Essays on Technological Progress and Economic Growth

Essays on Technological Progress and Economic Growth

Jakub Growiec

Thèse présentée en vue de l'obtention
du grade de docteur en sciences économiques
à l'Université catholique de Louvain

Composition du Jury:

Raouf Boucekkine (promoteur)
Jean-Pascal Bénassy
Claude d'Aspremont
David de la Croix
Tomasz Szapiro

Contents

List of Figures

List of Tables

Acknowledgments

This doctoral thesis summarizes the research I have been doing throughout the years 2004-2007, mostly at CORE in Louvain-la-Neuve but also at Warsaw School of Economics. Its completion in the current form would not be possible without the help, support, and encouragement I have received from several important people. I would like to express my gratitude for them in the following paragraphs.

First of all, I benefited greatly from the guidance of my supervisor, Raouf Boucekkine. Any one of his numerous comments and suggestions was able to push my work forward much further than long hours spent on solitary thinking. Even more astonishing, he would always know the answer to my question right in the moment I finished asking it, no matter how complex (or misleading, or plain stupid) my arguments were.

I am greatly indebted to Tomasz Szapiro, my tutor and boss at Warsaw School of Economics, who has always supported me, offered me helpful advice – not only on my research but also on the workings of the academic world in Poland, in Belgium, everywhere. He began to encourage me to apply for a visiting doctoral student position at CORE before I realized that CORE, or Louvain-la-Neuve, existed at all.

This thesis would be much worse if not for the constructive comments and criticisms I have obtained from the other members of my doctoral jury: David de la Croix, Claude d'Aspremont, and Jean-Pascal Bénassy. I am grateful for their help.

I would also like to thank Ingmar Schumacher for all the discussions, suggestions, and for our fruitful cooperation: he is the coauthor of chapters 3 and 6 of this thesis. I would also like to thank him, along with Luca, I-Ling, Gül, Alfonso, Agustín, Alexei, ..., ..., for making my stay at Louvain-la-Neuve so enjoyable, for creating a friendly Englishspeaking atmosphere in the middle of an area which I personally find quite francophone.

I acknowledge financial support from the EDNET Marie Curie fellowship program (mid-September 2005 until mid-July 2006); the Foundation for Polish Science (since January 2007); and the Dean of the Collegium of Economic Analysis, Warsaw School of Economics (February – March 2007).

Now I can finally express my personal thanks. I would like to thank my parents for their continuous support and encouragement. Above all, however, I am very, very, very grateful to my wife Kasia, for all her love, support, her tolerance for my outrageous behavior (including staying so far from home for so long); for her patience, understanding, and wonderful conversations.

Chapter 1

Introduction

The dissertation covers a broad range of topics in the general area of economic growth theory and economics of technological change. Few specific questions are addressed here, such as the prevalence of knife-edge conditions in growth models, overcoming the essentiality of exhaustible resources in production, endogenous fertility choice as a possible source of long-run growth, the microfoundations of the shape of production functions and the direction of technical change, as well as the impact of the heterogeneity of innovations on long-run growth and convergence. The dissertation is primarily about the ultimate sources and the ultimate limitations of growth, and its objective is thus to provide a new perspective on the very foundations of the process of long-run economic growth. We investigate the background and implications of several specifications of long-run growth "engines" which can be found across the theoretical literature, and then put forward their generalizations and extensions. We identify the circumstances under which balanced long-run growth is possible and study the alternatives accordingly.

Since the main objective of this doctoral thesis is to provide a thorough analytical treatment of the fundamental sources and limitations of economic growth, the discussion often abstracts from various important characteristics of real-world economies. While taking our broad "macro" perspective, we shall neglect issues like uncertainty, imperfect competition, heterogeneity in consumers' preferences, or multiple equilibria, as well as limit ourselves to social planner solutions. Thanks to these simplifications, we shall obtain a substantial gain in analytical tractability of our models and a clearer picture of our main results.

1.1 Outline

The articles that make up the dissertation are ordered according to the "fundamentality" of the questions under consideration, as perceived subjectively by the author. We descend gradually from the most "general" results to more "specific" ones.

At the highest level of generality, Chapter 2 provides a formal proof that balanced (i.e. exponential) growth requires knife-edge assumptions which cannot be satisfied by typical values of model parameters. This important new result can be applied to the discussion on the (in)validity of several linearity assumptions used in the literature, conducive for long-run growth (cf. Jones, 2003; Christiaans, 2004): it implies that at least one such knife-edge assumption *must* be made if a given model is supposed to deliver balanced growth over the long run. The generality of the point made in Chapter 2 implies that there is no way to circumvent this strict requirement.

Moving to a more specific question, Chapter 3 deals with the issue of resource-based limits to long-run growth. It is acknowledged there that the reliance of production on non-renewable resources whose stock is finite imposes an ultimate limit on economic activity (cf. Dasgupta and Heal, 1974). As a way to avoid this constraint, we propose to promote technological progress which would improve the substitutability between non-renewable and renewable resources. The elasticity of substitution between the two kinds of resources is identified as the decisive figure here: if it exceeds unity, production will not fall down to zero even after the non-renewable resources will have been completely depleted. Factor-augmenting technological progress can also be helpful, but its effects are much less pronounced and it must go on forever in order to assure sustainable production.

Another question asked in the dissertation is whether it is plausible that R&D-based growth, fueled by steady increases in total population (as the semi-endogenous growth theory requires, cf. Jones, 1995), can be extended into the long run. To address this concern, Chapter 4 considers endogenous fertility choice in a world where economic growth is semi-endogenous and driven by R&D activity. We argue that this set of assumptions makes it possible for the long-run growth rate of the economy to be again fully endogenously determined. We also discuss the new problems which may arise once the birth rate is taken as a control variable, and childrearing costs are accounted for in the households' optimization problem.

The next issue addressed here relates to the reasons for, and consequences of, relaxing the assumption of Cobb-Douglas aggregate production functions. Over the years, this assumption has become a standard in growth theory. Questioning this standard on a theoretical basis, Chapter 5 refers to the idea-based microfoundations of aggregate production functions. The correspondence between the shape of production functions, the direction of technical change, and the possibility of sustained endogenous growth is discussed there in detail. A broad class of production functions, nesting both the Cobb-Douglas and the CES function, is derived from a microfounded framework extending Jones (2005b).

Finally, we turn to the issue of heterogeneity of innovations. In Chapter 6, we discuss the impact of this heterogeneity on economic dynamics over the long run. To do so, we extend the standard semi-endogenous growth model with a distinction between two inherently different types of innovations, referred to as radical and incremental innovations. We also assume that total R&D output depends on technological opportunity which is depleted by incremental innovations but renewed by radical innovations (cf. Olsson, 2005). We thus reinterpret the standard R&D-based growth framework and draw novel conclusions. The dynamic interplay of the arrivals of the two types of innovations is shown to give rise to dampened oscillations along the transition to the economy's balanced growth path. Under more special parametrizations, the model can also give rise to limit cycles, arising endogenously without uncertainty.

1.2 Knife-edge conditions in growth models

Chapter 2 relates to the "linearity critique", a controversy raised by the endogenous growth theory literature (e.g. Lucas, 1988; Romer, 1990). Reliance of this class of models on linear differential equations for long-run growth is striking. However, actually *all* models that generate exponential growth contain either (i) a strictly linear differential/difference equation (cf. Jones, 2005a), or (ii) a system of differential equations whose determinant is exactly zero (Christiaans, 2004). This means that they all rely on "non-typical" assumptions for parameter values; in the "normal" case, such conditions would be violated.[1] It has been already acknowledged that not only endogenous growth models are hit by the linearity critique, but the literature lacked a complete elaboration of the issue.

Chapter 2 fills this gap in the literature. It extends the already known results (Christiaans, 2004; Laffargue, 2004) and proves in general that the necessity for knife-edge conditions is a direct consequence of the assumption of existence of a balanced growth path (BGP). The general result applies to all dynamic models based on differential or difference equations[2] of an arbitrary order. It is impossible to build a model that delivers balanced growth and at the same time does not rely on knife-edge conditions.

The main negative result has a number of consequences. First, it indicates how fragile balanced growth is: an arbitrarily small change in parameter values eliminates it, causing the growth rates either to fall gradually over time, or to rise in an implausible explosive manner. It requires the parameters to be set exactly at bifurcation values, separating from each other cases where no long-run growth is possible and the explosive cases where the economic variables reach infinity in finite time. Thus, we get a flavor of how curious the phenomenon of exponential growth is.

Another implication of the "impossibility" theorems in Chapter 2 is that the linearity critique is not a valid criterion for discriminating between different models of balanced growth. They are all equally affected by the critique. As a way to circumvent it, one could try to search for knife-edge conditions that look natural and intuitively acceptable (see the discussion in Jones, 2003 and in Chapter 4 of this thesis), or

[1] Parameter values are said to be "non-typical" when the set of values satisfying a given condition is of Lebesgue measure zero (or has an empty interior) in the set of all possible parameter values. A knife-edge condition is a condition that requires parameters to take non-typical values. A detailed discussion of these definitions is included in Chapter 2.

[2] That is, set up in continuous or discrete time, respectively.

explain historical growth patterns as transitional phenomena (e.g. Kremer, 1993).

The article in Chapter 2 has been published in *Economic Theory*.[3]

1.3 The substitutability between non-renewable and renewable resources

Chapter 3 (written together with Ingmar Schumacher) adds another brick to our understanding how fragile long-run balanced growth is. It brings into the picture an element often neglected in mainstream growth economics: exhaustible resources (cf. Solow, 1974; Stiglitz, 1974; Dasgupta and Heal, 1974). More precisely, we put forward a simplified "Ramsey–Hotelling" growth model where production requires labor and an intermediate resource input, which is a CES aggregate of non-renewable and renewable resources. We derive the long-run outcomes as well as short-term dynamics of the model allowing for three types of technical change: (i) exogenous increase in total factor productivity; (ii) factor-augmenting technical change which increases the unit productivity of a particular type of resource (cf. Acemoglu, 2003; Grimaud and Rouge, 2005); (iii) technical change in the elasticity of substitution σ between non-renewable and renewable resources.[4]

It is argued that if perpetual technological progress is guaranteed, the constraint on total production imposed by the finiteness of the non-renewable resource stock can be overcome in each of the three cases. However, only a successful increase in substitutability, such that $\sigma > 1$ from some time on, guarantees sustainable production even if further technological progress fails. In other words, we advocate the type of technical change that would render non-renewable resources inessential for production. Without this qualitative change in the production process, not only long-run growth but also maintaining constant output

[3]Exact reference: Jakub Growiec (2007), Beyond the Linearity Critique: The Knife-Edge Assumption of Steady-State Growth. *Economic Theory* 31(3), pp. 489–499.

[4]For a discussion of relevance of the elasticity of substitution as a measure of economic progress, or the width of the 'menu of choice' available to entrepreneurs, see Yuhn (1991) and Klump and Preissler (2000).

over time would be highly uncertain: if R&D stops, production will stop as well.

On the other hand, if $\sigma > 1$ then the economy is able to produce only with labor and renewable resources whose renewal process is natural and does not require investment costs.[5] In such case, balanced growth remains fragile (as in Chapter 2; note also that by the laws of physics, the available stock of renewable resources is necessarily bounded from above), but sustained consumption is guaranteed. This means that the renewable resource works as a *potential backstop technology*: it is a backstop technology since it guarantees sustainable production without non-renewable resources, and it is only potential because its emergence is conditioned by $\sigma > 1$.

A significant part of Chapter 3 is devoted to simulations. We analyze numerically the short- to medium-run dynamics of our Ramsey–Hotelling model under each of the three scenarios: (i) no technological progress, a case analogous to Dasgupta and Heal (1974); (ii) elasticity of substitution between renewable and non-renewable resources increases over time; (iii) factor-augmenting technical change. We discuss the possibility of finite-time depletion of the non-renewable resource stock and the question of optimal timing of resource depletion if the exhaustible resource is inessential for production.

The article in Chapter 3 has been published in the CORE Discussion Paper series (CORE DP 2006/63) and submitted to a journal.

1.4 Endogenous fertility in R&D-based growth models

Chapter 4 abstracts again from natural resources. Instead, it discusses the long-run impact of endogenous fertility choice on economic growth. The considered framework bases on the by now standard Jones' (1995) R&D-based semi-endogenous growth model and adds to it an endogenous fertility part.

Building upon the negative result in Chapter 2, the semi-endogenous

[5] Though it may possibly be accelerated by purposeful investment. A further discussion of this issue is beyond the scope of this thesis.

R&D-based growth model with endogenous fertility is proposed as a partial solution to the knife-edge conditions problem. Knife-edge conditions are inevitable if balanced economic growth is presupposed, but among the available options, the linear differential equation for population growth, $\dot{N} = (b - d)N$ where N denotes population, while b and d denote respectively the birth and death rate, is probably the most natural one. As Jones (2003) puts it, "it is a biological fact of nature, that people reproduce in proportion to their number". Thus, Chapter 4 proposes a way to build models where growth is fully endogenous (i.e. the long-run growth rate is ultimately determined by the endogenously set birth rate) but no unmotivated knife-edge assumptions are made. Local stability of the asymptotic balanced growth path of the economy is also discussed, and a few numerical examples are provided to illustrate our analytical findings.

Before presenting the actual model of Chapter 4, a number of methodological issues has to be resolved. We compare the formal treatments of models with endogenous and exogenous fertility, concluding that for many reasons, these two are usually very different. In the original Barro and Becker (1989) setup, population size enters the utility functional multiplicatively and thus alters the effective discount rate just as it does in the exogenous fertility setup. In most other papers, however (Barro and Sala-i-Martin, 1995, Chapter 9; Jones, 2003; Connolly and Peretto, 2003), it enters the utility functional additively and the dynamic tradeoffs attached to fertility decisions are switched off. Furthermore, in models with endogenous fertility one is forced to assume that flow utility is positive. Otherwise, it would always pay to lower fertility and thus increase total discounted utility by making it closer to zero. Finally, second order conditions for optimality are not automatically satisfied under endogenous fertility even if all production and utility functions are strictly concave.

Of course, the simplified model in Chapter 4 is subject to several limitations. It ignores the fact that production might rely on exhaustible resources. It also ignores the fact that the Earth is finite and therefore cannot provide shelter to arbitrarily large populations. In this context, it should be emphasized that the model predicts that if population growth is forced to stop, economic growth will consequently stop as well.

The article in Chapter 4 has been published in *Topics in Macroeconomics*.[6]

1.5 Deriving the shape of production functions

Chapter 5 relates to the observation that in a large class of models, endogenous growth is obtained because the marginal product of capital does not fall down to zero when capital goes to infinity. That is, one of the Inada conditions for neoclassical production functions is violated (cf. Barro and Sala-i-Martin, 1995; Jones and Manuelli, 1990). Hence, the shape of the production function can have a decisive impact on the long-run dynamics of a given economic growth model. Indeed, we can obtain convergence to a steady state, oscillations, balanced growth, or even explosive dynamics, and the result varies according to the assumed returns to scale to reproducible factors as well as to the assumed elasticity of substitution σ between the factors.

We focus our attention in Chapter 5 on production functions that take capital and labor as their only inputs. Nevertheless, our results could be applied to any pair of inputs, e.g. also to non-renewable and renewable resources, like in Chapter 3.

The first main contribution of Chapter 5 is to derive from idea-based microfoundations the "global" (firm-level) production function. Extending the recent work of Jones (2005b), we characterize a wide class of production functions, not mentioned previously in the literature, that nests both the Cobb-Douglas and the CES. In certain cases, we are able to derive an explicit formula for the elasticity of substitution between capital and labor. Capital and labor may be gross complements but also may be gross substitutes; if they are gross substitutes ($\sigma > 1$) then endogenous growth à la Jones and Manuelli (1990) is possible.

The fundamental building blocks of the microfounded framework are the assumptions that (i) firms do not only announce factor demands, but they also choose their preferred technology (unit factor productivities of capital and labor) from the available technology frontier (cf. Caselli and Coleman, 2006); (ii) the technology frontier is shaped by the se-

[6]Exact reference: Jakub Growiec (2006), Fertility Choice and Semi-Endogenous Growth: Where Becker Meets Jones. *Topics in Macroeconomics* 6(2), Article 10.

quential arrival of labor- and capital-augmenting innovations; (iii) these innovations (ideas) are Pareto-distributed (like in Kortum, 1997; Jones, 2005b); (iv) the distributions of labor- and capital-augmenting innovations are not independent as in Jones (2005b), but dependent according to the Clayton copula.

The second main contribution of Chapter 5 is to check under which conditions technical change is purely labor-augmenting (as required by the famous Uzawa's (1961) steady-state growth theorem), when it is purely capital-augmenting, and when it augments both factors of production. Since the direction of technical change is explicitly endogenous in our setup, we can derive precise predictions on this issue. It turns out that technical change is purely labor-augmenting only in two specific *knife-edge* cases: (i) if capital- and labor-augmenting developments are independent (the Jones', 2005b, case), or (ii) if "local" production techniques are Leontief and the derived production function is CES.

The article in Chapter 5 has been published in the CORE Discussion Paper series (CORE DP 2006/56) and submitted to a journal.

1.6 Incremental and radical innovations

In Chapter 6, written jointly with Ingmar Schumacher, we return to the familiar world of R&D-based semi-endogenous growth with Cobb-Douglas production functions. We also abstract from natural resources and assume exogenous population growth.

This time we take a closer look at the aggregate R&D output and observe that innovations are not alike. Building on the intuitive arguments and anecdotal evidence brought about by Olsson (2005), we distinguish between incremental and radical innovations as well as discoveries. Furthermore, we introduce the concept of *technological opportunity* which works as an upper limit for further incremental innovations. As opposed to models where innovations are homogenous (e.g. Jones, 1995), technological opportunity is here a necessary input to R&D. New incremental innovations increase the level of total factor productivity, but at the same time deplete technological opportunity. Radical innovations, on the other hand, provide no direct increases to productivity; they serve to renew technological opportunity. In a sense, thus, technological opportunity is modelled like a renewable resource here.

The main goal of Chapter 6 is to analyze the dynamics of R&D and growth in an economy where opportunity for further developments is disconnected from the technology as such. We find that balanced growth requires in our model one additional knife-edge assumption apart from the usual linear population equation. We have to assume that the elasticity of technological opportunity in the production of incremental innovations must be equal to the elasticity of the technology level in the production of radical innovations. This assumption implies that technology and technological opportunity grow at equal rates along the BGP. Otherwise, a BGP would not exist and either (i) technological opportunity would grow faster than technology, implying that there is an ever increasing pool of possibilities that will never be realized, or (ii) technology would grow faster than technological opportunity, meaning that the latter must get ultimately depleted, stopping further growth. In the latter case, insufficient technological opportunity proves to be a serious limitation to balanced growth which has not been recognized in previous literature.

Another finding is that for a wide range of assumed parameter values, our framework gives rise to dampened oscillations along the transition to the balanced growth path, and under more unusual parametrizations, even to limit cycles. These results arise due to the dynamic interplay between radical and incremental innovations as well as to the reallocations of R&D employment from one R&D sector to another.

Discussing a number of further related issues and empirical questions, Chapter 6 advocates the concept of technological opportunity as a useful tool for describing the process of economic growth.

The article in Chapter 6 has been published in CORE Discussion Paper series (CORE DP 2007/57) and submitted to a journal.

1.7 Summary

The current dissertation is a collection of five papers in the field of economic growth theory, oscillating around the two broad questions: "What are the ultimate sources of balanced economic growth?," and "What are the most fundamental limitations to economic growth?". We cannot provide final answers to these questions, unfortunately, but we put forward a number of arguments and models which could bring us closer to the final answers. In particular, Chapter 2 argues that balanced growth requires knife-edge conditions and is thus very fragile to all changes in the economic environment. Chapter 3 finds that the reliance of production on non-renewable resources which are subject to depletion can be fully overcome only if technological progress succeeds in making non-renewable and renewable resources gross substitutes. Chapter 4 identifies the conditions under which endogenous fertility choice, combined with R&D, can give rise to fully endogenous long-run growth. Chapter 5 derives from microfoundations a class of production functions which are consistent with Pareto distributions of ideas. Technical change is then shown to be purely labor-augmenting only in a specific knife-edge case. Finally, Chapter 6 argues how the distinction between radical and incremental innovations in modeling R&D modifies the standard results of the R&D-based growth literature.

Chapter 2

Beyond the Linearity Critique: The Knife-Edge Assumption of Steady-State Growth

2.1 Introduction

The starting point of this paper is the "linearity critique", in its appealing form due to Jones (2005a). Jones argues that all growth models ever discussed in literature which were able to deliver steady-state growth, necessarily contained at least one linear differential/difference equation. This is criticized, because strict linearity is a knife-edge assumption, equivalent to requesting that in an equation of form

$$\dot{X} = \alpha X^{\phi}, \tag{2.1}$$

the parameter ϕ value *exactly* equals unity. Apparently, the long-run dynamics of the X variable would be qualitatively different if $\phi < 1$ or $\phi > 1$.

For the sake of clarity, let us now and until the end of this paper, define a "knife-edge condition" as a condition imposed on parameter values, such that the set of values satisfying this condition has an empty interior in the space of all possible values. Parameter values that are requested to satisfy a particular knife-edge condition would also be referred to in text as "non-typical".

An important remark is that we will find knife-edge conditions in the form of equality constraints. Hence, if the parameter space is finite dimensional (a subset of $\mathbb{R}^n$), the set of values satisfying a knife-edge condition would automatically be of Lebesgue measure zero in the given space.

The main point of this paper is to inspect the following claim due to Jones (2005, p. 62):

> What is not sufficiently well-appreciated, however, is that *any* model of sustained exponential growth requires such a knife-edge condition. Neoclassical growth models [as opposed to endogenous growth models – *J.G.*] are not immune to this criticism; they just assume the linearity to be completely unmotivated.

This claim shall be supported by an explicit proof, that it is the steady-state growth property itself, that is "knife-edge" or requires "non-typical" parameter values.[1] We shall generalize the standard growth framework and prove that even introducing differential/difference equations of an arbitrary finite order does not solve the fundamental problem: sustained exponential/geometrical growth (i.e. existence of a balanced growth path with strictly positive growth rates) *necessarily* relies upon knife-edge assumptions.

Our finding supports the third of Temple's (2003) five "obvious" rules for thinking about long-run growth: "Do not dismiss a model of growth because the long-run outcomes depend on knife-edge assumptions". Indeed, we claim that there is no alternative if one wants the model to possess the steady-state growth property. Thus, this paper adds a further qualification to Temple's view that the issue of knife-edge assumptions in growth models does not deserve the attention it has recently received.[2] No discrimination between models can be made on these grounds.

---

[1] Exponential growth is strikingly consistent with some of the well-documented empirical data, however. See e.g. the figure on page 42 in Jones (2005a).

[2] Knife-edge conditions in economic growth models have been reviewed and classified by, among others, Eicher and Turnovsky (1999b), Jones (1999), Li (2000), and

However, given that all growth models share knife-edge assumptions, and therefore, smallest parameter shifts are enough to reverse their long-run dynamics and eliminate steady-state growth – we come to a pessimistic conclusion that growth is very fragile.

To reinforce our argumentation, let us also point out that the knife-edge character of the "$\phi = 1$ in equation (2.1)"-type assumptions consists not only in the fact that the set of parameter values satisfying them has an empty interior (or is of Lebesgue measure zero) in the set of all possible parameter values, but also in the fact that they bound away from each other cases of qualitatively different dynamic behavior of the model. Most likely, these would be explosive cases ($\phi > 1$), and cases where the growth rates gradually fall down to zero ($\phi < 1$).

To our best knowledge, the fundamental result of this paper has not yet been proven in its generality. Closely related literature includes Christiaans (2004) who provides the proof for the specific case of continuous-time models with three state variables (physical capital, technology, population) and differential equations of first order only; and Laffargue (2004) who proves it for discrete-time models with up to two lags, while concentrating on a completely different issue – solving macro-econometric models with perfect foresight – and thus neither attaches any particular weight to the mathematical result, nor gives an interpretation for it.

In the following section, we rephrase the "linearity critique" in the context of a generalized growth model. In section 3, we prove the general theorem that steady-state growth is a knife-edge assumption. Section 4 contains further discussion and concluding remarks.

2.2 The "singularity" critique

Let us consider a generalized continuous-time model of economic growth. The variant we are going to analyze is standard in the sense that its steady-state properties are determined by a system of *first order* autonomous differential equations of the form

$$\hat{X} = F(X), \qquad X(0) \text{ given.} \tag{2.2}$$

Christiaans (2004).

By $X = (X_1, X_2, ..., X_n)$ we denote a vector of n state variables. Each i-th variable X_i is assumed to be twice continuously differentiable with respect to time. By $\dot{X}$ we denote a vector of X_i's first order time derivatives, and by $\hat{X}$ we denote a vector of their first order log-time derivatives (growth rates). It is assumed that all X_i's are strictly positive.[3] We shall concentrate on autonomous differential equations only, since it is natural for economists to look for general laws that are valid irrespective of time.

A further remark is that in (2.2), we ignore control (choice, decision) variables. Although these are vital ingredients of economic models which include optimization – as most contemporary growth models do – they can be ruled out from present considerations, since we are interested in the long-run dynamics only.

Assuming that $F \in C^1(\mathbb{R}^n)$ and differentiating both sides of (2.2) with respect to time yields

$$\dot{\hat{X}} = DF(X) \cdot \dot{X}. \tag{2.3}$$

The right hand side of (2.3) is linear with respect to $\dot{X}$. We shall adopt the following definition of a steady state: it is a state in which all growth rates $\hat{X}_i$ are constant. Imposing a steady state is equivalent to setting all $\dot{\hat{X}}_i$'s equal to zero. Observing that in the steady state, $DF(X) \cdot \dot{X} = 0$, we state the "linearity critique":

$$\dot{X} = 0 \quad \text{or} \quad DF(X) \text{ is singular.} \tag{2.4}$$

From (2.4), we see that the "linearity" critique discussed in literature should rather be called "singularity" critique: it is *singularity* of the $DF(X)$ matrix that is inevitable if one wants to obtain positive steady-state growth in any state variable. Namely, for models propelled by first order autonomous differential equations, exponential growth requires fulfilling the following knife-edge condition: in the steady state, the

[3] One can argue that taking (2.2) already limits the analysis, since we assume that the general equation $\Phi(X, \dot{X}) = 0$ can be solved explicitly for $\dot{X}$. This problem can be resolved for almost all $\dot{X}$ using the Implicit Function Theorem, however. Since this procedure is not revealing and sometimes cumbersome, we purposefully limit ourselves to (2.2).

determinant of the $DF(X)$ matrix (i.e. the Jacobian of the F mapping) *exactly* vanishes.[4]

Note that here, as opposed to e.g. Jones (1999) who only considers Cobb-Douglas functions with a finite number of parameters (say p), so that the knife-edge conditions are imposed on a subset of $\mathbb{R}^p$, our parameter space is $C^1(\mathbb{R}^n)$ – which is an infinite dimensional function space. Lebesgue measure cannot be applied on such space, but it remains clear that the set $\mathcal{F} \subset C^1(\mathbb{R}^n)$ of functions which have a zero Jacobian everywhere along the time path of X, has an empty interior. To show this, let $F \in \mathcal{F}$. For every $\varepsilon > 0$ there exists a function $F_\varepsilon \in C^1(\mathbb{R}^n)$ such that $||F_\varepsilon - F||_{C^1(\mathbb{R}^n)} < \varepsilon$ and $\det(DF_\varepsilon(X)) \neq 0$ for some X along its time path $\{X(t)\}_{t=0}^{\infty}$.[5] Thus, $F_\varepsilon \notin \mathcal{F}$, so $\mathcal{F}$ has an empty interior.

Let us now manipulate (2.3) to obtain a direct formula for $\hat{X}$. We shall multiply all terms in each i-th column $(i = 1, 2, ..., n)$ of the $DF(X)$ matrix by $X_i > 0$ and denote the resultant matrix by $\Psi(X)$. Notice that

$$\det \Psi(X) = \left(\prod_{i=1}^{n} X_i \right) \det DF(X). \tag{2.5}$$

It follows that $\det \Psi(X) = 0 \Leftrightarrow \det DF(X) = 0$, i.e. the singularity/ non-singularity of $DF(X)$ is inherited by $\Psi(X)$. It is also obtained that

$$DF(X) \cdot \dot{X} = 0 \quad \Leftrightarrow \quad \Psi(X) \cdot \hat{X} = 0. \tag{2.6}$$

The $\Psi(X)$ matrix proves useful in the formulation of the following claim.

The claim goes as follows. If the matrix $\Psi(X)$ is singular, there exists a *continuum* of $\hat{X}$ vectors – steady-state growth rates – located

[4] A slightly less general result than this can be found in Christiaans (2004), Proposition 1.

[5] One may wonder how to construct the F_ε function. For example, one could apply the following procedure. First, fix an X from the time path such that $||X|| > 1$. This is always possible, because we have assumed exponential growth in at least one state variable. Denote $M = 2||X||$, and then take $F_\varepsilon(X) = F(X) + \frac{\varepsilon}{M} X$. Observe that $DF_\varepsilon(X) = DF(X) + \frac{\varepsilon}{M} I$. In result, $||F_\varepsilon - F||_{C^1(\mathbb{R}^n)} = ||F_\varepsilon - F||_\infty + ||D(F_\varepsilon - F)||_\infty = \frac{\varepsilon}{M}||X|| + \frac{\varepsilon}{M}||I||_\infty = \frac{\varepsilon}{2} + \frac{\varepsilon}{M} < \varepsilon$. Moreover, $\det(DF_\varepsilon(X)) = \det(DF(X) + \frac{\varepsilon}{M} I) \neq 0$ unless $-\frac{\varepsilon}{M}$ is an eigenvalue of $DF(X)$. Limiting the scope of our reasoning to ε's small enough (in the end, we are interested only in arbitrarily small ε's anyway) rules out this unwanted possibility and thus guarantees $\det(DF_\varepsilon(X)) \neq 0$.

along the eigenspace associated with the zero eigenvalue of $\Psi(X)$.[6]

If parameters of the model happen to fulfill multiple knife-edge conditions, so that the zero eigenvalue is multifold, then this eigenspace is of a dimension higher than 1.[7]

Let us again emphasize that the parameter values, satisfying the singularity condition are non-typical, that is, the set of such values has an empty interior in the space of all possible values. In consequence, we may say that if there is randomness in their determination, then the singularity condition is typically violated.

What is more, if the F function is not Cobb-Douglas, then $\Psi(X)$ depends on X. Thus, as the variables contained in X grow exponentially over time, the singularity condition still needs to be satisfied for *all* X along the balanced growth path. This accounts for an indeed stringent condition on parameter values and at the same time disqualifies many functional forms of F.

This corollary actually builds up to a version of Uzawa's (1961) Steady-State Growth Theorem: if the production function is not Cobb-Douglas, then while the state variables (X_i's) grow at a constant rate, partial derivatives in DF's columns should change proportionately, in a way that the determinant of $DF(X)$ remains zero at all times.

2.3 The knife-edge assumption of steady-state growth

The results of the previous section can be generalized. We shall consider models both in continuous and in discrete time. We shall also include differential/difference equations of an arbitrary finite order. Although this point may seem overly theoretical – economic growth models typically include only first-order equations[8] – it remains worth emphasizing

[6] As an example, take a family of Jones' models, indexed by the population growth rate $n \geq 0$.

[7] As an example, take a family of neoclassical growth models, indexed by the population growth rate $n \geq 0$ and the exogenous technology growth rate $g \geq 0$.

[8] It is straightforward to borrow an interpretation of second time derivatives from classical physics: we would not only be talking about the *pace* of accumulation of a certain stock variable, but also about its *acceleration*.

that there does not exist a way to circumvent knife-edge assumptions and yet obtain positive long-run growth rates.

Intuitively, the above point is true because of the distinguishing feature of an exponential function, or equivalently, a geometrical sequence. An exponential function is a linear function of its derivative; a geometrical sequence, multiplied by a constant, remains a geometrical sequence.

2.3.1 The theorem (continuous-time version)

We shall now prove the following general theorem.

Theorem 2.1 *It is impossible to construct a model, propelled by differential equations, in which there exists a steady state that*
(a) implies that some state variables grow exponentially at a constant positive rate,
(b) is obtained without knife-edge (singularity) assumptions.

Proof. We shall assume the former and show impossibility of the latter.

For the sake of the proof, let us consider a generalized type (2.2) differential model that includes higher order time derivatives of the state variables, contained in the vector $X = (X_1, X_2, ..., X_n)$. Each i-th variable X_i is assumed to be at least m times continuously differentiable with respect to time. By $X^{(k)}$, we shall denote the vector of k-th time derivatives of X. The model reads:

$$X^{(m)} = \Phi(X, \dot{X}, ..., X^{(m-1)}), \qquad X(0), \dot{X}(0), ..., X^{(m-1)}(0) \text{ given,} \tag{2.7}$$

where m is an arbitrary positive integer.[9] The mapping $\Phi : \mathbb{R}^{mn} \to \mathbb{R}^n$ is assumed to be at least once continuously differentiable. We shall consider the whole $C^1(\mathbb{R}^{mn}; \mathbb{R}^n)$ function space to be our "parameter space".

According to a theorem fundamental to differential equations (see Arnold, 1975), we can decompose our system (2.7) into a system of mn

[9] Again, it is possible to get an explicit solution for $X^{(m)}$ locally almost everywhere. The Implicit Function Theorem is used to prove this claim.

differential equations, $(m-1)n$ of which linear, and write

$$\begin{aligned} \dot{X} &= Y_1 \\ \dot{Y}_1 &= Y_2 \\ &\vdots \\ \dot{Y}_{m-2} &= Y_{m-1} \\ \dot{Y}_{m-1} &= \Phi(X, Y_1, Y_2, ..., Y_{m-1}). \end{aligned} \tag{2.8}$$

Now, we shall proceed with the derivations for each i-th state variable separately ($i = 1, 2, ..., n$). We observe that the exponential growth condition requires

$$\dot{\hat{X}}_i = 0 \quad \Rightarrow \quad \dot{Y}_{1,i} X_i = \dot{X}_i Y_{1,i} \quad \Rightarrow \quad \hat{Y}_{1,i} = \hat{X}_i, \tag{2.9}$$

since we assumed $X_i > 0, \dot{X}_i > 0$ and $\dot{X}_i = Y_{1,i}$. By forward recursion, it is automatically obtained that $\hat{X}_i = \hat{Y}_{1,i} = ... = \hat{Y}_{m-1,i}$. We shall now pass to the last equation.

Imposing that $\dot{\hat{Y}}_{m-1,i} = 0$ yields

$$\hat{X}_i \Phi_i = \hat{Y}_{m-1,i} \Phi_i = \frac{d\Phi_i}{dt} = D_X \Phi_i \cdot \dot{X} + D_{Y_1} \Phi_i \cdot \dot{Y}_1 + ... + D_{Y_{m-1}} \Phi_i \cdot \dot{Y}_{m-1}, \tag{2.10}$$

where Φ_i denotes the i-th coordinate function of the Φ mapping, $D_X \Phi_i$ denotes the vector of n first order partial derivatives of Φ_i with respect to all X_i's. The same notational convention applies to $Y_{k,i}$'s. Arguments of the functions have been omitted for convenience.

With all vectors $D_X \Phi_i, D_{Y_1} \Phi_i, ..., D_{Y_{m-1}} \Phi_i$, we redo the same algebraic manipulations as we did with DF in (2.3), that is, we multiply each j-th element of the vector by X_j, and $Y_{k,j}$ respectively, and denote the resulting vector by $\Psi_{k,i}$, where $k = 0, 1, 2, ..., m-1$. It is then obtained that

$$\begin{aligned} \hat{X}_i \Phi_i &= \Psi_{0,i} \cdot \hat{X} + \Psi_{1,i} \cdot \hat{Y}_1 + ... + \Psi_{m-1,i} \cdot \hat{Y}_{m-1} && (2.11) \\ &\Downarrow \qquad (\text{as } \hat{X} = \hat{Y}_1 = ... = \hat{Y}_{m-1}) \\ \Psi_{\Sigma,i} \cdot \hat{X} &= 0, && (2.12) \end{aligned}$$

where the vector $\Psi_{\Sigma,i} = \Psi_{0,i} + ... + \Psi_{m-1,i} - \Phi_i \mathbf{e}_i$, and $\mathbf{e}_i = (0, ..., 0, 1, 0, ..., 0)$ is the i-th unit vector.

We now combine these n symmetrical results to get the final result:

$$\Psi_\Sigma \cdot \hat{X} = 0, \tag{2.13}$$

where we have denoted

$$\Psi_\Sigma = \begin{pmatrix} \Psi_{\Sigma,1} \\ \Psi_{\Sigma,2} \\ \vdots \\ \Psi_{\Sigma,n} \end{pmatrix} = \begin{pmatrix} \Psi_{0,1} + ... + \Psi_{m-1,1} - \Phi_1 \mathbf{e}_1 \\ \Psi_{0,2} + ... + \Psi_{m-1,2} - \Phi_2 \mathbf{e}_2 \\ \vdots \\ \Psi_{0,n} + ... + \Psi_{m-1,n} - \Phi_n \mathbf{e}_n \end{pmatrix}. \tag{2.14}$$

Since we assumed $\hat{X} \neq 0$, the $n \times n$ matrix Ψ_Σ is necessarily singular.

To prove that singularity is indeed a knife-edge condition, we take $\mathcal{F} \subset C^1(\mathbb{R}^{mn}; \mathbb{R}^n)$ to be the set of functions Φ such that their associated $\det(\Psi_\Sigma) = 0$. Let $\Phi \in \mathcal{F}$. For every $\varepsilon > 0$ there exists a function $\Phi_\varepsilon \in C^1(\mathbb{R}^{mn}; \mathbb{R}^n)$, together with an associated matrix $\Psi_{\varepsilon,\Sigma}$ such that $||\Phi_\varepsilon - \Phi||_{C^1(\mathbb{R}^{mn};\mathbb{R}^n)} < \varepsilon$ and $\det(\Psi_{\varepsilon,\Sigma}) \neq 0$ for some X along the time path.[10] Thus, $\Phi_\varepsilon \notin \mathcal{F}$, so $\mathcal{F}$ has an empty interior. ■

Equation (2.13) is the central result of this paper, albeit the final "impossibility" conclusion in its core is very much alike (2.4). If the knife-edge condition that Ψ_Σ be singular is satisfied, there emerge a continuum of steady states along the space spanned by the eigenvectors associated with Ψ_Σ's zero eigenvalue. If the knife-edge singularity condition is *not* satisfied, which happens in the typical case, then the only steady state is the one with zero growth.

Let us also emphasize, that for non-Cobb-Douglas Φ functions, Ψ_Σ becomes a non-trivial function of the vector $(X, \dot{X}, ..., X^{(m-1)})$. In such case, the facts that the singularity condition imposed on Ψ_Σ is assumed to hold at *all* times, and that arguments of the Ψ_Σ function evolve in time, together constitute a very stringent condition on Φ's parameters and effectively rule out numerous functional forms of Φ.

As m was arbitrary, the main result applies to systems of differential equations of any finite order.

2.3.2 The theorem (discrete-time version)

We shall now prove that a version of the above theorem holds also if the model is set up in discrete time.

[10] One could again use the family of functions $\Phi_\varepsilon = \Phi + \frac{\varepsilon}{M} X$, as in footnote 5.

Theorem 2.2 *It is impossible to construct a model, propelled by difference equations, in which there exists a steady state that*
(a) implies that some state variables grow geometrically at a constant positive rate,
(b) is obtained without knife-edge assumptions.

Proof. We shall again assume the former and show impossibility of the latter.

Let us now consider a general discrete-time growth model. Values of the state variables at time $t = -m+1, -m+2, ..., -1, 0, 1, 2, ...$ are contained in the vector $X_t = (X_{1,t}, X_{2,t}, ..., X_{n,t})$. The model can be written as:

$$X_t = \Phi(X_{t-1}, X_{t-2}, ..., X_{t-m}), \qquad X_{-m+1}, X_{-m+2}..., X_0 \text{ given}, \quad (2.15)$$

where m is some arbitrary positive integer. We do not have to impose any particular assumptions on the Φ mapping. Thus, the space of all mappings $\Phi : \mathbb{R}_+^{mn} \to \mathbb{R}_+^n$ is going to be considered our "parameter space", and denoted by $\mathcal{P}$. We shall endow the space $\mathcal{P}$ with the usual supremum metric but without ruling out functions that are divergent with respect to this metric.

The model (2.15) is a discrete-time version of the continuous-time model (2.7).

The assumption of geometrical steady-state growth implies that for each i-th state variable X_i, there exists a constant $\gamma_i \in (0, 1]$ such that $X_{i,t-k} = \gamma_i^k \cdot X_{i,t}$ for all $k = 1, 2, ...$, and for at least one i, $\gamma_i < 1$. If we denote

$$\gamma = \begin{pmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_n \end{pmatrix}, \quad (2.16)$$

then we can write all n symmetrical equations simultaneously in the compact form $X_{t-k} = \gamma^k \cdot X_t$.

Write (2.15) for two consecutive time periods t and $t+1$ to see that for each i-th state variable, it is also true that

$$\Phi_i(\gamma X_t, \gamma^2 X_t, ..., \gamma^m X_t) = \gamma_i \Phi_i(X_t, \gamma X_t, ..., \gamma^{m-1} X_t), \quad (2.17)$$

where Φ_i denotes the i-th coordinate function of the Φ mapping.

Now collect all n equations (2.17), denote $\Psi(X_t) = \Phi(X_t, \gamma X_t, ..., \gamma^{m-1} X_t)$ and make use of (2.16) to see that

$$\Psi(\gamma X_t) = \gamma \cdot \Psi(X_t). \tag{2.18}$$

Since we assumed $\gamma \neq I$ and $\gamma \neq 0$, the Ψ mapping is necessarily homogenous of degree *exactly* one with regard to the given diagonal γ matrix.

Degree-one homogeneity is a knife-edge condition. To prove it, let $\mathcal{F} \subset \mathcal{P}$ be the set of functions Φ such that their associated Ψ functions are homogenous of degree one for all X_t along the time path $t = 1, 2, ...$ Take $\Phi \in \mathcal{F}$. For every $\varepsilon > 0$ there exists a function $\Phi_\varepsilon \in \mathcal{P}$, together with its associated Ψ_ε and γ_ε, such that $||\Phi_\varepsilon - \Phi||_{L^\infty(\mathbb{R}_+^{mn};\mathbb{R}_+^n)} < \varepsilon$ and $\Psi_\varepsilon(\gamma_\varepsilon X_t) \neq \gamma_\varepsilon \cdot \Psi_\varepsilon(X_t)$ for some X_t along the time path $\{X_t\}_{t=0,1,2,...}$[11] Thus, $\Phi_\varepsilon \notin \mathcal{F}$, so $\mathcal{F}$ has an empty interior. ■

Equation (2.18) is assumed to hold for *all* X_t's along the time path. For non-Cobb-Douglas Ψ functions, it becomes an indeed stringent condition on Φ's parameters and rules out numerous functional forms of Φ.

As m was arbitrary, the result applies to systems of difference equations of any finite order.

At this point, we shall give more intuition on how knife-edge conditions in the form of singularity and degree-one homogeneity correspond to each other. Let us assume that the coordinate functions of Ψ, denoted $\Psi_i(X_t) \equiv \Phi_i(X_t, \gamma X_t, ..., \gamma^{m-1} X_t)$ are homogenous of degree one and twice continuously differentiable for all i. Then, their partial derivatives $\frac{\partial \Psi_i}{\partial X_{j,t}}(X_t)$, $j = 1, 2, ..., n$ are homogenous of degree zero and by Euler's theorem on homogenous functions,

$$\sum_{k=1}^{n} \frac{\partial^2 \Psi_i(X_t)}{\partial X_{j,t} \partial X_{k,t}} X_{k,t} = 0. \tag{2.19}$$

When put in matrix notation, equation (2.19) reads $D^2 \Psi_i(X_t) \cdot X_t = 0$. Since $X_t \neq 0$ by assumption, this implies singularity of the Hessian of Ψ_i. Our reasoning is valid for all i, of course.

[11] Again, we refer back to footnote 5 for the idea how to construct such functions Φ_ε. For example, one could use $\Phi_\varepsilon = \Phi + \frac{\varepsilon}{M} X_t$.

Hence, we see that in the specific case of homogenous and twice continuously differentiable mappings Φ, *degree-one homogeneity implies singularity* of Hessians of the coordinate functions of Ψ. Please note, however, that the main theorem of this subsection applies to a more general class of mappings Φ than discussed in the above paragraph.

2.4 Discussion

We shall now briefly discuss some of the theoretical issues arising from the above "impossibility" theorem.[12] We shall identify the consequences of the theorem by confronting it with a few informal rules which seem to be commonly obeyed in contemporary economic growth theory. The list goes as follows:

- growth models are designed to explain (some) historical data on productivity and GDP;
- they should generate steady-state growth in productivity and GDP;
- exogenous technology growth (e.g. in the form of a $\dot{A} = gA$ equation) should be avoided;
- strong scale effects should be eliminated;
- models are appreciated more, if they are able to deliver optimistic predictions. Long-run growth that does not require continued population growth, and does not hinge upon some exogenously growing factor in production, is certainly a desirable outcome.

This list is admittedly simplified, but still able to capture some of the vital characteristics of the paradigm in contemporary growth theory. We shall now juxtapose these points with the fact that steady-state growth is a knife-edge assumption.

Consistency with empirical data is an obvious property of a good economic theory, and an approximately exponential growth in productivity and GDP in twentieth-century U.S.A. is part of the evidence. The

[12] I would like to thank Charles Jones for shifting my attention to some of these discussion points.

most popular (and natural) way to obtain long-run exponential growth of such variables in a formal model is to make it a steady-state property. "Long run" is then readily identified with the steady state, and "short run" – with the transitional dynamics. In such models, however, knife-edge assumptions are necessary.

The desire to avoid exogenous technology growth gave birth to a wide variety of models, in which growth is driven endogenously, i.e. via purposeful human capital accumulation, R&D expenditure, etc. The original linearity assumption has been shifted from the $\dot{A} = gA$ equation to other equations of the model, or disguised among multiple variables.

Building endogenous growth models without strong scale effects leads either to semi-endogenous theories in which the long-run growth rate is pinned down by the exogenous population growth rate (e.g. Jones, 1995), or to endogenous growth models "of the second generation" (e.g. Young, 1998) which require multiple knife-edge conditions.

The last point of the above list is affected by the theorem in the most obvious way. If one refrains from making knife-edge assumptions, she will no longer be able to obtain the optimistic prediction of sustained growth.

We shall also address the question whether taking demographics as exogenously given (external to the economy) can alleviate the problem of knife-edge assumptions. In the light of the main theorem of this paper, the answer is clearly 'No.' It is not a way to falsify the linearity critique, but only to circumvent it: one assumes exponential population growth to be an *outside* process that nevertheless drives the long-run dynamics of her model. And what seems realistic in empirical research – e.g. Jones' (1995) model explains historical data without imposing any linearity: $\hat{A} = \frac{\lambda \hat{L}}{1-\phi}$, where $\hat{L}$ is exogenous – is irrelevant when considering a *complete* economic theory. To this extent, semi-endogenous growth models are also not immune to the linearity critique. How can one say, that she has $\dot{L} = nL$ in her model, but it is *not her* assumption? That there is a linearity that drives the dynamics of the model, but it is *outside* of the model? Linearity critique applies to all models which contain $\dot{A} = gA$; hence, it also applies to all models which contain $\dot{L} = nL$. And the argument that "it is a biological fact of nature, that people reproduce in proportion to their number" (Jones, 2003) unfortunately does not stand the test of endogenization of fertility: an unmotivated knife-edge assumption that the intertemporal elasticity of substitution in consumption should be exactly one, is called for (see Jones, 2003, or Connolly and Peretto, 2003).[13]

If steady-state exponential growth is a knife-edge condition, then what alternatives do we have? Instead of looking for the least "painful" knife-edge assumption, we could try to explain long-run growth without relying on steady-state analysis at all. However, very little has been done in this area. We can only quote Kremer (1993) and Jones (2001), who argue that the world's economic history over the very long run (since one million BC in Kremer; since 25000 BC in Jones) can be explained like a transitional process, and not a long-run phenomenon in the traditional steady-state sense.

The result of this paper disqualifies the linearity critique as a method of rejecting growth models that contain linear differential equations. We have shown that all models that generate exponential/geometrical steady-state growth contain knife-edge assumptions – in some form of linearity, singularity or degree-one homogeneity.

On the other hand, the theorem emphasizes fragility of exponential/geometrical steady-state growth, arising due to its great sensitivity to disturbances in parameter values.

[13]One could conjecture, of course, that people at large do not optimize over the number of their children, but follow simple rules of thumb. There are at least two counter-arguments to this point, however. First, once one endows the agents of her model with utility-maximizing behavior, leaving some of their "propensities" exogenous for reasons other than simplification seems highly questionable. And second, the enormous fertility drop in the course of the Demographic Transition certainly does not look like a purely exogenous change in the applied rule of thumb, or in the propensity to have children.

Chapter 3

On Technical Change in the Elasticities of Resource Inputs[1]

3.1 Introduction

The sheer size of today's population coupled with its endless desires imposes a never known demand on the production of goods and services. The world's GDP has never been as high as now, and with the advent of China's and India's economic expansions, it seems clear that world production is bound to increase substantially. A noticeable problem is, however, that production depends on various inputs, some of which are known to be limited as well as non-renewable. This has raised concerns (also by major institutions such as RFF, World Bank) for the possibility of continuing production at today's levels to the further future. As pointed out in a special issue of the Review of Economic Studies already in 1974, the ability to extend our current opportunities to future generations when production utilizes limited non-renewable resources depends on a variety of factors, of which two have been singled out as the most crucial ones: substitutability and technical change.

Substitutability is vital because, as Dasgupta and Heal (1974) show, non-renewable resources are essential for production if other inputs are

[1]This chapter has been written jointly with Ingmar Schumacher.

poor substitutes for them. For example, insulation can reduce the amount of oil necessary for heating, but cannot be a perfect substitute for oil. Empirical evidence for low substitution possibilities is provided by Cleveland and Ruth (1997). If this is actually the case, this strand of research predicts a bleak future for generations to come.

Technical change, either exogenous as in Solow (1956) or endogenous as in Romer (1990), is another factor vital for further understanding of production processes. As an example, one could think of cars which are developed such that they can do the same mileage with less and less petrol. Equipped with these new growth-theoretic tools, resource economists (e.g. Scholz and Ziemes, 1999; Schou, 2000; Bretschger, 2005; Grimaud and Rouge, 2005) return with the hope that, even under initially unfavorable circumstances, production possibilities would not decline due to human ingenuity. These theories teach us that the general requirement for non-declining production is a fast enough technical change. However, it is again Cleveland and Ruth (1997, p. 217) who forcefully argue that "...technology and substitution have not been sufficiently strong to offset the effects of depletion at the macroeconomic scale in some nations."

Given the extensive theoretical research on a time-invariant elasticity of substitution as well as on factor-neutral technical change on the one side, but the rather bleak outlook from the empirical literature on the other side, we shall consider a different way of investigating the potential for sustainability. This we will do in the following way.

Unlike the classic works (Solow, 1974; Stiglitz, 1974; Dasgupta and Heal, 1974) which mainly analyze the relationship between capital and non-renewable resources, this article concentrates on the relationship between non-renewable and renewable resources in production. We use a production function of the constant elasticity of substitution (CES) type to allow for various degrees of substitutability between these two inputs. Thus, our paper is closely related to recent articles by André and Cerdá (2005) as well as Grimaud and Rouge (2005). Our analysis, though, unlike these contributions, allows for technical change in the actual elasticity of substitution (EoS hereafter). This permits us to gain new insights into short-run and long-run dynamics. To our knowledge, this approach is novel to the literature.[2] In addition, we will analyze biased technical change in the partial elasticities.

The idea of technical change in the EoS does not just come out of nowhere. Demands to analyze the effect of technical change on the elasticity of substitution have been raised time and again during the past 70 years, with stronger and clearer requests. Already Hicks, who invented the concept of the EoS in the 30's, points out that the elasticity of substitution might change because "(...) methods of production already known, but which did not pay previously, may come into use." Even more precise, he suggests that the increase in substitutability "(...) partly takes place by affording a stimulus to the invention of new types." (1932, p. 120) De La Grandville then proposes to think about the EoS as "a measure of the efficiency of the productive system" (1989, p. 479), which is something that is going to become clear during the following sections. Yuhn, who tests de La Grandville's proposition, suggests to view the elasticity of substitution as "a 'menu of choice' available to entrepreneurs." (1991, p. 344) All of these researchers believe the EoS to be a decisive component, if not a determinant, of growth. They furthermore suggest that the elasticity is by no means invariable over time. That there might exist technical innovations driving changes in the elasticity of substitution has recently been put forward by Klump and Preissler (2000, p. 52) who suggest that "[a]s far as invention of new methods of production is concerned, the elasticity of substitution as a measure of economic progress can, of course, be related to a society's capability to create and maintain a high rate of innovative activities." Finally, and most precisely, Bretschger (2005, p. 150) suggests that "all possibilities of substitution and, specifically, the effects technology exerts on promoting substitution have to be studied." Among others, these points quoted here provide a foundation for introducing technical change into the EoS.

In the subsequent section, we introduce technical change in the dis-

[2] There are papers which consider an EoS which changes over time, but they do not refer to non-renewable and renewable resources and also, the mechanisms therein do not have the "technological progress" flavour. These contributions are Miyagiwa and Papageorgiou (2005) – with a changing EoS between capital and labour, and Petith (2001) – with (broadly defined) land and labour.

tribution parameters (and thus, unit productivities) of the renewable and non-renewable resources. We show how this can be linked to the literature on biased technical change (e.g. Acemoglu, 2003). We assume that both resource inputs are subject to technical change and therefore increase their productivity, but one of the resources is subject to quicker technical change than the other resource. This allows us to compare our results to those present in the literature, notably in the papers which adapt Acemoglu's (2003) framework to renewable and non-renewable resources (e.g. Grimaud and Rouge, 2005) and in other related contributions (e.g. Di Maria and Smulders, 2004; Amigues *et al.*, 2006). To provide some slightly more intuitive perspective, if we take the Cobb-Douglas function as an example, then our extension permits to analyze the effect of changing Cobb-Douglas shares.

We introduce only exogenous technical change here. Several reasons can be put forward for this choice, apart from simplicity. Firstly, before trying to investigate under which circumstances a policy maker should invest more in the one type of technical change or the other, one has first to understand the generic dynamic effects of these changes. As we shall show, in many cases these generic effects are far from obvious.[3] Secondly, the exogenous technical change approach enables one to implicitly recover the value a benevolent policy maker would attach to the increases in the EoS and other technological parameters. This helps predict how much should be invested in e.g. R&D aimed at increasing the EoS in more sophisticated environments. For these reasons we believe that our simplified approach is not only justifiable, but also sufficient as a good first step.

The article is structured as follows. Section 3.2 lays out the simple Ramsey–Hotelling framework which we use for studying various types of technical change. Section 3.3 presents the benchmark case of no

[3] Our analysis reveals much more than a simple comparative-statics exercise would: not only are we able to assess the instantaneous impact of one-shot changes in technological parameters (easily inferred from our results), but also to quantify the relative power of these forces at different stages of development – with non-renewable resources being first relatively abundant and then increasingly scarce. Furthermore, our model, by assuming constancy of the rates of technological progress, provides a useful benchmark for further analyses which will have them endogenously determined and thus, presumably, varying over time.

technical change. In section 3.4 we introduce an increasing elasticity of substitution. In section 3.5 we compare these results to the results obtained when biased technological change is allowed for. Section 3.6 discusses the way in which the renewable resource works as a "potential backstop technology" in the model. Section 3.7 concludes.

3.2 The model

This section introduces the implications of a changing elasticity of substitution as well as technical change in the partial elasticities of a CES bundle consisting of a non-renewable and a renewable resource.

3.2.1 Technical change in the CES function

Throughout the analysis, we shall be using the standard constant-returns-to-scale CES production function, as derived in the seminal article by Arrow *et al.* (1961). We shall allow for technical change in its EoS, or alternatively, in its distribution parameter. The intermediate resource input R is produced with flows of non-renewable resource R_N and renewable resource R_R. They are combined according to:

$$R(t) = [\psi(t) R_N(t)^{\theta(t)} + (1 - \psi(t)) R_R^{\theta(t)}]^{\frac{1}{\theta(t)}}, \tag{3.1}$$

where the distribution factor is given by $\psi(t) \in [0, 1]$, and the elasticity of substitution, $\sigma(t) \in (0, +\infty)$, is related to the elasticity parameter $\theta(t)$ via

$$\theta(t) = \frac{\sigma(t) - 1}{\sigma(t)}. \tag{3.2}$$

The EoS $\sigma(t)$ and the distribution parameter $\psi(t) \in [0, 1]$ are allowed to change over time.

The non-renewable resource stock $0 < S_N(0) < \infty$ is extracted according to $\dot{S}_N(t) = -R(t) \leq 0$.[4]

For the renewable resource R_R, we assume that it arrives in *constant flows* over time, $R_R \equiv \text{const}$. One can thus interpret it as a fixed factor.

[4]Throughout the article we shall use $\dot{B}(t)$ to denote the time derivative of $B(t)$ and $\hat{B}(t)$ to denote its growth rate. For any function $G(B)$, $G'(B)$ denotes its first derivative, and $G''(B)$ its second derivative with respect to B.

This is a strong assumption which helps simplify the subsequent analysis. Yet, it is excusable: Dasgupta and Heal (1974) consider the renewable resource to be “a perfectly durable commodity which provides a flow of services at [a] constant rate” (Dasgupta and Heal, 1974, p. 19). Moreover, the renewable resource is a potential backstop technology: it becomes a backstop technology if the economy succeeds in shifting σ above unity.

Taking the growth rate of R and dropping time subscripts yields

$$\hat{R} = \epsilon_N \hat{R}_N + \epsilon_R \hat{R}_R + \epsilon_\sigma \hat{\sigma} + \epsilon_\psi \hat{\psi}, \tag{3.3}$$

where the partial elasticities, $\epsilon_i = \frac{\partial R}{\partial i}\frac{i}{R}$, for $i = R_N, R_R, \sigma, \psi$, are given by

$$\begin{aligned}
\epsilon_N &= \frac{\psi R_N^\theta}{R^\theta}, \\
\epsilon_R &= \frac{(1-\psi) R_R^\theta}{R^\theta}, \\
\epsilon_\sigma &= \frac{\psi R_N^\theta \ln R_N + (1-\psi) R_R^\theta \ln R_R - R^\theta \ln R}{(\sigma - 1) R^\theta}, \\
\epsilon_\psi &= \frac{\psi (R_N^\theta - R_R^\theta)}{\theta R^\theta}.
\end{aligned}$$

Both $\epsilon_N \in [0,1]$ and $\epsilon_R \in [0,1]$ function as shares in the traditional sense. This is due to the assumption of constant returns to scale in the CES function. $\epsilon_\sigma > 0$ implies that improvements in the EoS always increase output. We also have that $\epsilon_\psi > 0$ if $R_N > R_R$ ($\epsilon_\psi < 0$ if $R_N < R_R$) which suggests that technical change in the distribution factor should always go in the direction of the relatively abundant (i.e. more heavily used) resource input.[5]

This result is complementary to the one obtained by André and Cerdá (2005): optimizing subject to both resource inputs allows them to find a corner solution where it might be useful to delay the exploitation of the renewable resource in order to allow for its stock to grow

[5] One might be tempted to think that this finding is just artifact of our modeling assumption where a change in ψ changes the factor augmentation levels in opposite directions. This is, however, not the case. If one re-writes (3.1) as $R = [\alpha R_N^\theta + \beta R_R^\theta]^{1/\theta}$, with α and β being independent, then it is still more profitable to increase α than β ($\epsilon_\alpha > \epsilon_\beta$) if and only if $R_N > R_R$.

towards the maximum sustainable yield. Such result is not possible here, as we concentrate on the case with constant renewable resource flows. However, we are more precise in defining under which circumstances it is more profitable (in terms of total output) to increase the relative productivity of a given type of resource: it should be relatively abundant.

As can be easily inferred from the ϵ_ψ formula, decreases in non-renewable resource use (R_N) make the non-renewable resource-augmenting technical change relatively less profitable (ϵ_ψ falls), independently of the magnitude of the EoS. This result, calculated at the level of the production function, stands in contrast to some of the results found in the directed technical change literature, e.g. in Di Maria and Smulders (2004). There, depending on the EoS, this relationship may be reversed and reductions in resource use might actually be an incentive for firms to do *more* R&D on resource-saving technologies.[6]

CES and changing elasticity of substitution

The EoS σ gives the percentage change in relative quantities of used resources given a one percent change in their relative prices. If $\sigma = 0$ then the function is Leontief so the inputs are perfect complements, for $\sigma = 1$ we obtain the standard Cobb-Douglas form (as a limiting case), and if $\sigma = +\infty$ then the function is linear so the inputs are perfect substitutes. The EoS therefore gives information on the ease with which one can move along a given isoquant, and in that way it can either be understood as a measure of flexibility, efficiency (de La Grandville, 1989), or 'menu of choice' (Yuhn, 1991).

One cannot, however, claim that *all* new items on the menu of technology choice imply increases in the EoS: many new inventions would tend to shift the isoquant without altering its curvature, and some others might even pull the EoS down. A microfounded framework must be employed to delineate these types of innovations and identify the ones that effectively increase the EoS. In this respect, one can refer to the model built by Jones (2005b) and extended by Growiec (2006b; chapter 5 of

[6]Again, this result is robust to re-writing (3.1) as $R = [\alpha R_N^\theta + \beta R_R^\theta]^{1/\theta}$, with α and β independent. We obtain $\partial\epsilon_\alpha/\partial R_N > 0$, $\partial\epsilon_\beta/\partial R_N < 0$ irrespective of the EoS.

this thesis). The model allows the firms to select their production technology from the currently available menu, while each item on that menu is characterized by specific values of unit productivities of all production factors (there: capital and labour, here: non-renewable and renewable resources, respectively). Technological progress is then viewed as arrivals of new items on the technology menu, superior (Pareto-dominating in terms of respective factor productivities) to some of the items developed earlier. Jones (2005b) finds that if unit factor productivities of new developments augmenting the production factors are stochastically drawn from independent Pareto distributions, then the firm-level production function will converge to the Cobb-Douglas. In this benchmark case, ongoing technological change does not imply systematic increases in the EoS. Growiec (chapter 5 of this thesis), however, generalizes these results and finds that if factor-augmenting developments are drawn from Pareto distributions dependent according to the Clayton copula, and their mutual dependence is negative, then a CES production function with an EoS greater than one can be obtained.[7] Furthermore, the stronger is the negative correlation between factor-augmenting developments, the greater is the EoS. Within the same microfounded framework, Growiec (2006b) finds that the CES production function might also be derived when factor-augmenting developments are drawn from independent Weibull distributions. The relationship between the shape parameter of the Weibull distribution and the EoS is also negative: the heavier the tail of the distribution (the more probable it is to discover extremely useful ideas), the greater the EoS.

Unfortunately, we are not aware of any empirical studies attempting to quantify the scale of improvements in substitutability between inputs obtained thanks to some real-world technological developments. But there exists large anecdotal evidence for such innovations:[8] for exam-

[7]The Clayton copula illustrates the only pattern of dependence between Pareto-distributed factor-augmenting developments that is capable of generating the CES function. Negative correlation means that on average, highly labour-efficient technologies tend to be overly capital-consuming, and vice versa.

[8]One must be careful when interpreting these examples since most of the mentioned innovations resulted not only in changes in substitutability (on which we focus), but also in changes in the relative shares of non-renewable and renewable resources in production.

ple, the invention of (various sorts of) plastic wrappings improved the substitution possibilities between non-renewable resources (upon which plastics are based) and renewable resources (paper, cardboard, and textile wrappings) in the packaging industry. The invention of carbon fibre provided numerous industries with a durable material which could substitute out metals, in particular light alloys, in production. Chemical fertilizers can now be substituted for manure in agriculture. Plastic panels can be substituted for wood in the production of furniture. Solar panels and wind turbines can be substituted for the burning of fossil fuels in the production of electricity. In the production of shoes, we have been observing several increases in the substitutability between leather and rubber: initially, these two substances were Leontief complements, leather was used for the top and rubber for the sole of the shoe. Nowadays, leather can be substituted out by rubber and plastics in innumerable ways and there exists a vast multiplicity of shoe designs.

On the other hand, the introduction of new market segments, or of new products which are complementary to the hitherto existing ones, might lower the aggregate elasticity of substitution: usually just one way of assembling brand new products is known, implying the need to use resources in fixed proportions, and there exist no direct substitutes for them which could be produced taking the relevant inputs in different proportions. This is what happened, for example, when several household appliances were first introduced. Television sets had essentially the same architecture (picture tube, loudspeakers, antenna) and were produced from essentially the same substances for long decades before recent substitutes of LCD and plasma screens were introduced. Similarily, irons had always been made of iron (or steel) until the recent invention of teflon.

In the current paper, we view the EoS as a measure of technical efficiency. From the above discussion, it follows that the type of innovations we are considering here should be associated with lowering the correlation coefficient between non-renewable and renewable resource-augmenting ideas, or with increasing the skewness of their respective distributions. Only such developments effectively *widen the menu of useful technologies* and not just push the technological frontier forward. In result, not only are the respective unit factor productivities increased, but the EoS is improved as well.

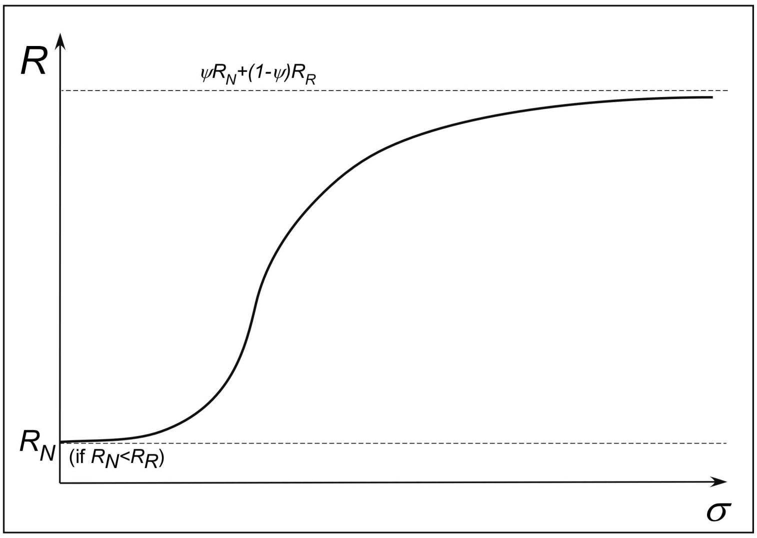

Figure 3.1: R as a function of σ

As can be seen in Figure 3.1, the higher the EoS, the larger the amount of output R for *any* given inputs R_N and R_R. The intermediate resource input production is an increasing function of the EoS ($\partial R/\partial\sigma > 0$), and follows a convex-concave shape.

The EoS pulls double duty: on the one hand it is a measure of efficiency, and on the other it determines the essentiality of the inputs. If $\sigma \leq 1$, then $R \to 0$ if any of the two inputs goes to zero. On the contrary, if $\sigma > 1$, neither of the inputs is essential any longer.

Bearing in mind the above discussion on relevant microfoundations, we shall assume exogenous technical change in the EoS here, as summarized by the following reference formula:

$$\sigma(t) = \sigma(0)e^{st}, \qquad \sigma(0) > 0, \tag{3.4}$$

which implies that the growth rate of σ is constant. Moreover, whatever the initial relationship between our inputs, they will become perfect substitutes in the limit. This assumption simplifies the subsequent analysis a lot but is admittedly strong, and in fact not necessary for at least one crucial qualitative results: If we take the EoS as a “dummy” for the essentiality of inputs then we only require σ to exceed unity from some point in time on. But, as is obvious from Figure 3.1, the consecutive efficiency gains from improvements in the EoS should not be neglected and, as we shall show, provide new qualitative insights for the short-run dynamics. Finally, allowing for the EoS to approach infinity allows our article to serve as a bridge to Tahvonen and Salo (2001) who assume perfect substitutability between non-renewable and renewable resources.[9]

CES and biased technical change

As an alternative to a changing elasticity of substitution, we shall introduce technical change in the distribution parameter ψ which captures the bias in factor-augmenting technical change.

The standard literature focuses mainly on efficiency improvements in the usage of the non-renewable resource (e.g. Scholz and Ziemes, 1999, Amigues et al., 2006). We are, however, going to focus on faster improvements in the usage of the renewable resource. Following this idea, we shall assume that technical change affects the relative share of the two inputs by changing the distribution factor as follows:

$$\psi(t) = \psi(0)e^{zt}, \qquad \psi(t) \in [0,1], \quad \psi(0) \in (0,1), \tag{3.5}$$

where the growth rate of ψ is a constant $z \leq 0$.

This simplistic assumption allows us to construct a useful benchmark for the analysis of biased technical change. If one would like to allow technological change in ψ to be determined endogenously, though, the case where a flip in the direction of technical change appears at the moment when $R_N = R_R$ should be of special interest, since $\epsilon_\psi > 0$ if $R_N > R_R$ and $\epsilon_\psi < 0$ if $R_N < R_R$.

The effect of changing ψ on R is

$$\frac{\partial R}{\partial \psi} = \frac{1}{\theta}\left(R_N^\theta - R_R^\theta\right)\left[\psi R_N^\theta + (1-\psi)R_R^\theta\right]^{\frac{1-\theta}{\theta}}, \tag{3.6}$$

[9] We have omitted the important issues of normalization here (cf. de La Grandville, 1989, Klump and Preissler, 2000). These issues have been neglected in most prior literature but are clearly important if one does not want to do inter-family comparisons of CES functions. A proof that our approach is in fact consistent with normalization is available from the authors upon request.

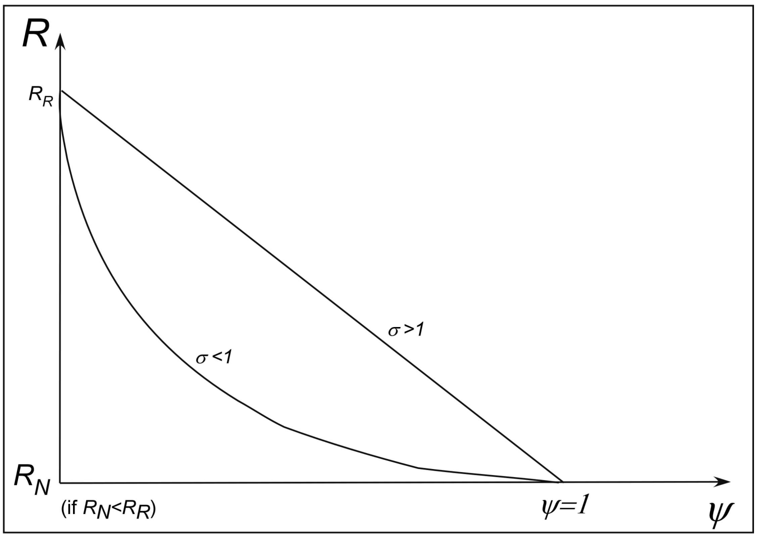

Figure 3.2: R as a function of ψ

implying that $\partial R/\partial\psi > 0$ if $R_N > R_R$, and $\partial R/\partial\psi < 0$ if $R_N < R_R$. Figure 3.2 illustrates the effect of changing ψ for the case of $R_N < R_R$ and two different values of σ. The lower the possibility to substitute the stronger the initial effect of changes in ψ. In the extreme, for Leontief inputs, a marginal change from $\psi = 0$ to $\psi > 0$ will result in a reduction in the intermediate output from R_R to R_N (given $R_N < R_R$). Biased technical change can thus be viewed as varying the relative productivity of R_R and R_N. The marginal rate of technical substitution is given by $MRTS = \frac{\psi}{1-\psi}\left(\frac{R_N}{R_R}\right)^{-1/\sigma}$: changes in ψ affect the slope of the isoquants of R leaving their curvature intact.

3.2.2 The Ramsey–Hotelling framework

We shall now embed our intermediate resource input R with technical change in σ or ψ in a simple Ramsey-type model with an infinite planning horizon, where a representative agent maximizes discounted utility subject to an equation of motion of the non-renewable resource stock.

Formally, this means that our infinitely-lived representative agent maximizes

$$\begin{aligned}
\max_{\{R_N(t)\}_{t=0}^{\infty}} & \quad \int_0^{\infty} U(Y(t))e^{-\rho t}dt, \qquad \text{subject to:} \\
\dot{S}_N(t) &= -R_N(t), \\
R_R &\equiv \text{const}, \\
Y(t) &= F(A(t), L, R_N(t), R_R) = A(t)L^{1-\beta}R(t)^{\beta}, \\
R(t) &= [\psi(t)R_N(t)^{\theta(t)} + (1-\psi(t))R_R^{\theta(t)}]^{\frac{1}{\theta(t)}}.
\end{aligned}$$

At this point we assume that $U(Y) = Y^{1-\gamma}/(1-\gamma)$, where $\gamma = -YU''(Y)/U'(Y) \in (0,\infty)$[10] is the inverse of the intertemporal elasticity of substitution. We abstract from population growth and normalize $L \equiv 1$.

Technical change is decomposed into three components: factor-neutral technical change $A(t) = A(0)e^{gt}$ with $g \geq 0$, technical change in the EoS, $\sigma(t) = \sigma(0)e^{st}$ with $s \geq 0$, and the bias in technical change due to $\psi(t) = \psi(0)e^{zt}$ with $z \leq 0$. On top of this, we assume that either $s = 0$ with $z < 0$ or $z = 0$ with $s > 0$ for the sakes of normalization and the transparency of results.

Since no savings decision is allowed for, all output is immediately consumed and thus $C(t) = Y(t)$ for all t.

The simple, "bare-bones" model which we use features a number of strong simplifying assumptions, such as constancy of the renewable resource flow over time, exogenous technical change, and the neglect of capital accumulation. On the other hand, we agree to pay such a price, because in return we get a clear understanding of the dynamics in all, even the most complex cases. This makes our article a good starting point for further analyses: relaxing several of the assumptions would provide interesting questions for future research. In particular, we point out that our results should be verified in an optimal growth framework which fully endogenizes research in both resource inputs. Also the mechanisms behind technical change in our basic model should be provided with a formal treatment.

[10] $\gamma > 1$ implies that utility is bounded from above. $\gamma \geq 1$ implies that the utility of zero consumption is minus infinity, enforcing positive consumption at all times. If $\gamma < 1$ then zero consumption at some (later) t is not automatically ruled out.

The first-order conditions give us (see Appendix 3.8.1 for more details):

$$\widehat{\left(\frac{\partial U(Y)}{\partial R_N}\right)} = \rho. \tag{3.7}$$

This is the Ramsey-Hotelling condition that characterizes the interior solution of the optimization problem. Rewriting, we obtain that

$$\hat{Y} = \frac{\hat{F}_N - \rho}{\gamma}, \tag{3.8}$$

with $F_N = \partial F / \partial R_N$, which states that the growth rate of income, $\hat{Y}$, is positive if the growth rate of the marginal product of the non-renewable resource, $\hat{F}_N$, exceeds the discount rate.

In its core, this result is similar to the outcome of the "cake-eating" model (see e.g. Dasgupta and Heal, 1974). In the cake-eating model, the marginal product of the non-renewable resource is a function of two variables only: the rate of depletion and the rate of factor-neutral technical change. Hence, without large enough factor-neutral technical change, the growth rate of the marginal product of the non-renewable resource will be negative for all times and income will decline continuously. Analogously, income growth can be positive in the Ramsey–Hotelling model only if technical change is fast enough to keep the growth rate of the marginal product of the non-renewable resource above the discount rate. However, as we decompose technical change into three different kinds, we obtain more possibilities to achieve sustainable production. The way in which the model's results are changed due to the introduction of technical change in the elasticity of substitution and biased technical change through the distribution parameter will become clear in the subsequent sections.

Solving equation (3.8) for $\hat{R}_N$, we obtain the optimal growth rate of the non-renewable resource extraction in terms of other variables:

$$\begin{aligned} \hat{R}_N &= \frac{(1-\gamma)g - \rho + [((1-\gamma)\beta - \theta)\epsilon_\sigma - \frac{1}{\sigma}(\ln R - \ln R_N)]s}{(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N} + \\ &+ \frac{[((1-\gamma)\beta - \theta)\epsilon_\psi + 1]z}{(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N}. \end{aligned} \tag{3.9}$$

As is proved in Appendix 3.8.2, the denominator of the above expression is always positive. We also notice that all three kinds of technical change affect the dynamic path of the optimal resource extraction rate R_N.

3.2.3 Comparative statics

We shall now turn to the comparative statics which provide further insights into the impact of certain parameters on the growth rate of income. Along the optimal growth path we obtain the following comparative statics:

$$\frac{\partial \hat{Y}}{\partial \rho} = -\frac{\beta \epsilon_N}{(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N} < 0, \tag{3.10}$$

$$\frac{\partial \hat{Y}}{\partial g} = \frac{(1-\theta) + \theta \epsilon_N}{(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N} > 0, \tag{3.11}$$

$$\frac{\partial \hat{Y}}{\partial s} = \frac{\beta(1-\theta)[\epsilon_\sigma - \epsilon_N(\ln R - \ln R_N)]}{(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N}, \tag{3.12}$$

$$\frac{\partial \hat{Y}}{\partial z} = \frac{\beta[\epsilon_N + \epsilon_\psi(1-\theta)]}{(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N}. \tag{3.13}$$

Along the optimal path, the discount rate affects the growth rate of income always negatively (Eq. (3.10)). Clearly, the less we care about the future the more resources we use up now. This brings about a larger current level of income but less growth potential since less resources are left for use later on.

The growth rate of income is unambiguously positively related to the rate of factor-neutral technical change g (Eq. (3.11)). This result also carries forward from the standard literature.

We have already shown that changes in the distribution factor should always go towards the kind of resource which is more abundant and thus more important for production. Equation (3.13) confirms it by saying that increases in z have a positive effect of income growth if $\epsilon_\psi > -\frac{\epsilon_N}{1-\theta}$. Hence, a sufficient condition for a positive effect of z on income growth is $\epsilon_\psi > 0$, i.e. if the more abundant resource is subject to relatively faster technical change.

The effect of s on income growth depends on whether more renewable or more non-renewable inputs are used in production. If more non-renewable inputs are used then the effect of s on income growth is unambiguously positive. If more renewable resources are used, then the effect of s on income growth depends on the importance of the non-renewable resource for production relative to the partial elasticity of σ and cannot be unambiguously signed. This observation seems then crucial for the transition period. This gives us another reason for not only looking at a fixed EoS, but especially points at the significance of looking at a changing EoS in the short-run.

3.3 The benchmark case

3.3.1 The long run

The properties of the benchmark model with no technical change in σ ($s = 0$) and ψ ($z = 0$) are similar to the ones established in previous literature (see e.g. Dasgupta and Heal, 1974). We distinguish three important sub-cases here: $\sigma > 1$, $\sigma = 1$, and $\sigma < 1$.

Case $\sigma > 1$. If non-renewable and renewable resources are gross substitutes, then the following asymptotic results hold:

$$Y = F(A, L, R_N, R_R)|_{R_N=0} > 0$$
$$\lim_{t\to\infty} \epsilon_N = 0 \qquad \lim_{t\to\infty} \epsilon_R = 1$$
$$\Rightarrow \quad \lim_{t\to\infty} \hat{Y} = g.$$

As the non-renewable resource is not essential for production, it becomes a virtually negligible input when it approaches zero. Hence, income growth can stay positive forever if only there is factor-neutral technical change, $g > 0$.

Case $\sigma = 1$. In the Cobb-Douglas case the non-renewable resource is essential for production and

$$Y = F(A, L, R_N, R_R)|_{R_N=0} = 0$$
$$\epsilon_N = \psi \qquad \epsilon_R = 1 - \psi$$
$$\lim_{t\to\infty} \hat{R}_N = \frac{(1-\gamma)g - \rho}{1-(1-\gamma)\beta\psi}$$
$$\Rightarrow \quad \lim_{t\to\infty} \hat{Y} = \frac{g}{1-(1-\gamma)\beta\psi} - \frac{\beta\psi\rho}{1-(1-\gamma)\beta\psi}.$$

We see that in this case, the gradual depletion of the non-renewable resource pulls down the long-run growth rate of the economy. This is especially apparent when one compares this case to the $\sigma > 1$ case.

Case $\sigma < 1$. If the two types of resources are gross complements, we have that

$$
\begin{aligned}
Y &= F(A, L, R_N, R_R)|_{R_N=0} = 0 \\
\lim_{t\to\infty} \epsilon_N &= 1 \qquad \lim_{t\to\infty} \epsilon_R = 0 \\
\lim_{t\to\infty} \hat{R}_N &= \frac{(1-\gamma)g - \rho}{1-(1-\gamma)\beta} \\
\Rightarrow \quad \lim_{t\to\infty} \hat{Y} &= \frac{g}{1-(1-\gamma)\beta} - \frac{\beta\rho}{1-(1-\gamma)\beta}.
\end{aligned}
$$

The partial elasticity of the non-renewable resource tends to one because R_N,as it approaches zero, drives the output if it is a gross complement to R_R. In this case the rate of resource depletion provides an even stronger drag on per capita income than in the Cobb-Douglas case.

The important transversality condition, guaranteeing finiteness of the objective integral, is for all three cases $(1-\gamma)g < \rho$. It is automatically satisfied if $\gamma > 1$ and $g \geq 0$.

As far as long-run results are concerned, the benchmark case confirms the decisive importance of the EoS in the production process, a result that carries forward from the classical literature.

The more patient the representative agent, i.e. the smaller is ρ, the more important is the future for her and therefore the larger the growth rate of resource extraction. If we assume away the possibility of depletion in finite time, then this implies that initially smaller amounts of resources will be utilized.

If zero consumption is not penalized by infinite disutility (which is the case if $\gamma < 1$), finite-time depletion of the non-renewable resource is possible irrespectively of whether the resource inputs are gross substitutes or gross complements. This case is quite counterintuitive and empirically implausible. If $\gamma \geq 1$, on the other hand, then finite-time depletion will be the case only if $\sigma > 1$.

3.3.2 The short run

We study the optimal path of income in the benchmark case using numerical simulations. The results are presented in Figure 3.3.[11]

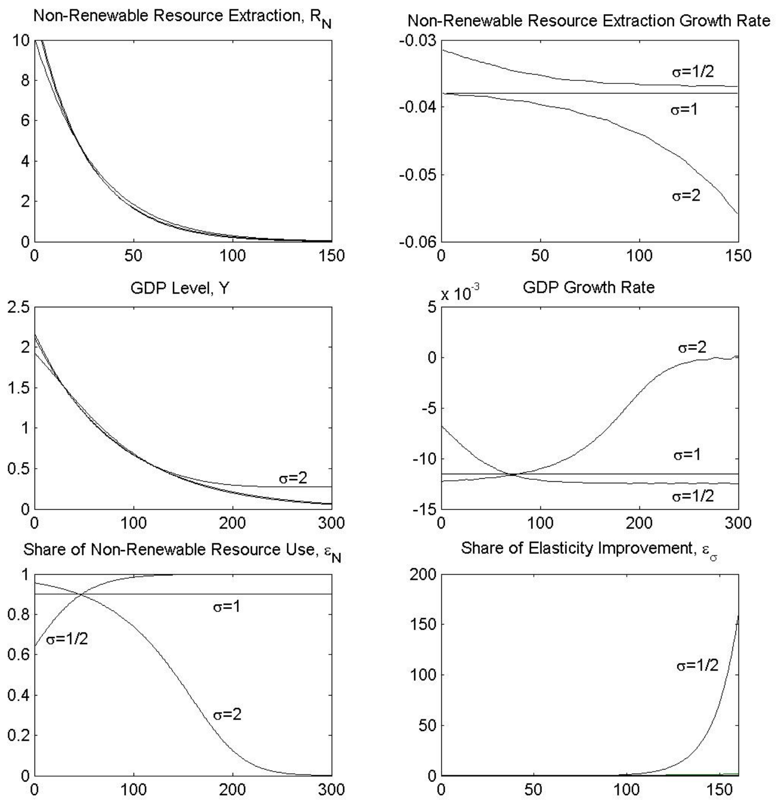

Figure 3.3: Benchmark case, $s = 0$, $z = 0$, $g = 0$

In the case $\sigma > 1$ or $\gamma < 1$, there is finite-time depletion of R_N but it happens late enough not to affect the short-run results substantially.

The interpretation of the above result is as follows. The discounted

[11]The simulation procedure is described in the Appendix 3.8.3. Unless stated otherwise, we use the following parameter choices for all simulations: $\beta = 1/3$, $\gamma = 2$, $R_R = 2$, $\rho = 0.05$, $\sigma(0) = 0.1$, $\psi(0) = 0.9$, $S(0) = 300$. Some of these parameters can be easily verified (γ, β, ρ), others have been chosen freely.

utilitarian criterion regards future consumption as less important than current consumption, wherefore most non-renewable resources will be extracted when utility seems most valuable. The simulations show that in case the non-renewable resource is not essential for production ($\sigma = 2$), the share of R_N in production is initially very large, suggesting that most non-renewable resources are used initially, and then declines to zero over time. Thus the level of GDP is initially larger for the case of high EoS in comparison to the case of $\sigma \leq 1$, as technical efficiency does not constrain the intermediate input R by complementarity problems. The GDP growth rate is initially higher in the case of complementarity as at first more non-renewable resources are conserved as production is depending on those resources later. Moreover, the amount of the non-renewable resource used in production is divided more equally over time the lower is σ, due to its essentiality for production.[12] Nevertheless though, if the non-renewable resource and the renewable one are gross complements in production, the decrease in the amount of non-renewable resources available decreases the overall resource bundle that can finally be used in production. In such case, the non-renewable resource is driving the size of the resource bundle R, implying that the share of the non-renewable resource in the resource bundle will go to one. Hence, the (negative) GDP growth rate will become smaller and smaller. In contrast to this, if the non-renewable resource is inessential ($\sigma > 1$) then, as less and less of it remains available, its share in the resource bundle will drop to zero. In this sense, the non-renewable resource allows greater GDP levels as long as it is available, but as soon as it gets depleted, GDP goes to the level that it would have had, had it been produced without the non-renewable resource.

3.3.3 The possibility of finite-time depletion

If the non-renewable resource is not essential for production ($\sigma > 1$), it will be depleted in finite time. Formally, one could write the finite-time

[12]For the case $\sigma = 2$, R_N gets depleted around $T^* = 160$ which is not visible to the naked eye.

depletion problem as a problem of choosing an optimal T in:

$$\begin{aligned}
\max_{\{R_N(t)\}_{t=0}^T, T} \quad & \mathcal{U}_0 = \int_0^T U(Y(t))e^{-\rho t}dt + \int_T^\infty U(\bar{Y}(t))e^{-\rho t}dt, \quad \text{s. t.} \\
\dot{S}_N(t) &= -R_N(t), \\
R_R &\equiv \text{const}, \\
Y(t) &= F(A(t), L, R_N(t), R_R) = A(t)L^{1-\beta}R(t)^\beta, \\
R(t) &= [\psi(t)R_N(t)^{\theta(t)} + (1-\psi(t))R_R^{\theta(t)}]^{\frac{1}{\theta(t)}}, \\
S_N(T) &= 0,
\end{aligned}$$

where $\theta > 0$, and $\bar{Y}$ is the exogenously given amount of output produced with renewable resources only: $\bar{Y}(t) = (1-\psi)^{\beta/\theta} R_R^\beta A(0)e^{gt}$.

The problem is solved in two steps: first, one chooses the optimal path of non-renewable resource extraction given T, and then one chooses T knowing these "response functions" such that the sum of both integrals, $\mathcal{U}_0$, is maximized. The first step gives the already discussed condition (3.9) subject to a boundary condition $S_N(T) = 0$. Because the dynamic equation (3.9) is not solvable, we shall again resort to numerical simulations here to approximate the optimal depletion time T^*. The results of the numerical exercise, which takes $\gamma = 2$ and $g = 0$, are summarized in Table 3.1. In particular, for the case $\sigma > 10$, the optimal depletion time is $T^* = 110$ as presented in Figure 3.4.

σ	1	1.1	1.5	2	10	40	100	$+\infty$
T^*	$+\infty$	>300	180	160	110	≈ 100	≈ 100	0

Table 3.1: Optimal exhaustion time T^* as a function of σ

3.4 Increasing flexibility

3.4.1 The long run

We shall now analyze the effect of allowing for technical change in the EoS σ, but not in the distribution parameter ψ, such that $s > 0$, $z = 0$.

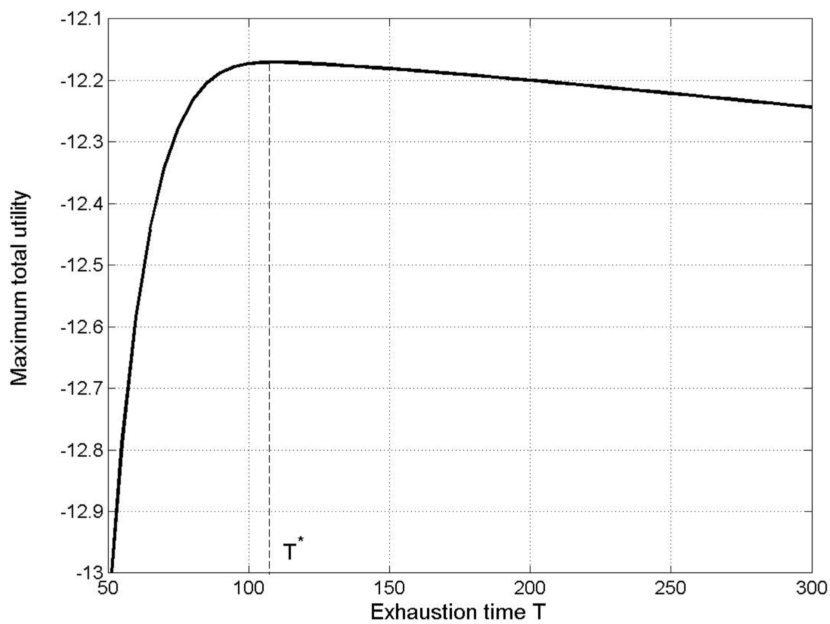

Figure 3.4: Exhaustion time in the case $\sigma = 10$

The long-run asymptotics are straightforward:

$$Y = F(A, L, R_N, R_R)|_{R_N=0} > 0$$

$$\lim_{t\to\infty} \sigma = \infty \Rightarrow \lim_{t\to\infty} \theta = 1$$

$$\lim_{t\to\infty} \epsilon_N = 0 \qquad \lim_{t\to\infty} \epsilon_R = 1 \qquad \lim_{t\to\infty} \epsilon_\sigma = 0$$

$$\Rightarrow \quad \lim_{t\to\infty} \hat{Y} = g$$

Furthermore, there is no need to assume unbounded growth in the EoS. We do this for simplicity here but we would have obtained the same asymptotic growth rate of income if we had assumed that σ grew until some time t_0, after which $\sigma(t) > 1$, i.e. non-renewable resources were not essential for production any more. Hence, these asymptotic properties hold as long as the EoS manages to cross the "magical" barrier of one.

The required transversality condition is again $(1 - \gamma)g < \rho$ which is automatically satisfied if $\gamma > 1$ and $g \geq 0$.

3.4.2 The short run

In the case of increasing EoS, we know that from a certain time $t_0 \geq 0$ onwards, given the EoS improved to exceed one, the resource inputs will be gross substitutes. Thus, the non-renewable resource will necessarily be depleted in finite time. This result is intuitive because perfect substitutability between non-renewable and renewable resources allows us to utilize the non-renewable resource without compromising the productivity of the renewable one in any way.

The results of the simulative exercise are presented in Figure 3.5. Vertical dashed lines indicate the moment in which the elasticity of substitution crosses one.

As the elasticity of substitution increases it is at first optimal to utilize non-renewable resources at an increasing rate. This is because initially, the improvements in technical efficiency allow for large increases in the intermediate input R.[13] Hence, this allows for an increasing growth rate of income. However, these initial exponential-like increases in R, due to the increasing EoS, level off quickly, such that consecutive improvements in σ increase the overall resource bundle only slightly. It is now optimal to slow down the extraction of the non-renewable resource, as it is still essential for production. At this time, the model is rather similar to the benchmark case with an essential resource and produces a negative growth rate of income. At one point in time, the technical efficiency will have improved substantially enough, so that the non-renewable resource will not be essential for production any more. The model then traces the benchmark case again, but with a non-essential non-renewable resource. However, continued improvements in the technical efficiency parameter result in further improvements in the intermediate good production process. Therefore, the economic growth rate can be positive. As the consecutive improvements in efficiency increase the intermediate resource by less and less, and as the non-renewable resource stock is depleted in some finite time after σ crosses unity ($T^* > t_0$; in the simulated case, we obtain $T^* = 170$),[14] the growth rate of income

[13]The initial level of $\sigma(t)$ is set at $\sigma(0) = 0.1$, which corresponds to the strongly increasing part of Figure 3.1.

[14]Again, this fact is not visible to the naked eye in the figure. R_N approaches zero smoothly and approaches the vicinity of zero much earlier than it actually takes the

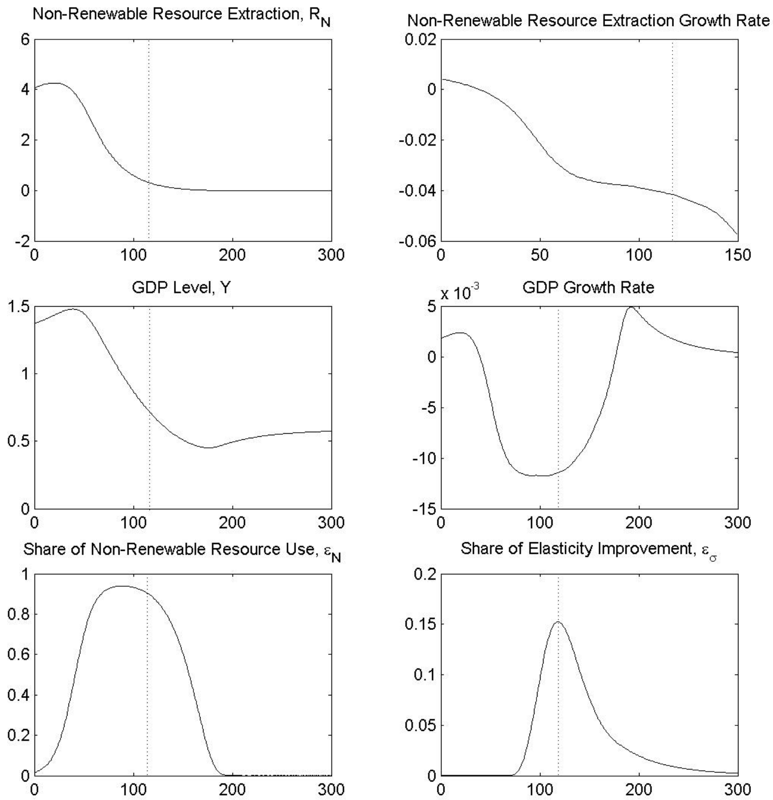

Figure 3.5: Increasing flexibility case, $s = 0.02$, $z = 0$, $g = 0$

tends to zero from above.

Generally speaking, one can notice that the changing EoS pulls double-duty. On the one hand, it reflects technical efficiency, and on the other hand, it reflects the essentiality of the resource. We find that improvements in technical efficiency through increases in the EoS are more relevant for the short-run, whereas essentiality determines long-run behaviour.

3.4.3 Finite-time depletion

As a consequence of the fact that R_N becomes eventually inessential, we obtain that finite-time depletion is certain. Numerical computations of the optimal depletion time T^* have been summarized in Table 3.2. We emphasize that all optimal depletion times are strictly (and significantly) greater than the time in which σ crosses unity, which is in our case $t_0 = 115$.

s	0	0.005	0.01	0.02	0.03	0.04
T^*	$+\infty$	>300	260	170	≈ 160	≈ 160

Table 3.2: Optimal exhaustion time T^* as a function of s

3.5 Biased technical change

3.5.1 The long run

We shall now compare the case of technical change in the EoS with the biased technical change case. We assume $s = 0$, $z < 0$, and $g \geq 0$ which implies that technical change improves the efficiency of the renewable resource more quickly than it improves the efficiency of the non-renewable resource. A priori, it may seem counterintuitive to suggest this as a way out of the long-run constraint imposed by the essentiality of a non-renewable resource. However, having Cleveland and Ruth's (1997) observation in mind, namely that technical change has not been sufficiently strong to offset the effects of depletion, we believe that looking at technical change from a different angle, through its effect on the distribution

zero value.

parameters, deserves some scrutinization. We shall show and explain why and when this kind of biased technical change is useful.

We shall be dealing with three cases: $\sigma > 1$, $\sigma = 1$, and $\sigma < 1$, where the last case is further divided into three sub-cases of fast, medium, and slow biased technical change, delineated by the expression $z - \theta(\rho - (1-\gamma)g)$ being negative, zero, and positive, respectively.

Case $\sigma > 1$. If the resource inputs are gross substitutes, then the following asymptotic properties hold:

$$Y = F(A, L, R_N, R_R)|_{R_N=0} > 0$$
$$\lim_{t\to\infty} \epsilon_N = 0 \qquad \lim_{t\to\infty} \epsilon_R = 1 \qquad \lim_{t\to\infty} \epsilon_\psi = 0$$
$$\Rightarrow \quad \lim_{t\to\infty} \hat{Y} = g.$$

As the renewable resource can be easily substituted for the non-renewable resource, its flow into the intermediate resource input alone can guarantee positive long-run output. In this case, biased technical change has no qualitative effect on the asymptotic results and the non-renewable resource will be depleted in finite time just as in the benchmark case.

The required transversality condition is again $(1-\gamma)g < \rho$ which is automatically satisfied if $\gamma > 1$ and $g \geq 0$.

Case $\sigma = 1$. For the Cobb-Douglas case, it turns out that biased technical change with the share of the non-renewable resource $\epsilon_N = \psi \to 0$ is enough to guarantee positive output forever even in the absence of factor-neutral technical change ($g = 0$) but in contrast to the $\sigma > 1$ case, R_N will be depleted only in infinite time:

$$Y = F(A, L, R_N, R_R)|_{R_N=0} = 0$$
$$\lim_{t\to\infty} \epsilon_N = 0 \qquad \lim_{t\to\infty} \epsilon_R = 1 \qquad \lim_{t\to\infty} \epsilon_\psi = 0$$
$$\lim_{t\to\infty} \hat{R}_N = (1-\gamma)g + z - \rho \qquad \lim_{t\to\infty} \hat{Y} = g$$

The transversality condition is again $(1-\gamma)g < \rho$ which is automatically satisfied if $\gamma > 1$ and $g \geq 0$.

Case $\sigma < 1$. If the resource inputs are gross complements, the speed of technical change is crucial for the long-run results. We find that three distinct regimes may emerge. In all of them, though, only infinite-time depletion is possible.

FAST TECHNICAL CHANGE: $z < \theta(\rho - (1-\gamma)g)$. This assumption implies that the technical change is quick enough to fully compensate for the declining flow of non-renewable resources. This condition is more likely to be satisfied the weaker the complementarity of resource inputs (the greater the negative θ). Analogously, the greater the degree of complementarity between the resource inputs, the faster must be the increase in the relative productivity of the renewable resource. We obtain:

$$Y = F(A, L, R_N, R_R)|_{R_N=0} = 0$$
$$\lim_{t\to\infty} \epsilon_N = 0 \qquad \lim_{t\to\infty} \epsilon_R = 1 \qquad \lim_{t\to\infty} \epsilon_\psi = 0$$
$$\lim_{t\to\infty} \hat{R}_N = \frac{(1-\gamma)g + z - \rho}{1-\theta} \qquad \lim_{t\to\infty} \hat{Y} = g$$

MEDIUM TECHNICAL CHANGE: $z = \theta(\rho - (1-\gamma)g)$. In this knife-edge case, the speed of technical change is just enough to guarantee that the share of the non-renewable resource, $\epsilon_N \to \bar{c}$ where $\bar{c} \in (0,1)$. The depletion of R_N provides a drag on the long-run growth rate of the economy. These results are similar to the less parsimonious case of slow technical change, described below. The details are available from the authors upon request.

SLOW TECHNICAL CHANGE: $z > \theta(\rho - (1-\gamma)g)$. If technical change is too slow to fully compensate for the shrinking flow of R_N, then its depletion exerts an unambiguously harmful effect on the long-run growth rate of the economy. The asymptotic results for this case are given by:

$$Y = F(A, L, R_N, R_R)|_{R_N=0} = 0$$
$$\lim_{t\to\infty} \epsilon_N = 1 \qquad \lim_{t\to\infty} \epsilon_R = 0 \qquad \lim_{t\to\infty} \epsilon_\psi = \frac{1}{\theta} < 0$$
$$\lim_{t\to\infty} \hat{R}_N = \frac{(1-\gamma)g + \frac{((1-\gamma)\beta)}{\theta} z - \rho}{1-(1-\gamma)\beta}$$
$$\Rightarrow \quad \lim_{t\to\infty} \hat{Y} = \frac{g - \beta\rho + \frac{\beta}{\theta} z}{1-(1-\gamma)\beta}.$$

For this sub-case, the transversality condition is different: the long-run impact of the bias in technical change coupled with the gradual depletion of the non-renewable resource has to be accounted for. It is now $(1-\gamma)(g+\frac{\beta}{\theta}z)<\rho$. However, it is still automatically satisfied if $\gamma>1$ and $g\geq 0$.

These results allow us to draw conclusions which are a little more precise than those of recent research. For example, André and Cerdá (2005) derive equations describing the optimal dynamic evolution of the resource input ratio and the optimal output path, and then conclude from these that the "equations (...) express, in a mathematical way, the interest (and, in the long run, the need) to promote the research and use of renewable energy sources (...) to substitute nonrenewable energies (...) from a sustainability perspective." Our analysis allows to find that biased technical change has its merits when it is directed to the *more abundant resource input*. This will certainly imply that in the short run it would be optimal to invest more in research promoting the non-renewable resource rather than the renewable one, and only much later should the bias be reversed.

One more finding should be emphasized here. In the case $\sigma<1$, there is one more "bifurcation" set of parameter values, which bounds away from the other cases of qualitatively different dynamic behaviour of the model. Indeed, authors working within the Acemoglu (2003) framework deal with (endogenously determined) biased technical change of the type which we call "slow"; their models typically do not allow for a jump to the regime where biased technical change is "fast".[15]

Summarizing, the long-run results for the biased technical change case stand in stark contrast to the results for the increasing flexibility case where technical change was unambiguously good and its desirable direction could be only towards higher substitutability. Moreover, we find here that the actual *value* of substitutability σ plays a more important role for the long-run dynamics even than the *growth rate* of the distribution parameter ψ, denoted as z. A one-shot improvement in the EoS could be more beneficient for the economy than perpetual growth in the productivities of resource inputs.

[15]To keep our results closer to this strand of literature, one could redefine $\psi\equiv a^{\theta}$. Then, $z=\theta\hat{a}$ with $\hat{a}>0$, and the condition for "fast" biased technical change becomes $\hat{a}>\rho-(1-\gamma)g$.

3.5.2 The short run

The results of the simulative exercise for the case of biased technical change are presented in Figures 3.6 and 3.7. In the first of these figures, we distinguish between the three major cases of $\sigma > 1$, $\sigma = 1$ and $\sigma < 1$. In the second one, we analyze the difference in dynamics stemming from the fact that the speed of biased technical change z may be either slow or fast (in figure 3.6, it is slow).

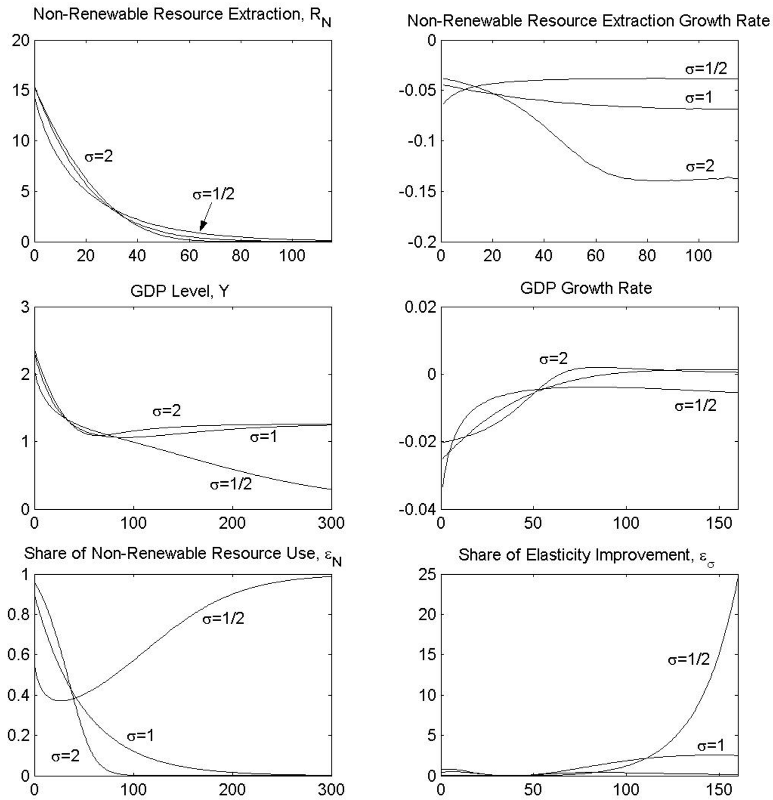

Figure 3.6: Biased technical change case, $s = 0$, $z = -0.02$, $g = 0$

We can interpret these simulation results as follows. The discounted utilitarian criterion leads the infinitely-lived agent to perceive her current income as more important for the creation of utility, so she initially utilizes most non-renewable resources. As in all previous cases, we observe the crucial role played by the EoS. If $\sigma < 1$ then the relatively scarce resource drives production. Conversely, if $\sigma > 1$, then the relatively more abundant resource will be decisive for the amount of the intermediate good. We see this in the picture of ϵ_N. Initially, more non-renewable resources are utilized in production. For the case of $\sigma < 1$, this implies that the renewable resource constrains the amount of intermediate output, and therefore relatively quicker improvements in the marginal product of the renewable resource (as captured by $z < 0$) reduce the drag on intermediate good output imposed by the renewable resource. In contrast to this, if the non-renewable resource is the largest input in the intermediate good and $\sigma > 1$, then the non-renewable resource drives the intermediate good output. Hence, if the importance of the renewable resource now increases relatively to the non-renewable resource, then the effect on GDP growth is negative. These initial effects are reversed once the relative importance of the resource inputs changes due to the depletion of the non-renewable resource stock. We also see that ϵ_σ explodes to infinity in the $\sigma < 1$ case, indicating that improvements in the EoS would be increasingly valued as R_N gets depleted.

Further simulations show that in case the resources are gross complements ($\sigma < 1$), the speed of z is vital in order to allow for a non-decreasing long-run income. For the case of $\sigma = 1/2$, our simulations demonstrate that the assumed speed of biased technical change, $z = -0.02$, is not sufficient to outweigh the decrease in the non-renewable resource flow. Hence, the GDP level will tend to zero, a result in line with the benchmark case. However, if the speed of z outweighs the reduction in the flow of non-renewable resources (in our case, $z = -0.06$), then a positive constant GDP level can be attainable. The speed of z must be such that the renewable resource becomes quickly enough more and more important to production in order to compensate for the decreasing quantities of the non-renewable resource available, $z < \theta(\rho - (1 - \gamma)g)$. Only then is it guaranteed that technical change wins the race against the gradual disappearance of the essential production input.

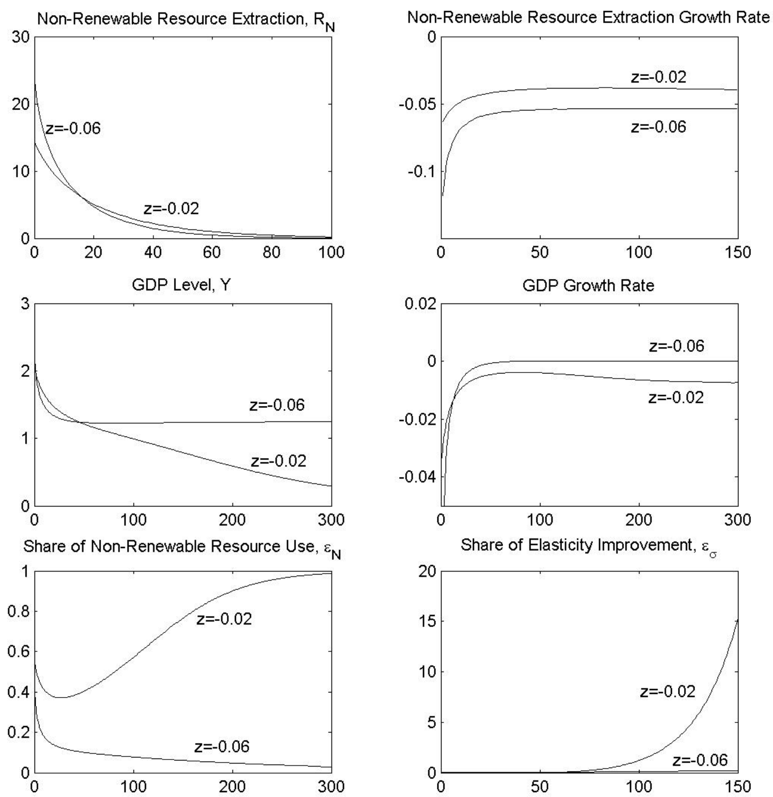

Figure 3.7: Biased technical change case, $s = 0$, $g = 0$. Dependence on the magnitude of the bias in technical change, z.

3.6 Renewable resource as a potential backstop technology

One further corollary may be drawn from our analysis. We see that the renewable resource works as a *potential* (or *conditional*) backstop technology in this model. It is a backstop technology (cf. Nordhaus, 1973)[16] because it is a technology which allows to produce positive output forever without the need to use non-renewable resources as well. But it is only potential, because the economy has first to assure that the EoS between the two resource inputs exceeds one. Of course, this need not be the case: even if we allow for factor-augmenting technical change in both types of resources, σ may well stay below unity. The benchmark case with $\sigma \leq 1$ and the biased technical change case with $\sigma < 1$ illustrate the trouble with such a situation.

Whereas in the benchmark case and in the biased technical change case, the renewable resource works as a conditional backstop technology (the condition being that $\sigma > 1$), an EoS increasing over time in a way which ensures crossing unity gives it the characteristic of a *dynamically emerging* backstop technology. In such case, this backstop technology is known from the start, the "application" of this technology is certain, but the timing depends crucially on the growth rate of EoS.

We also find that once the conditions for the usage of renewable resources as a backstop technology are satisfied, it becomes profitable to deplete the non-renewable resource in finite time.

3.7 Conclusion

This article investigates two non-standard ways of looking at technical change: technical change in the elasticity of substitution as well as technical change in the partial elasticities. We use a simple Ramsey–Hotelling model to compare the long- and short-run implications of the two types of technical change.

[16]Nordhaus (1973), pp. 547-548, writes: "The concept (...) is the backstop technology, a set of processes that (1) is capable of meeting the demand requirements and (2) has a virtually infinite resource base".

These approaches follow recent suggestions found in the literature, i.e. Bretschger (2005, p. 150) suggests that “all possibilities of substitution and, specifically, the effects technology exerts on promoting substitution have to be studied”.

Continued substitutability improvements have two effects. One, they improve upon the technical efficiency of the overall resource bundle. Two, they help render the non-renewable resource inessential for production from some time on. Our analysis shows that a policy maker will take the changes in the elasticity of substitution into account when choosing the optimal time path of non-renewable resource extraction. Here we focus on two contributions: the effect of an increasing EoS on the short-run where the main driving force are continuous improvements in the EoS, and on the long-run, where the most important factor is the EoS as a “dummy” for essentiality.

Biased technical change means that both resource inputs are subject to technical change, but that the productivity of one of the resources improves quicker than the other resource. We find that technical change is most useful when it is directed to the more abundant resource, and in the long run, is especially useful if it is quick enough to compensate for reductions in the flows of extracted non-renewable resources. Quantitative changes in the speed of biased technical change can bring about qualitative differences in the long-run outcome of the model.

Our strongest message is however that the elasticity of substitution plays a more important role for the long-run dynamics even than the *growth rate* of the distribution parameter between the resource inputs. In an extreme case, a one-shot improvement in the substitutability is more beneficial for the economy than perpetual factor-augmenting technical change.

In the light of these findings, consecutive research should address the following points. Firstly, exogenous technical change in the substitutability should be endogenized. It is one of the most important parameters, if not the most important, in models with non-renewable resources. This is by no means a simple task and valuable results might only be attainable for special cases. Secondly, one should attempt to compare the outcome of a model with endogenous technical change in the EoS with the outcomes of models featuring endogenous biased technical change in both resource inputs, already present in literature. Such results would help deepening our understanding as to which kind of technical change should be promoted by policy makers. Clearly, it is difficult to use the standard balanced growth approach to capture an optimal shift in research towards to more abundant resource. Therefore, how can one model the possibility of an optimal shift of R&D effort from one sector to the other.

Thirdly and finally, our analysis has (partly) been conducted in response to the bleak outlook painted by Cleveland and Ruth (1997), who suggest that the "traditional" types of technical change do not seem to be fast enough. Knowing the way in which introducing technical change in the EoS updates the existing results, it remains to be asked whether empirical evidence still gives the same bleak outlook as before.

3.8 Appendix

3.8.1 Optimization of the Hotelling model

We can write the Hamiltonian as follows (omitting time subscripts):

$$H(R_N, S_N, \lambda) = U\left(A[\psi R_N^{\theta} + (1-\psi)R_R^{\theta}]^{\frac{\beta}{\theta}}\right) e^{-\rho t} - \lambda R_N. \tag{3.14}$$

The Pontryagin maximum conditions are:

$$\frac{\partial H}{\partial R_N} = 0 \quad \Rightarrow \quad \lambda = \frac{\partial U}{\partial R_N} e^{-\rho t}, \tag{3.15}$$

$$\frac{\partial H}{\partial S_N} = -\dot{\lambda} \quad \Rightarrow \quad -\dot{\lambda} = 0. \tag{3.16}$$

Differentiating equation (3.15) with respect to time and substituting into equation (3.16) gives

$$\hat{Y} = \frac{\hat{F}_N - \rho}{\gamma}, \tag{3.17}$$

which is equation (3.8) in the main text. The marginal product of the non-renewable resource R_N is given by $F_N = A\beta\psi R_N^{\theta-1} R^{\beta-\theta}$. Its growth rate is given by

$$\hat{F}_N = g + z + (\theta-1)\hat{R}_N + (\beta-\theta)(\epsilon_N \hat{R}_N + \epsilon_\psi z + \epsilon_\sigma s) + \frac{1}{\sigma}[\log(R_N) - \log(R)]s. \tag{3.18}$$

As the growth rate of income is $\hat{Y} = g + \beta(\epsilon_N \hat{R}_N + \epsilon_\psi z + \epsilon_\sigma s)$, we can substitute these two growth terms into equation (3.17) to get

$$(1-\gamma)g + [1 + (\beta - \theta)\epsilon_\psi - \gamma\beta\epsilon_\psi]z + [(\beta-\theta)\epsilon_\sigma - \gamma\beta\epsilon_\sigma]s - \rho + \\ + \frac{1}{\sigma}[\log(R_N) - \log(R)]s = [\gamma\beta\epsilon_N + 1 - \theta - (\beta - \theta)\epsilon_N]\hat{R}_N.$$

Solving for $\hat{R}_N$ gives us the optimal growth rate of the non-renewable resource extraction, equation (3.9).

3.8.2 Proof that the denominator of $\hat{R}_N$ is always positive

The denominator of $\hat{R}_N$ is given by $(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N$. Rewriting this gives $(1-\theta) - (1-\gamma)\beta\epsilon_N + \theta\epsilon_N + \epsilon_N - \epsilon_N = (1-\epsilon_N)(1-\theta) - [(1-\gamma)\beta - 1]\epsilon_N$. As we know that $\epsilon_N \in [0,1]$ and $\theta \in (-\infty, 1]$, we also know that the first term is greater or equal to zero. As $\gamma \in (0, \infty)$, $0 < \beta < 1$, then $(1-\gamma)\beta - 1 < 0$, so the denominator turns out to be a sum of two non-negative expressions, one of them being strictly positive. This implies that $(1-\theta) - ((1-\gamma)\beta - \theta)\epsilon_N > 0$. ■

3.8.3 The simulation procedure

The simulations have been run using the Matlab routine ODE45. They have been done in two steps. In the first step, we implement the "shooting" method to find the initial amount of extracted resource $R_N(0)$, i.e. we run the routine for a range of initial values and pick the one that guarantees $S_N(t) \to 0$. Since of course, the simulations have to be stopped in finite time, we approximate this limit value by $S_N(300)$. This approximation is valid in the case of infinite-time depletion, because with our parameter choices, the remaining stock of non-renewable resource becomes negligible by the time $t = 300$. In the second step, we re-run the routine with the correct value of $R_N(0)$ as an initial condition. Having obtained the time path of R_N, we calculate the time paths of other variables by straightforward inserting.

In the case finite-time depletion is possible, we run the simulations for different values of T (the moment in time when the resource is exhausted) separately. The FOC remains the same, but we have to take into account the boundary condition $S_N(T) = 0$ (fixed terminal point).

After T, the economy uses for production the renewable resource flow R_R only. Other variables of the model continue their dynamic evolution without any interruption.

Chapter 4

Fertility Choice and Semi-Endogenous Growth: Where Becker Meets Jones

4.1 Introduction

The starting point of this chapter is the semi-endogenous growth theory where long-run growth is driven by technological progress created by the R&D sector. R&D output is a function of the number of active researchers as well as of the current level of technology. The distinguishing feature of semi-endogenous theories is that because they assume decreasing returns to R&D labor, a positive rate of technological progress cannot be sustained until the long-run limit unless the population of researchers will increase without bound (Jones, 1995). Semi-endogenous theories have gained substantive attention because they allow to reconcile the rapid increases in R&D expenditures, prevalent in the post-war U.S. data, with the remarkably stable growth rates of total factor productivity and GDP per capita.

One of the most often discussed features of R&D-based models of semi-endogenous growth is that they imply a strong link between long-run economic growth and population growth. This link has been questioned for many reasons (e.g. "people become skillful researchers by education rather than birth" – Strulik, 2005), but the semi-endogenous theory remains one of the most prominent contemporary growth theories nevertheless. However, ever since the Jones' (1995) article, authors of semi-endogenous growth models have been customarily assuming population growth to be exponential and *exo*genous. On the one hand, the linear population equation of form $\dot{N} = nN$ is a natural assumption – as opposed to other linearity assumptions put forward in literature which are much more difficult to interpret – since "it is a biological fact of nature, that people reproduce in proportion to their number" (Jones, 2003), and policy and the economy are believed not to affect the population growth rate n much. On the other hand, it effectively pushes the *endo*genous growth mechanism out of these models.

Moreover, the assumption of sustained exogenous population growth may soon be at odds with evidence. Modern demographic trends, notably the Second Demographic Transition, already present in all developed countries (see van de Kaa, 1997), put the exponential population growth assumption into severe doubt.

Several ideas on how population growth can be endogenized are already present in the literature. Becker was probably the first to doubt the Malthusian (1798) claim, that "the passion between the sexes has appeared in every age to be so nearly the same, that it may always be considered, in algebraic language as a given quantity", and his ideas have influenced many economic theories (see e.g. Becker, 1981; and more notably, the growth theory of Barro and Becker, 1989) – semi-endogenous growth theories as well.[1]

The main point raised in this paper is that semi-endogenous growth models with the "naturally linear" population equation $\dot{N} = nN$ and endogenous fertility have the potential to generate sustained endoge-

[1] Another question is whether perpetual population growth is, in fact, a desirable outcome. It has been estimated that the Earth's population will peak in the 21st century with a 75% probability. Mainstream semi-endogenous growth theories, though, take into account neither finiteness of Earth nor the fact that production of goods depends on various kinds of natural resources, *exhaustible* resources in particular. If one believes that finiteness of Earth will ultimately put a limit to population growth (see Pimentel et al., 1999, for a survey on this "interdisciplinary" strand of literature), she will probably be not too pleased with the prediction of the semi-endogenous theory that growth in per capita wealth will also cease. Here, however, we only point at this problem, and do not consider it any further.

nous growth without unmotivated knife-edge assumptions, as long as explicit limits such as finiteness of Earth or dependence of production on exhaustible resources are assumed out.

The article starts with a survey of the multiple ways to set up R&D-based growth models with endogenous fertility. For clarity of the discussion, we reduce to the necessary minimum the components of economic frameworks which do not influence their long-run dynamics. We present the decisive assumptions that various authors make, compute the weights attached by their optimizing agents to consecutive generations, and indicate the problems with second order conditions.[2]

It is argued here that there exists a particular feature of growth models with endogenous fertility which has not received the necessary attention yet: population size enters the utility functional multiplicatively. Thus, the *level* of flow utility enters the first order conditions (not just *marginal* utility), its sign matters, and the second order optimality conditions are no longer automatically satisfied. Hence, numerous authors preferred to get rid of this uneasy feature of endongenous-fertility models by writing the (logarithm of) population size as an additive component in the utility functional (Barro and Sala-i-Martin, 1995, Chapter 9; Jones, 2003; Connolly and Peretto, 2003)[3] or by considering a static optimization problem instead (Jones, 2001). By providing an analysis of the case where population size enters the utility functional multiplicatively, we fill a gap in the literature.

---

[2]The finding that optimizing over the growth rate of population might lead to problems with second order conditions has been first noticed by Deardorff (1976) in Samuelson's (1975) reasoning. In an exogenous fertility framework with overlapping generations and decreasing returns to accumulable inputs, Samuelson (1975) had derived a population growth rate which, according to him, "maximized" consumption at the steady state. It was baptized the "goldenest golden-rule state". Deardorff (1976) destroyed this result by showing that Samuelson's optimum is in fact a *minimum*, and that his model cannot be easily amended to offer an interior maximum. Recently Abio (2003) referred to these works and showed that when endogenous fertility is allowed for in these classic models, under certain parametrizations there emerges a possibility of existence of an interior maximum of the steady-state consumption level as a function of the population growth rate.

[3]Existence of long-run growth in these three models relies upon the weakly motivated knife-edge assumption that the intertemporal elasticity of substitution (IES) in consumption be exactly unity (the logarithmic case).

In the quest to reconcile all the problems outlined above, we put forward a bare-bones model with infinite-horizon ("dynastic") optimization, endogenous fertility choice, R&D-based semi-endogenous growth, and population size entering the utility functional multiplicatively. The dynasty's optimization problem where consumption and fertility are the only sources of utility (the Barro–Becker approach), has been given the most attention. A detailed discussion of first order, second order and auxilliary conditions, as well as the long-run dynamics of the model, is the most important contribution of this paper. In particular, we identify and characterize the model's asymptotic balanced growth path (analogous to the one in Jones, 2001). We also find that the restriction on the value of the intertemporal elasticity of substitution (IES hereafter) in consumption is crucial for the long-run outcome of semi-endogenous growth models with endogenous fertility.

The main results obtained in this article are expected to hold also if one expands our basic model so that it allows for physical or human capital accumulation, endogenous labor allocation, imperfect competition in the production sector, and the like. The long-run behavior of R&D-based semi-endogenous growth models where population growth is exogenous, is typically qualitatively different to the behavior of models with endogenous fertility. In addition, more stringent conditions have to be imposed.

In section 2, we present the specific features of growth models with endogenous fertility which are not shared by models with exogenous fertility. In section 3, we lay out and solve the basic model. In section 4, we discuss the requirements for the IES in consumption, imposed by this model as well as by the alternative ones put forward in literature. Section 5 concludes.

4.2 Exogenous vs endogenous fertility in growth models

4.2.1 Writing down the utility functional

Prior to reading any involved literature, one would expect the mathematical treatment of demographics in economic growth models to be the same regardless of whether fertility is exogenous or endogenous: after all, it is just the population growth rate n that is endogenized. However, it is not the case. Even more curiously, the difference strikes usually already in the first equation: the objective functional of the model.

The usual growth model with infinite-horizon optimization and *exo*genous fertility would have population size entering the objective functional *multiplicatively*: the social planner or representative household would maximize $\mathcal{U}_0 = \int_0^\infty N_t u(c_t)e^{-\rho t}dt$, where $N_t = N_0 e^{nt}$ denotes the population size at time t, n is the exogenous population growth rate, c_t denotes per capita consumption at time t, and u is the flow utility function. One can then rewrite the utility functional $\mathcal{U}_0$ as $\mathcal{U}_0 = N_0 \int_0^\infty u(c_t)e^{-(\rho-n)t}dt$, from which it follows immediately that population growth provides a drag on the discount rate ρ. The *effective* discount rate is computed as $\rho - n$: the difference between the pure time preference rate and the population growth rate. This benchmark case reflects *total utilitarianism* where utilities are summed over all members of the population, and nobody's utility has a priority.

If fertility is endogenous, the objective functional gets changed, for technical reasons rather than conceptual (as we shall see shortly). In particular, Barro and Becker (1989) replace N_t in $\mathcal{U}_0$ with N_t^θ, with $\theta \in (0, 1)$, thereby allowing for "incomplete altruism" – larger generations' ("vintages") utility is embedded in $\mathcal{U}_0$ with a decreasing marginal share. The effective discount rate becomes $\rho - \theta n$. Although the Barro and Becker's assumption is relatively easily interpretable – it reflects a compromise between total utilitarianism characterized by $\theta = 1$ and *average utilitarianism* where average person's utility is all that matters ($\theta = 0$) – one ought to know that its introduction is also necessary to assure that the second order conditions for optimality hold. This formulation will be our preferred one throughout the paper. We admit, however, that there is a number of drawbacks of this approach. We shall discuss one of them in the next subsection 4.2.2.

In principle, however, aggregating utilities across a population whose size is endogenous turns out to be a very difficult task. Several paradoxes and problematic conclusions appear in such case which have been discussed extensively by philosophers and welfare theorists. One crucial finding here is the *repugnant conclusion* due to Parfit (1984): "[f]or any possible population of (...) people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living". Total utilitarianism, imposing summation of individual utilities over all persons, arrives at the repugnant conclusion very quickly. The bad news is that, unfortunately, the problem is shared by all compromise versions between total utilitarianism and average utilitarianism, which we model by allowing for $\theta \in (0,1)$, as well as by all principles that assume thresholds or critical levels when aggregating utility, or condition the aggregation procedure on the satisfaction of a number of conditions (cf. Ryberg, Tänssjö, and Arrhenius, 2006). Average utilitarianism ($\theta = 0$) does not lead to the repugnant conclusion, but it has problems of its own, in particular because "[it] has implications generally regarded as highly counterintuitive. For instance, the principle implies that for any population consisting of very good lives there is a better population consisting of just one person leading a life at a slightly higher level of well-being" (Ryberg, Tänssjö, and Arrhenius, 2006).

There are strands of macro growth literature which deal with endogenous fertility differently than Barro and Becker do. In particular, it is widespread[4] to use the following trick: to redefine the objective functional $\mathcal{U}_0$ as $\mathcal{U}_0 = \int_0^\infty [u(c_t, b_t) + \epsilon v(N_t)] e^{-\rho t} dt$, where b_t is the per capita fertility rate (having children may be a source of utility). In this specification, population size enters $\mathcal{U}_0$ *additively.* We note that the above expression can be split into a sum of two integrals, $\mathcal{U}_0 = \int_0^\infty u(c_t, b_t) e^{-\rho t} dt + \epsilon \int_0^\infty v(N_t) e^{-\rho t} dt$.

Two facts should be mentioned here. First, the impact of births on agents' utility through an increasing population size is neglected. Thus, the channel of intertemporal substitution via the effective discount rate is switched off: the effective discount rate remains just ρ. One still treats population size as a normal good, but no longer considers it to be related to the marginal utility of consumption. Second, all involved models

[4]See Barro and Sala-i-Martin (1995), Chapter 9, Jones (2003), and Connolly and Peretto (2003). These three contributions constitute the core of the infinite-horizon literature on endogenous fertility and economic growth. The situation looks quite differently in the overlapping generations setups, though.

rely upon the knife-edge assumption that v is a logarithmic function: $v(N_t) = \ln N_t$ – the IES in population size is exactly unity.

Another way to simplify matters is to disregard *any* impact of births on population size whatsoever. One disposes then of all dynamic tradeoffs associated with the evolution of endogenously determined population size over time. This approach was taken in the previous, discussion-paper version of this article (Growiec, 2006a). Similarily, Jones (2001) prefers to exchange the original dynamic setup for a static one where by definition, no dynamic tradeoffs attached to fertility decisions are present.

4.2.2 Weighting the generations

Let us now write down the general form of the objective functional as

$$\mathcal{U}_0 = \int_0^\infty N_t \omega_t u(c_t, b_t) e^{-\rho t} dt. \tag{4.1}$$

By ω_t, we denote the *weight function* of generation t. We shall simply purge all differences between alternative setups into ω_t.

In the case of total utilitarianism ($\theta = 1$), characteristic for models with exogenous fertility, the weight function is $\omega_t \equiv 1$. The aggregate level of utility is just the sum of individuals' utility levels. We take this result as benchmark.

In the Barro and Becker's (1989) "incomplete altruism" case, $\omega_t = N_t^{\theta-1}$. Thus, individuals in larger generations are systematically less taken care of than individuals in smaller generations. This implies in particular, that if population is increasing over time, later generations are less taken care of than earlier ones.[5]

In the case where population size enters $\mathcal{U}_0$ additively, it can be shown that $\omega_t = \frac{1}{N_t}\left(1 + \frac{v(N_t)}{u(c_t,b_t)}\right)$ and thus the weight of each generation depends on: (i) its level of per capita consumption, (ii) its size, and (iii) its fertility rate. Moreover, it does so in a rather unintuitive and complex way. In particular, it matters if the levels of $u(c_t, b_t)$ and $v(N_t)$ – flow utilities – are positive or negative. We see that the "degree of altruism"

[5]Please note that we do not insist that population size be increasing over time. It may be well constant or decreasing.

in this case exhibits no systematic pattern,[6] which is a serious drawback of this methodology in comparison to the Barro and Becker's one.

4.2.3 Special features of models with endogenous fertility

Incorporating endongenous fertility in economic growth models with infinite-horizon optimization requires us to think about several of its characteristics, which would be of less significance if fertility were to remain exogenous. The list of such characteristics goes as follows.

1. The natural specification of the population equation of motion is the linear one: $\dot{N}_t = n_t N_t = (b_t - d_t)N_t$, where b_t is the birth rate and d_t is the death rate. This provides a potential for fully endogenous sustained growth (as opposed to other endogenous growth models where the linearity assumption is *ad hoc* rather than natural: see the detailed discussion in Jones, 2003).

2. The fact that population size enters the dynasty's utility functional multiplicatively, $\mathcal{U}_0 = \int_0^\infty N_t^\theta u(c_t, b_t)e^{-\rho t}dt$, implies that the flow utility needs to be always *positive*: $\frac{\partial \mathcal{U}_0}{\partial N_t} = \theta N_t^{\theta-1}u(c_t, b_t)e^{-\rho t} > 0$ only if $u(c_t, b_t) > 0$. Otherwise, utility (and in general, life) would be a painful burden rather than a blessing and the benevolent dynastic head would not impose such burden on her descendants. In other words, if flow utility is negative, it always pays to lower fertility: the utility functional is then closer to zero and thus greater.[7]

3. Another consequence of the fact that population size enters the utility functional multiplicatively is that levels of flow utility enter the first order and second order conditions. The utility function loses one of its "textbook" properties: its ordinal character. The utility function becomes a cardinal measure instead; the agent's decisions cease to be invariant to affine transformations of her utility

[6] For instance, it is obtained that $\omega_{c,t} \equiv \frac{\partial \omega_t}{\partial c_t} = -\frac{v(N_t)u'(c_t)}{N_t u^2(c_t, b_t)}$ and thus, $\omega_{c,t} \lesseqgtr 0$ if and only if $v(N_t) \gtreqless 0$. It is a mystery why the *direction* of the impact of per capita consumption on the weight attached to a generation should depend on whether flow utility from population size is positive or negative.

[7] I am grateful to Charles Jones for this point. The same result applies to models with health status which affects the survival probability (Hall and Jones, 2007).

function. Adding a constant to it and multiplying it by a positive constant changes the first order conditions.

4. Population growth provides a drag on the effective discount rate. Thus, a problem of endogenous fertility may be understood as a problem of endogenous discounting (see e.g. Becker and Mulligan, 1997; Das, 2003 on this issue).

5. There are substantial costs of childrearing, ranging from financial to time (opportunity) costs. In particular, time devoted to bringing up children has to be deducted from the total time spent on productive activities (in our case, production and R&D).[8]

6. The endogenous population growth rate adds to the depreciation rate of accumulated stocks (per capita capital, per capita stock of exhaustible resources, etc.), thus creating the "effective" depreciation rate. This is due to the fact that total stocks have to distributed among an increasing number of people.

7. It must be assumed that individuals receive utility from having children. In all endogenous fertility models discussed in literature, the birth rate b_t enters the flow utility function $u(c_t, b_t)$.

We deal with most of these points when solving our basic model in section 4.3.

4.2.4 Problems with second order conditions

Having population size as a multiplicative factor in the utility functional can cause problems with second order conditions. To explain this claim, let us first write down the following "generic" optimization problem:

$$\max_{N_t, b_t, \ldots} \int_0^\infty N_t^\theta u(c_t, b_t, \ldots) e^{-\rho t} dt \qquad \text{s.t.} \qquad \dot{N}_t = (b_t - d_t) N_t, \ldots \quad (4.2)$$

Three dots denote further variables and further equations which can be freely added to the optimization problem under the condition that the

[8]In the continuous-time setup, the simplest way to model childhood is to consider its duration to be infinitesimally short. We shall take this approach in the following sections. The childrearing costs are then deducted from parents' earnings only once, at the moment their child is born.

population size N_t does not enter any of these equations (in particular, it implies that the production function ought to have constant returns to scale). The resulting Hamiltonian reads:

$$\mathcal{H}(N_t, b_t, ...) = N_t^{\theta} u(c_t, b_t, ...)e^{-\rho t} + \Lambda_t(b_t - d_t)N_t + ... \tag{4.3}$$

We write N_t as the first argument of the Hamiltonian for convenience – the term $\frac{\partial^2 \mathcal{H}}{\partial N_t^2}$ becomes then the first minor of its Hessian. We shall maintain this order throughout the paper.

Now, let us remind ourselves that for the FOCs to describe a maximum of the Hamiltonian, it is necessary that its Hessian is non-positive definite. This implies that there are no positive elements on the main diagonal.

However, it is quickly verified that

$$\frac{\partial^2 \mathcal{H}}{\partial N_t^2} = \theta(\theta - 1)N_t^{\theta-2} u(c_t, b_t, ...)e^{-\rho t}, \tag{4.4}$$

and thus $\frac{\partial^2 \mathcal{H}}{\partial N^2}$ is positive if $\theta \in (0, 1)$ and the flow utility $u(c_t, b_t, ...)$ is negative. For the sake of optimality, we must rule this case out. Therefore, *flow utility must be positive in models with endogenous fertility.*

To emphasize the importance of this result, let us now restrict our attention to the class of CRRA utility functions, which is probably the one most often used by economists. The simplest CRRA functions are defined as $u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$. If $\gamma > 1$, then flow utility is always negative and thus cannot be used in models with endogenous fertility. However, $\gamma > 1$ implies that the IES in consumption is less than unity – a result endorsed by a multiplicity of empirical works (see the review in the section 4.4). Accordingly, the same negativity result holds for generalized CRRA utility functions of form $u(c_t, b_t) = \mu_0 + \mu_1 \frac{c_t^{1-\gamma}}{1-\gamma} + \mu_2 \frac{b_t^{1-\eta}}{1-\eta}$ where $\gamma, \eta > 1$, $\mu_0 \leq 0$, and $\mu_1, \mu_2 > 0$. Even adding an arbitrarily large positive constant $\mu_0 > 0$ to flow utility cannot help dispose of the negativity problem for small values of c_t, since $\lim_{c \to 0} u(c, b) = -\infty$ for *any* μ_0. We note that for CRRA utility functions, u can be everywhere positive only if $\gamma < 1$.

4.3 The basic model

We shall now analyze a simplified R&D-based semi-endogenous growth model with population size entering the utility functional multiplicatively. We shall keep in mind the general remarks of above subsections, in particular of section 4.2.3 where we have listed the special features of such models.

At this point, one must emphasize strongly the fact that agents in the model specified below optimize over an infinite planning horizon. The approach we are taking here is markedly different from the overlapping generations approach where individuals plan only one step ahead, facilitating the analysis of endogenous fertility, e.g. de la Croix and Doepke (2003, 2004). Unlike that literature, here we have to deal with the following problems explicitly: (i) we must choose the method for aggregating utility over an infinity of generations whose sizes are *endogenous*, and any choice will inevitably affect the outcome; (ii) dynamic inconsistency, arising due to the endogenous determination of the effective discount rate, rules out the possible use of recursive optimization methods; (iii) second order conditions may fail because of the objective function being not concave in the birth rate; (iv) the possible unstable dynamics may lead to a violation of transversality conditions. In exchange for the increased analytical complexity we obtain, however, a framework that is directly comparable to its exogenous-fertility counterpart, and well suited for illustrating the scope of long-run consequences of allowing for an endogenous determination of the birth rate.

4.3.1 Setup

Demographics

To give the demographics of the model an explicit treatment, we shall make use of a continuous-time overlapping-generations setup with indeterministic lifespan. For mathematical simplicity, we assume that N_t – population at time $t \geq 0$ – is in fact not the integer number of individuals, but rather the measure of an interval, populated by a continuum of agents. Thus, although the lifespan of each individual is random, the Law of Large Numbers enables us to treat the death rate at each instant of time as deterministic. For each individual, we shall introduce a survival function $m : \mathbb{R}_+ \rightarrow [0, 1]$ such that $m(0) = 1, \lim_{t\rightarrow\infty} m(t) = 0$, and m is decreasing. The total number of births at time t is denoted $B_t \equiv b_t N_t$, with b_t being the birth rate. The size of generation t at time $z \geq t$ is equal to

$$S_{z,t} = B_t m(z - t), \tag{4.5}$$

and the total population at time t is

$$N_t = \int_0^t B_z m(t - z) dz. \tag{4.6}$$

The population growth rate can be calculated as

$$n_t = \frac{\dot{N}_t}{N_t} = \frac{B_t m(0)}{N_t} + \frac{\int_0^t B_z m'(t - z) dz}{N_t} \equiv \underbrace{b_t}_{\text{birth rate}} - \underbrace{d_t}_{\text{death rate}} . \tag{4.7}$$

Instead of maintaining the general form of the survival function m throughout the paper, we shall simplify the analysis by limiting ourselves to the "perpetual youth" case. Namely, we shall take the exponential function m implying a constant probability of death d at all ages $x \geq 0$, conditional on having reached the age x. This assumption reads:

$$m(x) = e^{-dx} \quad \Rightarrow \quad d_x \equiv d, \quad \text{where } d > 0. \tag{4.8}$$

It is possible for some individuals to live forever, although such probability can well be neglected. For simplicity, we also neglect the impact of technological progress and increasing per capita wealth on the survival function m.[9]

Production and R&D

The production function of the single consumption good is assumed to be Cobb-Douglas with constant marginal returns to labor L, and increasing returns to scale, once the technology level A is included as well, which reflects the concept of non-rivalry of ideas (see Jones, 2005, for a discussion):

$$Y_t = A_t^\sigma L_t = A_t^\sigma (1 - \beta(b_t)) N_t, \qquad \sigma > 0, \tag{4.9}$$

[9]Provided that we rule out the possibility that the expected lifespan be growing without bound, this assumption does not change our results qualitatively.

where $\beta(b_t)$ describes the time cost of childrearing. For reasons which we shall describe later in more detail, we assume that the function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ is twice continuously differentiable, increasing and convex ($\beta' > 0$ and $\beta'' > 0$), and such that $\beta(0) = 0$. We are going to assume out all other forms of childrearing costs.

By taking (4.9) as our production function, we assume out physical capital accumulation. This assumption is made to ease the exposition – further derivations, focused on the endogenous fertility component, are more transparent without capital – but is not necessary. Indeed, one would make this model more realistic by considering capital accumulation as well.

Ideas are accumulated according to the Jones' (1995) R&D equation

$$\dot{A}_t = \nu L_t^{\lambda} A_t^{\phi}, \qquad 0 < \lambda < 1, \quad 0 < \phi < 1, \quad \nu > 0. \tag{4.10}$$

We assume that the spillovers in idea production are positive (the "standing-on-shoulders" effect) but not sufficiently strong for fully-endogenous R&D-driven growth ($\phi < 1$).

The above Cobb-Douglas assumptions are standard for the associated literature (see Jones, 2005, for a justification). At the same time, they greatly facilitate obtaining balanced growth. In this paper, we think of this property as desirable, because we shall emphasize *different* (population-side, not production-side) barriers to endogenous balanced growth.

It is assumed that the whole working population is employed both in R&D and in the production sector. To keep things as simple as possible, we do not consider the allocation of labor between these two sectors explicitly. In our setup, people receive remuneration for their production work but not for their research. Thus, R&D is considered here an inevitable side-effect of production, rather than a distinct sector of the economy.[10] Despite the fact that this model may be interpreted a

[10] This is an admittedly heroic assumption. However, it does not change the results qualitatively, because in the long run, the ratio of researchers is expected to approach a constant. We assure it by trivially setting it to a constant – i.e. endowing each individual with a constant amount of time for production work and for research. Then, we say that the long-run research/production effort ratio is already included in ν and A_0.

model of "learning by doing" in the Arrow's (1962) tradition (or better: "inventing by doing"), all the long-run results we are discussing would have clearly gone through if we had endogenized labor allocation and allowed for an explicit treatment of the R&D outlay. We abstract from these issues only to simplify exposition.

Households

The representative agent maximizes discounted utility of her dynasty, born at time 0. Her altruism is *imperfect*: the representative agent does not explicitly take into account the fact that some members of the dynasty are born, and some die at each instant of time, but she systematically attaches smaller weight to bigger generations. Utility is derived from consumption and the number of children (as in Barro and Becker, 1989):

$$\max_{\{b_t\}_{t=0}^{\infty}} \int_0^{\infty} N_t^{\theta} u(y_t, b_t) e^{-\rho t} dt, \qquad \rho > 0, \tag{4.11}$$

subject to the population equation of motion $\dot{N}_t = (b_t - d)N_t$ and the per capita production function $y_t = A_t^{\sigma}(1 - \beta(b_t))$. Since no savings are allowed, all production is immediately consumed and thus $c_t = y_t$ for all t.

Moreover, we shall assume that the flow utility function is of the argument-separable CRRA form with a non-negative constant:

$$u(y_t, b_t) = \mu_0 + \mu_1 \frac{y_t^{1-\gamma}}{1-\gamma} + \mu_2 \frac{b_t^{1-\eta}}{1-\eta}, \tag{4.12}$$

where $\mu_0 \geq 0$, $\mu_1, \mu_2 > 0$, and $\gamma, \eta \in (0, 1)$.[11] This closes the setup of our basic model.

[11] The assumption that $\gamma, \eta \in (0, 1)$ assures that flow utility is always positive. In fact, we could have been less stringent here. It would suffice if u were positive just for the values of variables actually realized along the time path of the economy. One would then have to be very careful with the choice of initial conditions and parameter values, though. We acknowledge that the case with $\gamma > 1$ and (despite that) $u > 0$ is a cumbersome but potentially very useful case to analyze. We leave it for future work.

4.3.2 The asymptotic balanced growth path

Before we start actually solving the model, we shall describe the necessary properties of the balanced growth path (BGP), or alternatively, the asymptotic BGP. These will be helpful in the subsequent derivations.

We shall define the balanced growth path as a sequence of time paths $\{A_t, N_t, y_t\}_{t=0}^{\infty}$, along which all economic variables grow at constant non-negative rates, possibly zero. It implies that the birth rate $\{b_t\}_{t=0}^{\infty}$ must be constant along the BGP.

Similarily, we shall say that the model approaches an *asymptotic* BGP if the growth rates of variables approach constant values. Please note that this happens only as the levels of A_t, N_t, y_t diverge to infinity. In such case no proper BGP exists – there is no exponential solution to the dynamical system at hand.

We shall claim that in our model no BGP exists, but under an appropriate parameter configuration, an asymptotic BGP is converged to as time goes to infinity. In the following subsections, we shall derive the conditions under which the birth rate approaches a constant $\bar{b}$ implying that the population growth rate also approaches a constant $\bar{n} = \bar{b} - d$. Consequently, the growth rate of knowledge stock and the economic growth rate approach constants as well.

From the R&D equation (4.10), it is obtained that along an asymptotic BGP, necessarily

$$\frac{\dot{A}_t}{A_t} = \frac{\lambda \bar{n}}{1 - \phi}, \tag{4.13}$$

where $\bar{n} = \bar{b} - d$ is the endogenously determined long-run population growth rate.

This implies that the economic growth rate approaches:

$$g \equiv \frac{\dot{y}_t}{y_t} = \sigma \frac{\dot{A}_t}{A_t} = \frac{\sigma \lambda \bar{n}}{1 - \phi}. \tag{4.14}$$

Please note that although the childrearing cost $\beta(b_t)$ imposes a significant level effect on production and R&D output, no long-run growth effects should be expected.

4.3.3 Optimization

To solve the households' optimization problem, we set up the Hamiltonian:

$$\mathcal{H}(N_t, b_t) = N_t^{\theta} u(y_t, b_t) e^{-\rho t} + \Lambda_t (b_t - d) N_t, \tag{4.15}$$

where Λ_t is the shadow price of population size. b_t is the only control variable in this setup, and N_t is the only state variable. Thanks to our prior simplifying assumptions, the accumulation of knowledge A_t is considered external to the household and no savings decision is allowed.

When dealing with transversality and summability conditions, we shall utilize the properties of the asymptotic BGP, derived in the previous subsection. For a time being, we assume that the asymptotic BGP is indeed approached; in the next subsection we shall prove it, and thus prove that the transversality and summability conditions hold. But of course, we shall begin with the FOCs.

First order conditions

The first derivatives of the Hamiltonian read:

$$\begin{aligned}
\frac{\partial \mathcal{H}}{\partial N_t} &= \theta N_t^{\theta-1} u(y_t, b_t) e^{-\rho t} + \Lambda_t (b_t - d) = -\dot{\Lambda}_t, & (4.16)\\
\frac{\partial \mathcal{H}}{\partial b_t} &= N_t^{\theta} [u_b(y_t, b_t) - u_y(y_t, b_t) A_t^{\sigma} \beta'(b_t)] e^{-\rho t} + \Lambda_t N_t = 0. & (4.17)
\end{aligned}$$

Substituting the suitable expression for the shadow price Λ_t from the second FOC into the first FOC ($\Lambda_t = -N_t^{\theta-1}[u_b(y_t, b_t) - u_y(y_t, b_t) A_t^{\sigma} \beta'(b_t)] e^{-\rho t}$), using additive separability of u as apparent in (4.12), and rearranging yields the following equation of motion of the birth rate:[12]

$$\begin{aligned}
\dot{b}_t \underbrace{\Big(u_{bb} + u_{yy} A_t^{2\sigma} (\beta'(b_t))^2 - u_y A_t^{\sigma} \beta''(b_t)\Big)}_{\text{Left-hand side } \equiv \Psi} &= \underbrace{\theta u - \Big(\theta(b_t - d) - \rho\Big) u_b +}_{\text{Right-hand side } \equiv \Phi} \\
\underbrace{+ u_y A_t^{\sigma} \beta'(b_t) \Big(\theta(b_t - d) - \rho + \sigma \frac{\dot{A}}{A}\Big) + u_{yy} A_t^{2\sigma} \sigma \frac{\dot{A}}{A} (1 - \beta(b_t)) \beta'(b_t)}_{\text{Right-hand side } \equiv \Phi}&. \quad (4.18)
\end{aligned}$$

[12] By u_x, we denote the derivative of u with respect to x. We omit the arguments of u and its derivatives for convenience.

We shall describe the properties of this rather complicated expression in subsection 4.3.4. We shall denote its left-hand side (apart from $\dot{b}_t$) as Ψ and the right-hand side as Φ.

Transversality condition

The transversality condition requires that the shadow value of population size tends to zero as time approaches infinity. For the condition $\lim_{t\to\infty} \Lambda_t N_t = 0$ to hold, it suffices that from some time on, the growth rate of $\Lambda_t N_t$ is uniformly negative. Solving for this growth rate and taking an asymptotic balanced growth path approximation (which implies that the *positive* growth rate of population is $\bar{n}$, the economy grows at a rate g, and the marginal utility of consumption declines at a rate γg), we can write the transversality condition as:

$$\frac{\dot{\Lambda}_t}{\Lambda_t} + \frac{\dot{N}_t}{N_t} = \theta \bar{n} + (1-\gamma)g - \rho = \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)\bar{n} - \rho < 0. \quad (4.19)$$

In other words, the discount rate should be large enough compared to the population growth rate. Condition (4.19) is identical to the restriction of a positive effective discount rate in exogenous fertility models.

Summability condition

The summability condition requires that the integral (4.11) converges. To satisfy this requirement, it suffices that along the asymptotic BGP, the growth rate of the product $N_t^\theta y^{1-\gamma} e^{-\rho t}$ is negative. This implies that in our setup, the transversality condition and the summability condition coincide.

Second order conditions

Let us now proceed to the second order (sufficiency) conditions which would guarantee that the FOCs describe an actual maximum of the Hamiltonian. We are going to be conscientious here since we have already shown that in models where endogenous fertility enters the utility functional multiplicatively, satisfaction of second order conditions is not automatic. This step is typically omitted in the literature, though.

It turns out that assuming $u > 0$ is necessary for optimality, but by no means sufficient. The following proposition holds.

Proposition 4.1 *In the vicinity of the asymptotic BGP, the first order condition* (4.18) *describes a maximum of the Hamiltonian* (4.15) *if*

$$\frac{\beta''(\bar{b})(1-\beta(\bar{b}))}{(\beta'(\bar{b}))^2} > \frac{1-\gamma-\theta}{\theta}, \tag{4.20}$$

where $\bar{b}$ denotes the asymptotic steady-state birth rate.

Proof. See Appendix. ■

In particular, a linear childrearing cost function $\beta(b) = \kappa b$ can be in agreement with the second order conditions provided that $\theta > 1-\gamma$. In section 4.3.5, where we specialize to the linear case, we will assume this inequality to hold.

4.3.4 Existence of the solution and dynamics

We shall now characterize the conditions which have to be satisfied for the birth rate b_t to approach a constant $\bar{b}$ as time goes to infinity, thus yielding an asymptotic BGP. First, we shall derive the steady-state birth rate $\bar{b}$ (in most cases it will be unique, but we will not rule out the possibility of multiple solutions). The next step will be to analyze the dynamics of the model around its asymptotic BGP.

The existence issue must be addressed first. The following propositions hold.

Proposition 4.2 *The steady-state birth rate $\bar{b}$ is defined as a solution to the implicit equation:*

$$\frac{\theta}{1-\gamma}\left(\frac{1-\beta(\bar{b})}{\beta'(\bar{b})}\right) = \rho - \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\bar{b}-d). \tag{4.21}$$

Proof. See Appendix. ■

Proposition 4.3 *A solution to the implicit equation* (4.21) *is guaranteed to exist if*

$$\beta'(0) < \frac{\theta}{(1-\gamma)\left(\rho + \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)d\right)} \tag{4.22}$$

and

$$\rho > \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\beta^{-1}(1) - d), \tag{4.23}$$

or if the signs in (4.22) *and* (4.23) *are simultaneously reversed.*

Proof. See Appendix. ■

Equation (4.21) in proposition 4.2 is central to this paper: by defining the endogenous asymptotic steady-state birth rate, it also pins down the long-run growth rate of the economy.

Please note that the parameters $\eta, \mu_0, \mu_1, \mu_2$ have disappeared in the course of taking limits. They play their roles during the transition, but not along the asymptotic BGP.

Imposing an upper bound on $\beta'(0)$ such as in (4.22) and a lower bound on ρ as in (4.23) closes the existence issue. The roots of (4.21) can be evaluated numerically, and we shall do so in the following subsection. Moreover, the sufficiency conditions (4.22)–(4.23) can be substantially weakened: it is enough to show that for some $b \in [0, \beta^{-1}(1)]$, the left-hand side of (4.21) is greater than the right-hand side, and for some other b, it is smaller, and the intermediate value theorem argument used in the proof of proposition 4.3 still goes through.

Let us pass on to the question of stability. The following proposition holds.

Proposition 4.4 *The steady-state birth rate $\bar{b}$, implicitly defined in* (4.21)*, is asymptotically stable if in the vicinity of $\bar{b}$,*

$$\frac{\theta}{1-\gamma}\left(\frac{1-\beta(\bar{b})}{\beta'(\bar{b})}\right) < \rho - \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\bar{b} - d) \tag{4.24}$$

for $b < \bar{b}$, and the inequality (4.24) *is reversed for $b > \bar{b}$. Otherwise, the steady-state birth rate is unstable.*

If the steady-state birth rate is unstable but $1 - \theta - \gamma > 0$, then no transitional adjustments in b_t are possible and the birth rate is always set so as to assure $\dot{b}_t = 0$. Otherwise, a solution that has $b_t \to \beta^{-1}(1)$ may also be chosen by the representative household (depending on initial conditions).

Proof. See Appendix. ■

Summarizing, the following conditions should be checked before one could be sure that a given $\bar{b}$ is the asymptotic steady-state birth rate delivered by our model, giving rise to an asymptotic BGP:

1. The necessary condition for optimality combined with the asymptotic steady-state requirement, summarized by the implicit equation (4.21).

2. The transversality/summability condition, given by (4.19).

3. The sufficient condition for optimality, given by (4.20).

Local dynamics around the asymptotic BGP may then be studied using proposition 4.4. In this respect, it must be noted that if the asymptotic steady-state birth rate proves to be unstable, then in many cases, there will be no transitional adjustments in b_t.[13] The logic is the following. The birth rate b_t is a control variable so it can always be set such that $\dot{b}_t = 0$; setting it at a different value might lead to an eventual violation of the transversality condition or of the second order optimality condition (as apparent in the proof of proposition 4.4). One has to bear in mind, however, that no proper BGP exists here. Thus, if the solution that has $b_t \to \beta^{-1}(1)$ is ruled out, then it must be the case that $\dot{b}_t = 0$ at all times, so that $\Phi(b_t) = 0$ at all times, but since $u_b/u \to 0$ and $\frac{u}{y^{1-\gamma}} \to \frac{\mu_1}{1-\gamma}$ only gradually, then also b_t will approach $\bar{b}$ only gradually. In the end, there will be some transition dynamics here even though the asymptotic birth rate is unstable and divergence is ruled out. These will not be driven by purposeful dynamic adjustments in b_t, though.

Let us also comment on the following: we have been using asymptotic steady-state approximations when deriving all above inequalities. Thus, we have been implicitly assuming that the initial population size and the initial knowledge stock are large enough for the approximations to be valid. Using the continuity argument, one could claim that what is quite important here is the *margins* by which our approximate inequality conditions are satisfied. The narrower is the margin, the larger has to be the initial population size as well as the initial knowledge stock (which

[13]Such instability results may also be found in endogenous-fertility growth models of Barro and Sala-i-Martin (1995), Chapter 9, and Connolly and Peretto (2003).

are our only stock variables) in order for a given condition in its precise form to be satisfied as well.

4.3.5 The case of a linear childrearing cost function

The case of a linear childrearing cost function is a particularly tractable one. Indeed, if such functional form is assumed, the steady-state birth rate may be derived explicitly: it is straightforward to show that if $\beta(b) = \kappa b$, with $\kappa > 0$, then there exists a unique asymptotic steady-state birth rate $\bar{b}$ given by:

$$\bar{b} = \frac{\frac{\theta}{1-\gamma}\frac{1}{\kappa} - \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right) d - \rho}{\frac{\gamma\theta}{1-\gamma} - (1-\gamma)\frac{\sigma\lambda}{1-\phi}}, \tag{4.25}$$

provided that κ is small enough to guarantee $\bar{b} \geq 0$. Consequently, the steady-state population growth rate $\bar{n} = \bar{b} - d$ equals:

$$\bar{n} = \frac{\frac{\theta}{1-\gamma}\left(\frac{1}{\kappa} - d\right) - \rho}{\frac{\gamma\theta}{1-\gamma} - (1-\gamma)\frac{\sigma\lambda}{1-\phi}}. \tag{4.26}$$

The model at hand is able to deliver the prediction that the steady-state birth rate $\bar{b}$ decreases with ρ, κ, and d, quite in line with intuition.[14] In order to obtain such plausible comparative statics, however, one must impose that the denominator in (4.25) be positive.

It can be checked that the population growth rate over the long run is positive – and thus the economic growth rate is positive as well – provided that:

$$\frac{\theta}{1-\gamma}\left(\frac{1}{\kappa} - d\right) > \rho. \tag{4.27}$$

As far as the auxilliary conditions are concerned, the transversality condition can be now rewritten, inserting (4.26) into (4.19), as:

$$\left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)\left(\frac{1}{\kappa} - d\right) < \rho. \tag{4.28}$$

[14] The last relationship is perhaps the least intuitive, but one has to bear in mind that in our model, a greater death rate means not only a shorter lifespan of the children but also of the parents, and that people here have children at a constant rate thoughout their lives. A shorter life leaves them with less time to be allocated to childrearing.

Since $\beta''(\bar{b}) = 0$, the second order condition for optimality simplifies radically in the linear case. It becomes:

$$\theta > 1 - \gamma. \tag{4.29}$$

For the sake of an illustration that for several plausible parameter values, the above described asymptotic steady-state birth rate indeed exists, we proceed by means of a numerical example. The parameter values are going to be fixed at some plausible benchmark levels. We require them to fall into appropriate intervals, but we do not intend to make any direct connections to the empirically observed values whatsoever. Our benchmark set of parameter values is summarized in table 4.1.

γ	κ	ϕ	σ	λ	ρ	d	θ
0.8	25	0.7	0.5	0.8	0.04	0.02	0.8

Table 4.1: The benchmark parameter values. The linear case.

Please note that since the values of η, μ_0, μ_1, and μ_2 do not play any role along the asymptotic BGP, we do not have to fix them.

It is straightforwardly checked that under this parametrization, conditions (4.27) through (4.29) are satisfied and the denominator in (4.25) is positive. Furthermore, the asymptotic steady-state value of the endogenous birth rate is equal to $\bar{b} = 0.0336$, which implies a steady population growth rate of $\bar{n} = 0.0136$. The economic growth rate approaches $g = 0.0182$, with a fraction of time $(1 - \beta(\bar{b})) = 0.1591$ allocated to productive activity and the remaining fraction $\beta(\bar{b}) = 0.8409$ to childrearing.

The asymptotic birth rate is unstable in the current case: $\dot{b}_t$ as a function of b_t crosses the zero line at $\bar{b}$ from below. This instability result implies here that there are either (i) no transitional adjustments in b_t apart from those necessary to converge to the asymptotic BGP (so that $u_b/u \to 0$, $u_{bb}/u \to 0$), or (ii) transitional dynamics implying gradual divergence of childrearing costs: $b \to \beta^{-1}(1)$.

4.3.6 Numerical example with a non-linear childrearing cost function

Let us now posit a different functional form for the increasing childrearing cost function β. We shall assume that

$$\beta(b_t) = \kappa b_t^{\psi}, \qquad \kappa > 0, \quad \psi \neq 1. \tag{4.30}$$

Using the definition of function β in (4.30), we note the following. First, the plausibility condition $\bar{b} \leq \beta^{-1}(1)$ becomes $b \leq (1/\kappa)^{1/\psi}$. Second, the sufficient condition for optimality (4.20) is automatically satisfied if $\psi > 1$; if $\psi < 1$, however, then (4.20) is satisfied only if

$$\bar{b} > \left(\frac{1}{\kappa + \frac{1-\gamma-\theta}{\theta} \frac{\psi}{\psi-1}} \right)^{\frac{1}{\psi}}. \tag{4.31}$$

The obvious next step is to fix the parameter values again. Just like in the linear case above, we shall fix them at some plausible benchmark levels without intending to make connections to the empirically observed values. This set of parameter values is summarized in table 4.2.

γ	κ	ψ	ϕ	σ	λ	ρ	d	θ
0.8	40	1.2	0.7	0.5	0.8	0.08	0.02	0.8

Table 4.2: The benchmark parameter values. The non-linear case.

Using a numerical routine to find the root of Φ defined in (4.21) (in this case, it is unique), we arrive at a result that the endogenous birth rate approaches an asymptotic steady-state value of $\bar{b} = 0.0296$, which implies a steady population growth rate of $\bar{n} = 0.0096$. The economic growth rate approaches $g = 0.0128$, with a fraction of time $(1 - \beta(\bar{b})) = 0.4140$ allocated to productive activity and remaining fraction $\beta(\bar{b}) = 0.5860$ to childrearing.

A simple numerical check of the transversality/summability condition (4.19) and the second order optimality condition (4.20) confirms that they both hold along the asymptotic BGP.

The dynamics of the birth rate are depicted in figure 4.1.[15] We see that the asymptotic steady-state birth rate is again unstable. Since in

[15] We have used the parameter choices given in table 4.2 when producing this figure.

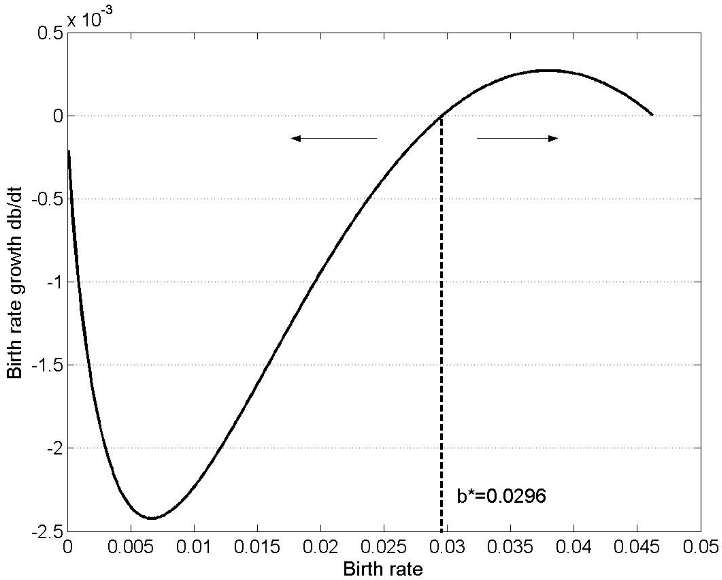

Figure 4.1: Unstable asymptotic steady-state birth rate.

the current example, $1 - \gamma - \theta < 0$, the solution implying $b_t \to \beta^{-1}(1)$ cannot be ruled out.

This numerical example ends the qualitative analysis of our simplified "Barro–Becker–Jones" model. We proceed now to discuss one of the major drawbacks of the methodology we used: it requires the IES in consumption to be greater than one ($\gamma < 1$) which is an empirically doubtful assumption.

4.4 The IES in consumption

Several growth models with endogenous fertility discussed in literature (notably in Barro and Sala-i-Martin, 1995, Chapter 9; Connolly and Peretto, 2003; Jones, 2003), as long as their long-run dynamics are concerned, rely on the knife-edge assumption that the IES in consumption $1/\gamma$ equals exactly unity. The $\gamma = 1$ assumption is necessary for the results obtained in those works, but it is quite questionable as well. In our setup (as well as in Jones', 2001), this assumption has been substituted with a much less stringent assumption that $\gamma < 1$ – that is, that the IES in consumption exceeds unity. Empirical studies suggest the opposite assumption to be much more plausible, however. We provide a short summary of this literature below. Please note that since the magnitude of the true IES in consumption has been estimated in a wide range of empirical works, we can name just a small percentage of all contributions here.

The discussion began with the work of Hall (1988) who concluded that the IES in consumption in the United States were very small, and possibly zero – that is, that for sure $\gamma > 1$.

Patterson and Pesaran (1992) upgraded Hall's methodology by assuming that the slope coefficient of the MA process, governing per capita consumption, is not known a priori, as Hall presumed. This modification helped them obtain the result of the American IES being around 0,213 and significantly different from zero.

Hall's results have also been criticized on other grounds. It has been argued that the Euler equation he estimated was misspecified, and unsuitable for the country-level aggregate data whatsoever. Beaudry and van Wincoop (1996) used a panel of U.S. states instead (spanning 1953-1991, or in the other estimation round, 1978-1991), and modified the Hall's original Euler equation. They have achieved a large improvement in the estimation precision, and obtained a result of IES being around 0.7-1.1 (depending on the estimation method and the set of instrumental variables): clearly different from zero and not significantly different from one.

Guvenen (2006) added another dimension to this discussion, pointing out that in reality, as opposed to most theoretical approaches, agents are heterogenous. In particular, a large fraction of households does not participate in stock markets at all. Moreover, the IES in consumption varies significantly across individuals, increasing with income. Consumption is much more evenly distributed than wealth. This asymmetry accounts, according to Guvenen, for a serious underestimation of the IES in all previous studies. He concludes, that among non-stockholders, IES is indeed around 0.1 (as e.g. Hall suggested); but among stockholders, it is rather expected to fall into the interval (0.8, 1.2). And it is the stockholders who effectively determine the real interest rate of the economy.

This makes the variant IES=1 again possible. However, Guvenen himself states that "a plausible range for this parameter is possibly (0,1)".

Favero (2005) merged the Euler equation with the (linearized) budget constraint of the households, and used the resultant equation to estimate the IES. He obtained IES=0.78, with a standard deviation of 0.11 (i.e. clearly smaller than one).

Harashima (2005) went in a different direction. He overthrew the assumption that in the estimated Euler equation, the real interest rate is taken as given (the "endowment" economy assumption). Instead, he proposed to consider a closed "production" economy, in which the IES is obtained directly from some version of Euler equation. He concluded, that in such case, the IES would be again very low, around 0.09.[16]

The above literature review is by no means complete; it is included here to point out, that although the assumption of $\gamma = 1$ is knife-edge and disputable, the alternative assumption of $\gamma < 1$ is also quite troublesome. Empirical investigations bring somewhat convincing evidence, that $\gamma > 1$.

A proposed solution to this problem would be then to get out of the CRRA framework and to disconnect the curvature of the flow utility function from its sign. The fact is that for CRRA functions, $\gamma > 1$ is associated with high curvature and negative sign, and $\gamma < 1$ – with low curvature and positive sign. However, we insist that this identification should not be taken too far: the research focus varies from study to study. In the empirical works, the main focus is on curvature; in our analysis, necessity for $\gamma < 1$ stems from the sign requirement imposed on the utility function rather than from its curvature.[17]

[16]Harashima just calibrated his theoretical growth model, and did not use any econometric methods.

[17]Several generalizations of CRRA utility have been already put forward in the literature, but these works concentrated mostly on disentangling the elasticity of intertemporal substitution from the coefficient of relative risk aversion and not on disentangling the curvature of the utility function from its sign. Most notable in this respect are Epstein and Zin (1989) and Weil (1990).

4.5 Conclusion

In this paper, we have studied the dynamic behavior of R&D-based semi-endogenous growth models with endogenous fertility. The number of knife-edge conditions required for sustained growth in this model is just one, and a naturally interpretable one: we just assume that people reproduce in proportion to their number, and thus that the population equation is linear. Per capita GDP growth is obtained here, however, thanks to R&D – in particular, thanks to constantly increasing numbers of researchers.

We have argued that in the involved literature, endogenous fertility is usually treated differently than exogenous fertility is. This is probably because the assumption of population size entering the utility functional multiplicatively brings about cumbersome complications which researchers would rather prefer to omit. We have provided a detailed treatment of the case at hand and thus filled a gap in the literature.

We have analyzed a bare-bones "Barro–Becker–Jones" model that features infinite-horizon ("dynastic") optimization, fertility choice and R&D-based semi-endogenous growth. We have shown that it has the capacity to produce an asymptotic steady-state birth rate, and thus an asymptotic BGP. The long-run growth rate of the economy is fully endogenous and the ultimate factor driving growth is fertility.[18]

The paper ends with a brief review of empirical literature quantifying the IES in consumption that is required to exceed unity in the class of R&D-based growth models with endogenous fertility and CRRA utility. Since the empirical evidence favors the opposite case of IES<1, we arrive at a puzzle that calls for a resolution.

[18]Of course, we have completely set aside the issue of policy effectiveness in our model. Without developing these arguments any further, we indicate here that by saying that growth is "fully endogenous," we mean that there is a possibility for public interventions (and more systematic public policies) to change the long-run growth rate of the economy. By saying that fertility is the "ultimate factor driving growth," we mean that the only way a policymaker can influence the long-run economic growth rate, is by influencing the long-run fertility rate.

4.6 Mathematical appendix

Proof of Proposition 4.1. The second derivatives of the Hamiltonian (4.15), after substituting the appropriate expression for Λ_t, read:

$$\frac{\partial^2 \mathcal{H}}{\partial N_t^2} = \theta(\theta-1)N_t^{\theta-2}ue^{-\rho t}, \tag{4.32}$$

$$\frac{\partial^2 \mathcal{H}}{\partial N_t \partial b_t} = (\theta-1)N_t^{\theta-1}[u_b - u_y A_t^\sigma \beta'(b_t)]e^{-\rho t}, \tag{4.33}$$

$$\frac{\partial^2 \mathcal{H}}{\partial b_t^2} = N_t^\theta [u_{bb} - u_y A_t^\sigma \beta''(b_t) + u_{yy}A_t^{2\sigma}(\beta'(b_t))^2]e^{-\rho t}. \tag{4.34}$$

The Hessian is negative definite if $\frac{\partial^2 \mathcal{H}}{\partial N_t^2} < 0$ (which is the case since $\theta \in (0,1)$ and $u > 0$) and its determinant $\det(D^2\mathcal{H})$ is positive. The latter statement can be rewritten as:

$$\begin{aligned} \det(D^2\mathcal{H}) = N_t^{2\theta-2}e^{-2\rho t}(\theta-1) \times \\ \times \Bigg(\theta u(u_{bb} - u_y A_t^\sigma \beta''(b_t) + u_{yy}A_t^{2\sigma}(\beta'(b_t))^2) - \\ - (\theta-1)(u_b - u_y A_t^\sigma \beta'(b_t))^2 \Bigg) > 0. \end{aligned} \tag{4.35}$$

Thus, the determinant is positive if the expression in big parentheses is negative. Taking the asymptotic BGP approximation: $\lim_{t\to\infty} \frac{u_y A_t^\sigma}{u} = \frac{1-\gamma}{(1-\beta(\bar{b}))}$; $\lim_{t\to\infty} \frac{u_{yy}A_t^{2\sigma}}{u} = -\frac{(1-\gamma)\gamma}{(1-\beta(\bar{b}))^2}$ and rearranging transforms this requirement to a condition that

$$\frac{\beta''(\bar{b})(1-\beta(\bar{b}))}{(\beta'(\bar{b}))^2} > \frac{1-\gamma-\theta}{\theta}. \quad \blacksquare \tag{4.36}$$

Proof of Proposition 4.2. Dividing (4.18) sidewise by $u > 0$ (which makes both sides of the equation asymptotically stationary), equating $\dot{b}_t$ to zero and taking the limits $\lim_{t\to\infty} u_b/u = \lim_{t\to\infty} u_{bb}/u = 0$ yields:

$$\begin{aligned} \theta - \frac{u_y A_t^\sigma \beta'(\bar{b})}{u}\left(\rho - \left(\theta + \frac{\sigma\lambda}{1-\phi}\right)(\bar{b}-d)\right) + \\ + \frac{u_{yy}A_t^{2\sigma}(1-\beta(\bar{b}))\beta'(\bar{b})}{u}\frac{\sigma\lambda}{1-\phi}(\bar{b}-d) = 0. \end{aligned} \tag{4.37}$$

Taking further limits: $\lim_{t\to\infty} \frac{u_y A_t^\sigma}{u} = \frac{1-\gamma}{(1-\beta(\bar{b}))}$; $\lim_{t\to\infty} \frac{u_{yy} A_t^{2\sigma}}{u} = -\frac{(1-\gamma)\gamma}{(1-\beta(\bar{b}))^2}$, makes us rewrite the steady-state equation (4.37) as an equation in a single variable $\bar{b}$. It reads:

$$\frac{\theta}{1-\gamma}\left(\frac{1-\beta(\bar{b})}{\beta'(\bar{b})}\right) = \rho - \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\bar{b}-d). \quad \blacksquare \tag{4.38}$$

Proof of Proposition 4.3. Note that both sides of (4.21) are continuous for $\bar{b} \in [0, \beta^{-1}(1)]$. Denote:

$$\Phi(\bar{b}) = \frac{\theta}{1-\gamma}\left(\frac{1-\beta(\bar{b})}{\beta'(\bar{b})}\right) - \rho + \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\bar{b}-d). \tag{4.39}$$

At the steady state, $\Phi(\bar{b}) = 0$. The boundary values of Φ are as follows:

$$\begin{aligned}\Phi(0) &= \frac{\theta}{1-\gamma}\frac{1}{\beta'(0)} - \rho - \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)d, & (4.40)\\ \Phi(\beta^{-1}(1)) &= -\rho + \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\beta^{-1}(1)-d). & (4.41)\end{aligned}$$

Thus, existence of at least one root follows from the intermediate value theorem if (i) $\Phi(0) > 0$ and $\Phi(\beta^{-1}(1)) < 0$ or if (ii) $\Phi(0) < 0$ and $\Phi(\beta^{-1}(1)) > 0$. In case (i), these two conditions are equivalent to (4.22) and (4.23), respectively. In case (ii), one has to reverse the signs of both inequalities. $\blacksquare$

Proof of Proposition 4.4. To assure local asymptotic stability in the single-dimensional setup, we have to show that $\dot{b}_t$ as a function of b_t crosses the zero line from above.

The left-hand side of (4.18), Ψ, divided by $u > 0$, reads:

$$\bar{\Psi}(\bar{b}) = -\frac{\gamma(\beta'(\bar{b}))^2 + (1-\beta(\bar{b}))\beta''(\bar{b})}{(1-\beta(\bar{b}))^2} < 0. \tag{4.42}$$

Negativity of $\bar{\Psi}$ at the steady state, equivalent to $\frac{(1-\beta(\bar{b}))\beta''(\bar{b})}{(\beta'(\bar{b}))^2} > -\gamma$, follows from the second order condition:

$$\frac{(1-\beta(\bar{b}))\beta''(\bar{b})}{(\beta'(\bar{b}))^2} > \frac{1-\gamma-\theta}{\theta} > -\gamma, \tag{4.43}$$

because the last inequality is equivalent to the trivial $(1-\theta)(1-\gamma) > 0$.

It follows that $\dot{b}_t$ crosses the zero line from above if the right-hand side of (4.18), Φ, divided by $u > 0$, crosses the zero line from below, as in (4.24).

One final remark is due here: if (4.22) and (4.23) hold, we know that Φ crosses the zero line from above at least once (it is surely true for the *first* and *last* time it crosses zero), but it doesn't have to cross the zero line from below. If the signs in (4.22) and (4.23) are reversed, then Φ crosses the zero line from below at least once so there exists at least one asympotically stable steady-state birth rate.

In the case the steady state is unstable, we obtain the following. Unless the knife-edge value of b_t implying $\dot{b}_t = 0$ is chosen, two possibilities may emerge. First, if $b \to 0$ implying $u_b \to \infty$, then $\Lambda < 0$ and $\dot{\Lambda} < 0$ (note that $\dot{\Lambda}/\Lambda + \dot{N}/N = \frac{\theta u}{u_b - u_y A_t^\sigma \beta'(b_t)} > 0$ as $u_b \to \infty$). In consequence, $\lim_{t\to\infty} \Lambda N < 0$ so the transversality condition is violated. Second, if $b \to \beta^{-1}(1)$ then the possibility that $y \to \infty$ (which happens if $\rho > \left(\theta + (1-2\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\beta^{-1}(1) - d))$ results in:

$$\lim_{b\to\beta^{-1}(1)} \frac{\gamma\beta'(b)}{1-\beta(b)}\dot{b} = \rho - \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\beta^{-1}(1) - d),$$

and thus

$$\frac{\dot{\Lambda}}{\Lambda} + \frac{\dot{N}}{N} \quad \to \quad \left(\theta + (1-\gamma)\frac{\sigma\lambda}{1-\phi}\right)(\beta^{-1}(1) - d) - \rho + \frac{\gamma\beta'(b)}{1-\beta(b)}\dot{b} \to 0.$$

Since $y \to \infty$ might happen under acceptable parameter configurations, this means that the possibility that $b \to \beta^{-1}(1)$ cannot be ruled out of the analysis on the basis of the transversality condition.

The second order condition can be analyzed as follows: dividing the determinant of $D^2\mathcal{H}$ as defined in (4.35) by $u_y A_t^\sigma$ and taking the limit $b \to \beta^{-1}(1)$ simplifies the formula inside the parentheses in (4.35) as $\beta'(\beta^{-1}(1))^2 \left(\frac{1-\theta-\gamma}{1-\gamma}\right)$. This means that the second order condition fails if $1 - \theta - \gamma$ is positive. In such case, the solution $b \to \beta^{-1}(1)$ must be ruled out of the analysis and thus (since $b \to 0$ is also impossible), no transitional adjustments in b_t are possible and the birth rate is always set so as to assure $\dot{b}_t = 0$. ■

Chapter 5

A New Class of Production Functions and an Argument Against Purely Labor-Augmenting Technical Change

5.1 Introduction

This paper accepts the Jones' (2005) view that the production function, commonly assumed by macroeconomists to be a primitive, is in fact only a reduced form which should be derived from microfoundations. It is also acknowledged that such economy-wide production function has to be viewed as an assembly of a multiplicity of production techniques, particular methods of producing the final good.

The intuition behind this is quite simple. As Jones (2005, p. 517-8) puts it, "Suppose production techniques are ideas that get discovered over time. One example of such an idea would be a Leontief technology that says, 'for each unit of labor, take k^* units of capital. Follow these instructions [omitted], and you will get out y^* units of output.' The values k^* and y^* are parameters of this production technique. If one wants to produce with a very different capital-labor ratio from k^*, this

Leontief technique is not particularly helpful, and one needs to discover a new idea 'appropriate' to the higher capital-labor ratio. (...) According to this view, the standard production function that we write down, mapping the entire range of capital-labor ratios into output per worker, is a reduced form. It is not a single technology, but rather represents the substitution possibilities across different production techniques."

Generalizing the Jones' setup, we derive from idea-based microfoundations a new class of production functions, baptized herein *Clayton–Pareto functions*. It is indeed a large class: it nests both the Cobb-Douglas and the CES function.[1] We think of this as a particularly desirable property because it is argued that both the Cobb-Douglas and the CES could be the plausible form for the real-world economy-wide production function (an in-depth discussion and an empirical investigation of this issue can be found in Duffy and Papageorgiou, 2000). Thus, if one carries out econometric tests to check whether it is the Cobb-Douglas or the CES function that is more useful for explaining the observed data, why couldn't one check if there exists an even more general class that provides more degrees of freedom and thus provides an even better fit? The Clayton-Pareto class could be useful in this regard.

To our knowledge, both the derivation of a class of production functions that nests the Cobb-Douglas and the CES and the provision of an idea-based microfoundation for the CES production function are novel to the literature.

Taking Jones' (2005b) results as our benchmark, we show that by the means of a slight modification of his assumptions, the "Cobb-Douglas global production function and purely labor-augmenting technical change" result can be overturned. We also build a formal link between Jones (2005b) and Caselli and Coleman (2006). We find an idea-based explanation for (a generalization of) the shape of the technology frontier Caselli and Coleman postulate and thus in a sense, we bring their model to a common denominator with Jones'. These two

[1]The parameters of our microfounded production function are interrelated. In consequence, it makes sense to indicate that "the Clayton-Pareto family nests both the Cobb-Douglas and the CES": saying that it nests the CES only (which itself nests the Cobb-Douglas as a limiting case) would not reflect the result properly. At the intersection of conditions guaranteeing the CES, and the Cobb-Douglas result, we would obtain a Cobb-Douglas function with exponents $(.5, .5)$ only.

models have been empirically justified on diverse bases,[2] but they are characterized by a profound mathematical unity which we would like to uncover.

As a by-product of our analytical method, we also enrich the literature by providing alternative proofs for Jones' results and finding explicit solutions for the firms' technology choices wherever it is possible.

Unfortunately, we are unable to provide closed-form formulae for all Clayton-Pareto functions. We carry out a detailed study of solvable special cases instead.

To discuss the long-run implications of the new class of production functions, we embed our microfounded framework in the standard neoclassical (Solow, 1956) growth model. We find that Jones', and Acemoglu's (2003), results of technical change being purely labor-augmenting in the long run do not hold in general, but do hold in certain important cases, among which independence of the capital- and labor-augmenting developments – the case studied by Jones – is probably the most prominent. In general, *technical change augments both capital and labor,* even in the long run. One simple intuition for this result is that if productivities of production factors are correlated, then it is impossible to achieve higher and higher productivity of labor without altering the productivity of capital as well. We shall study this issue in greater detail in section 4.

In that section, we shall also discuss the possibility of endogenous growth that arises if the global production function exhibits elasticity of substitution greater than one in the long run.[3]

It is beyond all doubt that the problems discussed within this paper are important for growth theory. Indeed, we are considering such fundamental questions as: "What is the true shape of the economy-wide production function?", "What direction of technical change should we expect in the long run?", and "What drives the elasticity of substitution

[2]The idea that countries with different factor endowments use different production technologies has been studied earlier, among others, by Atkinson and Stiglitz (1969) as well as Basu and Weil (1998).

[3]Such possibility has already been noticed in the seminal work of Solow (1956). There exists substantial literature that deals with these issues, including de La Grandville (1989), Jones and Manuelli (1990), Yuhn (1991), Klump and de La Grandville (2000), and Palivos and Karagiannis (2004).

between production factors?". We manage to get new insights here because we derive what most economists assume, and endowing the models with solid microfoundations lets one draw more refined macro-scale conclusions.[4]

As far as empirical relevance of Clayton-Pareto production functions is concerned, they should be tested in future against the null of a Cobb-Douglas or a CES function, perhaps in a manner similar to the literature on translog functions (Christensen, Jorgenson, and Lau, 1973; Griffin, 1977; Kim, 1992). The upside of the current analysis is that, as opposed to translog functions which are essentially a second order approximation to any arbitrary twice-differentiable function, our class of production functions is derived from sound microfoundations, and all our parameters are directly interpretable.

The remainder of the article is structured as follows. In section 2, we generalize the Jones' setup by allowing for dependence between capital- and labor-augmenting developments and we derive the class of Clayton-Pareto functions. In section 3, we discuss the cases of particular interest and describe the determinants of the elasticity of substitution. In section 4, we determine the long-run direction of technical change and check the conditions for endogenous growth. Section 5 concludes.

5.2 The microfounded framework

Our microfounded framework has a representative profit-maximizing firm. In order to produce, the firm picks a single production technique (henceforth a *local production function*, LPF) from the range of available ones. Intuitively, one could associate each LPF with a "recipe", a list of instructions to follow. In line with this intuition, LPFs should be rigid and not allow for much substitutability between the factors of

[4]Earlier literature that deals with closely related issues includes Houthakker (1955-56), Uzawa (1961), Kortum (1997) and Acemoglu (2003). Moreover, endogenous technology choice has common features with the assignment problem with heterogenous workers and tasks, analyzed by Sattinger (1975); and vintage capital theory where technological improvements are embodied in consecutive vintages of machines. This strand of literature dates back to Solow (1960), as well as Solow *et al.* (1966).

production.[5]

The second assumption is that factors can be utilized, according to a given LPF, with certain efficiency levels only. Following Jones (2005b), we identify the notion of an *idea* with these efficiency levels (unit factor productivities). Since we consider two factors of production only: capital K and labor L, an idea is consequently a pair (a, b), where b and a are unit productivities of capital and labor, respectively.[6] Technical change is then identified with the sequential arrival of new, better and better, ideas. The *global production function* (GPF) is the convex hull of LPFs: such an assembly of LPFs that for each K and L, productivities b and a are chosen optimally by the firm whose choice is constrained within the set of available ideas. The *technology frontier* is a subset of this set which contains only the ideas which could possibly be used by the profit-maximizing firm.

Following Jones (2005b), we assume that each dimension of an idea, be it a or b, is randomly drawn from a Pareto distribution.[7] However, when a new idea arrives, it is already a pair (a, b) chosen from some joint bivariate distribution. The individual Pareto distributions of factor productivities serve only as marginal distributions here.

As opposed to Jones, we allow for dependence between the marginal idea distributions. On the positive side, this poses a new problem: we do not have any empirical evidence on the pattern of dependence. On the normative side, however, the Clayton family of copulas[8] can be used to show that marginal Pareto distributions imply neither a Cobb-Douglas global production function, nor purely labor-augmenting technical change in the long run. On the contrary, they produce a wide va-

[5]An extreme but probably the most intuitive example is the Leontief LPF ("fixed coefficients") where the factors have to be used up in fixed proportions. More flexible functional forms are plausible as well, as we shall see shortly.

[6]We reverse the order of a and b in order to stick to the Jones' (2005) notation where possible.

[7]Empirical evidence for the prevalence of Pareto distributions in scientific productivity dates back to Lotka (1926). Theoretical literature linking Pareto distributions to exponential growth includes Kortum (1997) and Gabaix (1999).

[8]There exist many families of copulas different to the one we have chosen. The Clayton family belongs to the Archimedean class, as i.e. the Frank and the Gumbel family do. Another widely recognized class of copulas is the elliptic class. See Nelsen (1999) for an introduction to the copula theory.

riety of possibilities, the empirically relevant CES function being among them.

Our derivation procedure of the GPF from the LPFs and the technology frontier can be decomposed into three steps:

1. Derive the technology frontier from the joint distribution of ideas. This is done in Proposition 5.2.

2. Find the optimal unit factor productivities a^* and b^* (i.e. pick the optimal LPF) given the available technology level and the associated technology frontier as well as stocks of capital K and labor L. This task is accomplished in Proposition 5.3.

3. Build the convex hull of LPFs – i.e. insert the optimal pair of technologies (a^*, b^*) to the LPF, separately for each pair of endowments (K, L). The convex hull of LPFs is the GPF, as in Proposition 5.4.

Step 1. The technology frontier

It is assumed that when new ideas arrive, their levels are stochastic and drawn from Pareto distributions.

Assumption 5.1 *Unit factor productivities $\tilde{a}$ and $\tilde{b}$ are Pareto-distributed:*

$$P(\tilde{a} > a) = \left(\frac{\gamma_a}{a}\right)^{\alpha}, \qquad a \geq \gamma_a > 0, \qquad \alpha > 0, \tag{5.1}$$

$$P(\tilde{b} > b) = \left(\frac{\gamma_b}{b}\right)^{\beta}, \qquad b \geq \gamma_b > 0, \qquad \beta > 0. \tag{5.2}$$

We allow $\tilde{a}$ and $\tilde{b}$ to be mutually dependent. A multiplicity of joint (bivariate) ideas distributions, with (5.1) and (5.2) as marginal distributions, is produced using the Clayton family of copulas.

Assumption 5.2 *Dependence between the unit factor productivities $\tilde{a}$ and $\tilde{b}$ is represented by the Clayton copula, specified in* (5.3).

All members of the Clayton family of copulas are characterized by the following formula (Nelsen, 1999):

$$C(u, v) = \max \left\{ 0, (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta} \right\}, \tag{5.3}$$

where u and v are random variables, uniformly distributed over the unit interval, and $\delta \geq -1$ captures the degree and sign of dependence between the marginal idea distributions: if $\delta < 0$, they are negatively correlated; if $\delta > 0$, they are positively correlated. $\delta = 0$ denotes independence and calls for a replacement of (5.3) with $C(u, v) = uv$. We consider δ to be *the* crucial parameter here, because it is exactly the Jones' (2005) assumption of $\delta = 0$ that we relax and whose relaxation yields so interesting results.

The random variables u and v should be replaced with suitable cumulative distribution functions (CDFs). For the ease of exposition, we relegate this point to Appendix 5.6.1 which contains the proof of the following proposition.

Proposition 5.1 *Distribution of the two-dimensional random variable* $(\tilde{a}, \tilde{b})$ *is given by*

$$P(\tilde{a} > a, \tilde{b} > b) = \max \left\{ 0, \left(\left(\frac{\gamma_a}{a} \right)^{-\alpha\delta} + \left(\frac{\gamma_b}{b} \right)^{-\beta\delta} - 1 \right)^{-\frac{1}{\delta}} \right\}, \tag{5.4}$$

if $\delta \in [-1, +\infty) \backslash \{0\}$*; or*

$$P(\tilde{a} > a, \tilde{b} > b) = \left(\frac{\gamma_a}{a} \right)^{\alpha} \left(\frac{\gamma_b}{b} \right)^{\beta}, \tag{5.5}$$

if $\delta = 0$ *(the marginal distributions are independent).*

Proof. See Appendix 5.6.1. ■

The probability (5.4) stands a chance of being zero only if $\delta < 0$. We also note that imposing $\delta = 0$ (independence) leads to (5.5) which is equation (20) of the Jones' paper.

We shall now define formally one of the most important concepts of the paper: the technology frontier. Application of this definition will make the correspondence between the stochastic arrival of ideas and the deterministic GPF clearer.

Definition 5.1 *The* technology frontier *is a curve in the* (a, b) *space, such that the probability* $P(\tilde{a} > a, \tilde{b} > b)$ *is constant along this curve.*

This seemingly simple definition helps us move away from the tedious probabilistic considerations back to the deterministic world. It is consistent with the approach of Caselli and Coleman (2006), and section II of the Jones' article. However, its motivation turns out to be not so simple after all. In quest of such, we build a little model of research where each invention needs to be understood and connected with the existing stock of knowledge by a given percentage of researchers before it comes into industrial use. An outline of this model has been relegated to Appendix 5.6.4 so that the main line of reasoning is not obstructed by this digression.

Using Definition 5.1 and equations (5.4) – (5.5), we can write the formula for the technology frontier.

Proposition 5.2 *Assume* $\delta \in [-1, +\infty)\backslash\{0\}$. *Then, the technology frontier* $H(a, b)$ *is given by*

$$H(a, b) = \gamma a^{\alpha\delta} + b^{\beta\delta} = N, \tag{5.6}$$

where $\gamma = \gamma_b^{\beta\delta}/\gamma_a^{\alpha\delta} > 0$, *and* $N = \gamma_b^{\beta\delta}[P(\tilde{a} > a, \tilde{b} > b)^{-\delta} + 1] > 0$. *See that* $N \in [\gamma_b^{\beta\delta}, 2\gamma_b^{\beta\delta}]$ *if* $\delta < 0$, *and* $N \geq 2\gamma_b^{\beta\delta}$ *if* $\delta > 0$.

If $\delta = 0$, *then the technology frontier becomes*

$$H(a, b) = a^{\alpha} b^{\beta} = N, \tag{5.7}$$

where $N \equiv \frac{\gamma_a^{\alpha}\gamma_b^{\beta}}{P(\tilde{a}>a, \tilde{b}>b)} \geq \gamma_a^{\alpha}\gamma_b^{\beta}$.

Proof. It is easy and requires only algebraic manipulations. ■

Equation (5.7) above is Jones' equation (8). Please note the following consequence of Proposition 5.2: if there is negative dependence between the unit factor productivities ($\delta < 0$), then perspectives for technological progress are limited: N is bounded. In such case, we have a strong "fishing-out" effect: with $N = \gamma_b^{\beta\delta}$, the probability $P(\tilde{a} > a, \tilde{b} > b)$ becomes zero and further progress is impossible. Intuitively, because of the negative dependence, all further labor-(capital-)augmenting innovations would have to be unambiguously capital-(labor-)impairing.

On the other hand, if there is positive dependence between individual factor productivities ($\delta > 0$), then R&D can well go on forever. We will need these results for our considerations on the long-run direction of technical change in section 4.

Step 2. Endogenous technology choice

We shall now make an important assumption that assures analytical tractability of our model. Namely, we assume the LPFs to be CES.

Assumption 5.3 *The local production function (LPF) $\tilde{Y}$ is given by*

$$\tilde{Y}(K, L; a, b) = \tilde{A}\Big(\psi(bK)^{\theta} + (1-\psi)(aL)^{\theta}\Big)^{\frac{1}{\theta}}, \tag{5.8}$$

where $\tilde{A} > 0$, $\theta \in (-\infty, 1]\backslash\{0\}$ and $\psi \in (0,1)$, or

$$\tilde{Y}(K, L; a, b) = \tilde{A}(bK)^{\psi}(aL)^{1-\psi}, \tag{5.9}$$

where $\tilde{A} > 0$ and $\psi \in (0,1)$, if $\theta = 0$.

This assumption is taken from Caselli and Coleman (2006). At this point, we do not impose $\theta < 0$ on top of it, as they do, but doing so would be natural for anyone who sticks to the Jones' "recipe" interpretation of the LPF anyway. We allow for more flexibility here, keeping in mind the interesting limiting case of Leontief LPFs ($\theta = -\infty$).[9]

In the following analysis, we shall also assume that $\alpha\delta > \theta$ and $\beta\delta > \theta$, so that the curvature of the LPF is always greater than the curvature of the technology frontier. This would guarantee an interior solution to the optimization problem presented below. These two inequality constraints have been derived from the second-order conditions in Appendix 5.6.3.

The competitive firm faces a continuum of LPFs given by (5.8) or (5.9), and indexed by a and b along the available technology frontier

[9] Assuming $\theta < 0$ is equivalent to saying that the elasticity of substitution of the LPF, $\sigma = 1/(1-\theta)$, lies below unity. The "recipe" understanding of a LPF leads to $\theta = -\infty$, so that the LPFs are Leontief. See Jones (2005, p. 517-8) for an intuitive clarification of his understanding of a LPF. On the other hand, if the LPF denotes a country's production function as in Caselli and Coleman (2006), one would expect $\theta \approx 0$ rather than $\theta \approx -\infty$.

(5.6) or (5.7). It chooses one of them in order to maximize its profit. It also optimally chooses its demand for capital and labor.

The problem of the competitive firm can be decomposed into two separate phases: first, choosing the optimal technology (a, b) for given endowments K, L and prices r, w; and then choosing the demand for both factors of production, taking their prices as given. Because we are only interested in finding the shape of the GPF and not in the general equilibrium of the economy, we shall skip the second phase.[10]

The first phase of the competitive firm's optimization problem in the typical case $\delta \neq 0$, $\theta \neq 0$ can be written as

$$\max_{a,b} \left\{ \tilde{A} \left(\psi (bK)^\theta + (1-\psi)(aL)^\theta \right)^{\frac{1}{\theta}} \right\} \text{ s.t. } \gamma a^{\alpha\delta} + b^{\beta\delta} = N. \tag{5.10}$$

First-order conditions imply that[11]

$$\frac{b^{\beta\delta-\theta}}{a^{\alpha\delta-\theta}} = \frac{\gamma\psi}{1-\psi}\frac{\alpha}{\beta}k^\theta, \tag{5.11}$$

where we denoted $k \equiv K/L$ for convenience. Solving this equation for b yields

$$b = ck^{\frac{\theta}{\beta\delta-\theta}} a^{\frac{\alpha\delta-\theta}{\beta\delta-\theta}}, \tag{5.12}$$

which, plugged into (5.6), gives

$$\Phi(a, k; N) = \gamma a^{\alpha\delta} + c^{\beta\delta} k^{\frac{\theta\beta\delta}{\beta\delta-\theta}} a^{\frac{\beta\delta(\alpha\delta-\theta)}{\beta\delta-\theta}} - N = 0, \tag{5.13}$$

where $c \equiv \left(\frac{\gamma\psi}{1-\psi}\frac{\alpha}{\beta} \right)^{\frac{1}{\beta\delta-\theta}}$.

Proposition 5.3 *Let $\alpha\delta > \theta$ and $\beta\delta > \theta$. Then, the optimization problem* (5.10) *allows a unique positive solution* $(a^*(k; N), b^*(k; N))$.

Proof. See Appendix 5.6.2. ■

Equation (5.13) suffices to infer existence and uniqueness of the solution; unfortunately, it contains a sum of two arbitrary powers of a

---

[10] Caselli and Coleman (2006) close the model by assuming that the produced good is the numeraire and in the whole economy, stocks of available capital K and labor L are fixed. Jones (2005b) skips the second phase.

[11] Exact derivations have been relegated to Appendix 5.6.3.

and thus, unless some particular equality constraint holds, no explicit formula for a as a function of capital per worker k and technology level N can be obtained from it. If $\delta = 0$ (the Jones' case) or $\theta = 0$ (Cobb-Douglas LPFs), then a Cobb-Douglas GPF is obtained. The most interesting case is nevertheless the one with $\alpha = \beta$ (equal Pareto exponents) – it yields the CES result.

Comparing our results to Jones', we see that here the optimal technology levels a^* and b^* depend on θ, the curvature parameter of the LPF. Hence, the Jones' (2005, p. 528) claim, that

> [o]f course, the intuition regarding the global production function suggests that it is determined by the distribution of ideas, not by the shape of the local production function

does not hold in general. Equation (5.13) clearly suggests that in the typical Clayton-Pareto case, the shape of the LPF *matters* for the shape of the GPF.

Step 3. The global production function

We shall now find the convex hull of LPFs, with $a^* \equiv a^*(k; N)$ and $b^* \equiv b^*(k; N)$ defined as above.

The most straightforward (and underestimated) way to do it is to insert a^* and b^* directly into the LPF formula (5.8). Another one is to characterize it in terms of partial elasticities, following Jones' equation (7):

$$\frac{\varepsilon_K}{1 - \varepsilon_K} = \frac{\eta_b}{\eta_a}, \tag{5.14}$$

where $\varepsilon_K = (\partial Y / \partial K)(K/Y)$ is the partial elasticity of the GPF with respect to capital; and η_a, η_b are partial elasticities of the technology frontier H with respect to a and b, respectively. Deriving equation (5.14) requires usage of the Envelope Theorem and the degree-one homogeneity of the GPF.

We shall point out that although both methods give the same results for the *shape* of the GPF, they are not equivalent. The first method is probably computationally more demanding, but it does not incur a loss of information due to differentiation as the second one does. Thus, only thanks to the first method all parameters of the GPF, including its intercept term, can be recovered.

Summing up, we obtain the following characterization of a GPF which belongs to the Clayton-Pareto family:

Proposition 5.4 *Given Assumptions 5.1, 5.2 and 5.3, Propositions 5.1 and 5.2, and the firms' optimizing behavior summarized in* (5.10)*, the GPF is given by*

$$Y(K, L; N) \equiv \tilde{A}\left(\psi(b^*K)^\theta + (1-\psi)(a^*L)^\theta\right)^{\frac{1}{\theta}}, \tag{5.15}$$

where a^* and b^* satisfy (5.11)*,* (5.12) *and* (5.13)*. This implies that*

$$\frac{\varepsilon_K}{1-\varepsilon_K} = \frac{\psi}{1-\psi}\left(\frac{b^*K}{a^*L}\right)^\theta = \frac{\beta}{\gamma\alpha}\frac{(b^*)^{\beta\delta}}{(a^*)^{\alpha\delta}}. \tag{5.16}$$

Proof. Already given in text. ■

Proposition 5.4 implicitly defines all functions that belong to the Clayton-Pareto class. A few of them have been depicted in Figure 5.1.

The most important advantage of the Clayton-Pareto class of functions is its generality. We have shown that from a specific setup that is based on Pareto distributions of factor productivities, a much wider class of functions can be derived that it had been previously acknowledged.

Unfortunately, not all Clayton-Pareto functions are really *production* functions: some of them are not concave and thus violate the diminishing marginal utility requirement. Such situation is most likely to emerge if $\alpha\delta \approx \theta$, i.e. if the second-order conditions are satisfied with a very narrow margin. In Figure 5.1, we present such a case for $\delta = -.09$. See also that for $\delta = -.01$, the Clayton-Pareto function is concave, albeit its implied marginal product of capital decreases *very* slowly. As we will see in the next section, in the CES case the concavity condition can be put in a simple analytical form, $\alpha\delta - \theta - \alpha\delta\theta > 0$.

5.3 Special cases

Let us now dwell more on some special cases that provide explicit solutions. They all have meaningful economic interpretations. We find that the Clayton-Pareto family nests the Cobb-Douglas function and the CES function.

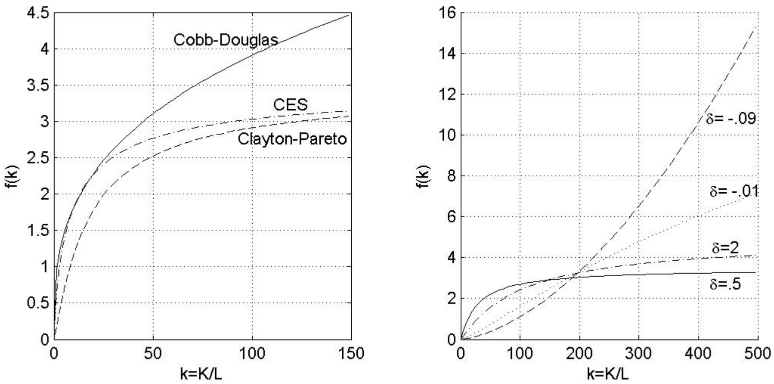

Figure 5.1: Left panel: a representative Clayton-Pareto production function compared to the Cobb-Douglas and the CES. Right panel: impact of the degree of dependence δ on the resultant Clayton-Pareto production functions. Assumed parameter values (unless indicated otherwise): $\tilde{A} = 1, \gamma_a = 1, \gamma_b = .2, \theta = -1, \alpha = 5, \beta = 2.5, \psi = .5, \delta = .5$, $P(\tilde{a} > a, \tilde{b} > b) = .1$. For the CES case, we equalized $\alpha = \beta = 5$; for the Cobb-Douglas case, $\delta = 0$.

5.3.1 Independent Pareto distributions

Independence of the marginal idea distributions, i.e. $\underline{\delta = 0}$, yields a particularly nice and interpretable result – the benchmark result due to Jones (2005b). It links *independent* Pareto idea distributions with Cobb-Douglas production functions.

With $\delta = 0$, the technology frontier is given by equation (5.5), which implies that the firm's optimality condition equivalent to (5.13) can be solved explicitly as

$$\begin{cases} a^*(k;N) &= N^{\frac{1}{\alpha+\beta}}\tilde{c}^{\frac{\beta}{\alpha+\beta}}k^{\frac{\beta}{\alpha+\beta}} \\ b^*(k;N) &= N^{\frac{1}{\alpha+\beta}}\tilde{c}^{-\frac{\alpha}{\alpha+\beta}}k^{-\frac{\alpha}{\alpha+\beta}}, \end{cases} \tag{5.17}$$

where $\tilde{c} = \left(\frac{\alpha}{\beta}\frac{\psi}{1-\psi}\right)^{1/\theta}$ and $\theta < 0$. As Jones finds out, the GPF comes to satisfy:

$$\frac{\varepsilon_K}{1-\varepsilon_K} = \frac{\beta}{\alpha}, \tag{5.18}$$

so it needs to be Cobb-Douglas with constant returns to scale, and exponents proportional to α and β, i.e. it needs to be

$$Y(K, L; N) = A(N) K^{\frac{\beta}{\alpha+\beta}} L^{\frac{\alpha}{\alpha+\beta}}. \tag{5.19}$$

Further calculations yield that the factor-neutral productivity term $A(N)$ is given by

$$A(N) = \tilde{A} N^{\frac{1}{\alpha+\beta}} \left[\psi^{\frac{\beta}{\alpha+\beta}} (1-\psi)^{\frac{\alpha}{\alpha+\beta}} \left(\left(\frac{\alpha}{\beta}\right)^{-\frac{\alpha}{\alpha+\beta}} + \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} \right) \right]^{\frac{1}{\theta}}.$$

Moreover, if we take $\theta \to -\infty$, so that the LPFs are approximately Leontief, then we obtain $A(N) \to \tilde{A} N^{\frac{1}{\alpha+\beta}}$. This precisely corresponds to Jones' equation (28).

In the case of independent Pareto distributions, we could have missed the second step (finding a^* and b^*), and yet we would have arrived at the same result. This is because in (5.18), the GPF is required to exhibit *constant* partial elasticities with respect to its arguments, which is a defining property of Cobb-Douglas functions. So here, shapes of the LPFs have no impact on the shape of the GPF.[12]

5.3.2 Cobb-Douglas local production functions

It turns out if local production functions are Cobb-Douglas (i.e. $\underline{\theta = 0}$), then the GPF is Cobb-Douglas as well, irrespective of the shape of the technology frontier.[13] This is fairly confusing a result, taking into account the Jones' assertion which we cited in section 5.2.

Namely, once we assume that the LPFs are given by (5.9), we obtain that the shape of the technology frontier takes no part in determining

[12]We have to maintain the assumption that local production functions exhibit constant returns to scale, though. Moreover, to remain in the interior solution, we still have to assume that the elasticity of substitution of the LPFs is everywhere lower than unity. Finally, shapes of the LPFs – described by the parameters θ and ψ – do have an impact on $A(N)$.

[13]Cobb-Douglas LPFs are not plausible if one maintains a "recipe" understanding of a LPF, because they are characterized by unitary elasticity of substitution between factors. On the other hand, for the Caselli and Coleman's understanding of a LPF as a country-wide production function, they are perfectly plausible.

the shape of the GPF. To see this implication, note that in the last step of our procedure, we obtain

$$\frac{\varepsilon_K}{1-\varepsilon_K} = \frac{\psi}{1-\psi}, \tag{5.20}$$

implying a GPF of a Cobb-Douglas form:

$$Y(K,L;N) = A(N)K^{\psi}L^{1-\psi}, \tag{5.21}$$

irrespective of the shape of the technology frontier, provided that the two marginal distributions of unit productivities are positively dependent ($\delta > 0$) which is necessary for an interior solution.

Further calculations yield $A(N) = \tilde{A}\bar{c}^{\frac{\psi}{\beta\delta}}\left(\frac{N}{\gamma+\bar{c}}\right)^{\frac{\psi}{\beta\delta}+\frac{1-\psi}{\alpha\delta}}$, where $\bar{c} = \frac{\gamma\psi}{1-\psi}\frac{\alpha}{\beta}$. This means that technology choice is irrelevant for the shape of GPF, but it is not irrelevant for the level of $A(N)$. To facilitate comparisons, we shall point out here that in the optimum, firms choose technologies a^* and b^* according to:

$$\begin{cases} a^*(k;N) &= \left(\frac{N}{\gamma+\bar{c}}\right)^{\frac{1}{\alpha\delta}}, \\ b^*(k;N) &= \left(\frac{N\bar{c}}{\gamma+\bar{c}}\right)^{\frac{1}{\beta\delta}}. \end{cases} \tag{5.22}$$

Note that both expressions are independent of k.

5.3.3 The CES production function

Let us now assume that the shape parameters of both Pareto idea distributions are equal, i.e. $\underline{\alpha = \beta}$. In this important knife-edge case, an explicit formula for the GPF is obtained. This function exhibits a constant elasticity of substitution, being a function of α, δ and θ. The CES result is preserved for Leontief LPFs as well.

With $\alpha = \beta$, the technology frontier (5.6) becomes:

$$H(a,b) = \gamma a^{\alpha\delta} + b^{\alpha\delta} = N, \tag{5.23}$$

which is apparently Caselli and Coleman's equation (5). By writing it down as a special case of our model, we provide a "Clayton-Pareto" microfoundation for it. Please note that this result is obtained already in Step 1 of our derivation procedure: the link between the setups of Jones as well as Caselli and Coleman is valid regardless of the shape of the LPFs which enters the analysis at a later stage.

The first order condition (5.11) for the representative firm's optimization problem (5.10) can be now solved explicitly as:

$$\begin{cases} a^*(k;N) &= \left(\frac{N}{\gamma+c^{\alpha\delta}k^{\frac{\alpha\delta\theta}{\alpha\delta-\theta}}}\right)^{\frac{1}{\alpha\delta}}, \\ b^*(k;N) &= \left(\frac{Nc^{\alpha\delta}k^{\frac{\alpha\delta\theta}{\alpha\delta-\theta}}}{\gamma+c^{\alpha\delta}k^{\frac{\alpha\delta\theta}{\alpha\delta-\theta}}}\right)^{\frac{1}{\alpha\delta}}. \end{cases} \tag{5.24}$$

Thus, the GPF satisfies the equation:

$$\frac{\varepsilon_K}{1-\varepsilon_K} = \frac{c^{\alpha\delta}}{\gamma}k^{\frac{\alpha\delta\theta}{\alpha\delta-\theta}}. \tag{5.25}$$

In consequence, it takes the CES form (see Arrow *et al.*, 1961):[14]

$$Y(K,L;N) = A(N)(\zeta K^\xi + (1-\zeta)L^\xi)^{1/\xi}, \tag{5.26}$$

where

$$\xi = \frac{\alpha\delta\theta}{\alpha\delta-\theta}, \tag{5.27}$$

$$\zeta = \frac{c^{\alpha\delta}}{c^{\alpha\delta}+\gamma}, \tag{5.28}$$

$$A(N) = \tilde{A}N^{\frac{1}{\alpha\delta}}\left((1-\psi)^{\frac{\alpha\delta}{\alpha\delta-\theta}}\gamma^{-\frac{\theta}{\alpha\delta-\theta}} + \psi^{\frac{\alpha\delta}{\alpha\delta-\theta}}\right)^{\frac{\alpha\delta-\theta}{\alpha\delta\theta}}, \tag{5.29}$$

and $c = \left(\frac{\gamma\psi}{1-\psi}\right)^{\frac{1}{\alpha\delta-\theta}}$. The assumption $\alpha\delta > \theta$ guarantees an interior solution to the competitive firm's optimization problem.

---

[14]To prove this, note that $\frac{\varepsilon_K}{(1-\varepsilon_K)} = \frac{K/L}{MRS}$, where MRS is the marginal rate of substitution between capital and labor. Straightforward algebraic manipulations yield that the elasticity of substitution, $\sigma = \frac{\partial(K/L)}{\partial MRS}\frac{MRS}{K/L} = \frac{1}{1-\xi}$, where $\xi = \frac{\alpha\delta\theta}{\alpha\delta-\theta}$. Now use the theorem due to Arrow *et al.* (1961) to get (5.26). To obtain the coefficients A and ζ, more algebra is necessary. They have been computed by equalizing $\tilde{Y}(K,L;a^*,b^*) = Y(K,L;N)$, and then gradually simplifying the resultant expression.

Let us also note that if one takes $\theta = -\infty$, i.e. assumes that the LPFs are Leontief (as in the reference case discussed by Jones, 2005), then the above result simplifies greatly while still implying a CES GPF: $\xi = -\alpha\delta$, $A(N) = \tilde{A}N^{\frac{1}{\alpha\delta}}\left(\frac{1}{\gamma+1}\right)$, $\zeta = \frac{1}{\gamma+1}$, and $c = 1$.

The exponent ξ of the Pareto-microfounded CES function, derived in (5.27), implies that the elasticity of substitution σ of the GPF is equal to

$$\sigma = \frac{1}{1-\xi} = \frac{\alpha\delta - \theta}{\alpha\delta - \theta - \alpha\delta\theta}. \tag{5.30}$$

For Leontief LPFs we obtain $\xi = -\alpha\delta$ and accordingly, $\sigma = \frac{1}{1+\alpha\delta}$. This clear-cut result clarifies that we have obtained the CES GPF here not because we have assumed CES LPFs, but because we have assumed Pareto idea distributions with equal exponents, dependent according to the Clayton copula. Indeed, for Leontief LPFs which have zero elasticity of substitution, the CES result is preserved; if any of the latter assumptions is relaxed, the CES result is *not* preserved.

One important remark is due here. To assign a serious economic interpretation to the above formulae, i.e. to assure that the elasticity of substitution σ is positive so that the function is concave and the marginal products of production factors are decreasing, we need to assume $\alpha\delta - \theta - \alpha\delta\theta > 0$.

Having this assumption in mind, we find that the elasticity of substitution σ of the GPF and the degree of dependence between the marginal Pareto distributions δ are inversely related. Moreover, we also find that for the smallest possible values of δ, σ explodes to infinity, and for $\delta \to \infty$, it approaches the (low) elasticity of substitution of the LPF:

$$\lim_{\delta \to \left(\frac{\theta}{\alpha(1-\theta)}\right)_+} \frac{\alpha\delta - \theta}{\alpha\delta - \theta - \alpha\delta\theta} = +\infty \quad \text{and} \quad \lim_{\delta \to +\infty} \frac{\alpha\delta - \theta}{\alpha\delta - \theta - \alpha\delta\theta} = \frac{1}{1-\theta}. \tag{5.31}$$

The inverse relation between δ and σ gets the most apparent once θ is evaluated at $-\infty$. Indeed, in the case of Leontief LPFs, the relation between δ and ξ is linear, the elasticity of substitution σ exceeds unity if and only if $\delta < 0$, and it approaches zero as $\delta \to +\infty$.

We conclude that what drives the elasticity of substitution of the GPF is the actual *difference* between the curvature of the LPF, described by θ, and the curvature of the technology frontier, described by $\alpha\delta$: the greater the difference, the lower the elasticity of substitution of the GPF.

The elasticity of substitution between capital (which is an accumulable factor) and labor (which cannot be accumulated) has sizeable effects on the long-run performance of the economy (see e.g. the two theorems by Klump and de La Grandville, 2000). In particular, $\sigma > 1$ may imply endogenous growth in the absence of technical progress and explosive growth when technical progress is present (see Jones and Manuelli, 1990; Palivos and Karagiannis, 2004). We shall discuss this issue in the following section.

5.4 The direction of technical change

Let us now embed our microfounded model in the neoclassical growth framework (Solow, 1956) to draw conclusions about the long-run direction of technical change.

From now on, we shall assume exogenous technical progress in the form of a decline in $P(\tilde{a} > a, \tilde{b} > b)$: as time passes, more and more efficient technologies are invented. Due to Definition 5.1, so defined an R&D activity is equivalent to pushing the technology frontier further and further. The choice of an optimal technology pair (a, b) from the technology frontier which is available at each instant of time is again left to firms. This means that we shall explicitly take advantage of the fact that in our framework, the direction of technical change is endogenous.

Capital evolves according to the usual equation of motion:

$$\dot{K} = sY - \delta_K K, \tag{5.32}$$

where $\delta_K > 0$ is the capital depreciation rate, and $s \in (0, 1)$ is the exogenous savings rate. Labor force grows at a constant rate $\hat{L} \equiv \dot{L}/L = n \geq 0$. Equation (5.32) implies that along the balanced growth path (BGP) – if there exists one, which is a very stringent condition – the product/capital ratio stays constant.

In further derivations, we shall dwell on the three particular cases, $\delta = 0$, $\theta = 0$, and $\alpha = \beta$, that offer closed-form solutions only. They already give a taste of the vast multiplicity of long-run outcomes of our microfounded model. In particular, the CES case turns out to be the most revealing: it allows us to draw preliminary conclusions about the dynamic properties of Clayton-Pareto production functions in general.

5.4.1 Independent Pareto distributions

We start with the case of independent Pareto distributions ($\underline{\delta = 0}$), studied in detail by Jones (2005b). Having assumed $\hat{N} = g$ and $y = Y/L$, and used $\hat{Y} = \hat{K}$, we find that the growth rate of the economy along the BGP is given by

$$\hat{Y} = \frac{g}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}\hat{K} + \frac{\alpha}{\alpha+\beta}\hat{L} \quad \Rightarrow \quad \hat{y} = \frac{g}{\alpha}. \tag{5.33}$$

From (5.17), we also have that $\hat{a} = \frac{g}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}\frac{g}{\alpha} = \frac{g}{\alpha}$ and $\hat{b} = \frac{g}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}\frac{g}{\alpha} = 0$. This is exactly the Jones' benchmark result of purely labor-augmenting technical change in the long run. Despite the fact, that for each given K and L, technical progress is factor-neutral in this case, proportions of factors actually used in production change along the BGP in such a way that the unit productivity of capital – b – stays exactly constant.

5.4.2 Cobb-Douglas local production functions

We can now proceed to the case of Cobb-Douglas local production functions ($\underline{\theta = 0}$).

In such case, we still obtain existence of a BGP and a Cobb-Douglas GPF. However, we observe that technical change ceases to be purely labor-augmenting in the long run. Namely, along the BGP,

$$\hat{Y} = \Big(\frac{\psi}{\beta\delta} + \frac{1-\psi}{\alpha\delta}\Big)g + \psi\hat{K} + (1-\psi)\hat{L} \quad \Rightarrow \quad \hat{y} = \Big(\frac{\psi}{(1-\psi)\beta\delta} + \frac{1}{\alpha\delta}\Big)g. \tag{5.34}$$

Now, the growth rate of the economy $\hat{y}$ depends not only on α, as in equation (5.33), but also on β (as well as ψ and δ). Redoing the same exercise as in the previous subsection, we show that in this case, along the BGP, $\hat{a} = \frac{g}{\alpha\delta} > 0$ and $\hat{b} = \frac{g}{\beta\delta} > 0$. This means that technical change augments *both* factors of production in the long run.

5.4.3 The CES production function

The case of the CES global production function ($\alpha = \beta$) turns out to be more revealing than the two Cobb-Douglas ones discussed just above, and it also gives a hint on what one can expect in the general Clayton-Pareto case.

First of all, a balanced growth path (with a positive growth rate of capital per worker) does not exist in the CES case. Thus, to get some meaningful outcome concerning the long run, we have to dwell on the asymptotics: by "in the long run" we mean "as $t \to \infty$" rather than "along the balanced growth path". Moreover, we have to consider the two cases separately: $\xi > 0$ where capital and labor are gross substitutes, and $\xi < 0$ where they are gross complements.

Case $\xi > 0$. Capital and labor are gross substitutes if $0 > \alpha\delta > \theta$ or if $\alpha\delta > \theta > 0$. Simple calculations yield that in such case,

$$\hat{k} = sA(\zeta + (1-\zeta)k^{-\xi})^{1/\xi} - \delta_K - n \quad \Rightarrow \quad \lim_{k\to\infty} \hat{k} = sA\zeta^{1/\xi} - \delta_K - n. \tag{5.35}$$

Hence, there is endogenous growth via capital accumulation if $sA\zeta^{1/\xi} - \delta_K - n > 0$. Such endogenous growth is possible here thanks to the high elasticity of substitution of the analyzed GPF.[15] When capital and labor are gross substitutes, the marginal product of capital never declines to zero, and the Inada condition in infinity does not hold.[16]

Note that equation (5.35) makes economic sense only if $\delta < 0$ (so N is bounded) or if we exogenously fix N (assume out all R&D activity). If $A = A(N)$ were to grow exogenously on top of the elasticity-driven endogenous growth, we would have arrived at an implausible explosive case of an economy that reaches infinite production in finite time. What is more, in the gross-substitutes case $\xi > 0$, with $\delta > 0$ and positive growth in $A(N)$, the condition for endogenous growth $sA\zeta^{1/\xi} - \delta_K - n > 0$ will sooner or later be satisfied. Thus, such economy is *bound* to explode to infinity in finite time. For obvious reasons, we rule this case out.

[15] $\xi > 0$ implies $\sigma > 1$. Further reasoning follows Palivos and Karagiannis (2004), and Jones and Manuelli (1990), in that order.

[16] In the general Clayton-Pareto case, σ is not constant. However, one can clearly expect that endogenous growth via capital accumulation can, and should, appear only if $\lim_{k\to\infty} \sigma(k) > 1$ (see Palivos and Karagiannis, 2004).

Within the case of an endogenously and yet non-explosively growing economy, individual factor productivities approach certain limits as $k \to \infty$ with time: $b^* \to N^{\frac{1}{\alpha\delta}}$, and $a^* \to 0$ if $\delta > 0$ or $a^* \to +\infty$ if $\delta < 0$. Hence, although there is no actual technical change, it is clearly a^*, the productivity of labor that drives long-run endogenous growth.

Case $\underline{\xi < 0}$. If capital and labor are gross complements, endogenous growth is impossible. On the other hand, since $\xi < 0$ implies $\delta > 0$ and $\theta < 0$, *exo*genous R&D-driven growth (represented by perpetual growth in N) is both possible and plausible.

We find that in the gross-complements case there exists an asymptotic BGP – a BGP that cannot be reached in finite time but is gradually converged to as $k \to \infty$ with time. Precisely speaking, we have that

$$\hat{Y} = \frac{g}{\alpha\delta} + \varepsilon_K \hat{K} + (1 - \varepsilon_K)\hat{L} \quad \Rightarrow \quad \hat{y} = \frac{g}{\alpha\delta} + \varepsilon_K \hat{k}. \tag{5.36}$$

Let us now calculate the limit of $\hat{y}$ as $t \to \infty$. Because of exogenous growth in N, we have that $A \to \infty$ and $k \to \infty$. In consequence, the partial elasticity ε_K disappears in the limit: $\varepsilon_K = A^\xi \zeta(k/y)^\xi \to 0$. This suffices to show that $\hat{y} \to \frac{g}{\alpha\delta}$. And by the virtue of the capital's equation of motion, we know that $\hat{k} \to \frac{g}{\alpha\delta}$ as well.

As for the growth rates of unit factor productivities, we use (5.24) to show that $\hat{a} \to \frac{g}{\alpha\delta}$ and $\hat{b} \to \frac{g}{\alpha\delta} + \left(\frac{\theta}{\alpha\delta - \theta}\right)\frac{g}{\alpha\delta} = \frac{g}{\alpha\delta - \theta} > 0$. Hence, we see that in the CES case, technical change *augments both factors of production in the long run*, not only labor as it was in the Jones' case.[17]

There is one important twist to this reasoning, however. For $\theta = -\infty$ (so the LPFs are Leontief), we get that $\hat{b} \to 0$. Thus, in the Leontief limit, technical change is purely labor-augmenting in the long run, a result that carries forward from Jones (2005b) and Acemoglu (2003). The difference is that in their setups this property was general, and here it relies upon the assumption $\theta = -\infty$ which is both knife-edge and extreme.

[17] One should expect this result in the general Clayton-Pareto case as well, if only the long-run elasticity of substitution is less than unity (if $\lim_{k\to\infty} \sigma(k) < 1$).

5.5 Conclusion

In this article, we have derived from microfoundations the "Clayton-Pareto" class of global production functions (GPFs). We have discussed the properties of some of its most distinguished members. We have also analyzed the long-run implications of such production functions for the direction of technical change, i.e. we checked when it is labor-augmenting, capital-augmenting, or both.

Our "Clayton-Pareto" class of production functions has been obtained assuming that each of the unit factor productivities is Pareto-distributed, that dependence between these marginal distributions is captured by the Clayton copula, and that local production functions are CES. This class has been shown to nest both the Cobb-Douglas functions and the CES (both being empirically relevant, see Duffy and Papageorgiou, 2000). Contrary to Jones' presumptions, the shape of the GPF typically depends on the shapes of local production functions.

Embedding our microfounded model in the neoclassical (Solow) growth framework, we have proven that in general, technical progress tends to augment both factors of production in the long run. We could find only two exceptions to this rule, which would lead to purely labor-augmenting technical change: the case of independent marginal distributions, which implies a Cobb-Douglas GPF; and the case of Leontief local production functions, while the GPF is either CES or Cobb-Douglas.

We have also proven that in some cases, the neoclassical growth model with Clayton-Pareto production exhibits endogenous growth via capital accumulation. Moreover, assuming exogenous technological progress on top of it leads to explosivity.

Summing up: the goals achieved in this paper have been to critically re-examine the assumptions of Cobb-Douglas production functions and purely labor-augmenting technical change, ubiquitous in contemporary growth theory, and to show that even a slight departure from the original Jones' (2005) framework can lead to significantly different results. For instance, the CES production function can also be derived from Pareto distributions of unit factor productivities.

This research can be extended in the following ways. First, the setup can be generalized to include further factors of production, such as human capital or non-renewable and renewable resources. Second, alternative copulas may be used to capture the dependence between marginal idea distributions. Third, one may wish to analyze the behavior of our microfounded model under R&D-based endogenous growth. Fourth, one may want to relax the assumption that the LPFs be CES. And last but not least, empirical studies in this field would be extremely valuable. The class of Clayton-Pareto functions generalizes both the CES and the Cobb-Douglas classes and is derived from technology choice-based microfoundations, as opposed to e.g. translog functions.

5.6 Appendix

5.6.1 Proof of proposition 5.1

Recall that the marginal idea distributions of $\tilde{a}$ and $\tilde{b}$ are Pareto, given by (5.1) and (5.2), respectively.

Because the Pareto distribution is nicely defined in terms of its survival function and not the cumulative distribution function (CDF) itself, we find it useful to substitute $X = -\tilde{a}$, $Y = -\tilde{b}$.[18]

CDFs of X and Y satisfy:

$$P(\tilde{a} > a) = P(X < -a) \equiv F_X(-a) = \left(\frac{\gamma_a}{a}\right)^{\alpha},$$

$$P(\tilde{b} > b) = P(Y < -b) \equiv F_Y(-b) = \left(\frac{\gamma_b}{b}\right)^{\beta},$$

where $a \geq \gamma_a > 0$ and $b \geq \gamma_b > 0$. Applying the Clayton copula formula given in (5.3) to the CDFs of X and Y yields:

$$F(a,b) = C(F_X(a), F_Y(b)) = \max\left\{0, \left(\left(\frac{\gamma_a}{-a}\right)^{-\alpha\delta} + \left(\frac{\gamma_b}{-b}\right)^{-\beta\delta} - 1\right)^{-\frac{1}{\delta}}\right\},$$

where $a \leq -\gamma_a < 0$ and $b \leq -\gamma_b < 0$.

Finally, we apply $P(\tilde{a} > a, \tilde{b} > b) = F(-a, -b)$ to get (5.4).

If $\delta = 0$, then

$$P(\tilde{a} > a, \tilde{b} > b) = F_X(-a)F_Y(-b) = \left(\frac{\gamma_a}{a}\right)^{\alpha}\left(\frac{\gamma_b}{b}\right)^{\beta},$$

[18] Instead of inverting the random variables, we could also rotate the Clayton copula.

which is directly (5.5). ■

5.6.2 Proof of proposition 5.3

We apply the Implicit Function Theorem to the Φ function, defined in (5.13). First of all, we notice that $\Phi \in C^1(\mathbb{R}^3_+)$. Second, we observe that

$$\frac{\partial \Phi}{\partial a} = \gamma\alpha\delta a^{\alpha\delta-1} + \Big(\frac{\beta\delta(\alpha\delta-\theta)}{\beta\delta-\theta}\Big)c^{\beta\delta}k^{\frac{\theta\beta\delta}{\beta\delta-\theta}}a^{\frac{\beta\delta(\alpha\delta-\theta)}{\beta\delta-\theta}-1}, \tag{5.37}$$

and so, using $\alpha\delta-\theta>0$ and $\beta\delta-\theta>0$, we obtain $\frac{\partial\Phi}{\partial a}>0$ if and only if $\delta>0$, and $\frac{\partial\Phi}{\partial a}<0$ if and only if $\delta<0$. Hence (since $\delta\neq 0$),[19] clearly $\frac{\partial\Phi}{\partial a}\neq 0$ for all $k>0$ and $N>0$, so for all $k>0$ and $N>0$, there locally exists an implicit function a^* such that $\Phi(a^*(k;N),k;N)=0$.

To obtain uniqueness of a^*, we shall note that for each given configuration of exogenous parameter values, and given $a,k,N>0$, both partial derivatives of the implicit function, $\frac{\partial a}{\partial k}$ and $\frac{\partial a}{\partial N}$ have a constant sign. By the Implicit Function Theorem, we have $\frac{\partial a}{\partial k} = -\frac{\frac{\partial\Phi}{\partial k}}{\frac{\partial\Phi}{\partial a}}$ and $\frac{\partial a}{\partial N} = -\frac{\frac{\partial\Phi}{\partial N}}{\frac{\partial\Phi}{\partial a}}$. Constancy of the sign of $\frac{\partial\Phi}{\partial a}$ we have already proven above. $\frac{\partial\Phi}{\partial N}=-1$ so it is obviously always negative. Since

$$\frac{\partial\Phi}{\partial k} = \left(\frac{\beta\delta\theta}{\beta\delta-\theta}\right)c^{\beta\delta}a^{\frac{\beta\delta(\alpha\delta-\theta)}{\beta\delta-\theta}}k^{\frac{\beta\delta\theta}{\beta\delta-\theta}-1},$$

we get $\frac{\partial\Phi}{\partial k}>0$ if and only if $\delta<0,\theta<0$ or $\delta>0,\theta>0$; and $\frac{\partial\Phi}{\partial k}<0$ if $\delta>0$ and $\theta<0$. Thus, signs of $\frac{\partial a}{\partial k}$ and $\frac{\partial a}{\partial N}$ never change. Uniqueness of a^* follows from a juxtaposition of this fact with global differentiability of Φ.

From (5.12), we have that if a positive $a^*(k;N)$ exists and is unique, then automatically so is $b^*(k;N)$. ■

5.6.3 Second-order conditions

To save on algebraic manipulations, we shall simplify the competitive firm's optimization problem (5.10). Solving it is a task equivalent to

[19]The $\delta=0$ (independence) case allows a unique, closed-form solution to the discussed problem. No sophisticated reasoning is necessary in this case. See section 5.3.1.

finding extreme values of the following Lagrangian:

$$\mathcal{L}(a,b,\lambda) = \tilde{A}^{\theta}[\psi(bK)^{\theta} + (1-\psi)(aL)^{\theta}] - \lambda(\gamma a^{\alpha\delta} + b^{\beta\delta} - N).$$

The Lagrangian $\mathcal{L}$ should be maximized (if $\theta > 0$), or minimized (if $\theta < 0$).

We have

$$\begin{cases} \mathcal{L}_a(a,b,\lambda) & = \tilde{A}(1-\psi)\theta a^{\theta-1}L^{\theta} - \lambda\gamma\alpha\delta a^{\alpha\delta-1} = 0, \\ \mathcal{L}_b(a,b,\lambda) & = \tilde{A}\psi\theta b^{\theta-1}L^{\theta} - \lambda\beta\delta b^{\beta\delta-1} = 0, \\ \mathcal{L}_\lambda(a,b,\lambda) & = \gamma a^{\alpha\delta} + b^{\beta\delta} - N = 0. \end{cases}$$

Moving the terms with λ in the first two equations to the RHS, and dividing sidewise yields (5.11).

In the optimum, we also have $\lambda = \frac{\tilde{A}^{\theta}(1-\psi)\theta L^{\theta}a^{\theta-\alpha\delta}}{\gamma\alpha\delta} = \frac{\tilde{A}^{\theta}\psi\theta K^{\theta}b^{\theta-\beta\delta}}{\gamma\beta\delta}$.

The second derivatives of the Lagrangian are (after inserting appropriate expressions for λ):

$$\begin{cases} \mathcal{L}_{aa}(a,b) & = \tilde{A}^{\theta}(1-\psi)\theta a^{\theta-2}L^{\theta}(\theta-\alpha\delta), \\ \mathcal{L}_{ab}(a,b) & = \mathcal{L}_{ba}(a,b) = 0, \\ \mathcal{L}_{bb}(a,b) & = \tilde{A}^{\theta}\psi\theta b^{\theta-2}K^{\theta}(\theta-\beta\delta). \end{cases}$$

If $\mathcal{L}_{aa}$ and $\mathcal{L}_{bb}$ are both negative, then we have a maximum; and if they are both positive, we have a minimum. It is straightforward to see that we need $\alpha\delta > \theta$ and $\beta\delta > \theta$ for our optimization criteria to be satisfied. ■

5.6.4 Model of research

This appendix presents a model of research which supports Definition 5.1.

We assume there is a continuum of researchers, located along the unit interval $I = [0,1]$. Each researcher $i \in I$ draws one technology pair (a_i, b_i) from the joint idea distribution $(\tilde{a}, \tilde{b})$. The Law of Large Numbers implies that for all values of $a \geq \gamma_a, b \geq \gamma_b$, there must exist a researcher who has drawn a technology pair arbitrarily close to (a,b) in terms of the Pythagorean metric on $\mathbb{R}^2$.

Let us now define the dependence between all the individual draws of ideas and the technology that becomes the output of the whole research process. We assume that a technology pair (a,b) can be made available for production only after a given fraction $x \in (0,1)$ of researchers understands it and helps incorporate it into the existing stock of knowledge. Put formally, a technology pair (a,b) becomes available only if a given fraction x of researchers have drawn a technology pair that is not inferior to (a,b).

Let us explain what we mean by *not inferior* in a two-dimensional setup. We first consider a single ray from the origin: a semi-straight line $\{(a,b) \in \mathbb{R}^2_+ : b = \omega a\}$ where $\omega \in (0,+\infty)$ is given. Along such a ray, all points are indexed by a single number $a \geq \gamma_a$, and can be easily ordered since on the real line on which a lies, there exists a natural ordering. Using the Law of Large Numbers again, we obtain that there exists a unique $\bar{a} > \gamma_a$ such that for all $\gamma_a < a < \bar{a}$, technology $(a, \omega a)$ is understood by at least x per cent of researchers. The "frontier" technology $(\bar{a}, \omega\bar{a})$ is characterized by $P(\tilde{a} > \bar{a}, \tilde{b} > \omega\bar{a}) = x$, and all technologies $(a, \omega a)$ for which $P(\tilde{a} > a, \tilde{b} > \omega a) > x$ are inferior to the technology $(\bar{a}, \omega\bar{a})$ because *both* their coordinates are lower. Thus, the "frontier" technology pair $(\bar{a}, \omega\bar{a})$ is the best one attainable within the given ray.

We carry out this procedure for all rays from the origin, letting the index ω go from 0 to $+\infty$. In the final outcome, we obtain that the best attainable ("frontier") technologies are located along the curve $\{(a,b) \in \mathbb{R}^2_+ : P(\tilde{a} > a, \tilde{b} > b) = x\}$. We refer to the x-th contour of the joint idea tail probability function as to the *technology frontier.*

Until now, there has been no technological progress (and no time dimension either) in this model. In this respect, we are going to assume that as time goes on, researchers (whom we assume to be infinitely-lived for simplicity) gradually accumulate knowledge, and so it is easier and easier for them to understand and accomodate further innovations. Hence, further and further technology frontiers become attainable in the passage of time. As we want to simplify our analysis as much as possible, we do not model the knowledge accumulation process explicitly here. Instead, we proxy it with an exogenous decline in x which yields equivalent results.

Employing our little model of research helps simplify the original

Jones' framework greatly. First, we do not have to use Fréchet distributions or Poisson processes to obtain our results. Second, we are able to achieve more generality, which would not be possible if we insisted on maintaining full stochasticity until the end of analysis: one can easily imagine that if the simplest case requires handling Fréchet distributions, then the more complex ones would simply be intractable.

Chapter 6

Technological Opportunity, Long-Run Growth, and Convergence[1]

6.1 Introduction

Ideas are not alike. Some ideas provide answers to certain questions, solve some puzzles, or at least add qualifications to the answers already known; others pose more questions than they actually answer. Some inventions provide the mankind with readily useful technologies; others need centuries of hard intellectual work to develop. Some ideas leave loose threads hanging – they open up the opportunity for further developments – while others tie these hanging ends together, creating useful knowledge by taking advantage of the opportunities created previously.

It is clear that this type of heterogeneity in ideas must have an impact on the rate of technological progress. Nonetheless, it has not been considered in the R&D-based growth literature until the recent articles by Olsson (2000, 2005) which laid the foundations for the distinction between *incremental* and *radical* innovations, and offered a formal description for the concept of *technological opportunity.* On the other hand, these two Olsson's papers seem to be quite detached from standard theories of long-run economic growth and convergence because they are

[1]This chapter has been written jointly with Ingmar Schumacher.

focused primarily on developing a set-theoretic model of ideas and expaining why technological breakthroughs tend to cluster in time. This creates a gap in the literature which this article is intended to fill.

The crucial contribution of this article is to span a bridge between the intuitive Olsson's theory and the mainstream literature on R&D-based (semi-)endogenous growth (e.g. Romer, 1990; Jones, 1995). We provide a common denominator for both these frameworks: our model is general enough to make it possible to analyze the impacts of technological opportunity as well as incremental and radical innovations on long-run growth and convergence. Specifically, we demonstrate that:

(i) technological opportunity must be continuously renewed in order to sustain technological progress over the long run;

(ii) a balanced growth path exists in the economy only in a specific knife-edge case which implies that technology and technological opportunity grow at equal rates: if the flows of radical innovations are too small in comparison to the flows of incremental innovations, technological opportunity will be gradually depleted and economic growth will eventually come to a halt, and if radical innovations are relatively abundant, then the volume of unused technological opportunity would tend to explode;

(iii) the transition to the balanced growth path may be oscillatory: in addition to the standard result of monotonic convergence, we obtain the possibility of complex dynamics for a wide variety of parameter values. These complex dynamics can come in the form of oscillations, converging, stable or diverging, and Andronov–Hopf bifurcations. As opposed to standard models in business cycle theory, our model does not require uncertainty to generate endogenous cycles.

It is important to contrast our framework featuring heterogenous innovations with the ones present in the literature up to now: unlike here, it is casually assumed that the amount of newly created knowledge depends primarily on the current stock of knowledge (e.g. Romer, 1990; Aghion and Howitt, 1992; Jones, 1995). Obviously, in all cases it depends also on the number of active researchers; the list of other factors appearing in the knowledge accumulation functions found in the literature includes however, among others, human capital (e.g. Strulik, 2005), spillovers (e.g. Howitt, 1999), R&D difficulty (e.g. Segerstrom, 2000), but does not include technological opportunity. In contrast, this paper assumes that once technological opportunity is depleted, no further developments can be made.

This article follows Olsson (2005) in assuming that the R&D sector of the economy produces two inherently different kinds of innovations, namely incremental and radical innovations, which are related as follows. Incremental innovations provide direct increases to the economy's productivity and utilize technological opportunity opened up by radical innovations. Radical innovations extend the existing technological opportunity by combining previous discoveries (abstract ideas which are initially useless) with existing knowledge. Technological opportunity behaves like a renewable resource here: it is exhausted by incremental innovations and renewed by radical innovations. As opposed to Olsson (2005), however, our model does not rely on a linearity assumption giving rise to bang-bang solutions and non-smooth dynamics. In the Olsson's model, all researchers work either in the incremental or in the radical innovation sector, but never in both at the same time. Intuitively, this is not a convincing result: even while this incremental innovation paper is being written, radical innovations across the globe take place. Our analysis confirms, however, Olsson's results signifying the importance of labor re-allocations for oscillatory dynamics.

When introducing different kinds of innovations into growth models, several researchers have already suggested the possibility of cycles to occur (Aghion and Howitt, 1992; Cheng and Dinopoulos, 1992; Amable, 1996; Francois and Lloyd-Ellis, 2003; Phillips and Wrase, 2006; Jovanovic and Rob, 1990; Bresnahan and Trajtenberg, 1995; Freeman, Hong and Peled, 1999). However, their approaches and therefore their implied sources of cyclical growth differ from ours. In particular, in none of these approaches does technological progress depend on an exhaustible factor, such as technological opportunity is in our case.

The remainder of the article is structured as follows. In section 6.2 we lay out the foundations of our framework. We introduce our principal concepts and present the system of equations determining the temporal evolution of knowledge. In section 6.3 we set up and solve our basic growth model. We discuss both the balanced growth path and transition dynamics. In section 6.4 we discuss some generalizations of our basic model and other related issues, in particular the prospects for an empirical measurement of technological opportunity. Section 6.5 concludes.

6.2 Technological opportunity and the evolution of knowledge

New developments vary not only in their impact on factor productivity, but also in their impact on the opportunity for further developments. This difference between two innovations can manifest itself in the magnitudes of productivity increments, but also in the magnitudes *and the direction* of change in the extent and efficiency of follow-up research. There are *radical* developments that open new avenues of thought and there are *incremental* developments which only "fish out" *technological opportunity*. Intuitively, one could look at this distinction through the lens of the dichotomy between basic and applied research. Basic research is often carried out in fields rather detached from day-to-day economic activity, such as plasma physics, neurobiology, or experimental psychology, and it would typically be of little business interest but potentially of immense interest to scientists working in related areas, able to build up on these developments. The principal role of basic research is thus to facilitate further research, both basic and applied. Applied research, on the other hand, would typically *base* on previous scientific developments and bring about new opportunities to business rather than to science. It would tend to capture the opportunities created by basic research while offering little scientific stimuli in return. Applied research is vacuous without basic research, but basic research alone cannot guarantee technological progress whose benefits everyone can experience: applied research is the necesseary means for making productive use of the conquests of basic science.[2]

[2]This is apparent already in the Wikipedia entry for *research*: "Basic research (also called fundamental or pure research) has as its primary objective the advancement of

Radical innovation as viewed in this paper resembles basic research in the sense presented above, while incremental innovation could be associated with applied research. For precise definitions and the relationship between all these terms and the set-theoretic setup put forward by Olsson (2000, 2005), please consult the appendix.

6.2.1 The laws of motion

Let us now clarify the notation of variables which shall be used in our growth model. By $B(t)$ we shall denote the amount of technological opportunity at time t, by $R(t)$ – the flow of radical innovations, by $C(t)$ – the flow of incremental innovations, and by $A(t)$ – the current stock of knowledge. As stated clearly in the appendix, technological opportunity is increased by radical innovations (basic research), whereas incremental innovations (applied research) add to the stock of knowledge but diminish technological opportunity. We write $\dot{A}(t) = C(t)$ and $\dot{B}(t) = R(t) - C(t)$.

The set-theoretic setup specified in the appendix as well as the intuitive rationale given above equip us with the basic understanding of how innovations and knowledge evolve over time, and we shall proceed to provide a more specific characterization which would then allow us to solve for the precise dynamic paths. To assure analytical tractability

---

knowledge and the theoretical understanding of the relations among variables. (...) **It is conducted without any practical end in mind**, although it may have unexpected results pointing to practical applications. The terms "basic" or "fundamental" indicate that, through theory generation, **basic research provides the foundation for further, sometimes applied research.** As **there is no guarantee of short-term practical gain**, researchers may find it difficult to obtain funding for basic research. (...) **Traditionally, basic research was considered as an activity that preceded applied research, which in turn preceded development into practical applications.** Recently, these distinctions have become much less clear-cut, and it is sometimes the case that all stages will intermix. This is particularly the case in fields such as biotechnology and electronics, where fundamental discoveries may be made alongside work intended to develop new products, and in areas where public and private sector partners collaborate in order to develop greater insight into key areas of interest." (emphasis added) More specific economic consequences of the differences between basic and applied research have been investigated in numerous works. Among others, one can name here: Griliches (1986), Cockburn, Henderson and Stern (1999), Malla and Gray (2005).

as well as comparability to the semi-endogenous growth literature which evolved from Jones (1995), we use standard Cobb-Douglas functional forms. Omitting time arguments for convenience, these read:

$$\dot{A} = \delta(u\ell_A L)^\beta B^\mu, \tag{6.1}$$

$$\dot{B} = \underbrace{-\delta(u\ell_A L)^\beta B^\mu}_{\text{incremental innovations}} + \underbrace{\gamma((1-u)\ell_A L)^\beta A^\nu}_{\text{radical innovations}}, \tag{6.2}$$

We denote total population by L and assume it to be equal to the total amount of working time available in the economy at time t. Population is assumed to grow exogenously at a constant rate $n > 0$. Then, ℓ_A is the proportion of working time devoted to R&D, with $u\ell_A$ being the proportion of time spent on working in the incremental innovation sector and $(1-u)\ell_A$ the time spent in the radical innovation sector. The parameter $\delta > 0$ is proportional to the rate at which incremental innovations come about, whereas $\gamma > 0$ relates to the rate at which radical innovations arrive. The exponents μ and ν are crucial for the dynamic behaviour of our model. From our preceding argumentation, we know that they ought to be strictly positive. We shall further assume that $\mu, \nu \in (0,1)$ (DRTS) which assures non-explosive semi-endogenous growth in the long run. The relative size of these exponents is decisive for the long-run evolution of innovations, which we investigate in the next subsection. The parameter $\beta > 0$ measures the degree of returns to scale in the R&D sectors. Our assumption that $\beta \in (0,1)$ implies decreasing returns to scale in R&D activity which are required for positive shares of both kinds of innovation to be pursued in optimum and stands in contrast to Olsson (2005) who has $\beta = 1$ and thus constant returns to scale, which leads to bang-bang solutions.

Let us contrast the setup of our model with several standard R&D-based models of (semi-)endogenous growth, on the one hand, and with Olsson (2000, 2005) on the other. We shall see shortly that this article spans a bridge between these two frameworks and can be reduced to either one by neglecting certain assumptions.

Romer's (1990) specification of technical change would correspond to equation (6.1) only, with $B = A$ and $\mu = 1$. This of course leads to the scale effect discussed by Jones (1995) and implies explosive dynamics if $n > 0$. However, even the Jones's R&D equation, which avoids scale effects, is only a subcase of our specification, corresponding to equation (6.1) only, with $B = A, \mu < 1$ and $n > 0$. It is thus clearly visible that the system (6.1)–(6.2) can be reduced to the standard specifications of R&D equations found throughout the economic growth literature by identifying technological opportunity with technology itself.

In comparison to Olsson (2005), on the other hand, we allow for both incremental and radical innovations happening at the same time here. We neglect the Olsson's linearity assumption $\beta = 1$ that gives rise to bang-bang solutions and thus abrupt reallocations of R&D workers between radical and incremental R&D. We are also more specific on discoveries fueling radical innovation by linking them to the current technology level A, with elasticity ν.

6.2.2 Evidence supporting the setup

There exists a large amount of anecdotal evidence which can be used to support the view that technological opportunity is a relevant concept and that the distinction between incremental and radical innovations is helpful for the proper understanding of technological change across centuries (cf. Olsson, 2005). Here, we only add a few more examples.

Example 1. In 1814, Joseph Niépce invented the first photo camera. It took 8 hours to take one picture. This certainly must be viewed as a radical innovation but it was not yet a useful technology for increasing the productivity of the economy. However, in 1851 the exposure time was reduced to 2-3 seconds, in 1888 the first roll-film was developed, and in 1941 – the color film. One and a half century later, we have digital cameras which are available at prices accessible to the general public and pictures can be printed at home. It is clear that the small improvements to the photo camera, in our terminology the incremental innovations, ought to be viewed as the crucial steps for spreading the technology into the economy, whereas the radical innovation of Joseph Niépce was the one which opened up the opportunity for these incremental innovations.

Example 2. The first locomotive was developed in 1804 by Richard Trevithick. It was the first steam-powered locomotive, and therefore ought to be considered a radical innovation. However, it was too heavy and even broke the very own rails it was supposed to travel on. Compared to this, the incremental innovations following that radical innovation were tremendous. In 1814 came the first steam locomotive that was actually able to travel, although at only 6 km/h; today the Maglev, the high-speed magnetic train, travels at more than 550 km/h. Again, the initial idea of Richard Trevithick was the one opening up the possibilities for the incremental innovations, whereas the radical innovation proved useless for improving productivity.

Example 3. In 1928, Alexander Fleming had, by accident, left a Staphylococcus plate culture lying in the warm cellar. Several days later, upon reminding himself of the forgotten plate culture, he noticed that there was a blue-green colored mold destroying the bacteria. He called this mold Penicillin. It was however too weak and unstable to provide a useful means of destroying bacterial infections in humans. Only subsequent research by Chain, Florey and Heatley developed the kind of Penicillin which now saves human lives throughout the world. As Sir Henry Harris had aptly put it: “Without Fleming, no Chain; without Chain, no Florey; without Florey, no Heatley; without Heatley, no Penicillin.”

Similar stories can be told about the invention of the first battery by Alessandro Volta in 1799, the first champagne by Dom Pérignon in 1670, nylon by DuPont in 1928, the steam engine, the airplane, electricity, and many more. What can be seen here is that we view radical innovations as spanning a broader class of innovations than for example General Purpose Technologies (e.g. Jovanovic and Rousseau, 2003). They all have one thing in common: opening up opportunities for small improvements, for incremental innovations that help make the technology accessible, practical and operative. Without the radical innovations being able to open up new opportunities, there would be no place for incremental innovations. And without doubt, in a considerable if not exhaustive number of cases, it were the incremental innovations which really proved to be useful for economic purposes.

A radical innovation is usually the first one in a series of innovations. It opens up a new line of research. It should not be mistaken, however, with an *early* innovation: for example, the early invention of gunpowder in ancient China did not open up many opportunities for developments, while the invention of MP3 music encoding did so despite being quite recent.

Our understanding of the evolution of technological knowledge differs from the articles basing on combinatorial calculations (Romer, 1993; Weitzman, 1998). We do not consider the potential for technological change as the number of possible ways to combine ideas. For example, 20 objects may be combined in $2^{20} = 1,048,576$ ways. Given an enormous number like this, papers in this vein conclude that there are no practical limits to technological change. However, why should we combine every possible idea? Intuitively, it seems more likely that just the "non-dominated" technologies, those on the technological frontier (Caselli and Coleman, 2006), are improved. As Poincaré observed: "To create consists precisely in not making useless combinations." So, we suggest that only ideas on the technology frontier may be usefully improved, an assumption which this article shares with Kortum (1997), Jones (2005b), and Olsson (2005). The implication is that incremental innovations would by themselves come to a halt if the technological frontier were not constantly pushed ahead by radical innovations.

So, how does technological opportunity increase? What qualifies as a radical innovation? We suggest that a radical innovation arises whenever a new class of known phenomena (physical, biological, chemical, etc.) is found to have economically useful applications. For example, "[f]or thousands of years, silicon dioxide provided utility mainly as sand on the beach, but now it delivers utility through the myriad of goods that depend on computer chips" (Jones, 2005a). Conclusively, we believe that the flows of radical innovations are related to the stock of existing knowledge. The larger the existing knowledge base, the more discoveries will be transformed into radical innovations.

6.2.3 Long-run evolution

We shall now explain the direct implications of considering the additional feedback loop through technological opportunity. Generally, it gives two new results. Firstly, the long-run results of e.g. Jones (1995) are preserved only when the specific knife-edge condition $\mu = \nu$ holds, and secondly, the optimal path can be subject to oscillations. Our model is thus able to span a bridge from the cyclical paths of Olsson to the monotonic dynamics of Jones or Romer.[3]

Long-run implications must be analyzed first. We shall thus first focus on the properties of the balanced growth path (BGP) under different relative sizes of the exponents in the radical and incremental innovations, μ and ν (or of the asymptotic balanced growth path, if a proper BGP does not exist).[4] This will be done by finding the necessary conditions under which the growth rates of all economic variables are constant. These conditions imply in particular that the sectoral allocation of labor does not change over time.

Let us take equations (6.1)–(6.2) and solve for the BGP. It is easy to see that the growth rate of technological opportunity, $\dot{B}/B = R/B - C/B$ is constant only if both the radical and incremental innovation flows grow at the same rate as technological opportunity itself does. This is the case if $\mu = \nu$. When this condition is satisfied, it implies that both knowledge and technological opportunity grow at the same rate, too. As can be observed from equation (6.2), the growth rates of C and R are different if the exponents μ and ν are not equal. The implications are summarized in the following table.

Long-Run Behavior of A and B		
Case 1	$\mu = \nu$	$\hat{A} = \hat{B} = \frac{\beta n}{1-\mu}$
Case 2	$\mu < \nu$	$\lim_{t\to\infty} \hat{B} = \frac{1+\nu}{1+\mu}\hat{A}$
Case 3	$\mu > \nu$	$\lim_{t\to\infty} \dot{A} = \dot{B} = B = 0$

The intuition is as follows. In Case 1 we observe that if our knife-edge condition $\mu = \nu$ is satisfied, and in consequence, the extents of external returns from technology to radical innovations and from technological opportunity to incremental innovations are equal, then this gives rise to standard semi-endogenous growth along the lines of Jones (1995). Unfortunately, this knife-edge condition is crucial for the result. This can be seen from Case 2 and Case 3. In Case 2, the external returns to radical innovations are larger than to incremental ones, and thus techno-

[3]Dynamics of the Jones' (1995) model have been analyzed in detail by Arnold (2006). Perhaps a bit surprisingly, he finds that the possibility of oscillatory convergence to the BGP cannot be ruled out there.

[4]When we refer to semi-endogenous growth we mean growth ultimately driven by population growth, see e.g. Jones (1995).

logical opportunity can grow at a faster rate than knowledge. Clearly, in the limit, technological opportunity will be exhaustively driven by radical innovations in that case. This is still not so bad news for the economy in comparison to Case 3. In such case, even though technological opportunity is renewed by radical innovations, the effect of its depletion due to more efficient incremental innovations is dominant in the long run. Growth comes to a halt in the limit.

Let us remark here that if one allows for different degrees of decreasing returns to scale in the two R&D sectors, such that $\dot{B} = -\delta(u\ell_A L)^\beta B^\mu + \gamma((1-u)\ell_A L)^\xi A^\nu$, the crucial knife-edge condition $\mu = \nu$ translates into a different knife-edge condition: $\beta/(1-\mu) = \xi/(1-\nu)$. This condition has exactly the same interpretation: flows of incremental and radical innovations are supposed to grow at equal rates. Further on, the counterpart for Case 2 is $\beta/(1-\mu) < \xi/(1-\nu)$, and the counterpart for Case 3 is $\beta/(1-\mu) > \xi/(1-\nu)$. All further implications are preserved. We shall abstract from this unnecessary complexity throughout the remainder of the article.

It is now clear that if technological opportunity is accepted to be the driving force behind the possibility to innovate, then the long-run behavior of the economy depends crucially on the relative size of external returns to different kinds of innovations. In consequence, long-run predictions along the lines of Jones (1995), Kortum (1997), Segerstrom (1998) or Peretto (1998) hold only in the case where the extents of external returns to incremental and radical innovations are exactly *equal*. Since current literature compares the model predictions with historical time paths of R&D expenditures and factor productivities, with the single dynamical equation of technology in mind, our knife-edge result here suggests one should be more cautious with those predictions. Given one accepts technological opportunity as the driving force behind actual technology, then it could very well be that the empirical literature misses the influence of the evolution of technological opportunity on effective technological progress and thus runs into systematic error.

As our main interest lies in elaborating the implications of introducing technological opportunity into a standard semi-endogenous growth framework, we shall concentrate mostly on the case $\mu = \nu$ which allows for a BGP. As we will show, even in the knife-edge case we will obtain novel results, this time for the transition.

6.3 The social planner problem

Let us now embed the dynamical equations analyzed above in a semi-endogenous growth model and find its social planner solution. Our interest is to understand in which way will the planner allocate consumption across time and labor across sectors in order to achieve the social optimum. We shall base on the standard Ramsey framework, where the infinitely-lived representative agent obtains utility from the discounted stream of a single consumption good. The utility function takes the CRRA form: $u(c) = c^{1-\theta}/(1-\theta)$, with $\theta > 0$ being the inverse of the intertemporal elasticity of substitution in consumption. Moreover, we shall assume a standard Cobb-Douglas production function which takes as inputs: technology A with elasticity σ, physical capital K with elasticity α, and labor $(1-\ell_A)L$ with elasticity $1-\alpha$. Technology accumulates faster the more labor is allocated to research, but its underlying possibility to accumulate is constrained by technological opportunity.

As suggested previously, we shall concentrate on the case of $\mu = \nu$ here. We impose this restrictive condition for analytical tractability and comparability to the semi-endogenous growth literature. We shall first solve for the BGP of the system. The maximization problem looks as follows.

$$\begin{aligned}
\max_{\{c,u,\ell_A;k,A,B\}_{t=0}^{\infty}} \quad & L_0 \int_0^{\infty} \frac{c^{1-\theta}}{1-\theta} e^{-(\rho-n)t} dt \qquad \text{s. t.} && (6.3)\\
y &= A^{\sigma} k^{\alpha} (1-\ell_A)^{1-\alpha}, && (6.4)\\
\dot{k} &= y - c - (d+n)k, && (6.5)\\
\dot{A} &= \delta (u \ell_A L)^{\beta} B^{\mu}, && (6.6)\\
\dot{B} &= [-\delta u^{\beta} B^{\mu} + \gamma (1-u)^{\beta} A^{\mu}](\ell_A L)^{\beta}, && (6.7)\\
L &= L_0 e^{nt}, \qquad n > 0, && (6.8)\\
& L_0, k_0, A_0, B_0 \quad \text{given.}
\end{aligned}$$

The parameter restrictions are $0 < n < \rho$, necessary to guarantee a positive effective discount rate, and $\sigma, \alpha, \beta, d \in (0,1)$ as well as $\theta, \mu, \delta, \gamma > 0$. Finally, $\mu < 1$ is required to guarantee positive semi-endogenous growth in the long-run. d is the instantaneous depreciation rate of physical capital.

6.3.1 The balanced growth path

Maximizing the Hamiltonian associated with the above optimization problem and solving for the BGP yields the following long-run growth rate of the economy:

$$g \equiv \hat{y} = \hat{k} = \hat{c} \quad = \quad \frac{\sigma}{1-\alpha}\frac{\beta n}{1-\mu}, \tag{6.9}$$

whereas the long-run growth rates of technology and technological opportunity are (as in Jones, 1995):

$$\hat{A} = \hat{B} \quad = \quad \frac{\beta n}{1-\mu}. \tag{6.10}$$

As expected from the previous section, technological opportunity and technology grow at a common rate, and consumption, income and capital grow at a rate being a multiple of this rate. Along the BGP, shadow prices our state variables evolve according to:

$$\hat{\lambda}_k \quad = \quad n - \rho - \theta g, \tag{6.11}$$

$$\hat{\lambda}_A = \hat{\lambda}_B \quad = \quad n - \rho - (\theta - 1 + \frac{1-\alpha}{\sigma})g, \tag{6.12}$$

and thus all transversality conditions boil down to the single requirement that $n < \rho + (\theta - 1)g$.

Denoting the steady-state ratio of technology to technological opportunity A/B by X, we find that the optimal share of incremental research effort relative to radical research effort is:

$$u^* = \frac{1}{1 + F^{\frac{1}{1-\beta}} X^{\frac{1+\mu}{1-\beta}} \left(\frac{\gamma}{\delta}\right)^{\frac{1}{1-\beta}}}, \tag{6.13}$$

where $X \equiv A/B$ solves the following implicit equation:

$$\left(\frac{\gamma}{\delta}\right)^{\frac{1}{1-\beta}} F^{\frac{\beta}{1-\beta}} (X)^{\frac{1+\mu}{1-\beta}} = 1 + X, \tag{6.14}$$

with $F = -\frac{\mu\hat{A}}{\hat{\lambda}_A} > 0$. It is straightforward to show that (6.14) always has a unique positive solution. The argument goes as follows. The right-hand side of equation (6.14) as a function of the ratio X increases linearly from $\lim_{X\to 0} = 1$ to $\lim_{X\to\infty} = \infty$. The left-hand side increases in a strictly convex manner from $\lim_{X\to 0} = 0$ to $\lim_{X\to\infty} = \infty$. Since both sides are continuous, there necessarily exists a unique, positive point in which they intersect.

Furthermore, we find that the optimal share of labor allocated to R&D along the BGP is

$$\ell_A^* = \frac{1}{1 + u^* \frac{(1-\alpha)\mu}{\beta\sigma}\left(\Phi X - \frac{1-u^*}{u^*}\right)}, \tag{6.15}$$

where

$$\Phi = 1 + \frac{\gamma}{\delta}\left(\frac{1-u^*}{u^*}\right)^{\beta-1} X^{\mu}.$$

The interpretation of these results is as follows. The determinants of the optimal sectoral allocation of labor are, among others:

(i) the growth rate of technology relative to the growth rate of its shadow price. The higher the growth rate of technology, the more worthwhile it is to increase technological opportunity by radical innovations;

(ii) the size of technology relative to the size of technological opportunity. The larger is technological opportunity, the more researchers should be allocated to incremental innovations and the more workers should be allocated to the overall R&D sector;

(iii) the ratio of the exogenous intensities of producing incremental and radical innovations, captured by γ and δ respectively. The higher is the chance of coming up with a radical innovation, the more researchers should be allocated to radical innovations, and vice versa;

(iv) the elasticity parameters. The smaller is the technology share in production (smaller σ), the less efficient researchers are (smaller β), or the smaller is the extent of external effects in R&D (smaller μ), the less workers will be employed in research;

(v) the effect of radical innovations creating an additional demand for research. The more researchers are allocated in the radical innovation sector (the smaller u^*), the more researchers are employed in overall R&D (larger ℓ_A^*).

6.3.2 Transition

In this section we shall analyze the transition dynamics of our model around the balanced growth path. We derive the dynamical equations for variables which are stationary along the BGP. Hence, our system is going to be rewritten in terms of the six following variables: $\{c/k, u, \ell_A, y/k, A/B, L^\beta/A^{1-\mu}\}$. We shall again denote $X \equiv A/B$, and use also the notation $Y \equiv L^\beta/A^{1-\mu}$. There are three choice-like variables, $c/k, u, \ell_A$, and three state-like variables: $y/k, X, Y$.

The transition dynamics are fully characterized by the following dynamical system.

$$\widehat{c/k} = \left(\frac{\alpha}{\theta} - 1\right)\frac{y}{k} - \frac{d+\rho}{\theta} + \frac{c}{k} + d + n \tag{6.16}$$

$$\widehat{y/k} = \sigma\delta(u\ell_A)^\beta X^{-\mu}Y + (\alpha-1)(\frac{y}{k} - \frac{c}{k} - d - n) - -(1-\alpha)\left(\frac{\ell_A}{1-\ell_A}\right)\hat{\ell}_A \tag{6.17}$$

$$\hat{X} = \delta(u\ell_A)^\beta X^{-\mu}Y + \delta(u\ell_A)^\beta X^{1-\mu}Y - \gamma((1-u)\ell_A)^\beta XY \tag{6.18}$$

$$\hat{Y} = \beta n - (1-\mu)\delta(u\ell_A)^\beta X^{-\mu}Y \tag{6.19}$$

$$\hat{u} = -\left\{\mu\hat{X} + \left[\frac{\beta\sigma}{1-\alpha}\left(\frac{1-\ell_A}{u\ell_A}\right) - \mu X\left(1 + \frac{\gamma}{\delta}\left(\frac{1-u}{u}\right)^{\beta-1} X^\mu\right) + +\mu\left(\frac{1-u}{u}\right)\right]\delta(u\ell_A)^\beta X^{-\mu}Y\right\}\frac{1-u}{1-\beta} \equiv -\frac{1-u}{1-\beta}\cdot\Xi, \tag{6.20}$$

$$\hat{\ell}_A = \left\{\alpha\frac{c}{k} - (1-\alpha)(d+n) + \beta n + (\mu-\sigma)\delta(u\ell_A)^\beta X^{-\mu}Y + -\mu\left(\frac{u}{1-u}\right)\gamma((1-u)\ell_A)^\beta XY - u\cdot\Xi\right\}\frac{1}{1-\beta+\alpha\frac{\ell_A}{1-\ell_A}}. \tag{6.21}$$

In the remainder of this section, we shall resort to numerical approximations because the implied analytical formulas, although readily attainable, are too large to be informative.

We assign baseline values to all parameters in our model. To bring our numerical example as close to reality as possible, we shall pick these values within a calibration exercise. This we do by drawing on a similar calibration exercise carried out by Steger (2005) for a generalized R&D-based semi-endogenous growth model. Following him, we choose $\sigma = 0.3, \alpha = 0.36, d = 0.04, \rho = 0.05, n = 0.015, \mu = 0.5, \beta = 0.512,$

parameters which have also been used in Eicher and Turnovsky (1999a) and other articles. We also pick $\theta = 2$ which roughly corresponds to the empirical estimates of the intertemporal elasticity of substitution in consumption for developed economies, even though the debate around this value has not yet been settled.

Further parameters, γ and δ, have clearly never had the opportunity to be discussed in the literature. Thus, we pick them at arbitrary plausible values and then discuss what happens if we manipulate them. All the calibrated parameters are listed in Table 6.1.

δ	γ	β	n	θ	d	α	σ	ρ	μ
2	2	.512	.015	2	.04	.36	.3	.05	.5

Table 6.1: The baseline calibration.

For the baseline calibration of our model, we obtain the following set of BGP characteristics (Table 6.2). Please note that below the actual steady-state values, we included comparative statics. These are helpful for the understanding of the workings of the analyzed model. A "+" means that an increase in a given parameter raises the given steady-state value, whereas a "–" means it lowers it. 0 denotes no impact. Please note that some of these comparative statics need not hold for all parameter values; what we assure is that they must hold in the vicinity of our baseline calibration.

	g	$\hat{A}$	$\frac{y}{k}$	$\frac{c}{k}$	s	$\frac{A}{B}$	$\frac{L^{\beta}}{A^{1-\mu}}$	u	ℓ_A
	.0072	.0154	.2900	.2278	.2145	3.163	.0674	.6429	.0688
β	+	+	+	+	-	-	+	-	+
μ	+	+	+	+	-	-	+	-	+
α	+	0	-	-	+	+	-	+	+
σ	+	0	+	+	-	+	-	+	+
n	+	+	+	+	+	-	+	-	+
θ	0	0	+	+	-	+	-	+	-
γ	0	0	0	0	0	-	-	+	+
δ	0	0	0	0	0	+	+	-	-
ρ	0	0	+	+	-	+	+	+	-

Table 6.2: The social planner allocation and comparative statics.

Our intention now is to emphasize some of the particularly noteworthy facts.

(i) Technological opportunity plays a similar role to savings. In particular, increases in the discount rate ρ and reductions in the intertemporal elasticity of substitution in consumption $1/\theta$ reduce both the savings rate and technological opportunity relative to technology.

(ii) The β parameter, measuring returns to scale in R&D, influences the steady-state variables in the same way as the μ parameter, measuring the extent of external effects in R&D.

(iii) Increases in the population growth rate raise long-run growth rates, but reduce the level of de-trended technology. The ratio of technology to technological opportunity is decreased as well. Relatively more technological opportunity goes then together with an increase in overall R&D labor share.

(iv) Apart from the cases of parameters γ and δ which are directly related to the arrival rates of incremental and radical innovations, technological opportunity and radical research intensity $1-u$ move together. Apart from the case of a changing discount rate ρ, they move in the opposite direction than the de-trended stock of knowledge.

(v) Increased dependence on physical capital (higher α) triggers both a higher steady-state R&D intensity, and a higher intensity of incremental research as compared to radical research. This lowers technological opportunity but, on balance, increases the de-trended stock of knowledge.

The most important reason why we are so explicit about the calibrated parameter values is that we would like to show that even under these very standard parameter choices, our second main result – various kinds of transition along the (unique) optimal path – is easily obtained. This can be seen in table 6.3 which presents the eigenvalues of the dynamical system after its linearization around the steady state.

The two complex eigenvalues with negative real parts imply that under our benchmark calibration, dampened oscillations are observed.

Furthermore, we confirm that our balanced growth path is saddle-path stable: there are three unstable eigenvalues – having positive real parts – and three stable eigenvalues (see table 6.3). The number of unstable roots in the benchmark parametrization of our model is exactly equal to the number of choice-like variables, and hence there exists a unique time path of the economy approaching its balanced growth path (or in the case of the de-trended system (6.16)–(6.21), approaching its steady state).

Indeterminacy is unambiguously ruled out here despite the fact that y/k is a function of ℓ_A which is a choice variable. The reasoning is the following: y/k is in fact not a separate choice variable since once ℓ_A is counted as a choice variable itself, there are no degrees of freedom left for choosing y/k. In sum, once $c/k, u$ and ℓ_A have been chosen, everything else is predetermined. This means that there are exactly three degrees of freedom and this is precisely the same number as the count of unstable roots to be "neutralized" by jump variables.

In other words: we could have used another variable in the de-trended system instead of y/k without altering any of the model implications and the count of stable roots, that variable being $Z = A^\sigma k^{\alpha-1} = (y/k)(1-\ell_A)^{\alpha-1}$. Z is a function of predetermined variables A and k only and is also stationary along the BGP. We would then have a system in 6 variables, $c/k, \ell_A, u, X, Y, Z$, three choice-like variables and three state-like variables, completely independent of the choice-like variables, which would be completely equivalent to the system at hand, (6.16)–(6.21).

0.1233	+ 0.0176i
0.1233	- 0.0176i
0.0557	
-0.0811	+ 0.0176i
-0.0811	- 0.0176i
-0.0135	

Table 6.3: Eigenvalues of the linearized system.

Furthermore, we confirm by the means of a sensitivity analysis that such oscillations indeed occur under a large variety of parameter choices. In contrast, occurrence of oscillatory dynamics is impossible in setups with semi-endogenous growth but without technological opportunity, such as the one of Jones (1995).

Conclusively, if one adopts the technological opportunity approach presented here, then his model can easily generate business-cycle dynamics along the transition to the BGP. The source of these cycles is found in the relative rates at which radical and incremental innovations arrive.

The intuition for the oscillatory dynamics is as follows. Incremental innovations at the same time reduce technological opportunity and improve actual technology. This technology then feeds back into radical innovations which increase technological opportunity. If radical innovations come at amounts lower than that of incremental innovations, then the increases in technological opportunity are small, and smaller than the reductions through incremental innovations. This leads to a monotonic convergence of the optimal ratio A/B to the steady-state value from above. On the contrary, if radical innovations come at relatively large amounts, then the same mechanism leads to convergence of A/B from below. In the case radical innovations arrive at some intermediate number, then the improvements in technological opportunity will be counterbalanced by the reductions through incremental innovations at an amount which makes the ratio of actual technology to technological opportunity fluctuate and converge to the steady-state value of A/B in a non-monotonic manner.

We can therefore conclude that growth in our model is subject to fluctuations around a trend which is ultimately driven by population growth.

We wish to provide further analysis of the complex dynamics by doing some comparative statics with respect to the decisive parameters. As is visible in the subsequent Figure 6.1, complex eigenvalues and thus oscillations occur for an intermediate range of arrival rates of the radical innovations relative to incremental innovations (ceteris paribus). Moreover, the oscillation frequency and thus the length of the cycles depends on the γ/δ ratio, with a maximum frequency appearing around $\gamma/\delta = 1$ (in the Figure, this corresponds to $\gamma = \delta = 2$). Even though not decisive

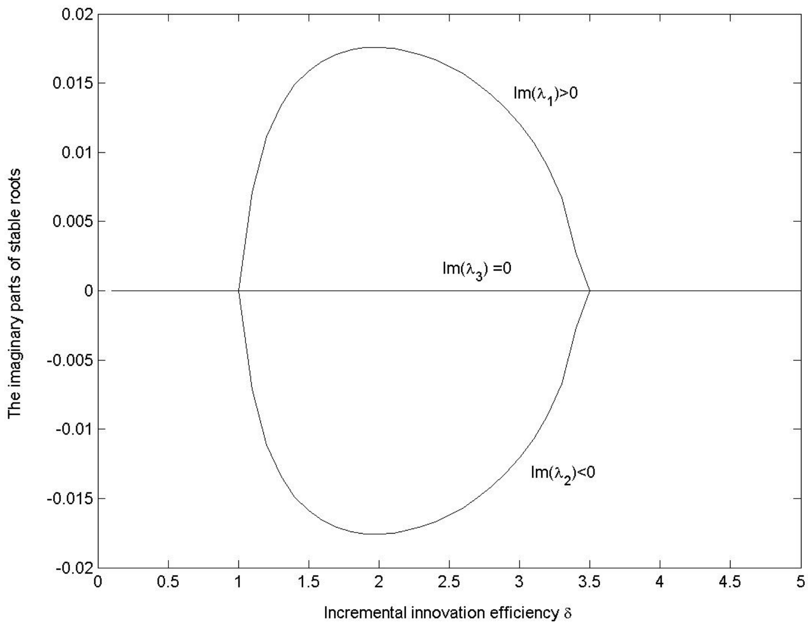

Figure 6.1: Oscillatory dynamics. The effect of varying δ, holding $\gamma = 2$ fixed.

for the growth rate, these two arrival rates are decisive for the transition.

To take the analysis of the model's dynamics a step further, we shall now discuss the consequences of varying the returns-to-scale parameter β in its range $(0, 1]$, as presented in Figure 6.2.[5] In such case, not only dampened oscillations appear but also limit cycles. We identify an Andronov–Hopf bifurcation.

As can be seen in Figure 6.2, as long as $\beta < 0.96569$, greater returns to scale in R&D imply faster convergence to the BGP but also oscillations of higher frequency. When β crosses the threshold value of 0.96569 from below, the relevant conjugate eigenvalues have their real parts crossing zero from below and there emerges a limit cycle. This means that we observe an Andronov–Hopf bifurcation. The intuition for this result is the following. If returns to scale in the R&D sector are reduced fast (low β), convergence to the BGP is monotonic, because

[5]All other parameters are set at their benchmark values.

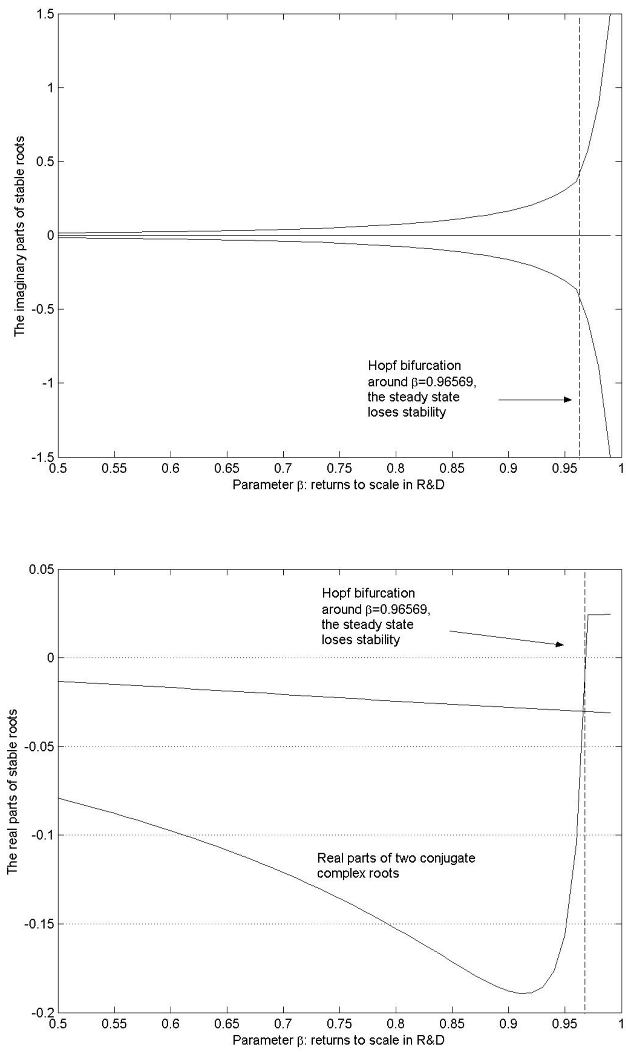

Figure 6.2: Changes in dynamics following changes in β. The Andronov–Hopf bifurcation appears around $\beta \approx 0.96569$.

R&D output is quickly becoming less and less responsive to labor reallocations. This curbs the incentives to constantly reallocate researchers across the two R&D sectors and thus eliminates fluctuations. If β is larger then dampened oscillations appear, and their frequency increases with β. If returns to scale in R&D are high or close to constant, then incentives to reallocate labor are very strong, and business cycles become persistent. Obviously, as we noticed above, the intuition for cycles does not only depend on the returns to scale to labor in the R&D sector, but indeed requires the two kinds of innovations to come in approximately similar amounts, too.

We also notice that the Olsson's (2005) initial intuition for obtaining persistent cycles, namely the linearity assumption which he imposes (here it corresponds to $\beta = 1$), was correct. However, as we demonstrate, $\beta = 1$ is not a necessary condition for cycles in our setup. Indeed, optimal permanent cycles occur already for less-than-constant returns to scale in R&D, $\beta < 1$, although given our baseline parameters, β needs to be close to 1.[6] Obviously, the higher is population growth, the lower the β which leads to permanent cycles.

As suggested, even though oscillations may occur for rather low values of β, it seems that a necessary condition for limit cycles is a high value of β. That this is not sufficient will be shown now. We assign β a value of 0.97, close to the bifurcation value discussed above, and check the consequences of varying the R&D spillover parameter μ as in Figure 6.3. The larger the spillover parameter μ the slower are the oscillations, and for large values (here $\mu > 0.85$) oscillations disappear completely. The Andronov-Hopf bifurcation occurs, of course, again and is identified at $\mu \approx 0.56$. A smaller μ suggests a relatively higher importance of labor for the creation of innovations, which implies that the preferences of the social planner have a larger influence over the path of innovations (which we confirm below). On the contrary, the larger is μ the more important become the relative amounts of technology and technological opportunity.

We have also performed a similar analysis for changes in the discount rate, as in Figure 6.4. At $\rho \approx 0.053$ we observe a similar Andronov–Hopf bifurcation. When $\rho < 0.053$, we have converging oscillations, and smaller values of ρ lead to oscillations of smaller frequency, as intuition would suggest. This result suggests that not only technology matters, but the preferences of the planner play a role, too. So, if the

[6] This finding, of course, does not exclude the possibility that other parameter values lead to bifurcations for values of β significantly below 1.

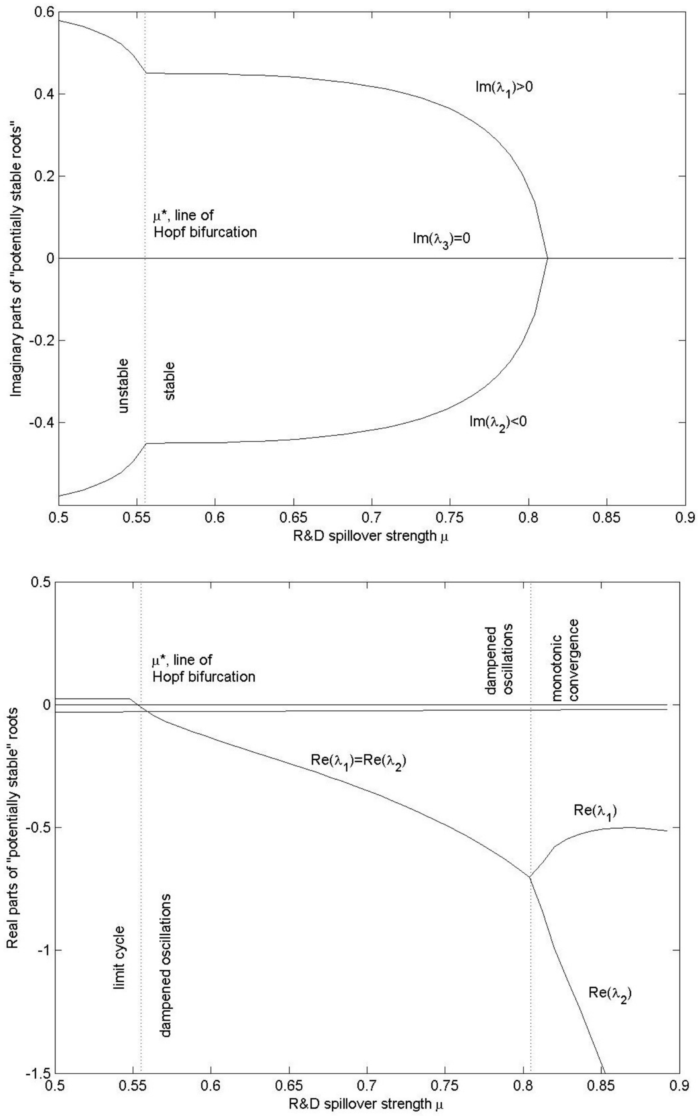

Figure 6.3: Changes in dynamics following changes in μ. When $\beta = 0.97$, the Andronov–Hopf bifurcation appears around $\mu \approx 0.56$.

planner is sufficiently impatient (large ρ) then he will initially allocate more labor to incremental innovations, allowing faster growth now and therefore more consumption. However, because radical innovations will then come in smaller amounts, and technological opportunity will be

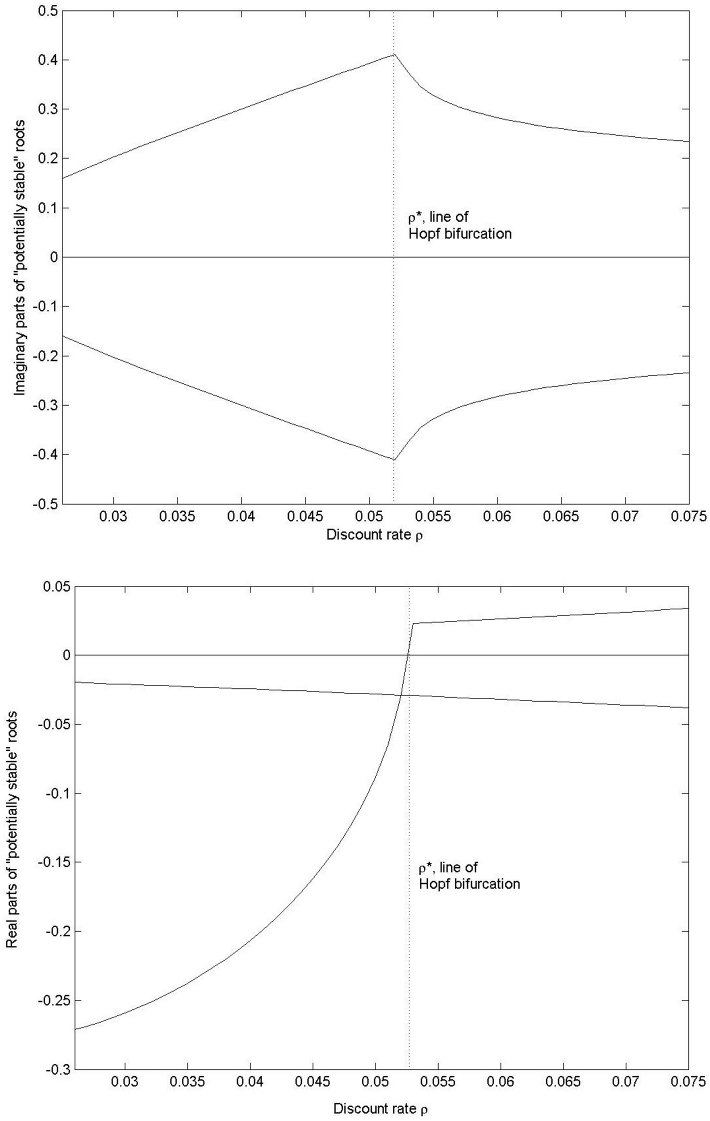

Figure 6.4: Changes in dynamics following changes in the discount rate ρ. When $\beta = 0.97$, the Andronov–Hopf bifurcation appears around $\rho \approx 0.053$.

gradually exhausted, at some point the planner will have to increase the number of researchers in the radical innovations sector in order to prevent economic stagnation. In the moment that enough technological opportunity will have been created, the planner will shift the workers again to the incremental innovations sector to satisfy his impatience for consumption. In case the discount rate is too large, the economy will never converge to the BGP because the planner will be too quick in reallocating labor across the two R&D sectors.

6.3.3 Cycles: a comparison to the literature

Our result of oscillatory dynamics ought to be compared to the predictions found in previous literature. The following list of contributions is obviously not exhaustive; it only serves to illustrate the differences between our approach and the other ones. We have also grouped together articles where cycles are generated through channels which are broadly the same. In these articles, the main sources of oscillatory dynamics are:

(i) the relationship between wage costs, population growth, and profits (Goodwin, 1967; Francois and Lloyd-Ellis, 2003). If growth is high, then unemployment falls, raising wages and decreasing profits. When growth is lower than population growth, this leads to a recession, increases unemployment and starts the cycle again;

(ii) creative destruction (Aghion and Howitt, 1992; Cheng and Dinopoulos, 1992; Amable, 1996; Francois and Lloyd-Ellis, 2003; Phillips and Wrase, 2005);

(iii) uncertainty (Kydland and Prescott, 1982). This is the basic assumption of the real business cycle theory which utilizes productivity shocks and adjustment lags to generate aggregate fluctuations;

(iv) monopoly profits accrued from the distinction between fundamental and secondary innovations (Jovanovic and Rob, 1990; Cheng and Dinopoulos, 1992). These models are based on a similar distinction as our model is. Effectively though, the fluctuations in monopoly profits are the actual source of cycles;

(v) General Purpose Technologies – GPTs (Bresnahan and Trajtenberg, 1995; Helpman and Trajtenberg, 1994; Freeman, Hong, and Peled, 1999). GPTs are assumed to have the property that before they can be applied, costly adaptation to them is required.

This reduces economic growth in the short run but increases it afterwards;

(vi) linearity (Olsson, 2005). Due to the Olsson's linearity assumption, labor is either fully allocated to radical or to incremental research. The interplay between these two types of innovations leads to cycles just like in our model, but it does so through the bang-bang labor allocation and discrete jumps in technological opportunity.

Unlike all these works, our model generates smooth endogenous cycles through the interplay between currently available knowledge and technological opportunity. Because of the sources as well as the nature of cycles in our model, it cannot be considered a member of any of the groups (i)-(vi).

6.4 Further questions

Since the article so far has focused on the need for a knife-edge condition for the technological opportunity-based model to replicate the results of a standard semi-endogenous growth model, and on the oscillatory transition dynamics, there are some questions which we left unanswered. An interesting issue is the contribution of technological opportunity to growth. If this contribution proves to be large, then further empirical investigations should try to address this issue. Another of the essential unanswered questions is the problem of finding a tractable decentralization of the technological opportunity-based model which would enable welfare analysis. Yet another related issue is concerned with capturing technological opportunity empirically. The final question deals with the social planner outcome when the external returns to radical and incremental innovations are not equal, $\mu \neq \nu$. Since most of these aspects are beyond the scope of this article, we shall only hint at how they might be approached in future research.

6.4.1 The contribution of technological opportunity to growth

The contribution of technological opportunity to growth can be assessed as follows. At the highest level, there is of course no direct impact. Indeed, we obtain the following growth rate decomposition (on the BGP): $\hat{y} = \sigma\hat{A} + \alpha\hat{k}$. With our calibrated parameters, it follows that the technology share accounts for $\sigma\hat{A}/\hat{y} = (1-\alpha) = 64\%$ of total growth; the capital accumulation share is $\alpha\hat{k}/\hat{y} = \alpha = 36\%$. Having accounted for the total accumulation of physical capital, resulting from an increase in GDP, and taken the BGP identity $\hat{y} = \hat{k}$, we can readily attribute all GDP growth to technology growth. Going one step deeper, though, one realizes that the increments in technology can also be attributed to various variables. We have $\hat{A} = \mu\hat{B} + \beta\hat{L}$. The central issue is that incremental innovations, increasing the stock of knowledge, cannot go on without technological opportunity. However, the share of technological opportunity growth in technology growth is $\mu = 50\%$. The remaining 50% is attributed to pure population growth, the ultimate source of growth in our model. We can thus conclude that if technology growth is decomposed into its constituent parts, $\sigma\mu\hat{B}/\hat{y} = \mu(1-\alpha) = 32\%$ of total GDP growth can be attributed to the growth in technological opportunity, $\sigma\beta n/\hat{y} = (1-\mu)(1-\alpha) = 32\%$ to pure population growth, and the remaining $\alpha = 36\%$ to capital accumulation. Summing up, given that our baseline parameter choices are accepted, our growth accounting exercise implies that 32% of GDP growth should be attributed to growth in technological opportunity, which is a large number.

6.4.2 Decentralizing the allocation

The decentralization of our model ought to handle the value of radical innovations with care. Their value should approximated by the technological opportunity they open up, but the problem is that in direct terms, they have *no* value. In the decentralized economy, who would then reap the profits from incremental innovations following a given radical innovation? Are patents able to span the whole addition to technological opportunity? How could the profits from technology improvements be shared between radical and incremental innovators in an interior equilibrium? Should the radical innovation sector be, at least partially, funded by the state to obtain a result comparable to the social optimum? We leave these questions open with the hope that they will be addressed in the subsequent research.

6.4.3 An empirical approach to technological opportunity

Another difficult issue is how to capture technological opportunity empirically. It is clear that if one accepts technological opportunity as the driving force behind technological advances, then to understand the evolution of economic variables it is not sufficient to look at the evolution of actual technology. One has to look at the evolution of technological opportunity as well. But how would one be able to capture technological opportunity, which is a forward-looking variable?[7]

A hint for a promising approach can be found in the works of Hall, Jaffe and Trajtenberg (2001) as well as Trajtenberg, Henderson and Jaffe (1994). Hall, Jaffe and Trajtenberg (2001) construct a comprehensive U.S. patent citation dataset. Two indexes, called *originality* and *generality*, are of particular interest here. Generality measures how often a patent is cited by subsequent patents that belong to different fields, whereas originality estimates how large the range of fields is that a given patent cites. The idea would then be then that patent citations and the measures of generality and originality can help find a method of discriminating between patents and estimating their impact on technological opportunity.

Patent citations come in two forms, citations made and received. Citations made give information on the amount of prior knowledge a given patent is based on. Citations received give information on how useful the innovation was for other patents. Thus, if a patent receives a citation, then this means the innovation underlying that patent had opened up technological opportunity. The more citations a patent receives, the larger is the amount of technological opportunity opened up. What Hall, Jaffe and Trajtenberg (2001) show is that around 80% of all citations are made during the first 20 years after a patent was accepted. Thus, it seems evident that when an innovation opens up technological opportunity, this opportunity is always finite. This finding seems to provide some support for our approach.

It remains an important concern, though, whether radical innova-

[7] Thompson (1996) attempts to estimate the elasticities of R&D production functions and suggests to think about these as measures of technological opportunity. Obviously, this interpretation is not applicable in our case and thus the results obtained are not extendable to the framework presented here.

tions, offering no direct applications for business, are patentable at all. In fact, it could also be the case that radical innovations are presented primarily in scientific literature whereas patents are the domain of incremental innovations only (Franz, 2007).

The open question is still how to estimate the degree to which an innovation is radical, and to which it is incremental. We conjecture that the information on patent citations in combination with the indexes of originality and generality might possess enough information to provide an empirical justification to the proposed division of innovations. We feel that the dataset on U.S. patent citations ought to be utilized for this purpose.

Quantifying technological opportunity is an important question for future research: this concept may be used to discriminate between competing growth theories. In particular, the technological complexity-based literature (e.g. Greenwood and Yorukoglu, 1997; Pintea and Thompson, 2007) which considers the recent slowdown in U.S. GDP growth to be a transitional effect due to the extensive application of information technologies which are inherently complex, could be given a new basis for empirical comparisons against the semi-endogenous growth literature that links this slowdown to the slowdown in population growth (e.g. Jones, 2002) and assumes returns to scale in R&D to be not constant but decreasing.

6.4.4 What if $\mu \neq \nu$?

Here we shall discuss what happens if the magnitudes of external returns to incremental and radical innovations are not equal. The knife-edge condition $\mu = \nu$ forces technology and technological opportunity to grow at equal rates along the BGP which somewhat blurs the clear-cut distinction between these two variables and suggests that if for some reason the condition $\mu = \nu$ was automatic, it would be sufficient to use the reduced-form technology accumulation framework along the lines of Jones (1995) for the assessment of long-run economic growth (but not for convergence to the BGP).

Let us now investigate whether a social planner can allocate factors of production across sectors and time in a way that guarantees long-run growth even if $\mu \neq \nu$. The answer to this question is a clear yes, and one possible tool is continuous labor re-allocations. Researchers should be continuously re-allocated from the sector that has higher returns to the sector that has lower returns, making up for the arising deficiency in output of a particular type of innovations.

To help with the intuition, let us take the ratio of incremental to radical innovations, which is given as

$$\frac{C}{R} = \frac{\delta}{\gamma}\left(\frac{u^*}{1-u^*}\right)^{\beta}\frac{B^{\mu}}{A^{\nu}}, \tag{6.22}$$

which implies a growth rate of $\widehat{(C/R)} = \mu\hat{B} - \nu\hat{A} \neq 0$ if labor re-allocations are assumed out. So, if the spillovers in incremental innovations are larger than those in radial innovations ($\mu > \nu$), technological opportunity will be depleted over time and the economy will experience stagnation afterwards. Intuitively, this means that there will be not enough radical innovations (like new product lines or general purpose technologies) to incrementally improve upon. Hence, a social planner who notices that there are insufficient radical innovations ought to re-allocate more researchers towards the sector creating radical innovations. From equation (6.22), one can directly deduce that continuous shifts of researchers towards radical innovations can, if fast enough, offset the depletion effect due to incremental innovations.

On the other hand, if $\mu < \nu$, then incremental innovations will arrive at a lower pace than radical innovations,[8] making technological opportunity expand over time faster than technology. This means that more and more opportunities for technological improvements will emerge, which cannot be utilized. Hence, a planner who allocates more and more researchers to incremental innovations over time, will be able to do better than a planner who sticks to a static labor allocation.

Please note that for $\mu \neq \nu$, but with both parameters sufficiently close to each other, the model does not offer a BGP, but preserves its major characteristics by continuity: growth remains semi-endogenous, both kinds of innovations are produced simultaneously, and the potential for cyclical dynamics (dampened oscillations) is preserved. What is

[8]Of course, initially incremental innovations might arrive faster, but over time radical innovations will for sure overtake.

changed is the fact that labor allocations and the ratio of technology to technological opportunity are no longer stationary in the long run.

Of course, we did not evaluate the optimal social planner solution for the general case with $\mu \neq \nu$ but merely hinted at the requirements which are sufficient to place the economy on a long-run (unbalanced) growth path. Whether the social planner would optimally chose this kind of re-allocations is a question that requires further investigation.

6.5 Conclusion

In this article we have derived an R&D-based semi-endogenous growth model where technological advances depend on the available amount of technological opportunity. We distinguish between two types of innovations, which have different impacts on the evolution of knowledge. Incremental innovations provide direct increases to the stock of knowledge but reduce the technological opportunity which is required for further incremental innovations. Radical innovations serve to renew this opportunity. Hence, technological opportunity behaves like a renewable resource. Even though the basic idea follows Olsson (2005), we generalize his framework in order to incorporate Jones' (1995) and Romer's (1990) models and to compare under what subset of parameter conditions these models predict the same outcomes as ours.

We analyze the model for its growth implications, leading to two novel observations. One, it predicts long-run growth along the lines of Jones (1995), Kortum (1997), and Segerstrom (1998) only if a specific knife-edge condition holds. This condition requires incremental and radical innovations to arrive at the same rates, and thus imply that technology and technological opportunity grow at the same speed. If one expects that this condition is satisfied, then the analytically more tractable model of e.g. Jones (1995) might be sufficient to estimate the growth effects of technological progress. If, however, one presumes that it is unlikely that this condition holds, then, of course, our model can help derive the long-run predictions. For example, as we have shown, economic growth can easily come to a halt if technological opportunity is not renewed sufficiently fast because e.g. the technology spillovers in the radical innovations sector are too small.

The second novel result is with respect to the transition. We focus on the particular knife-edge conditions in order to demonstrate that, even though the long-run implications of this model will then be the same as those in Jones (1995) or Segerstrom (1998), the transition need not. We work out the transition from the social planner solution of our model and show the conditions when it need not be monotonic; on the contrary, we obtain endogenous oscillations and limit cycles for a wide range of plausible parameter values. We therefore suggest that technological opportunity, as characterized here, can be a source of endogenous cycles without needing uncertainty.

Finally, we present the possible directions for future research. These include decentralizing the technological opportunity-based setup and quantifying technological opportunity empirically. Clearly, these extensions give rise to several complications which ought to be discussed in detail.

As a final note we would like to advocate the idea of technological opportunity as a concept which is extremely useful for the description of the evolution of technology. This is a rather novel idea which still lacks a sound empirical justification, but we are convinced that it will gradually earn its deserved place in the theory of economic growth and technical change.

6.6 Appendix

6.6.1 The set-theoretic setup: outline

This appendix demonstrates our version of the set-theoretic approach to modeling ideas put forward by Olsson (2000, 2005). This approach should be understood as a metaphor of the evolution of human thought in reality. Consequently, all the concepts defined precisely in this metaphorical world (such as technological opportunity) are supposed to have real-world analogues.

A more rigorous derivation follows in the next subsection. Figure 6.5 illustrates the set-theoretic approach in a two-dimensional case.

We start with the comforting assumption that any kind of idea can potentially be utilized. The space of ideas (known and unknown) is the

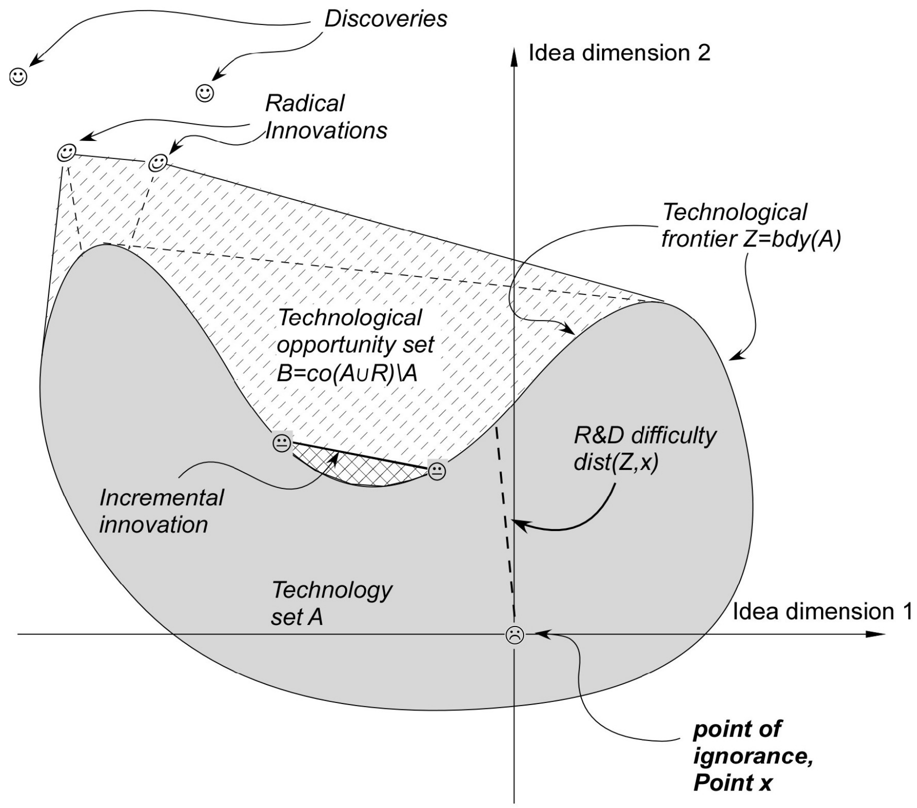

Figure 6.5: The set-theoretic approach illustrated.

whole $\mathbb{R}^n$ space, where $n \in \mathbb{N}$ reflects the (large) number of dimensions in which ideas are characterized. In this world of ideas, the standard Pythagorean distance metric $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is assumed to apply: $d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$. We use the Lebesgue measure λ to describe the volume of all Lebesgue measurable subsets of $\mathbb{R}^n$.

The **technology set** $\mathcal{A} \subset \mathbb{R}^n$ is assumed to be a compact and connected set encompassing all ideas that are already known. We write $\lambda(\mathcal{A}) = A$.

Within the technology set, there is the set of ignorance, called $\mathbf{0} \in \mathcal{A}$, which can be thought of as consisting of some natural ability and instinct, devoid of any systematic knowledge. The set of ignorance has a small but nevertheless positive measure, $\lambda(0) > 0$. It is assumed that in the beginning of time, the technology set $\mathcal{A} \subset \mathbb{R}^n$ is comprised of the set $\mathbf{0}$ only.

Discoveries form a set $\mathcal{D} = \{d_1, d_2, ..., d_m\}$, $\mathcal{D} \in \mathbb{R}^n \backslash \text{co}(\mathcal{A})$ with $m \in \mathbb{N}$, of isolated points away from (the convex hull of) the technology set. At this point, we choose to link the amount of discoveries made at a given moment in time to the current technology level A – i.e. to the current measure of the technology set. In principle, one should allow either for increasing, constant, or decreasing returns to scale here (IRTS, CRTS, DRTS, respectively). It turns out however that only the DRTS case does not lead to explosive dynamics. Therefore, we write $D \propto \#\mathcal{D}$, where $D = g(A) \geq 0$, with $g'(A) \geq 0$ having DRTS.

A **radical innovation** is a one that connects points on the boundary of the technology set $\mathcal{A}$ with discoveries, and so the **radical innovation set** is given by $\mathcal{R} = \bigcup_{i=1,2,...,r} I[x_i, d_i]$, $\forall i(x_i \in \text{bdy}(\mathcal{A}), d_i \in \mathcal{D})$, and $r \in \mathbb{N}$ with $r \leq m$. Please note that the measures of the radical innovation set as well as of the discoveries set are always zero, and thus radical innovations and discoveries cannot provide additions to the measure of the technology set: like in Olsson (2005), they do not improve technology directly.

In the subsequent sections, we will be also interested in the measure of the convex hull of the radical innovation set. We shall write $R = \lambda(co(\mathcal{R}))$.

The possibilities for further technological developments are collected in the **technological opportunity set** $\mathcal{B} \subset \mathbb{R}^n$. Formally, this set is comprised of all ideas that belong to the convex hull of the technology set but do not belong to the technology set itself: $\mathcal{B} = \text{co}(\mathcal{A}) \backslash \mathcal{A}$. Of course, if the technology set is convex ($\mathcal{A} = \text{co}(\mathcal{A})$), then the technological opportunity set is empty. We shall denote the extent of technological opportunity by $B = \lambda(\mathcal{B})$.

Developments combining ideas from the boundary of the technology set, are grouped together in the **incremental innovation set** $\mathcal{C} \subseteq \mathcal{B}$. The inclusion $\mathcal{C} \subseteq \mathcal{B}$ means that one can work out gradual developments only if the opportunity for them already exists. As already indicated before, this is an assumption where Olsson (2005) and this article clearly depart from the usual modeling practice à la Romer (1990), Aghion and Howitt (1992), or Jones (1995).

The "incremental", gradual nature of incremental innovations is captured by assuming that the distance between two combined ideas $i_m, i_n \in \text{bdy}(\mathcal{A})$ cannot exceed some given constant $\bar{d}$: $d(i_m, i_n) \leq \bar{d}$. In addition, a new idea can only be a convex combination of already known ideas, $i_\epsilon = \alpha i_m + (1-\alpha) i_n$ with $0 < \alpha < 1$, where the new idea incrementally extends the technology set, such that $i_\epsilon \in \text{co}(\mathcal{A}) \backslash \mathcal{A}$. We denote the measure of the incremental innovation set by $C = \lambda(\mathcal{C})$.

As derived in the next appendix, technological opportunity gets depleted by incremental innovations but renewed by radical innovations: $\dot{B} = -C + R$. The technology set is increased by incremental innovations and is not altered by radical innovations: $\dot{A} = C$.

One interesting question that remains is whether with larger technological opportunity it becomes increasingly easier to produce incremental innovations, or, on the contrary, it becomes harder to produce them. We are not aware of any empirical evidence on this, but for the sake of plausibility, we shall concentrate on the DRTS case here. Again, it is the unique case that does not lead to explosive dynamics.

When constructing our model, we shall also make the usual assumption that each individual in the population of size L splits her fixed time endowment $(1 = \ell_Y + \ell_A)$ between working, ℓ_Y, searching for radical innovations, $(1-u)\ell_A$, and doing incremental R&D, $u\ell_A$. This assumption allows us to clarify the evolution of the measures of the sets as follows: $\dot{B} = -C(u\ell_A L, B) + R((1-u)\ell_A L, A)$; $\dot{A} = C(u\ell_A L, B)$.

6.6.2 The set-theoretic setup: formal presentation

We shall now present the details of our set-theoretic approach in a more rigorous and compact form. This approach has been first formulated by Olsson (2000) and then refined by the same author in Olsson (2005). Here we have added in more structure (Lebesgue measure, Pythagorean metric, point of ignorance, R&D difficulty).

Definition 6.1 *The* idea space *is the $\mathbb{R}^n$ space, where $n \in \mathbb{N}$ is predefined, together with the Lebesgue measure λ defined on all Borellian subsets of $\mathbb{R}^n$, and the usual Pythagorean metric $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by $d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.*

Definition 6.2 *The* technology set $\mathcal{A} \subset \mathbb{R}^n$ *is a compact and connected set such that* $0 \in \mathcal{A}$.

Please note that the origin $0 \in \mathbb{R}^n$ is labeled also the *point of ignorance*; the set of ignorance is a ball of some small diameter $\bar{\epsilon} > 0$, surrounding the point 0.

Definition 6.3 *The* technology level *is the Lebesgue measure of the technology set,* $\lambda(\mathcal{A}) > 0$.

Definition 6.4 *The* paradigm $\mathcal{P} \subset \mathbb{R}^n$ *is the convex hull of the technology set:* $\mathcal{P} \equiv \mathrm{co}(\mathcal{A})$.

The paradigm is a compact, connected and convex set. We also have that $0 \in \mathcal{P}$, $\mathcal{A} \subset \mathcal{P}$, and $\lambda(\mathcal{P}) \geq \lambda(\mathcal{A}) > 0$.

Definition 6.5 *The* technological opportunity set $\mathcal{B} \subset \mathbb{R}^n$ *is (the closure of) the set of ideas that belong to the paradigm but do not belong to the technology set:* $\mathcal{B} = cl(\mathcal{P} \backslash \mathcal{A})$.

The technological opportunity set is a compact set, and $0 \notin \mathcal{B}$.

Definition 6.6 *The level of* technological opportunity *is the measure of the technological opportunity set* $\lambda(\mathcal{B})$.

Definition 6.7 *The* technology frontier $\mathcal{Z}$ *is the set of ideas that belong both to the technology set and to the technological opportunity set:* $\mathcal{Z} \equiv \mathcal{A} \cap \mathcal{B}$.

The technology frontier is a compact boundary set ($int\mathcal{Z} = \emptyset, \lambda(\mathcal{Z}) = 0$).

Definition 6.8 *The* R&D difficulty *is the distance of the technology frontier to the origin:* $\bar{\delta} \equiv dist(\mathcal{Z}, 0)$. *If the technology frontier is empty, we put* $\bar{\delta} = dist(bdy\mathcal{A}, 0)$.

Definition 6.9 *The* incremental innovation set *is a subset of the technological opportunity set,* $\mathcal{C} \subset \mathcal{B}$, *with* $\mathcal{C} \cap \mathcal{Z} \neq \emptyset$ *if* $\mathcal{Z} \neq \emptyset$.

Definition 6.10 *The* discoveries set *is a finite set of isolated points outside the paradigm:* $\mathcal{D} = \{d_1, d_2, ..., d_m\}$, $\mathcal{D} \in \mathbb{R}^n \backslash \mathcal{P}$, *and* $m \in \mathbb{N}$.

The discoveries set is a compact boundary set, with $int\mathcal{D} = \emptyset, \lambda(\mathcal{D}) = 0$.

Definition 6.11 *The* radical innovation set *is a set of closed intervals, connecting points on the boundary of the technology set with discoveries:* $\mathcal{R} = \bigcup_{i=1,2,\ldots,r} I[x_i, d_i]$*, where* $\forall i(x_i \in bdy\mathcal{A}, d_i \in \mathcal{D})$*, and* $r \in \mathbb{N}$ *with* $r \leq m$*.*

The radical innovation set is a compact boundary set, $int\mathcal{R} = \emptyset, \lambda(\mathcal{R}) = 0$.

6.6.3 Evolution of the sets over time

We assume that in a period of time of an infinitesimal length $\varepsilon > 0$,

$$\mathcal{A}_\varepsilon = \mathcal{A} \cup \mathcal{C} \cup \mathcal{R},$$

i.e. newly invented technologies add to the technology set. It follows that the paradigm is shifted by radical innovations; technological opportunity is diminished by incremental innovations and extended by radical innovations, and the technology frontier is pushed forward:

$$\begin{aligned}
\mathcal{P}_\varepsilon &= \mathrm{co}(\mathcal{A}_\varepsilon), \\
\mathcal{B}_\varepsilon &= cl(\mathcal{P}_\varepsilon \backslash \mathcal{A}_\varepsilon) = cl\left(\underbrace{(\mathcal{B}\backslash\mathcal{C})}_{\text{extraction}} \cup \underbrace{(\mathcal{P}_\varepsilon\backslash\mathcal{P})}_{\text{paradigm shift}}\right), \\
\mathcal{Z}_\varepsilon &= \mathcal{A}_\varepsilon \cap \mathcal{B}_\varepsilon.
\end{aligned}$$

Bearing in mind that the radical innovation set is of measure zero but its convex hull is (typically) of positive measure, we find that the technology level and the level of technological opportunity evolve in the following way:

$$\begin{aligned}
\lambda(\mathcal{A}_\varepsilon) &= \lambda(\mathcal{A}) + \lambda(\mathcal{C}), \\
\lambda(\mathcal{B}_\varepsilon) &= \lambda(\mathcal{B}) - \lambda(\mathcal{C}) + \lambda(\mathcal{P}_\varepsilon \backslash \mathcal{P}).
\end{aligned}$$

The last "paradigm shift" term is driven by radical innovations in $\mathcal{R}$ only. Please note that only incremental innovations have the power to increase the measure of the technology set.

R&D difficulty evolves as follows:

$$\bar{\delta}_\varepsilon = dist(\mathcal{Z}_\varepsilon, 0) \geq \bar{\delta}.$$

These findings lie at the foundation of the dynamic equations of our model, summarized in the system of equations (6.1)–(6.2).

Bibliography

[1] Abio, G. (2003), Interiority of the Optimal Population Growth Rate with Endogenous Fertility, *Economics Bulletin* 10(4), 1-7.

[2] Acemoglu, D. (2003), Labor- and Capital-Augmenting Technical Change, *Journal of the European Economic Association* 1, 1-37.

[3] Aghion, P. and P. Howitt (1992), A Model of Growth Through Creative Destruction, *Econometrica* 60(2), 323–351.

[4] Amable, B. (1996), Croissance et cycles induits par les innovations radicales et incrémentales, *Annales d'Economie et Statistique* 44, 91–109.

[5] Amigues, J.-P., M. Moreaux, F. Ricci (2006), Overcoming the Natural Resource Constraint Through Dedicated R&D Effort with Heterogenous Labor Supply. Université Toulouse 1 and LEERNA, *mimeo.*

[6] André, F., E. Cerdá (2005), On Natural Resource Substitution, *Resources Policy* 30(4), 233-246.

[7] Arnold, L.G. (2006), The Dynamics of the Jones' R&D Model, *Review of Economic Dynamics* 9, 143-152.

[8] Arnold, W.I. (1975), *Równania różniczkowe zwyczajne.* PWN, Warszawa.

[9] Arrow, K.J., H.B. Chenery, B.S. Minhas, R.M. Solow (1961), Capital-Labor Substitution and Economic Efficiency, *Review of Economics and Statistics* 43, 225-250.

[10] Arrow, K.J. (1962), The Economic Implications of Learning by Doing, *Review of Economic Studies* 29(3), 155-173.

[11] Atkinson, A.B., J.E. Stiglitz (1969), A New View of Technological Change, *Economic Journal* 79, 573-578.

[12] Barro, R.J., G.S. Becker (1989), Fertility Choice in a Model of Economic Growth, *Econometrica* 57(2), 481-501.

[13] Barro, R.J., X. Sala-i-Martin (1995), *Economic Growth.* McGraw-Hill, Inc.

[14] Basu, S., D.N. Weil (1998), Appropriate Technology and Growth, *Quarterly Journal of Economics* 113, 1025-1054.

[15] Beaudry, P., E. van Wincoop (1996), The Intertemporal Elasticity of Substitution: An Exploration Using a U.S Panel of State Data, *Economica* 63, 495-512.

[16] Becker, G.S. (1981), *A Treatise on the Family.* Harvard University Press, Cambridge.

[17] Becker, G.S., C. Mulligan (1997), The Endogenous Determination of Time Preference, *Quarterly Journal of Economics* 112, 729-758.

[18] Bresnahan, T. F., M. Trajtenberg (1995), General Purpose Technologies: Engines of Growth? *Journal of Econometrics* 65(1), 83–108.

[19] Bretschger, L. (2005), Economics of Technological Change and the Natural Environment: How Effective are Innovations as a Remedy for Resource Scarcity?, *Ecological Economics* 54 (2-3), 148-163.

[20] Caselli, F., W.J. Coleman (2006), The World Technology Frontier, *American Economic Review* 96(3), 499-522.

[21] Cheng, L.K., E. Dinopoulos (1992), Schumpeterian Growth and International Business Cycles, *American Economic Review Papers and Proceedings* 82(2), 409-414.

[22] Christensen, L.R., D.W. Jorgenson, L.J. Lau (1973), Transcendental Logarithmic Production Frontiers, *Review of Economics and Statistics* 55(1), 29-45.

[23] Christiaans, T. (2004), Types of Balanced Growth, *Economics Letters* 82(2), 253–258.

[24] Cleveland, C., M. Ruth (1997), When, Where, and By How Much Do Biophysical Limits Constrain the Economic Process? The Contribution of Nicholas Georgescu-Roegen to Ecological Economics, *Ecological Economics* 22, 203-223.

[25] Cockburn, I., R. Henderson, S. Stern (1999), Balancing Incentives: The Tension Between Basic and Applied Research, NBER Working Paper 6882.

[26] Connolly, M., P. Peretto (2003), Industry and the Family: Two Engines of Growth, *Journal of Economic Growth* 8(1), 114-148.

[27] Das, M. (2003), Optimal Growth With Decreasing Marginal Impatience, *Journal of Economic Dynamics and Control* 27, 1881-1898.

[28] Dasgupta, P., G. Heal (1974), The Optimal Depletion of Exhaustible Resources, *Review of Economic Studies* 41, 3-28.

[29] de la Croix, D., M. Doepke (2003), Inequality and Growth: Why Differential Fertility Matters, *American Economic Review* 93, 1091-1113.

[30] de la Croix, D., M. Doepke (2004), Public versus Private Education When Differential Fertility Matters, *Journal of Development Economics* 73, 607-629.

[31] de La Grandville O. (1989), In the Quest of the Slutsky Diamond, *American Economic Review* 79(3), 468-481.

[32] Deardorff, A. V. (1976), The Optimum Growth Rate for Population: Comment, *International Economic Review* 17(2), 510-515.

[33] Duffy, J., C. Papageorgiou (2000), A Cross-Country Empirical Investigation of the Aggregate Production Function Specification, *Journal of Economic Growth* 5, 87-120.

[34] Eicher, T.S., S.J. Turnovsky (1999a), Convergence in a Two-Sector Nonscale Growth Model, *Journal of Economic Growth* 4(4), 413–428.

[35] Eicher, T.S., S.J. Turnovsky (1999b), Non-Scale Models of Economic Growth, *Economic Journal* 109, 394-415.

[36] Epstein, L., S. Zin (1989), Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework, *Econometrica* 57, 937-969.

[37] Favero, C.A. (2005), Consumption, Wealth, the Elasticity of Intertemporal Substitution and Long-Run Stock Market Returns, CEPR Working Paper No. 5110.

[38] Francois, P., H. Lloyd-Ellis (2003), Animal Spirits Through Creative Destruction, *American Economic Review* 93(3), 530–550.

[39] Franz, J.S. (2007), The Economics of Ideas: Theory and Evidence. University of Illinois at Chicago, *mimeo.*

[40] Freeman, S., D. Hong, D. Peled (1999), Endogenous Cycles and Growth with Indivisible Technological Developments, *Review of Economic Dynamics* 2(2), 402–432.

[41] Gabaix, X. (1999), Zipf's Law for Cities: An Explanation, *Quarterly Journal of Economics* 114, 739-767.

[42] Goodwin, R.M. (1967), A Growth Cycle [in:] C.H. Feinstein, ed.: *Socialism, Capitalism, and Economic Growth.*

[43] Greenwood, J., M. Yorukoglu (1997), 1974, *Carnegie-Rochester Conference Series on Public Policy* 46, 49-95.

[44] Griffin, J.M. (1977), Inter-Fuel Substitution Possibilities: A Translog Application to Intercountry Data, *International Economic Review* 18(3), 755-770.

[45] Griliches, Z. (1986), Productivity, R&D, and the Basic Research at the Firm Level in the 1970's, *American Economic Review* 76(1), 141-154.

[46] Grimaud, A., L. Rouge (2005), Polluting Non-Renewable Resources, Innovation and Growth: Welfare and Environmental Policy, *Resource and Energy Economics* 27(2), 109-129.

[47] Growiec, J. (2006a), Fertility Choice and Semi-Endogenous Growth: Where Becker Meets Jones, CORE Discussion Paper 2006/23.

[48] Growiec, J. (2006b), Production Functions and Distributions of Unit Factor Productivities: Uncovering the Link, Warsaw School of Economics, *mimeo*.

[49] Guvenen, M.F. (2006), Reconciling Conflicting Evidence on the Elasticity of Intertemporal Substitution: A Macroeconomic Perspective, *Journal of Monetary Economics* 52(7), 1451-1472.

[50] Hall, B.H., A.B. Jaffe, M. Trajtenberg (2001), The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools, NBER Working Paper No. 8498.

[51] Hall, R.E. (1988), Intertemporal Elasticity of Substitution in Consumption, *Journal of Political Economy* 96, pp. 339-357.

[52] Hall, R.E., C.I. Jones (2007), The Value of Life and the Rise in Health Spending, *Quarterly Journal of Economics* 122(1), 39-72.

[53] Harashima, T. (2005), An Estimate of the Elasticity of Intertemporal Substitution in a Production Economy, mimeo, University of Tsukuba.

[54] Helpman, E., M. Trajtenberg (1994), A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies, NBER Working Paper No. 4854.

[55] Hicks, J. (1932), *The Theory of Wages*, 2^{nd} edition [1963]. London, Macmillan.

[56] Houthakker, H.S. (1955-56), The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis, *Review of Economic Studies* 23, 27-31.

[57] Howitt, P. (1999), Steady Endogenous Growth with Population and R&D Inputs Growing, *Journal of Political Economy* 107(4), 715–730.

[58] Jones, C.I. (1995), R&D-Based Models of Economic Growth, *Journal of Political Economy* 103, 759-784.

[59] Jones, C.I. (1999), Growth: With or Without Scale Effects? *American Economic Review Papers and Proceedings* 89, 139-144.

[60] Jones, C.I. (2001), Was an Industrial Revolution Inevitable? Economic Growth Over the Very Long Run, *Advances in Macroeconomics* 1(2), Article 1.

[61] Jones, C.I. (2002), Sources of U.S. Economic Growth in a World of Ideas, *American Economic Review* 92(1), 220-239.

[62] Jones, C.I. (2003), Population and Ideas: A Theory of Endogenous Growth, [in:] P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds., *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps.* Princeton University Press.

[63] Jones, C.I. (2005a), Growth and Ideas, [in:] P. Aghion and S.N. Durlauf, eds., *Handbook of Economic Growth.* North-Holland, Amsterdam.

[64] Jones, C.I. (2005b), The Shape of Production Functions and the Direction of Technical Change, *Quarterly Journal of Economics* 120(2), 517–549.

[65] Jones, L.E., R.E. Manuelli (1990), A Convex Model of Equilibrium Growth: Theory and Policy Implications, *Journal of Political Economy* 98, 1008-1038.

[66] Jovanovic, B., P.L. Rousseau (2003), Two Technological Revolutions, *Journal of the European Economic Association* 1(2-3), 419–428.

[67] Jovanovic, B., R. Rob (1990), Long Waves and Short Waves: Growth through Intensive and Extensive Search, *Econometrica* 58(6), 1391–1409.

[68] Kim, H.Y. (1992), The Translog Production Function and Variable Returns to Scale, *Review of Economics and Statistics* 74(3), 546-552.

[69] Klump R., O. de La Grandville (2000), Economic Growth and the Elasticity of Substitution: Two Theorems and Some Suggestions, *American Economic Review* 90, 282-291.

[70] Klump, R., H. Preissler (2000), CES Production Functions and Economic Growth, *Scandinavian Journal of Economics* 102, 41-56.

[71] Kortum, S.S. (1997), Research, Patenting, and Technological Change, *Econometrica* 65(6), 1389–1419.

[72] Kremer, M. (1993), Population Growth and Technological Change: From One Million BC to 1990, *Quarterly Journal of Economics* 108(3), 681-716.

[73] Kydland, F.E., E.C. Prescott (1982), Time to Build and Aggregate Fluctuations, *Econometrica* 50(6), 1345–1370.

[74] Laffargue, J.-P. (2004), A Sufficient Condition for the Existence and the Uniqueness of a Solution in Macroeconomic Models with Perfect Foresight, *Journal of Economic Dynamics and Control* 28, 1955-1975.

[75] Li, C.W. (2000), Endogenous vs. Semi-Endogenous Growth in a Two-R&D-Sector Model, *Economic Journal* 110, C109-C122.

[76] Lotka, A.J. (1926), The Frequency Distribution of Scientific Productivity, *Journal of the Washington Academy of Sciences* 16, 317-323.

[77] Malla, S., R. Gray (2005), The Crowding Effects of Basic and Applied Research: A Theoretical and Empirical Analysis of an

Agricultural Biotech Industry, *American Journal of Agricultural Economics* 87(2), 423-438.

[78] Malthus, T.R. (1798), *An Essay on the Principle of Population.* Printed for J. Johnson, in St. Paul's Church-Yard, London.

[79] Miyagiwa, K., C. Papageorgiou (2005), The Elasticity of Substitution, Hicks' Conjectures, and Economic Growth, Emory University & Louisiana State University, *mimeo.*

[80] Nelsen R.B. (1999), *An Introduction to Copulas.* Springer, New York.

[81] Nordhaus, W. (1973), The Allocation of Energy Reserves, *Brookings Papers on Economic Activity* 3, 529-570.

[82] Olsson, O. (2000), Knowledge as a Set in Idea Space: An Epistemological View on Growth, *Journal of Economic Growth* 5, 253-275.

[83] Olsson, O. (2005), Technological Opportunity and Growth, *Journal of Economic Growth* 10(1), 31–53.

[84] Palivos T., G. Karagiannis (2004), The Elasticity of Substitution in Convex Models of Endogenous Growth, University of Macedonia at Thessaloniki, *mimeo.*

[85] Parfit, D. (1984), *Reasons and Persons.* Clarendon Press, Oxford.

[86] Patterson K.D., B. Pesaran (1992), The Intertemporal Elasticity of Substitution in Consumption in the United States and the United Kingdom, *Review of Economics and Statistics* 74(4), 573-584.

[87] Peretto, P. F. (1998), Technological Change and Population Growth, *Journal of Economic Growth* 3(4), 283–311.

[88] Petith, H. (2001), Georgescu-Roegen Versus Solow/Stiglitz and the Convergence to the Cobb-Douglas, Departament d'Economia i d'Història Econòmica W.P. 489.01.

[89] Phillips, K.L., J. Wrase (2006), Is Schumpeterian "Creative Destruction" a Plausible Source of Endogenous Real Business Cycle Shocks? *Journal of Economic Dynamics and Control* 30, 1885–1913.

[90] Pimentel, D., O. Bailey, P. Kim, E. Mullaney, J. Calabrese, L. Walman, F. Nelson, X. Yao (1999), Will Limits of the Earth's Resources Control Human Numbers?, [in:] B. Nath, L. Hens and D. Pimentel, eds., *Environment, Development and Sustainability*, 19-39. Springer Science+Business Media B.V.

[91] Pintea, M., P. Thompson (2007), Technological Complexity and Economic Growth, *Review of Economic Dynamics* 10, 276-293.

[92] Romer, P.M. (1990), Endogenous Technological Change, *Journal of Political Economy* 98(5), S71–S102.

[93] Romer, P.M. (1993), Two Strategies for Economic Development: Using Ideas and Producing Ideas, *Proceedings of the Annual World Bank Conference on Development 1992*, Supplement, Washington, D.C., World Bank Economic Review.

[94] Ryberg, J., T. Tänssjö, G. Arrhenius (2006), The Repugnant Conclusion, [in:] *Stanford Encyclopaedia of Philosophy*. Available at http://plato.stanford.edu/entries/repugnant-conclusion/

[95] Samuelson, P.A. (1975), The Optimum Growth Rate for Population, *International Economic Review* 16(3), 531-538.

[96] Sattinger, M. (1975), Comparative Advantage and the Distributions of Earnings and Abilities, *Econometrica* 43(3), 455-468.

[97] Scholz, C., G. Ziemes (1999), Exhaustible Resources, Monopolistic Competition, and Endogenous Growth, *Environmental and Resource Economics* 13, 169-185.

[98] Schou, P. (2000), Polluting Non-Renewable Resources and Growth, *Environmental and Resource Economics* 16, 211-227.

[99] Segerstrom, P.S. (1998), Endogenous Growth without Scale Effects, *American Economic Review* 88(5), 1290–1310.

[100] Segerstrom, P.S. (2000), The Long-Run Growth Effects of R&D Subsidies, *Journal of Economic Growth* 5(3), 277–305.

[101] Solow R.M. (1956), A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics* 70, 65-94.

[102] Solow R.M. (1960), Investment and Technological Progress, [in:] Arrow K., Karlin S., Suppes P. (eds.), *Mathematical Methods in Social Sciences 1959.* Stanford, Stanford University Press.

[103] Solow, R.M. (1974), Intergenerational Equity and Exhaustible Resources, *Review of Economic Studies* 41, 29-45.

[104] Solow R.M., J. Tobin, C. von Weizsäcker, M. Yaari (1966), Neoclassical Growth with Fixed Factor Proportions, *Review of Economic Studies* 33, 79-115.

[105] Steger, T.M. (2005), Non-Scale Models of R&D-Based Growth: The Market Solution, *Topics in Macroeconomics* 5(1), Article 3.

[106] Stiglitz, J. (1974), Growth with Exhaustible Resources: Efficient and Optimal Growth Paths, *Review of Economic Studies* 41, 123-138.

[107] Strulik, H. (2005), The Role of Human Capital and Population Growth in R&D-Based Models of Economic Growth, *Review of International Economics* 13, 129-145.

[108] Tahvonen, O., S. Salo (2001), Economic Growth and Transitions Between Renewable and Nonrenewable Energy Resources, *European Economic Review* 45, 1379-1398.

[109] Temple, J. (2003), The Long-Run Implications of Growth Theories, *Journal of Economic Surveys* 17(3), 497-510.

[110] Thompson, P. (1996), Technological Opportunity and the Growth of Knowledge: A Schumpeterian Approach to Measurement, *Journal of Evolutionary Economics* 6, 77–97.

[111] Trajtenberg, M., R. Henderson, A.B. Jaffe (1997), University Versus Corporate Patents: A Window on the Basicness of Inventions, *Economics of Innovation and New Technology* 5, 19-50.

[112] Uzawa, H. (1961), Neutral Inventions and the Stability of Growth Equilibrium, *Review of Economic Studies* 28(2), 117-124.

[113] van de Kaa, D.J. (1997), Options and Sequences: Europe's Demographic Patterns, *Journal of the Australian Population Association* 14 (1), 1-30.

[114] Weil, P. (1990), Nonexpected Utility in Macroeconomics, *Quarterly Journal of Economics* 105(1), 29-42.

[115] Weitzman, M.L. (1998), Recombinant Growth, *Quarterly Journal of Economics* 113(2), 331–360.

[116] Young, A. (1998), Growth without Scale Effects, *Journal of Political Economy* 106, 41-63.

[117] Yuhn, K.-H. (1991), Economic Growth, Technical Change Biases, and the Elasticity of Substitution: A Test of the de La Grandville Hypothesis, *Review of Economics and Statistics* 73(2), 340-346.