A review of innovation-based growth models

Christian Groth

Department of Economics and EPRU*, University of Copenhagen,
Oester Farimagsgade 5, Building 26, DK-1353, Copenhagen K, Denmark
chr.groth@econ.ku.dk

This version: April 27, 2010.

Abstract

To be written.

——

I am grateful to Carl-Johan Dalgaard, Claus Thustrup Kreiner, Poul Schou, Sjak Smulders, and Holger Strulik for useful discussions and helpful comments.

——

* The activities of EPRU (Economic Policy Research Unit) are supported by a grant from The Danish National Research Foundation.

1 Introduction

This paper reviews the development of innovation-based growth theory with an emphasis on the recurrent issue of debatable linearity restrictions. This issue is important for robustness of the conclusions and for the potential of growth policy. In the discussion I also consider how the relationship between growth and scarce natural resources has been addressed by innovation-based growth theory.

Economists agree that ideas are different from most economic goods in that they are *nonrival*: their usage by one agent does not in itself limit their usage by other agents. This is likely to generate increasing returns to scale when knowledge is included in the total set of inputs. Yet, there is scope for differing assumptions about the size of the returns to scale with respect to only *producible* inputs, including knowledge.

First-generation endogenous growth theory, as put forward in Romer (1996, 1987, 1990), Lucas (1988), and Aghion and Howitt (1992), suspended the neoclassical presupposition of diminishing returns to producible inputs and replaced it with the assumption of exactly constant returns to producible inputs. This has far-reaching implications. It is possible for economic policy not only to lift the level of the path along which growth occurs, but also to tilt the path. This nourished many economists' belief that knowledge creation is likely to overcome the limits to growth implied by limited natural resources (Smulders 1995).

Without implicating the limits-to-growth debate, other economists (Solow 1994, Jones 1995a, 1995b) argued that the presumption of non-diminishing returns to producible inputs lacks empirical support as well as theoretical plausibility. This criticism did not deny that the systematic incorporation of the creation of new "ideas" (with their distinctive properties compared to other economic goods) into dynamic general equilibrium models with imperfect competition opened up for valuable new insights, whatever the returns to producible inputs presumed. But the motivation for addressing the returns to scale question was the concern about robustness of the results and the risk of counterfactual theoretical implications.

The robustness issue arises because an assumption of constant returns (or just asymptotically constant returns) to producible inputs has a knife-edge character in the sense that a hair's breadth deviation from the assumption implies completely different dynamics, depending on the sign of the deviation. A counterfactual implication is for example the "strong scale effect" that arose in first-generation innovation-based growth models. When a non-rival good – like technical knowledge – is one of the producible inputs, and non-diminishing returns to producible inputs are present and not offset by other factors, then a counterfactual scale effect on growth arises: the larger the economy is, ceteris paribus, the larger the long-run per-capita growth rate tends to be.[1] As a consequence, sustained growth in population should lead to a forever rising per-capita growth rate, a prediction not supported by the evidence for the industrialized world over the last century.

Fortunately, the whole debate about these matters gave rise to *second-generation* endogenous growth theory (see below) and further developments helping us to climb up the "quality ladder" of growth models. The attempts at combining robustness and empirical relevance has lead to better understanding of an "increasing variety" of growth mechanisms. Yet, the controversy over specific linearity assumptions and constant vs. diminishing returns to producible inputs did not stop.

It is well-known that any model which is capable of rendering balanced growth must contain at least one *linear* differential equation (Romer 1995, Growiec 2006). Sometimes, criticism of knife-edge conditions in growth theory is dismissed on the basis of that observation. The counter argument has of course been that not all linearity assumptions are equally defendable. Moreover, taking non-renewable natural resources into account illustrates (cf. Section 6 below) that the linearity required for balanced growth may be delivered by a differential equation which merely reflects accounting definitions and is therefore not controversial. In fact, incorporating non-renewable resources in the discussion puts several of the issues in a new light: no problematic knife-edge condition is any longer needed for fully endogenous balanced growth; there is no scale effect on growth; population growth can easily be integrated; the scope for various policy tools changes; and, as expected, more narrow limits to growth arise.

The rest of the paper discusses these issues within a unified framework for the presentation of the different "generations" of innovation-based growth models. The next section provides the necessary background: the first-generation models. In Section 3, the "Jones critique" is presented along with the semi-endogenous growth alternative (or "moderation") as well as the different responses to the Jones critique. Section 4 portrays the second-generation models which consider innovations along *two* dimensions. Section

[1] I follow the terminology in Jones (2005) and use "strong scale effect" as synonymous with "scale effect on growth". In contrast, "weak scale effect" refers to "scale effect on levels", a less controversial phenomenon arising naturally from the non-rival character of technical knowledge.

5 introduces recent developments with an emphasis on the attempts at endogenizing both the rate and *direction* of technical change. Repercussions of the presence of essential non-renewable resources are depicted in Section 6. Finally, Section 7 concludes.[2]

2 First-generation models

It is common to divide the models of the endogenous growth literature − also called new growth theory − into two broad classes: accumulation-based models and innovation-based models. The first class of models is based on the idea that the combination of physical capital and human capital accumulation may be enough to sustain long-run productivity growth (Lucas 1988, Rebelo 1991). The second class of models, which will be my focus here, attempts to explain how technological change comes about and how it shapes economic growth.

The origin of the innovation-based growth models goes back to Romer (1986, 1987, 1990), where growth is driven by specialization and increasing division of labor.[3] That is, here the focus is on *horizontal innovations:* the invention of new intermediate or final goods giving rise to new branches of trade. The invention of micro-processors is an example. Shortly after the Romer papers came out, Aghion and Howitt (1992) and others proposed theories in which growth is driven by *vertical innovations.* This strand of endogenous growth theory concentrates on the invention of better qualities of existing products and better production methods which make previous qualities and methods obsolete. Improvement in the performance of microprocessors provides an example. The two strands of models are often called "increasing variety" models and "increasing quality" models (or "quality ladder" models), respectively. We begin with an account of the increasing variety models.

2.1 Horizontal innovations

The following is a simplified version of Romer (1990). There are two production sectors. Sector 1 is the manufacturing sector which supplies "basic goods" under perfect

[2] In several respects, this review is inspired by Jones (1999) and Jones (2005). However, we discuss the non-robustness arising from arbitrary parameter values and parameter links in a broader context, including directed technical change, non-renewable resources, and limits-to-growth considerations.

[3] There are forerunners though: Arrow (1962a), Phelps (1966), Nordhaus (1979) and Shell (1973) are examples.

competition. The representative firm in the sector produces

$$Y = \left(\sum_{i=1}^{N} x_i^{\alpha} \right) L_Y^{1-\alpha}, \qquad 0 < \alpha < 1, \tag{1}$$

where x_i = input of capital good variety i, N is the number of different varieties of capital goods, and L_Y is labor input. The output of basic goods is used for consumption, C, and investment in "raw capital". The stock of raw capital, K, grows according to

$$\dot{K} = Y - C - \delta K, \qquad \delta \geq 0. \tag{2}$$

Here δ is a constant physical capital depreciation rate. A dot over a variable denotes the time derivative.

In Sector 2, the innovative sector, two activities take place. Firstly, there is investment in R&D in the sense that labor, L_R, is applied in research to invent technical designs for new capital varieties, i.e., new kinds of specialized capital goods. The sum of L_Y and L_R makes up the total labor force (= population),

$$L_Y + L_R = L, \tag{3}$$

where L is assumed constant (labor supply inelastic). There is free entry to the R&D activity. The number of new varieties invented per time unit is assumed proportional to R&D input. Ignoring indivisibilities, we have

$$\dot{N} = \tilde{\mu} L_R. \tag{4}$$

The individual research lab, which is "small" relative to the economy as a whole, takes R&D productivity, $\tilde{\mu}$, as given. At the economy-wide level, however, this productivity is an increasing function of the stock of technical knowledge in society, proxied by N.[4] Here Romer (1990) assumes linearity:

$$\tilde{\mu} = \mu N, \qquad \mu > 0, \tag{5}$$

where μ is a constant.

[4]Specific knowledge may be partially *excludable* in the sense that patents, concealment etc. can for a while exclude other firms from the commercial use of a given innovation. Yet the general engineering principles behind the innovation are likely to diffuse rather quickly and add to the stock of common technical knowledge in society. On the basis of - and inspired by - this knowledge stock, further innovations are made.

Secondly, once the technical design (blueprint) of a new variety has been invented, the inventor takes out (free of charge) an infinitely-lived patent and starts supplying the services of the new capital good under conditions of monopolistic competition. Given the invented technical design, say design no. i, the inventor can effortlessly transform any number of raw capital goods into the same number of specialized capital goods of type i simply by pressing a button on a computer, thereby activating a computer code. That is,

it takes x_i units of raw capital to supply x_i units of the specialized capital good i.

In view of the symmetric cost structure and the concavity in (1), profit maximizing firms in the basic-goods sector choose $x_i = x = K/N$ for all i, so that the aggregate production function becomes

$$Y = K^{\alpha}(NL_Y)^{1-\alpha}. \tag{6}$$

We see that increased N implies increased productivity: variety is productive. This is how Romer formalizes the idea that specialization and division of labor increase productivity. A further key feature of endogenous growth theory is apparent from (6) and (4) (combined with (5)). *Total* knowledge, N, enters both in the production function for Y and that for $\dot{N}$. This is because ideas - sets of instructions - are *non-rival:* their usage by one agent does not in itself limit their usage by other agents. This is in contrast to the *rival* goods: capital and labor. For example, a given unit of labor or capital can be used no more than one place at a time. Hence, only a fraction of the labor force enters the Y sector, the remaining fraction entering the innovative sector.

At the societal level there are two allocation problems: how to divide the labor force into L_Y and L_R and how to divide Y into consumption and investment. Adding perfect competition in the labor market and a specification of the household sector, including preferences, the model can be solved.

For any positive variable x let $g_x \equiv \dot{x}/x$ and let $y \equiv Y/L$. Then, along a balanced growth path[5] we have

$$g_y = g_N = \mu s_R L, \qquad \text{where } s_R \equiv \frac{L_R}{L}. \tag{7}$$

Here, the share of the labor force employed in R&D, s_R, will depend on parameters such as α, μ, and those describing the household sector (here left out for brevity). When parameter values are such that $0 < s_R < 1$, we get:

[5] A balanced growth path is defined as a path along which the quantities Y, C, K, and N change at constant proportionate rates (not necessarily positive).

(i) There is *endogenous growth* in the sense that the source of positive long-run growth is some internal mechanism in the model (in contrast to exogenous technology growth as in the Solow model). More specifically, growth is *fully endogenous* in the sense that the long-run growth rate in per-capita output is positive without the support of growth in any exogenous factor.[6] The key to this feature is the assumption of non-diminishing returns to the producible input, N, in the "growth engine": $\dot{N} = \mu N L_R$.[7]

(ii) Via influencing incentives, policy can affect s_R (e.g., by a research subsidy) and thereby the long-run growth rate.

(iii) Under laissez-faire, the market economy always does too little R&D.

It is generally recognized that at least result (iii) is not robust. This result arises due to two positive externalities, the intertemporal spillover in knowledge creation, apparent from (4) combined with (5), and the surplus appropriability problem, that is, the inability of the innovator to capture the whole contribution to output by the innovation. As Benassy (1998) and Groot and Nahuis (1998) argue, however, there is an arbitrary link in the Romer model between gains to specialization (i.e., the exponent, $1 - \alpha$, to N in (6)) and the output elasticity wrt. capital, α. With a more general specification of (1), this link is disentangled and this may open up for a negative externality to appear, a reminiscence of "creative destruction". If this negative externality is strong enough (gains to specialization low enough), then "too much R&D" in the market economy is possible. Further, Alvarez-Pelaez and Groth (2005) argue that there is yet another arbitrary parameter link, that between market power, $1/\alpha$, in the supply of specialized capital goods and the output elasticity wrt. capital, α. When this link is removed, Romer's original "too little R&D" conclusion is in fact vindicated as the empirically relevant case.[8] In addition, this parameter separation is needed in order not to blur the growth enhancing

[6] An alternative name for the same is "strictly endogenous" growth. Within the broad class of "endogenous growth" models the complementary sub-class is that of "semi-endogenous" growth models, where sustained (exponential) growth can not be maintained without the support of growth in some exogenous factor (usually the labor force).

[7] The *growth engine* of a model is defined as the set of sectors that produce goods used as input within the set.

[8] The generalized production function in the basic-goods sector is $Y = N^{\omega} X^{\alpha} L_Y^{1-\alpha}$, where $\omega > 0$, and X is a CES aggregate of specialized capital goods, i.e., $X = N(N^{-1} \sum_{i=1}^{N} x_i^{\varepsilon})^{1/\varepsilon}$, $0 < \varepsilon < 1$. With $x_i = x$ $= K/N$ for all i we get $X = Nx = K$. So at the aggregate level the difference compared to (6) is only that gains to specialization are now represented by the independent parameter ω, and ε is the independent parameter measuring the substitutability between the specialized capital goods. Thus $1/\varepsilon$ measures the market power of the suppliers of specialized capital goods.

effect of increased monopoly power by offsetting effects from decreased output elasticity wrt. capital.

Even the results (i) and (ii) turn out to be non-robust, but before going into that, we shall take a look at the "increasing quality" or "quality ladder" models.

2.2 Vertical innovations

We consider a simplified version of Aghion and Howitt (1992, 1998) comparable to the above version of Romer (1990). Again, there are two sectors (broadly defined): Sector 1, the basic-goods sector, and Sector 2, the innovative sector.

The representative firm in Sector 1 operates under perfect competition and produces

$$Y = \left(\sum_{i=1}^{\bar{N}} Q_i x_i^{\alpha} \right) L_Y^{1-\alpha}, \qquad 0 < \alpha < 1, \tag{8}$$

where x_i = input of capital good variety i, and Q_i is (essentially) the productivity ("quality") of the latest design of capital good i.[9] There is a *fixed* number, $\bar{N}$, of different capital good varieties. The output of basic goods is used for consumption, C, and investment in "raw capital" in the same way as in (2).

Sector 2 consists of $\bar{N}$ product lines, one for each capital good variety, $i = 1, 2, \ldots, \bar{N}$. In each product line two activities take place. The first is investment in R&D. In product line i, L_{Ri} units of labor per time unit is applied in research to invent an improved quality of capital good type i. There is free entry to research. Successful outcomes, i.e., innovations, are assumed to arrive randomly with a Poisson arrival rate, λ_i, proportional to current R&D labor in the product line, i.e.,

$$\lambda_i = \tilde{\mu} L_{Ri}, \qquad \tilde{\mu} > 0.$$

An innovation is an invention of a new design for capital good i, raising its productivity by a constant factor $\gamma > 1$. The τ_i'th innovation in product line i thus raises the quality to

$$Q_i = \gamma^{\tau_i}, \qquad \gamma > 1, \;\; \tau_i = 1, 2, \ldots.$$

Fig. 1 illustrates these "quality ladders" in the different product lines.[10]

[9] Strictly speaking, Q_i, is the productivity of the transformation, x_i^{α}, of x_i units of the specialized capital good i.

[10] In the quality ladder model of Grossman and Helpman (1991) each innovation consists of the invention of a better quality of *consumption* good i.

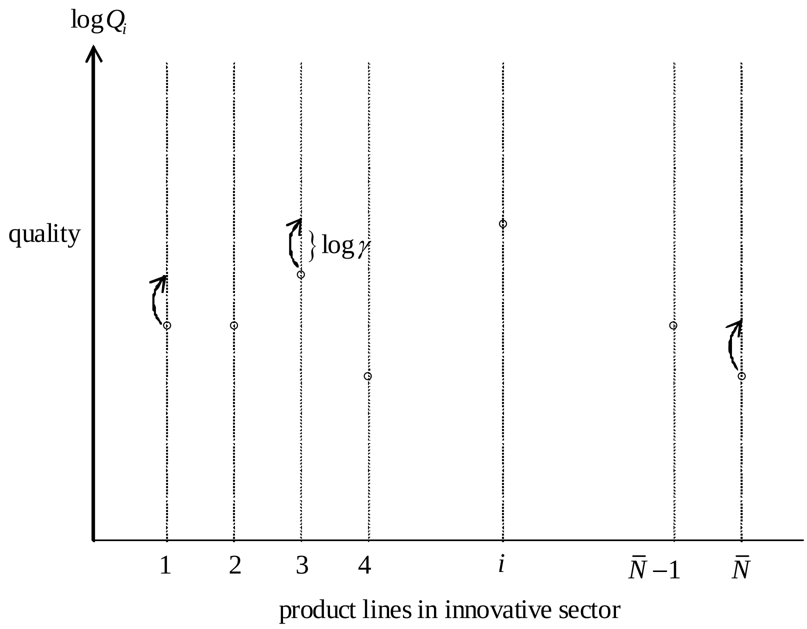

Figure 1: Quality ladders.

The second activity in each product line of the innovative sector is to supply the services of the most recent quality of the capital good in question under conditions of monopolistic competition. The innovator takes out (free of charge) an infinitely-lived patent on the commercial use of the new design. Because of its superiority, the new design outperforms previous designs; if necessary, the innovator uses limit pricing, rather than unconstrained monopoly pricing, to eliminate the previous "leading-edge" producer. This idea, that innovations render old products obsolete, was coined by Joseph Schumpeter (1912) in the notion of "creative destruction".

Leaving the sunk costs of making the innovation out, the only cost associated with supplying one unit of capital good i in its currently best quality is that:

it takes Q_i units of raw capital to supply 1 unit of capital good i in quality Q_i.

So raw capital in the amount K_i can effortlessly be converted into the current quality of capital good i according to

$$x_i = \frac{K_i}{Q_i},$$

simply by pressing a button on a computer. Succeeding vintages of the specialized capital good are thus increasingly capital intensive.

Under cost minimization, the symmetry in the setup leads to $x_i = K_i/Q_i = x$ for all

i and therefore

$$x = \frac{\sum_i K_i}{\sum_i Q_i} = \frac{K}{\bar{N}Q},$$

where Q is average quality across the different capital good varieties:

$$Q \equiv \frac{\sum_{i=1}^{\bar{N}} Q_i}{\bar{N}}.$$

Hence, by (8),

$$Y = \left(\sum_{i=1}^{\bar{N}} Q_i \left(\frac{K}{\bar{N}Q} \right)^{\alpha} \right) L_Y^{1-\alpha} = K^{\alpha}(\bar{N}QL_Y)^{1-\alpha}. \tag{9}$$

Given that the Poisson arrival rate of quality improvements in product line i is $\lambda_i = \tilde{\mu}L_{Ri}$ and given stochastic independence across product lines, the expected aggregate number of quality improvements per time unit is $\tilde{\mu}\sum_{i=1}^{\bar{N}} L_{Ri} \equiv \tilde{\mu}L_R$. Appealing to the law of large numbers, we take this to be the actual number of quality improvements per time unit. In view of the symmetric structure, if the number of product lines, $\bar{N}$, is large, the aggregate outcome of R&D can be approximately described by

$$\dot{Q} = (\frac{\tilde{\mu}}{\bar{N}} \log \gamma) QL_R \equiv \mu QL_R, \tag{10}$$

which is analogous to (4) combined with (5).

Along a balanced growth path, again

$$g_y = g_Q = \mu s_R L, \tag{11}$$

but now growth is driven by the increasing quality of a fixed spectrum of inputs.

Results (i) and (ii) from the previous section go through, but result (iii) is modified. Indeed, there are now three market failures, two of which tend to generate too little R&D in the market economy (the positive intertemporal spillover in research and the surplus appropriability problem). But a third market distortion works in the opposite direction, namely the "business stealing" effect, i.e., the failure of the innovator to internalize the loss to the previous innovator caused by inventing a better quality.[11] Because of this effect, under laissez-faire the market economy *may* generate too *much* research. Empirically, this does not seem to be the case, rather the contrary seems true (Jones and Williams 1998).

[11] On the one hand, the incorporation of business steeling is one of the strengths of the quality ladder model. On the other hand, a weakness - from an empirical point of view - of the first-generation versions of the model is that *all* R&D is done by outsiders. This is because outsiders have (by assumption) immediate access to the knowledge contained in the incumbent technology and can therefore straight away start their search for the next quality jump; and the net value of this next innovation is higher to them than to the incumbent who looses the profit generated by the previous innovation. For versions where also the incumbent does research, see, e.g., Aghion and Griffith (2005).

3 Non-robustness and the semi-endogenous "moderation"

Two features of the common conclusion $g_y = \mu s_R L$ stand out:

(a) There is a scale effect on growth: $\partial g_y/\partial L > 0$.[12]

(b) By affecting incentives, policy can affect s_R and thereby the long-run growth rate.

The feature (a) is discomforting, because the industrialized world has over the last century experienced sustained growth in L but not in g_y. The feature (b) makes economic policy very powerful, because increasing a growth rate just a little but permanent bit, one increases future productivity dramatically because of the compounding effect.

3.1 The Jones critique

In two important papers, Charles Jones (1995a, 1995b) claims:

(i) Both features are rejected by time-series evidence for the industrialized world.

(ii) Both features are theoretically non-robust (i.e., they are very sensitive to small changes in parameter values).

The empirical point, (i), is supported by, e.g., Evans (1996), Romero-Avila (2006), and Lau (2008), although partially challenged by McGrattan (1998), Kocherlakota and Yi (1997) and Dajin Li (2002). As to the theoretical point, (ii), take the Romer model as an example.[13] From (4) and (5), the aggregate invention production function is $\dot{N} = \mu N L_R$. A more general specification would be

$$\dot{N} = \mu N^{\varphi} L_R, \qquad \varphi \leq 1, \tag{12}$$

where the parameter φ is the elasticity of research productivity with respect to the level of technical knowledge. In the Romer model, the value of this parameter is arbitrarily made

[12] Indeed, $\partial g_y/\partial L = \mu(s_R + L\partial s_R/\partial L)$, which is probably not significantly less than μs_R, since increasing L is not likely to have a sizeable diminishing effect on relative research effort s_R. To be definite about this requires specification of the household sector. In the most common specifications s_R is in fact increasing in L.

[13] An analogue argument goes through for the vertical innovations model.

equal to one. One could easily imagine, however, this parameter being negative ("the easiest ideas are found first", also called the "fishing out" case). Even when one assumes $\varphi > 0$ (i.e., the case where the subsequent steps in knowledge accumulation requires less and less research labor), there is neither theoretical nor empirical reason to expect $\varphi = 1$. The standard replication argument for constant returns with respect to the complete set of *rival* inputs is not usable.[14] Even worse, $\varphi = 1$ is a knife-edge case. If φ is just slightly above 1, then explosive growth arises - and does so in a very dramatic sense: infinite output in finite time.[15] This was pointed out by Solow (1994a). The numerical example he gave corresponds to $\varphi = 1.05$, $s_R = 0.10, \mu L = 1$ and $N_0 = 1$. In this case, the Big Bang - the end of scarcity - is only 200 years ahead. The fact that this occurs only a hair's breadth from the presumed unit value of φ tells us something about how strong and optimistic that assumption is. To paraphrase Solow (1994b), it is too good to be true. The Big Bang pertaining to $\varphi > 1$ is mathematically analogous to the fact that the solution to the differential equation $\dot{x} = 1 + x^2$, with initial condition $x(0) = 0$, is $x(t) = \tan t$ and approaches infinity for $t \to \pi/2$.

On the other hand, with φ just slightly less than 1, productivity growth peters out, unless assisted by growth in some exogenous factor, say population. Indeed, let $L = L_0 e^{nt}$, where $n \geq 0$ is a constant. Then, deriving from (12) an expression for $\dot{g}_N/g_N$ we find that in a steady state (i.e., when $\dot{g}_N = 0$),

$$g_y = \frac{n}{1-\varphi}. \tag{13}$$

There are a number of observations to be made on this result. *First*, the unwelcome scale effect on growth has disappeared. *Second*, from (12) it can easily be shown that through *temporary* effects on growth, a scale effect on the *level* of $y(t)$ remains (i.e., $\partial y(t)/\partial L_0 > 0$ along a balanced growth path). From the point of view of theory, this is exactly what we should expect. The non-rival character of knowledge implies that output per capita depends on the total stock of ideas, not on the stock per person. A larger population breeds more ideas, leading to higher productivity. Empirically, the very-long run history of population and per capita income of different regions of the world gives evidence in favour of scale effects on levels (Kremer 1993); econometric evidence is provided by, e.g., Alcalá and Ciccone (2004). *Third*, scale effects on levels also explain why

[14] I therefore disagree with Dinopoulos and Sener (2003) who regard an assumption like $\varphi = 1$ as parallel to the constant returns to scale assumption often invoked in competitive equilibrium theory. This is a linearity w.r.t. the set of rival inputs and is *endogenous*, in view of the replication argument.

[15] This knife-edge feature pertains to not only innovation-based endogenous growth models, but also accumulation-based models, e.g., Lucas (1988).

the rate of productivity growth should be an increasing function of the rate of population growth, as implied by (13). However, in view of cross-border technology diffusion, this trait should not be seen as a prediction about individual countries in an internationalized world, but rather as pertaining to larger regions, perhaps the global economy. *Finally,* unless policy can affect φ or n (often ruled out by assumption[16]), the long-run growth *rate* is independent of policy, as in the traditional neoclassical story. Yet, this statement should not be interpreted as excluding that the general social, political, and legal environments can be *barriers* to growth or that policy can affect s_R (via influencing incentives by a research subsidy, say) and thereby affect the *level* of the time path of y.

The case $\varphi < 1$ constitutes an example of *semi-endogenous growth.* We say there is semi-endogenous growth when (a) per capita growth is driven by some internal mechanism (as distinct from exogenous technology growth), but (b) sustained per capita growth requires support from growth in some exogenous factor. In innovation-based growth theory, this factor is normally population size. The diminishing returns to knowledge is offset by a rising number of researchers; a constant positive growth rate of knowledge results.[17] Note that in our terminology, the distinction between fully endogenous growth (defined in Section 2.1) and semi-endogenous growth need not coincide with the distinction between policy-dependent and policy-invariant growth; this becomes important below. Moreover, albeit the Jones (1995b) model modifies Romer's increasing variety model, we may still classify the Jones model as belonging to the first-generation models of endogenous growth. Indeed, whether an analysis concentrates on the robust case $\varphi < 1$ or the non-robust (but analytically much simpler) case $\varphi = 1$, is in our terminology not decisive for to what generation the applied model framework belongs. The defining characteristic of the first-generation models is in our terminology rather the one-sided concentration on *either* horizontal innovations *or* vertical innovations.[18]

[16]There are exceptions though. Cozzi (1997) develops a model in which R&D can follow different directions and where short-term gains may conflict with long-term growth prospects. With taxes and subsidies it is possible to shift research to directions with higher growth potential.

[17]In Jones (1995b, (12) takes the extended form, $\dot{N} = \mu N^{\varphi} L_R^{\lambda}$, $0 < \lambda \leq 1$. The case $1 - \lambda$ represents a likely congestion externality of simultaneous research (duplication of effort). But for our discussion here this externality is not crucial. An early example of a semi-endogenous growth model is Arrow (1962a).

[18]For more elaborate variants of the semi-endogenous approach, with focus on vertical innovations within a fixed spectrum of product lines, see Kortum (1997) and Segerstrom (1998). A somewhat different way to aleviate or eliminate scale effects on growth is based on *adoption costs* (Jovanovic 1997).

3.2 Different responses to the Jones critique

The Jones critique provoked at least four different kinds of responses.

3.2.1 The knife-edge models are handy approximations

The knife-edge models are useful as simplifying devices. The assumption $\varphi = 1$ should be seen as an approximation to the generic case of φ less than, but perhaps close to, one (McCallum 1996, Temple 2003). Then it depends on the circumstances whether this yields an acceptable approximation and for how long. It should be emphasized that the length of the period for which such an approximation is acceptable may be more limited than usually believed. To get a flavour, consider the Cobb-Douglas version of the well-known Learning-By-Investing model without scale effect (Barro and Sala-i-Martin 2004, pp. 235, 237). Let σ be a subsidy to purchases of capital services. Departing from steady state, consider an unanticipated increase in σ from 0.40 to 0.56. Let the "true" learning parameter λ be as high as 0.8, and compare the effect of the shock to that in the simplified (knife-edge) model where $\lambda = 1$. For standard parameter values one may end up with, after 60-70 years, an aggregate capital intensity in the knife-edge model *double* to that in the "true" model, a difference that may be important, for example, to the evaluation of welfare effects.

3.2.2 Anyway, *some* linearity is needed for steady growth

As noted by Romer (1995), in order for balanced growth to be possible, a growth model must yield at least one differential equation that is linear:

$$\dot{x} = \text{constant} \cdot x. \tag{14}$$

Growth models differ according to a) what variable takes the role of x, and b) what determines the constant.[19] The key to having policy impinging on long-run growth is to have the constant determined such that policy can affect it. In Solow (1956), we have $\dot{T} = \tau T$, where T is the technology level and τ is exogenous; hence, the long-run growth rate is policy-invariant. In the one-sector AK model of Rebelo (1991), $Y = AK$, so that $\dot{K} = (A - C/K - \delta)K$, where A is an exogenous constant and C/K is constant in equilibrium and can be affected by tax policy. The human-capital model of Lucas (1988)

[19] In multi-sector models with more than one state variable, the simple proportionality in (14) generally takes the form of a vanishing determinant.

has $\dot{h} = \mu s h$, where h is per capita human capital and s is per capita educational time. Romer (1990) and Aghion and Howitt (1992) are described above. With a proviso to which we come back at the end of Section 4, it seems that no convincing explanation has as yet been given for any of these candidates.

Jones (2005) proposes $\dot{L} = (b - d)L \equiv nL$, where b is the birth rate and d is the death rate. He argues that this demographic candidate is less arbitrary than the other candidates. After all, people reproduce in proportion to their number.[20] Anyway, a strength of the semi-endogenous growth approach seems to be that no *other* knife-edge condition than exponential population growth is needed for explaining balanced growth. In contrast, the fully endogenous growth approach assumes both that n is constant (sometimes nil to avoid growth explosion) *and* that an additional debatable linearity is present. In Section 6, where the role of non-renewable resources is considered, we meet yet another casting of the role of x in (14).

3.2.3 One can do with only *asymptotic* linearity

Larry Jones and Rodolfo Manuelli (1990, 1997) point out that *asymptotic* linearity with respect to capital can be enough for fully endogenous growth to arise. This is because what in this context really matters is whether the marginal productivity of capital, when the capital intensity goes to infinity, is bounded away from zero.[21] Dalgaard and Kreiner (2003) follow up by applying the same principle to the invention production function. Jones and Manuelli as well as Dalgaard and Kreiner suggest a CES production function with elasticity of substitution above 1 as an example. However, in the Jones and Manuelli case (a CES in terms of K and L) one faces the problem that it is not clear that the theoretical or empirical basis for asymptotic linearity is essentially better than that for exact linearity. In the Dalgaard and Kreiner case (a CES in terms of L and N), there is the additional problem that the replication argument can no longer be used to defend the assumed exact constant returns to scale with respect to L and N, the latter being a non-rival input.

[20] Most of the Jones papers take n as given, but Jones (2003) deals with the very long run and provides a theory of endogenous fertility. Thus population growth is endogenized and by combining these demographic elemets with economic elements as in Section 3.1, the growth process becomes fully endogeneous. Yet, the implied predictions are very different than those from the fully endogenous growth models of Section 2.

[21] If we look for fully endogenous growth to be not only technologically feasible, but also actually induced by incentives, the lower bound has to be "sufficiently" above zero.

An important point from these contributions is, however, that what appears as a knife-edge condition depends on the class of functional forms allowed. In this perspective the constant elasticity class considered by Jones and others seems reasonable. It is analytically convenient and it spans a broad range of possibilities of empirical interest.

The fourth kind of response to the Jones critique is more elaborate and constitutes what may be called the *second-generation* innovation-based models.

4 Second-generation models

Young (1998), Peretto (1998), Aghion and Howitt (1998, Ch. 12), Dinopoulos and Thompson (1998) and Howitt (1999) (the Y/P/AH/DT/H models) try to establish that it is possible to get rid of the scale effect on growth and at the same time maintain that policy affects the long-run growth rate. The following is only a rough description that does not do justice to all the interesting insights and differing details of these papers.

The basic idea is to combine the quality ladder approach with the increasing variety approach by letting innovations occur along *both* dimensions, the vertical as well as the horizontal. Aggregate output of basic goods is

$$Y = K^{\alpha}(NQL_Y)^{1-\alpha} \ , \tag{15}$$

where now the number of varieties, N, is increasing over time. Indeed, total research effort is, as before, $L_R = s_R L$, but a fraction, s_N, of this is devoted to horizontal innovations,

$$\dot{N} = \mu s_N s_R L, \qquad \mu > 0, \tag{16}$$

and the remaining part is devoted to vertical innovations within the existing N product lines,

$$\dot{Q} = \xi Q \frac{(1 - s_N)s_R L}{N}, \qquad \xi > 0. \tag{17}$$

These equations give $g_N = \mu s_N s_R L/N$ and $g_Q = \xi(1 - s_N)s_R L/N$, respectively. So, along a balanced growth path, where s_R, s_Q, g_N, and g_Q are constant, *both* equations imply $g_N = n$. Taking growth rates in (15), we get

$$g_Y = \alpha g_K + (1 - \alpha)(g_N + g_Q + n).$$

In balanced growth, where $g_K = g_Y = g_y + n$, we thus have

$$g_y = g_N + g_Q = n + g_Q. \tag{18}$$

Given the technology and demography, represented by n, the balanced growth requirement pins down only g_N, thus leaving room for g_Q to depend on household preferences and economic policy. In turn this leaves room for g_y, the growth rate of productivity, to similarly depend on household preferences and economic policy.

Indeed, in a fully articulated model, combining (15), (16), and (17), with a specified household sector, these claims come true:

(i) Since the equilibrium values of s_R and s_N depend on R&D incentives and households' saving, which in turn depend on subsidies and taxes, economic policy can affect the long-run productivity growth rate, g_y.

(ii) There is no scale effect on growth, hence, population growth does not imply accelerating growth.

The latter conclusion hinges on the so-called *dilution effect* of expanding the number of product lines. Although this product line proliferation may imply gains by specialization, it also *thins down* the quality improving research effort in *each* product line. This results in less average quality improvement than otherwise. When the population increases, so does the number of product lines in the economy (since, as we saw, $g_N = n$ in balanced growth) and the dilution effect holds the productivity increases in check.[22]

Jones (1999), Chol-Won Li (2000), and others rejoin that, though interesting,

- these results rely on *several* arbitrary knife-edge conditions;
- a generic model with innovations along two dimensions tends to make the long-run growth rate policy-invariant.

Indeed, the above model is special: there are no knowledge spillovers *within* horizontal innovations, there are no knowledge spillovers *between* horizontal innovations and vertical innovations, and the parameter for the spillovers within vertical innovations happens to be exactly unity - a very specific value.

[22] Similar results are obtained if the vertical innovations take the form of cost-reducing process innovations, as in Peretto (1998).

It may be noted that at a purely formal level we have $g_y = (1 + \frac{\xi}{\mu}\frac{1-s_N}{s_N})n$ in balanced growth (only "purely formal" because both s_N and $g_N = n$ are not independent). Thus, population growth is here necessary to sustain positive per capita growth in the long run. So the setup implies semi-endogenous growth.Yet, within the second generation models there are also contributions that generate *fully* endogenous growth, e.g., Young (1998) and Dinopoulos and Thompson (1998).

A more general (and symmetric) model would be (essentially from Chol-Won Li 2000):

$$\dot{N} = \mu N^{\varepsilon_1} Q^{\varphi_1} \frac{(1-s_Q)s_R L}{Q}, \qquad \varepsilon_1 \geq 0, \varphi_1 \geq 0, \tag{19}$$

$$\dot{Q} = \xi N^{\varepsilon_2} Q^{\varphi_2} \frac{s_Q s_R L}{N}, \qquad \varepsilon_2 \geq 0, \varphi_2 \geq 0. \tag{20}$$

In balanced growth (s_R, s_Q, g_N, and g_Q constant) the numerator and the denominator in expressions for g_N and g_Q, derived from (19) and (20), must grow at the same rate. This implies

$$(1-\varepsilon_1)g_N + (1-\varphi_1)g_Q = n, \tag{21}$$

$$(1-\varepsilon_2)g_N + (1-\varphi_2)g_Q = n. \tag{22}$$

Solving for g_N and g_Q, and using (18), yields

$$g_y = \frac{(\varphi_1 - \varphi_2 + \varepsilon_2 - \varepsilon_1)n}{D}, \tag{23}$$

in balanced growth, presupposing $D \neq 0$, where $D \equiv (1-\varepsilon_1)(1-\varphi_2) - (1-\varphi_1)(1-\varepsilon_2)$. We see that in the generic case ($D \neq 0$) long-run growth can not be affected by standard policy tools.

On the other hand, in the Y/P/AH/DT/H models the spillover parameters happen to be:

$$\varepsilon_1 = \varepsilon_2 = 0 \qquad \text{(spillovers from horizontal innovations)}, \tag{24}$$

$$\varphi_1 = \varphi_2 = 1 \qquad \text{(spillovers from vertical innovations)}, \tag{25}$$

so that $D = 0$. In this knife-edge case, (21) implies $g_N = n$, leaving room for g_Q to be determined by policy, thus confirming the logic behind (i) above.

Chol-Won Li (2002) generalizes the model and shows that if intermediate goods have k quality attributes, which can be improved through R&D, then policy-dependent growth requires at least $k+1$ knife-edge conditions to be satisfied. Otherwise the long-run growth rate is independent of policy. Yet, the *level* of the growth path can depend on policy, which may still have sizeable welfare effects.

These observations do not preclude that a richer model might draw attention to economic mechanisms *affecting* the spillover coefficients. The models proposed by Cozzi (1997) and Peretto and Smulders (2002) are steps in this direction. In the Peretto and Smulders paper, the vertical innovations are "in-house" (no business-stealing effect) and the horizontal innovations raise technological distance. This reduces the effective spillovers originating in horizontal innovations. In this way the Peretto and Smulders paper can be interpreted as the case: $\varepsilon_1 \rightarrow 0, \varepsilon_2 \rightarrow 0$ for $N \rightarrow \infty$. This still leaves open how to argue for the parameter restrictions (25).

5 Later developments

5.1 Integration with IO theory and empirics

In the discussion of the first-generation vertical-innovations models it was mentioned that they share the unrealistic feature that all R&D is done by outsiders. There are newer models taking into account that the incumbent firm is likely to have a knowledge and cost advantage in R&D and may gain from striving towards the next innovation, thereby avoiding being overtaken by an outsider. This amounts to an integration of growth theory with industrial organization theory and empirics. Unlike the first-generation horizontal and vertical innovations models that unambiguously imply a positive relationship between monopoly power and innovations, a more nuanced understanding of the roles of competition and monopoly for growth is pursued. For overviews, see Aghion and Griffith (2005) and Aghion and Howitt (2009).

5.2 Directed technical change

Whatever the source of a potential for sustained technical progress, a question remains: technical progress in what "direction"? Will it tend to be purely labor-augmenting or not? The answer to this question matters for whether approximate balanced growth can be maintained.

Until recently, almost all growth models, whether with endogenous or exogenous technical change, just assumed either that the elasticity of substitution between capital and labor is 1 (the Cobb-Douglas production function) or that all technical change is purely labor-augmenting – capital-augmenting technical change being, by assumption, excluded. With an elasticity of substitution less than 1 (as the empirical evidence suggests, see Antras 2004), technical change must be purely labor-augmenting in order that balanced growth paths with constant income shares of labor and capital can exist. Fortunately, recent theory points at mechanisms that can possibly explain that technical change should be purely labor-augmenting, i.e., Harrod-neutral, in the long run.

In a series of papers, Acemoglu (1998, 2002, 2003) succeeds in integrating the somewhat ad-hoc theory from the 1960s about the "innovation possibility frontier"[23] with the microfounded endogenous growth theory of the late 1980s. The outcome is a theory in which the same economic forces – profit incentives – that affect the *rate* of technical change also shape the *direction* of technical change. Here I outline how the theory works in the horizontal innovation framework.[24]

Assume output of basic goods is given by the CES function

$$Y = \left[\alpha(MK)^{(\sigma-1)/\sigma} + (1-\alpha)(NL_Y)^{(\sigma-1)/\sigma}\right]^{\sigma/(\sigma-1)}, \tag{26}$$

where $\sigma > 0$ is the constant elasticity of substitution between K and L_Y. There are now *two* technology terms: M, which measures the range of capital-enhancing intermediate goods, and N, which measures the range of labor-enhancing intermediate goods. The technologies for invention of new varieties of the two kinds of intermediate goods are

$$\dot{M} = \hat{\psi} s_M L_R - \eta M, \quad \text{and} \quad \dot{N} = \hat{\mu}(1-s_M)L_R - \eta N, \tag{27}$$

where s_M denotes the fraction of research effort devoted to invention of new varieties of capital-enhancing intermediate goods, and $\eta > 0$ represents the rate of evaporation of varieties.[25] At the economy-wide level the research productivities are given as

$$\hat{\psi} = \psi M(s_M L_R)^{\varepsilon-1}, \quad \text{and} \quad \hat{\mu} = \mu N\left[(1-s_M)L_R\right]^{\varepsilon-1}, \tag{28}$$

where $\varepsilon \in (0,1)$ captures crowding effects not internalized by the individual R&D firm.[26]

Embedding this production structure in a standard representative-agent framework, Acemoglu (2003) essentially shows the following. An income share of capital above its long-run equilibrium value makes capital-augmenting innovations more favorable, i.e., s_M is increased. Thereby, the "effective" capital intensity, $k \equiv MK/NL_Y$, increases, and with $\sigma < 1$ this implies decreasing income share of capital. Similarly, a rate of interest above its long-run equilibrium value spurs capital accumulation, thereby decreasing the rate of interest. The system approaches a balanced growth path with constant k and constant rate of interest. The constancy of the remuneration to capital is obtained when s_M is at

[23] For a summary and critical assessment, see Nordhaus (1973).

[24] Fitting the theory to vertical innovations is also possible, but is slightly more complicated.

[25] The assumption $\eta > 0$ is invoked in order to avoid multiplicity of balanced growth paths.

[26] By having $\varepsilon < 1$, inconvenient discontinuities in the behaviour of s_M are avoided.

a level low enough to just maintain a constant M. On the other hand, N, and thereby the real wage, keeps growing along with K/L_Y, without stimulating labor supply, which is not a function of wages. In this way, technical change becomes purely labor-augmenting in the long run.

Although here formulated as a fully endogenous growth model, i.e., the spillover parameters in (28) are exactly one, seemingly the theory works just as well in the semi-endogenous growth case, where the spillover parameters are less than one (Acemoglu 2002, p. 795). The essential point is that the knife-edge condition of Harrod-neutral technical progress is replaced by a theory of induced Harrod-neutrality in the long run. On the other hand, when one problem is resolved, new problems appear. As recognized by Acemoglu, his theory relies on the knife-edge condition that there are no knowledge spillovers *between* the M and N promoting endeavours, cf. (28). A next task will be to either relax this assumption or provide a microfoundation for it.

The theory of endogenous directed technical change shows its usefulness in many applications. Different elaborations embrace topics such as skill-biased technical change (Acemoglu 1998, Kiley 1999), the long-run constancy of the capital income share despite large changes in fiscal policy and labor market policy (Acemoglu 2003) and cross-country differences in pollution (Di Maria and Smulders 2004). Of particular interest in relation to the limits-to-growth debate is the modelling of induced energy-saving technical change in André and Smulders (2004).

A problem with the Acemoglu directed-technical-change framework is its quite abstract nature. As noted by Stokey (2003), it is not easy to identify what N, M, and S_M correspond to empirically. Jones (2005) offers an alternative approach to the problem of the shape of the aggregate production function, implying that the question about Harrod-neutrality loses its importance. Indeed, Jones provides a microfoundation for the production function being Cobb-Douglas in the long term, though the short-term elasticity of substitution is likely to be less than one.

6 Non-renewable resources and growth

This section considers how the presence of scarce natural resources ("resources" for short) has been incorporated in new growth theory. In particular we will keep an eye on what light natural resources throw on knife-edge problems in new growth theory. As alluded to