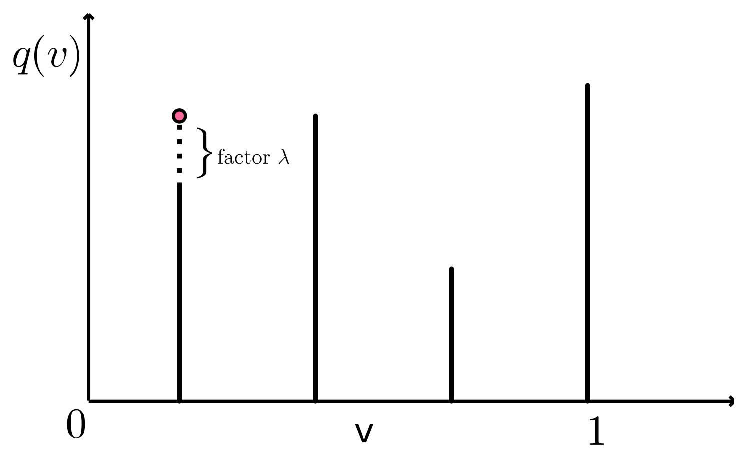

Figure 13.12: Quality ladders

Finally, the representative agent's utility function takes the standard CRRA form:

$$U(0) = \int_0^\infty e^{-\rho t} \frac{C(t)^{1-\sigma} - 1}{1 - \sigma} dt$$

Decentralized equilibrium The final goods sector is competitive and producers maximize current profits taking all prices as given (the final good whose price is normalized to one, the intermediate good prices, and the wage):

$$\Pi(t) = \frac{A}{1-\phi} L(t)^{\phi} \int_0^1 q(\nu,t)x(\nu,t)^{1-\phi} d\nu - \int_0^1 p(\nu,t)x(\nu,t)\, d\nu - w(t)L(t).$$

The first-order conditions for this problem imply

$$w(t) = \phi \frac{Y(t)}{L}, \tag{13.62}$$

and

$$x(\nu,t) = \left[\frac{A\, q(\nu,t)}{p(\nu,t)}\right]^{1/\phi} L, \tag{13.63}$$

where we already used labor market clearing $L(t) = L$. The wage is equal to the marginal product of labor, and intermediate input demand is inversely related to a variety's *quality-adjusted* price. Intermediate demand is increasing in production labor, a complementary input.

Intermediate good producers are monopolists over a particular quality level. We assume innovations are "drastic" in that

$$\lambda \geq \left(\frac{1}{1-\phi}\right)^{\frac{1-\phi}{\phi}}, \tag{13.64}$$

so that producers are not constrained by competition from firms who can produce at lower rungs on the quality ladder for their variety. Instead, their closest competitors are other varieties.[40] With a continuum of varieties they are monopolistic competitors. They maximize current profits:

$$\pi(\nu,t) = p(\nu,t)x(\nu,t) - \psi q(\nu,t)x(\nu,t)$$

Intermediate good monopolists take downstream demand as given by (13.63) and choose prices and quantities. This results in the profit-maximizing price

$$p(\nu,t) = \frac{\psi}{1-\phi} q(\nu,t).$$

The marginal cost is $\psi \cdot q(\nu,t)$ and the markup is $1/\phi/(1/\phi - 1)$, as $1/\phi$ is the elasticity of substitution across intermediate good varieties in final goods production. Using the profit-maximizing price, input demand becomes

$$x(\nu,t) = \left(\frac{A(1-\phi)}{\psi}\right)^{1/\phi} L, \quad \forall\, \nu, t.$$

and in turn maximized profits are

$$\pi(\nu,t) = \phi A \left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1-\phi}{\phi}} q(\nu,t)L, \quad \forall\, \nu, t. \tag{13.65}$$

[40]See the Appendix 13.A.9 which derives condition (13.64) and illustrates how limit pricing might prevail in the case of "incremental" innovations with step sizes below the monopoly markup if condition (13.64) does not hold.

We can use the intermediate good quantities x (which do not depend on v or t) to find an expression for aggregate output:

$$Y(t) \;=\; \frac{A}{1-\phi}\left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1-\phi}{\phi}} Q(t)L,$$

where $Q(t)$ is defined as the average quality of intermediate goods

$$Q(t) \equiv \int_0^1 q(\nu,t)\,d\nu.$$

In turn, the amount of aggregate output devoted to intermediate goods is

$$X(t) = \int_0^1 \psi q(\nu,t)x(\nu,t)d\nu = \psi^{1-1/\phi}(A(1-\phi))^{1/\phi}Q(t)L.$$

Meanwhile, there is free entry into the research sector. Research firms choose how much research to devote to each variety:

$$\max_{Z(\nu,t)} \frac{\eta Z(\nu,t)}{q(\nu,t)}\lambda V(\nu,t) - Z(\nu,t),$$

where $V(\nu,t)$ is the expected present discounted value of profits for the current leader. The innovator takes over the entire profit stream (creative destruction), and earns profits that are scaled up by $\lambda > 1$ relative to that of the previous leader because profits are proportional to the level of quality in (13.65). The first-order condition for researchers is

$$\frac{\eta}{q(\nu,t)}\lambda V(\nu,t) = 1.$$

Note that, if there is positive research on all varieties, then the ratio of the value of a variety to its quality level is the same for all varieties.

Would incumbents put effort into improving their own products to stave off creative destruction? If they did so, they would only reap the *increment* to their value, and this would not cover the cost of research to them:

$$\frac{\eta}{q(\nu,t)}(\lambda-1)V(\nu,t) < \frac{\eta}{q(\nu,t)}\lambda V(\nu,t) = 1.$$

This is called the **Arrow Replacement Effect**, which says that incumbents have less to gain from innovation than entrants because $\lambda - 1 < \lambda$.[41]

The expected present discounted value of profits from being the current leader of a quality ladder is:

$$V(\nu,t) = \int_t^\infty \left[e^{-\int_t^s r(s')ds'} \cdot e^{-\int_t^s z(\nu,s')ds'}\right]\pi(\nu,s)ds$$

[41] In order to explain why incumbents very much do improve their own products in the world, one could alter the model to have incumbents enjoy lower costs of innovating. If such costs are convex, however, then there could be both incumbent innovation on their own products and creative destruction of competitor products in equilibrium.

where $z(\nu, t)$ is the instantaneous rate of innovations on line ν, $z(\nu, t) \triangleq \eta \cdot Z(\nu, t)/q(\nu, t)$. The stream of profits is discounted by the real interest rate and the probability of creative destruction.

Finally, a representative household maximizes the presented discount value of its utility from consumption:

$$\max_{\{C(t), \mathcal{A}(t)\}_{t=0}^{\infty}} \int_0^{\infty} e^{-\rho t} \frac{C(t)^{1-\sigma} - 1}{1-\sigma} dt$$

subject to

$$\dot{\mathcal{A}}(t) = r(t)\mathcal{A}(t) + w(t)L - c(t), \, \forall t,$$

and a no-Ponzi game condition. Note that assets are simply the combined value of all intermediate good monopolists, who are owned by the household:

$$\mathcal{A}(t) \equiv \int_0^1 V(\nu, t) \, d\nu.$$

The current value Hamiltonian for the household is therefore

$$\mathcal{H}(c, a, \mu) = \frac{c^{1-\sigma} - 1}{1-\sigma} + \mu \left(r \cdot a + w - c \right)$$

where lower case c and a denote per-capita consumption and assets, respectively. The solution to this problem yields the standard Euler equation

$$\frac{\dot{c}}{c} = \frac{r - \rho}{\sigma},$$

and the usual transversality condition is

$$\lim_{T \to \infty} a(T) \exp \left[\int_0^T -r(s) ds \right] = 0.$$

As with the Romer model, one can show that the unique decentralized equilibrium of the quality ladder model entails no transition dynamics. That is, the economy is always on a BGP with growth rate

$$g^\star = \frac{\dot{Y}}{Y} = \frac{\dot{C}}{C} = \frac{\dot{X}}{X} = \frac{\dot{Q}}{Q}.$$

Note that, because L is constant, $c \equiv C/L$ grows at the same rate as C does. The Euler equation then gives us one equation in the two unknowns

$$r^\star = \sigma g^\star + \rho,$$

where $r^\star$ is the BGP real interest rate. And free entry into research gives us

$$\frac{V(\nu, t)}{q(\nu, t)} = \frac{1}{\lambda \eta}, \quad \forall \, \nu, t.$$

To obtain a constant value of each variety relative to its quality, we must have $z(\nu,t) = z^\star, \ \forall\, \nu, t$. This in turn implies

$$\frac{V(\nu,t)}{q(\nu,t)} = \frac{\phi A \left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1-\phi}{\phi}} L}{r^\star + z^\star} = \frac{1}{\lambda\,\eta}$$

$$\Rightarrow \quad r^\star + z^\star = \lambda\,\eta\,\phi A \left(\frac{A(1-\phi)}{\psi}\right)^{\frac{1-\phi}{\phi}} L.$$

To get an expression for average quality growth, it is useful to first difference average quality:

$$Q(t) \equiv \int_0^1 q(\nu,t)d\nu \rightarrow \dot{Q}(t) = \int_0^1 \dot{q}(\nu,t)d\nu.$$

Then

$$\frac{\dot{Q}(t)}{Q(t)} = \int_0^1 \frac{\dot{q}(\nu,t)}{Q(t)}d\nu = \int_0^1 \frac{\dot{q}(\nu,t)}{q(\nu,t)} \cdot \frac{q(\nu,t)}{Q(t)}d\nu = \int_0^1 \frac{\dot{q}(\nu,t)}{q(\nu,t)}dv \cdot \underbrace{\int_0^1 \frac{q(\nu,t)}{Q(t)}d\nu}_{=1},$$

where the last equality follows from the independence of quality growth across varieties. By Uhlig (1996), the integral of a continuum of iid random variables with finite variance is equal to the expectation of the variables:

$$\frac{\dot{Q}(t)}{Q(t)} = \int_0^1 \frac{\dot{q}(\nu,t)}{q(\nu,t)}d\nu = E\left[\frac{\dot{q}(\nu,t)}{q(\nu,t)}\right] = (\lambda-1)z^\star.$$

Thus average quality $Q(t)$ grows at a smooth rate despite the stochastic evolution of the individual $q(\nu,t)$s:

$$g^\star = (\lambda-1)z^\star.$$

Combining the three BGP equations yields

$$g^\star = \frac{\lambda\eta\phi A\,(A(1-\phi)/\psi)^{\frac{1-\phi}{\phi}}\,L - \rho}{\sigma + 1/(\lambda-1)}, \tag{13.66}$$

$$r^\star = \rho + \frac{\lambda\eta\phi A\,(A(1-\phi)/\psi)^{\frac{1-\phi}{\phi}}\,L - \rho}{1 + 1/(\sigma(\lambda-1))},$$

and

$$z^\star = \frac{\lambda\eta\phi A\,(A(1-\phi)/\psi)^{\frac{1-\phi}{\phi}}\,L - \rho}{1 + (\lambda-1)\sigma}.$$

Note that the growth rate is increasing in the innovation step size λ, the R&D efficiency parameter η, and the scale of the workforce L. The growth rate is also decreasing in ρ, σ, and ψ. All of these comparative statics are intuitive. For example, a higher population size increases profits for each variety, thereby encouraging more research and faster growth. There is the same strong scale effect as in the Romer model described above. Just as with that model, one could entertain diminishing returns in innovation to make growth semi-endogenous, so that sustaining growth requires population growth.

Appendix 13.A.10 provides the social planner's solution to the quality ladder model, which yields the growth rate

$$g^{SP} = \frac{(\lambda - 1)\eta\phi A(1-\phi)^{-1} (A/\psi)^{\frac{1-\phi}{\phi}} L - \rho}{\sigma}.$$

The social planner (SP) growth rate differs from the decentralized equilibrium (DE) growth rate $g^{\star} = g^{\mathrm{DE}}$ for three reasons. First, the planner sees only the incremental gain $\lambda - 1$ to innovation, whereas private innovators reap λ. This business stealing is a force for $g^{\mathrm{SP}} < g^{\mathrm{DE}}$. Second, the planner uses intermediates more intensively than intermediate goods monopolists (who discourage downstream purchases by charging a markup over marginal cost). This is a force for $g^{\mathrm{SP}} > g^{\mathrm{DE}}$ because the planner values a new variety more given it will be used more. Third, the planner internalizes the positive knowledge externalities from past to future innovation. Equivalently, the planner sees an innovation as lasting forever in that all future innovations will build upon it. Private profits, in contrast, are truncated by creative destruction. This is a force for $g^{\mathrm{SP}} > g^{\mathrm{DE}}$. To recap, the SP growth rate could be higher or lower than the DE growth rate. That said, calibrations typically find the SP growth rate to be far higher than the DE growth rate. See Jones and Williams (1998, 2000) for early calculations, and Atkeson and Burstein (2019) for a more recent analysis.

13.6 Conclusion

The topic of this chapter—economic growth—is of the utmost importance for welfare. The models and empirical results presented here have shaped our discipline on an unprecedented scale. Virtually all macroeconomic models have a neoclassical core as a backbone irrespective of whether they aim to study business cycles, long-run trends, or topics of international macroeconomics.

The proximate causes of cross-country income differences that follow from a development accounting exercise are crucial in directing research on development economics. The literature has aimed to shed light on the sizable residual called TFP, be it with randomized control trials, the sophisticated exploitation of quasi-natural experiments, or more structural approaches that rely on a combination of theory and (often microeconomic) data.

The literature on endogenous growth is vibrant. The combination of theoretical building blocks outlined in this chapter connect the aggregate growth rate to firm dynamics (entry, exit, and the life-cycle firm growth) and market structure (concentration and market power). Markups play a key role both as a source of misallocation as well as an incentive for entry and innovation. Chapter 22 presents a model analysis in that direction. Recent advances link the literature on growth with the one on industrial organization and derive subtle policy recommendations for promoting research and allocative efficiency. Mapping such theories to available microeconomic data on firms and products fell on particularly fertile ground. This provides exciting research opportunities for the years to come.