Working Party on Global and Structural Policies

DEVELOPMENTS IN GROWTH LITERATURE AND THEIR RELEVANCE FOR SIMULATION MODELS
FOREWORD

Developing a better understanding of efforts to incorporate endogenous growth into the mainstream of economic discourse is a necessary undertaking for policy analysts providing economic advice to decision-makers. Incorporating endogenous growth into numerical models used for policy initiatives, as well as for – both for providing a careful qualitative consideration of alternative policy initiatives, as well as for quantifying the relative importance of factors affecting growth versus those affecting other economic phenomena. For environmental policy, this is potentially very important, inasmuch as endogenous growth is largely founded on technological change in all forms. Technological change is a key variable in the search for environmental improvement, in both the short-and long-terms. If certain types of environment-related policy decisions generate not only direct effects on the targets of these policies, but also engender consequences for other sectors of the economy (via technology effects), traditional analysis may be missing some important channels of action.

This document explores the many facets of developments in endogenous growth modelling – both theoretical and empirical – and discusses links to the specific needs of computable general equilibrium (CGE) models. The backdrop to this work is that, given the potential importance of growth theories, there is also some desire to incorporate elements of endogenous growth into numerical models in use within the OECD. There is also, however, considerable uncertainty as to how – and even whether – to proceed. Nonetheless, since the OECD uses models for analysis in a variety of contexts, there is a potentially important gap between elements of the mainstream economics literature and the applied analysis that the OECD presents to policy makers. The work undertaken with this study is intended to be a first step in rectifying this situation.

The discussion given in this document is presented in the form of a literature review, combines with a range of technical conclusions that emerge from that review. This document is an Annex to a smaller non-technical document which more succinctly summarises some of the main themes found herein and outlines some relevant considerations for the application of these concepts into OECD Environment work.

The views expressed in the report are those of the authors, and do not necessarily reflect the opinions of either the OECD or its Member countries. The report is published under the responsibility of the Secretary-General.
# TABLE OF CONTENTS

1. INTRODUCTION ................................................................................................................... 5

2. FROM EXOGENOUS TO ENDOGENOUS GROWTH THEORY .................................................. 7
   2.1 The limits of the Solow-Swan model .................................................................................... 7
   2.2 The foundations of endogenous growth models ................................................................. 9
   2.3 A common representation of endogenous growth theory for CGE models ......................... 11

3. AN EMPIRICAL AND THEORETICAL ASSESSMENT OF ENDOGENOUS GROWTH MODELS FOR CGE MODELS ........................................................................................................... 16
   3.1 The accumulation of capital stock ...................................................................................... 16
   3.2 Accumulation of human capital stock ............................................................................... 25
   3.3 The accumulation of new ideas: endogenous growth with research and development ........... 28

4. CONCLUDING REMARKS ...................................................................................................... 52
1. INTRODUCTION

This paper examines the developments of the new growth literature and their relevance for computable general equilibrium (CGE) models. An important direction for present and future research involves the operationalisation of endogenous growth concepts within environment/energy models for more practical purpose; beginning with their empirical testing and calibration. Some applied researchers have already begun to introduce elements from this literature into their models. Examples include Fougeyrollas et al. (2001), Van Bergeijk et al. (1997) and Den Butter and Wollmer (1996). Recent conferences have also attempted to look at climate change policy when technological change is accounted for in an analytical or numerical modelling framework. Notable are the conferences: Economic Modelling of Environmental Policy and Endogenous Technological Change (November, 2000, http://www.vu.nl/ivm); and, the Energy Modelling Forum’s Workshop on Modelling Technological Change (June 2001, EMF 19: Stanford University). Use of endogenous growth concepts in general equilibrium models of trade has also given interesting policy insights.

Within the OECD, there has also been considerable interest in the policy implication of growth theories for Member countries (Bassanini and Scarpetta, 2001, Guellec and van P. de la Potterie, 2001) Given the potential importance of these theories, there is also some desire to incorporate elements of endogenous growth into computable general equilibrium models already in use within the OECD. Since the OECD uses models for analysis in a variety of contexts, there is a potentially important gap between elements of the mainstream economics literature and the applied analysis we present to policy makers. However, there is considerable uncertainty as to how -- and even whether -- to proceed. This study is intended to explore the current state of endogenous growth models with an objective eye to examining their applicability to numerical policy analysis.

Long run growth was first introduced by Solow (1956) and Swan (1956) into the traditional neoclassical macroeconomic model by considering a growing population, coupled with a more efficient labour force. The direct consequence of this approach is that the long-run growth rate is ultimately tied to demographic factors and its productivity growth, all of which were typically taken to be exogenously determined. Hence, macroeconomic policy had no influence on long run growth performance. As pointed out by Barro and Sala-I-Martin (1995), this resulted in a model of growth explaining everything but the long run growth. More particularly, it seemed reasonable that at least the exogenous productivity changes in the Solow-Swan model are, in fact, the result of the decisions of economic agents’ decisions.

Probably due to the lack of empirical evidence and the difficulties of including a theory of technical change within the neo-classical framework, research interest in these models tapered off around 1970. Instead, economists focused on short-run fluctuations and the design of macroeconomic policies to deal with them. However, the seminal contributions of Romer (1986, 1987) and Lucas (1988) triggered a renewal of interest in economic growth theory. This new growth literature has been motivated by several issues. Among others, Romer (1994) points out the following: (i) a more satisfactory explanation of long run differences in performances of different countries, (ii) an attempt to explain aspects of the empirical puzzle not addressed by the neoclassical model, such as the importance of the Solow’s "residual”, (iii) a more central role for the accumulation of knowledge or the economy of ideas and (iv) a larger role for macroeconomics, e.g. fiscal and public policies, in explaining long run growth process. To go beyond the neoclassical framework, one has to escape from diminishing marginal product of capital and the tight link between long run per capita income growth rate and the exogenous technical progress. Ideally, the
balanced growth path would be based on per capita variables growing at a positive constant growth rate, independently of the rate of some exogenous variables. Basically, two ideas have been developed, one that emphasised factors counteracting diminishing returns to scale, and another that considered technological progress as an output from a technology-sector in the economy. All models of the new growth theory share at least one common property: they seek to explain the growth rate as an endogenous equilibrium outcome of the behaviour of rational optimising agents, reflecting the structural characteristics of the economy, such as technology and preferences, as well as macroeconomic policy. But there is an important difference between the two approaches in the sense that in the “first generation” of endogenous growth models, technological progress is assumed to be external to firms while the “second generation” of models assumes technological progress to derive from intentional activity (Fagerberg; 1994; Solow; 2001). The new growth literature has been analysed in both closed and open economies. In fact, one of the characteristics of this literature is that it has more of an international perspective, with several concepts only relevant in an international context or in an open economy, as for instance research and development spillovers.

This newer literature places a greater emphasis on empirical issues and the reconciliation of the theory with the empirical evidence. In this respect, an open debate concerns the so-called "convergence hypothesis", i.e., to what extent countries converge to the same per capita level of income or growth rate. In addition, the important theoretical advances of endogenous growth models have led to new empirical research, as for instance the measurement of externalities, the importance of research and development and the relevance of their assumptions. A survey of the empirical evidence relating to new growth literature displays, however, mixed results (Temple, 1999; Aron, 2000). While the theoretical literature has made considerable advances in identifying the means by which growth may occur, the empirical literature has not yet been able to clearly distinguish between competing models. These developments also have implications for computable general equilibrium models. In fact, introduction of endogenous growth rates in computable general equilibrium models gives rise to several issues: calibration, selection of equilibrium, the nature of expectations and the mechanisms driving endogenous growth rate etc. Early attempts to address these questions have been made by Rutherford and Tarr (1998), Diao and Roe (1997), Bayoumi et al. (1998), Diao et al. (1999), Ambler et al. (1999), Datta and Mohtadi (2000), and Roe (2001). At the same time, endogenous growth theories have become an important part of mainstream study. In particular, these theories have permeated into fields of economic analysis such as environmental and resource economies (Reis, 2001). In order to more strongly underpin the policy advice given in that field. This has, itself, led to a higher demand for endogenous technical progress-based CGE models. It is clear, however, that such CGE models cannot be built without a good comprehension of endogenous growth models and the empirical support for their crucial assumptions and main results.

This paper is organised as follows. The next section reviews the emergence of the new growth literature. The third section describes the different generations of endogenous growth models, as well as computable general equilibrium models that take into account recent developments in the endogenous growth literature. In particular, we separate that literature into three strands: models in which the source of growth is the accumulation of capital stock, the accumulation of human capital, and ideas. Section 4 concludes.
2. FROM EXOGENOUS TO ENDOGENOUS GROWTH THEORY

2.1 The limits of the Solow-Swan model

To set the stage for discussing recent theoretical development of growth models, it is useful to start with the standard neoclassical growth model. Solow (1956) and Swan (1956) focused on the process of capital accumulation. For the sake of simplicity, we consider a simplified version with no technological progress. The Solow-Swan model can be described as a set of three formulations. **First**, production involves the use of labour and capital according to a production function that exhibits: constant returns to scale; diminishing marginal product in each essential factor of production; and some regularity conditions (Inada’s conditions) which insure the existence of a positive steady-state. **Second**, the growth process is the result of the accumulation of capital: the variation of the capital stock is the result of the difference between the amount of new investment and the depreciation of the existing stock of capital. **Third**, aggregate savings (defined as a constant fraction of national income) are equal to investment and finance additions to the national capital stocks. Over time, capital per worker will rise, which will generate declines in the marginal product of capital per worker, since the production function exhibits constant returns to scale and the technology is fixed (stationary). Thus, the savings generated by the income accruing to new capital will also fall and will eventually be only sufficient to replace existing machines (technological depreciation) and equip new workers (the labour force grows at a constant rate n). At this point, the economy enters a steady state with an unchanging standard of living, (that is, a zero growth rate of per capita variables).

Hence, the Solow-Swan model has the properties that (1) in the long-run, output per capita converges to its steady state independently of initial conditions (non-hysteresis property); and (2) that the only potential sources of growth are sustained exogenous increases in primary factors (e.g. population growth), and exogenously given technological change (Solow, 1956 and 1957). In the latter, when there is no technical change the Solow-Swan model leads to stagnation of per capita growth, which contradicts the first two stylised facts enounced by Kaldor (1961). In addition, the long run growth rate is unaffected by the rate of savings (or investment) – an increase of the saving rate has only a level effect (the steady state value of capital per worker grows) and not a growth effect. Thus, growth is "exogenous" in the sense that the behaviour of economic agents does not alter the steady-state growth rate of the economy. With the explicit assumption of exogenous technology (and an implicit assumption of full appropriability of

---

1. It must be underlined that much of the characteristic flavour of the neoclassical model derives from the role of labour as a primary factor of production, (that is, an essential factor of production, not producible within the model. It should also be noted that the neoclassical model can easily accommodate increasing returns to scale, as long as there are diminishing returns to capital and labour separately. In this case, a special functional form is needed to generate the possibility of steady-state growth.

2. With an exogenous rate of technical progress, capital per worker grows at constant positive rate given by this trend and the level of capital stock grows at a constant positive rate equal to the sum of the exogenous trend of technical progress and the growth rate of population. These trends in the capital per worker and the stock of capital are inherited by output, consumption and investment. Accordingly, with technological progress, the rate of growth of employment in efficiency terms replaces the rate of growth in natural units.

3. Kaldor (1961) listed six stylised facts: (i) output per worker shows continuing growth « with no tendency for a falling rate of growth of productivity » (ii) capital per worker also shows continuing growth « with no tendency for a falling rate of growth of productivity »; (iii) the rate of return on capital is steady; (iv) the ratio of capital to output is steady; (v) labour and capital shares are constant and (vi) there are wide differences in the rate of growth of productivity across countries. These stylised facts have been extended by Romer (1989, 1994) and Easterly and Levine (2000).
investment, more on this later), the neoclassical growth model delivers an unequivocal answer: the government needs to do nothing to promote accumulation and growth – the equilibrium growth path is socially efficient since resources of the economy are efficiently utilised.

The neoclassical model has been criticised in three distinct areas:

1. As observed by Solow (1970), the neoclassical growth model with technical progress can explain five of the six stylised facts described by Kaldor. The sixth, – concerning observed income differences between rich and poor countries or the wide dispersion in growth rates across countries – is something a problem for that model. Moreover, the Solow-Swan model predicts that poor countries grow faster than rich countries (absolute convergence). This convergence is not supported by available empirical evidence – at least not for non-homogenous group of countries or regions (Temple, 1999). Some variants of the neoclassical model relax the absolute convergence hypothesis and give rise to the concept of “conditional convergence”, that is, each country converges to its own steady state. For this latter case, studies have found empirical evidence of conditional convergence at a ratio of about two per cent per year (Barro, 1991, 1997; Levine and Renelt, 1992). But calibration of the Solow-Swan model leads to implausible rates of convergence to the steady-state. Although the Solow-Swan model does predict the conditional convergence found in the empirical literature on economic growth, it does not predict the rate of convergence that these studies estimated. A third empirical critique of the Solow-Swan model emphasises the predicted difference in rates of return. The real interest rates must be larger in developing countries than in developed countries as it falls when the economy converges to the steady state. In contrast, empirical evidence suggests that we observe higher profit rates and higher real interest rates in “rich” countries than in “poor” countries. In this respect, Romer (1994) concludes that the neoclassical growth model with decreasing returns to capital, perfect competition and exogenous technology can not fully explain the cross-country variation in per capita incomes and national growth rates. As we will see later, this conclusion has been challenged by Mankiw et al. (1992), Mankiw (1995), Jones (1995a). In addition, to explain long-run differences in performances of countries, one needs to explain the importance of the total factor productivity (Solow’s residual) and to identify other sources of growth (Barro and Lee, 1994; Stokey, 1994 and Barro, 1997).

2. On the other hand, as pointed out by Barro and Sala-I-Martin (1995) and others (Jones and Manuelli, 1990; Romer 1986, 1989; Sala-I-Martin, 1991a and 1991b), we end up with a model of growth that explains everything but long run growth. In particular, it seems reasonable that the exogenous productivity changes are in fact, the result of decisions on the part of economic agents. As pointed out by Solow (1994), “no one could even have intended to deny that technological progress is at least partially endogenous to the economy”.6

---

4 In the case of homogenous group of countries, Dowrick and Nguyen (1989) show that convergence takes place among OECD countries. Barro and Sala-I-Martin (1991a, 1991b, 1995) obtain similar results for the US States, the regions of Europe and Japan. For further details, see Temple (1999), Bassanini et al. (2001), Bernanke and Gürkaynak (2001).

5 It should be noted that the capital share plays a key role in each of these problems. In particular, if we assume that the capital share is not one-third but rather two-third in rich countries, the Solow-Swan model could explain variations in income of the magnitude observed. This remark is further developed when we study the empirical relevance of endogenous growth models.

6 As is noted by Solow (1999), most neoclassical growth theory has treated it as exogenous. Some authors, e.g. Fellner (1961) and Von Weizsäcker (1966) have discussed the possibility that internal economic factors might influence the factor-saving bias of innovations.
3. Finally, the neoclassical model states that the equilibrium growth path is socially efficient and policies do not matter. The recent contributions to growth theory cast doubt on this conclusion. For example, when growth is driven by endogenous innovations, some obstacles stand in the way of market efficiency.

2.2 The foundations of endogenous growth models

Earlier authors considered the possibility of endogenous growth theory and recognised that stagnating per capita incomes need not be an inevitable outcome. For example, as noted by Solow (1970), an economy could grow indefinitely if the marginal product of capital remained above a certain level. In such a case, the theory continues to predict capital accumulation as the driving force behind economic growth. But, for the marginal product of capital per worker to be bounded from below, it would have to remain non-zero, even as the capital per worker grows unboundedly.\(^7\) However, this contradicts one of the assumptions of the Solow-Swan model (Inada’s conditions), which explains why it was abandoned. The “first” generation of endogenous growth models went back to this root and developed a more elaborate synthesis base on it. Solow (1990) also showed that a balanced growth path exists in the presence of increasing returns to scale. In this case, the rate of growth depends on the growth rate of the labour force and the technological parameters, as for instance the capital share in the economy. The implication is that a positive long run growth path is feasible that is still exogenous; that is, increasing returns are not necessarily the foundations of endogenous growth models (Solow, 1994, 2001). In contrast, the roots of endogenous growth models lie in the nature of knowledge. The rest of this section explains this statement.

The neoclassical model of growth succinctly framed the notion that the long run growth rate of income per capita was determined by the growth rate of technology. The problem was the difficulty in representing technological progress within a model in which perfectly-competitive price-taking firms had access to production functions with constant returns to scale in capital and labour. From a theoretical point of view, the argument goes as follows. The Solow-Swan model treats technology as a pure public good, that is, technology is non-rival and knowledge is non-excludable. Since the technology is non-rival, a replication argument (Romer, 1986, 1987, 1990a) suggests that a firm should be able to double its size by simply replicating itself and by using the same technology. This means that the concept of constant returns to scale should only apply to capital and labour. This implies, in turn, that perfectly competitive neoclassical firms will pay rental prices at the level of their marginal products. However, once the firm has paid for its inputs, there is no output left. Hence, the firm cannot devote resources to improving technology and there is no incentive to undertake research and development (R&D). The only possibility is thus to model an exogenous technological progress. In this respect, the question of interest is how to make technological progress endogenous. To do this, one has to take into account the nature of knowledge or technical progress.

As explained by Romer (1994), the evidence that economists have long taken for granted and that poses challenge for researchers in growth theory can be summarised in five basic facts: (i) there are many firms in a market economy; (ii) technological discoveries differ from other inputs, in the sense that many people can use them at the same time; (iii) it is possible to replicate physical activities; (iv) technological advance comes from things that people do; and (v) many individuals and firms have market power and earn monopoly rents on discoveries. The neoclassical model captured facts (i), (ii) and (iii), but postponed considerations of facts (iv) and (v). These limitations come from the public character of technology, that is, a non-rival and non-excludable good. In the latter case, non-excludability is inconsistent with the existence of rents in the activity of R&D. Overall, fact (v) involves that technology is partially excludable. In effect, the main physical characteristic of technology is that it is a non-rival good – the same formula and the same blueprint may be used by many agents simultaneously. In contrast, factors of production, as for

\(^7\) This means that the one of the Inada’s conditions is not fulfilled.
instance labour and capital, are "rival". The main difference being that non-rival goods involve an important fixed cost of production for the first unit and a subsequent small marginal cost for additional units. As noted by Romer (1983, 1986), this concept should be distinguished from that of non-excludability. This distinction between rivalry and excludability is important because non-rivalry is inextricably linked to non-convexities whereas non-excludability is not (Romer, 1990a). These non-convexities give rise to increasing returns to scale and thus to imperfect competition. This result stems from the fact that the average cost is always larger than the marginal cost in the presence of increasing returns – the equilibrium condition between prices and marginal costs engenders a negative profit in the case of perfect competition. Hence, if the firm cannot fix a price above the marginal cost, there is no incentive to support the fixed cost associated with non-rival goods (e.g. ideas). The prospect of a positive profit is a necessary condition in order to produce non-rival goods. In other words, one needs to abandon the perfect competition environment. To make technological progress endogenous, one has to model fact (iv), fact (v), or both. These are the two new alternatives offered by the new growth literature. On the one hand, in order to describe fact (iv), one does not need to abandon the perfect competitive environment. On the other hand, in order to describe fact (v), one needs to abandon the assumption of price-taking competition (due to non-convexities in the production function).

The hypothesis of perfectly competitive price-taking firms can be maintained by invoking external economies. This is the second approach of the first class of endogenous growth models. Following the seminal contributions of Marshall (1890) and Arrow (1962), these models assume that each firm’s investment in either physical capital (Romer, 1986) or human capital (Lucas, 1988) leads to technological progress in the form of "learning-by-doing". Growth can be sustained by continuing accumulation of the inputs that generate positive externalities or “spillovers”. When agents invest in human capital or physical capital, increasing returns to scale allow an increase of income and thus a growth of aggregate savings in the form of capital. The growth process is then unbounded, it depends of the saving behaviour of economic agents (e.g. is endogenous). In both of these models, the technology is endogenously provided as a side effect of private investment. However, technology is still treated as a pure public good. As a result, firms can be considered as price takers and an equilibrium with many firms can exist. In this respect, externalities are assumed to be external to the firm in order to support the equilibrium with a set of

---

8 A good is excludable if its utilisation can be prevented.

9 It should be noted that the extent of rivalry is entirely determined by the technology, whereas both the technology and the legal institutions determine the excludability.

10 Romer (1983, 1986) formalised the interaction between ideas and growth as follows: Ideas $\rightarrow$ non rivalry $\rightarrow$ increasing returns $\rightarrow$ imperfect competition.

11 An important comment is the following: the modelling of imperfect competition is made necessary by the appearance of increasing returns to scale, but the presence of increasing returns to scale is not the essence of these models. What is essential is the assumption of constant returns to capital at the aggregate level (see paragraph 12).

12 Lucas (1988, 1993) develops different versions of an endogenous growth model with human capital accumulation. This accumulation can be either intentional or non-intentional.

13 If we compare the nature of knowledge in the neoclassical approach and this class of endogenous growth models, then we will obtain the following distinction. The neoclassical model only considers the nonrival aspects of knowledge, that is, knowledge is not privately provided. In contrast, in Arrow (1962), the nonrival knowledge is privately provided by a non-intentional result of other activities. In the model of Romer (1986), knowledge is partly excludable and rival, whereas knowledge is partly nonexcludable and non-rival in Lucas-model. These two models allow for positive spillovers, that is, effects relating to the excludability of knowledge. But there is no conflict between nonrivalry and the private creation of a good in a competitive market.
competitive prices, so that they are constant return to scale at the firm level, but increasing returns to scale at the aggregate level.

In order to model fact (v), one has to abandon the idea of preserving perfect competition. The acquisition of knowledge is costly and its use is partly excludable. Most technological progress requires, at some stage, an intentional investment of resources by profit-seeking firms or entrepreneurs. The partial excludability of knowledge or ideas requires that its private counterpart must be protected, as for instance with patents. Patents are supposed to create monopolies and are necessary in order to engage in R&D activities. In this respect, Romer (1990b), Grossman and Helpman (1991) and Aghion and Howitt (1992, 1998) explicitly formalise the R&D sector and the growth process is based on the discoveries of new ideas. The production of knowledge consists of implementing a technological innovation of various forms: (i) new intermediate goods (Romer, 1990b), (ii) new consumption goods (Grossman and Helpman, 1991, 1995) and (iii) new processes of production (Aghion and Howitt, 1998). These inventions contribute to augmenting the existing stock of patents. But, at the same time, the stock of patents constitutes a positive externality: researchers can use the existing stock of knowledge or ideas to develop new innovations. This means that the variation of ideas are based on the present stock of available knowledge but their production is protected. However, as pointed out by Aghion and Howitt (1992), the monopolistic rent could be not perpetuated. This permits modelling of the idea of Schumpeter (1934) that access to a new technology confers at least temporary monopoly power on the first user. The appearance of a new technology makes earlier technology obsolete: the “creative destruction” process of Schumpeter. Thus, innovation-based endogenous growth models are based on three different mechanisms. Built on the Dixit and Stiglitz (1977) and Ethier (1982) formulations of horizontal product differentiation in industrial economics, R&D leads to the development of new varieties of the differentiated product (Romer, 1990b), or the final consumption product. Other specifications assign R&D the role of product improvement, “quality ladders” (Grossman and Helpman, 1991 and 1995), or cost reduction (Aghion and Howitt, 1992, 1998).

R&D based growth models have been criticised both at the theoretical and empirical levels, on the grounds of the existence of scale effects and the special functional forms used. These considerations led to the emergence of the “semi-endogenous” growth literature (Jones, 1995, 1998, 2000; Segestrom, 1998). These latter models are in many aspects a hybrid form of the endogenous and neoclassical growth models. In effect, technology is endogenous and the outcome of agents’ optimising behaviour, (as in Romer-model), but the long-run growth rate is determined as in the Solow-Swan model.

2.3 A common representation of endogenous growth theory for CGE models

These developments in the new growth literature have important implications for numerical models, especially computable general equilibrium models. Broadly speaking, a CGE model is a system of linear and non-linear equations, which is solved to simulate market equilibrium. Its structure includes equations describing consumer and producer supply and demand behaviour that are explicitly derived from conditions of utility or profit maximisation, and market clearing. Until recently, most CGE models assumed that technical progress was exogenous. This may be explained by many factors. First, some model solution techniques require the existence of a steady-state balanced growth path invariant over time. Thus, if the steady-state path is pegged out by the rate of technological progress, then algorithms, which solve forward for decades, can be perfectly determined. In this case, one has only to use the exogenous growth rate to determine the baseline dynamics of each state variable. Second, while the new growth literature has made considerable advances in identifying the means by which growth occurs, the empirical literature has not yet been able to clearly distinguish between competing models – a survey of the empirical evidence

14 For simplicity this paper will limit its treatment of CGE model to those that assume representative agents who optimise over a long time period. We do not discuss “overlapping generations” CGE models which feature agents with limited-duration lifespans.
shows mixed results (Temple, 1999). Third, even if one takes the new growth literature as a given, the introduction of endogenous technical change into CGE models gives rise to a number of challenges, methodological and implementation issues. Despite the predominance of exogenous technical change in CGE models, there is a widespread consensus that an important direction for present and future research in CGE models involves the operationalisation of endogenous growth.

In order to introduce endogenous technical progress in CGE models, it might be useful to characterise the different generations of endogenous growth models in a comparable setting, and to discuss the relevance of both the underlying hypotheses and their empirical support. The different generations of endogenous growth models can be characterised in terms of (i) the assumptions they make about preferences, the technology, and the equilibrium concept; (ii) policy implications; and (iii) the mechanisms that generate the “endogenous growth rate”.

- General assumptions (i) on the equilibrium concept for the models described here are either a complete markets competitive equilibrium with or without externalities or a Nash equilibrium with both externalities and imperfect competition. In terms of technology, all models assume the existence of an aggregate production function that depends on a set of inputs and a measure of technology or knowledge. In terms of preferences, all models described here use the standard discounted Ramsey preferences.

- Relating to policy implications (ii), some models of endogenous growth preserve the Pareto-optimality and reject the intervention of the government while models with externalities and innovations do not. In the first case, this only allows to use static comparison in CGE models, that is, one may compare the state of the economy after a shift of some exogenous parameters of the model. At the same time, it has to be noted, however, that macro policies have an impact on the long-run growth path of the economy because of the dependence to the saving rate. For example, any policy, which affects the saving rate, will alter the growth rate of the economy (Rebelo, 1991). In the second case, one may directly analyse the implications of different policies in terms of welfare, equity, etc. This is particularly relevant when one would like to examine the consequences of R&D policies for the environment.

- In the case of mechanisms generating endogenous growth (iii), as explained by Romer (1994), Jones (1995) and Barro (1998), endogenous growth models can be summarised by a set of two equations, the production function for final output and the production function for ideas. Let $A$ denote the stock of ideas in the economy. New ideas, the change in $A$, are produced by:

$$\dot{A} = \frac{dA}{dt} = G(A, R)$$

(1)

where $R$ represents resources devoted to research and the function $G$ exhibits some properties to be described latter. The consumption/ final output is produced using:

$$Y = F(X, A)$$

(2)

where $F$ is increasing in each of its arguments and can exhibit either increasing returns or constant returns to scale and $X$ is a set of inputs, (e.g. capital, raw labour, or non-renewable resources).
New growth models differ primarily either in the way they specify the F and G functions, or the way they label the A variable. In the former, models tell different stories in order to complete the two functions. In the latter, it is possible to construct a variety of models with different “engines” for growth, ranging from physical capital accumulation to human capital accumulation and/or the discovery of new ideas.

If we separate the different models according to the source of growth: (i) accumulation of physical capital, (ii) accumulation of human capital and (iii) the creation of new ideas, the different models of endogenous growth theory may be characterised as follows, in the light of equations (1) and (2).

- In the case of accumulation of physical capital (i), the “first generation” of endogenous growth models (Rebelo, 1991; Jones and Manuelli, 1990) are only concerned with the production function. The knowledge function is not taken into account, i.e. the result of endogenous growth rate is linked with some properties of the production function. The “second generation” of endogenous growth model (Romer, 1986) introduces an externality into the production function and does not model the knowledge function. This externality can be the average level of physical or human capital per capita (Romer, 1986; Lucas, 1988), the average level of public expenditures per capita (Barro, 1991), etc. The key point is that the externality is related to another accumulable factor in the model. The property of the endogenous rate is, however, still linked to some properties of the production function.

- In the case of accumulation of human capital (ii), models can be divided into two broad categories. Indeed, models with non intentional accumulation of human capital are only concerned with the production function, whereas models with intentional accumulation use a knowledge function. In the latter, the A variable is thus the human capital stock and the endogenous outcome of the economy is the result of both equations (1) and (2).

- In the case of the creation of new ideas, both equations (1) and (2) are present in the models. In the expanding varieties based model (Romer, 1990), the A variable in equations (1) and (2) is the number of intermediate goods in the economy. In the quality ladder based model (Grossman and Helpman, 1991), the A variable is an index of the quality of intermediate goods. In the cost reduction based model (Aghion and Howitt, 1992), the A variable is an index of the existing intermediate goods. In all these models, the property of endogenous growth comes from both the production and knowledge function.

---

15 This classification is standard in the endogenous growth literature. It should be noted that the growth rate of the economy is driven by the accumulation of both human and physical capital as well as R&D in some endogenous growth models. For the sake of simplicity, we use this presentation but the different sources of growth can be linked together.
Box 1. An introduction to computable general equilibrium models

- General characteristics of a CGE model

Following the terminology of Ginsburgh and Kuzzer (1998), computable general equilibrium models are a special case of the literature of applied general equilibrium models. A CGE model is a system of linear and non-linear equations that is solved to simulate market equilibrium. CGE models usually, though not always, have the following characteristics: (i) multiple interacting agents, (ii) individual behaviour derived from optimisation, (iii) multiple markets, (iv) multi-sector economy, (v) a decentralised equilibrium rather than optimising a planner’s objective function and (vi) designed for policy analysis. In the latter, CGE models have often been used in the studies of trade, environment and fiscal policies.

- Basic representation of a CGE model

We present here the different steps to build a simple CGE model in the case of a static and perfect competitive environment.

The first step is to determine the overall structure of the economy, that is, the number of agents and type of agents (regions, households, firms, government), the level of disaggregation (the number of sectors), the agents’ interactions as well as the links between sectors, and the set of exogenous and endogenous variables.

The second step usually specifies each agent’s optimisation problem and solves the behavioural equations. This requires choosing appropriate mathematical specification of production and consumption functions. Different functional forms are available in the literature. For example, the consumption function can be written as a (generalised) Leontief, a Cobb-Douglas, a constant elasticity of substitution (henceforth CES), transcendental logarithm indirect utility, etc. These functions have different properties, and one has to be careful of their individual implications.

The third step requires adding market-clearing conditions and accounting equations. Because of Walras’ Law, one market-clearing condition can be dropped and a price has to be chosen as a numeraire.

The fourth step involves parameterisation of the model. Several approaches have been used for parameter estimation in CGE models: the micro-study, the calibration method, the econometric method and semi-econometric method. Micro-econometric method in parameter estimates are usually partial-equilibrium based. The calibration method focuses on the selection of a particular functional form. The parameters to be calibrated are determined by forcing the model to replicate the data of a chosen benchmark year. The latter is summarised in a social accounting matrix in which all the transactions of agents are taken into account. While the calibration method requires limited data, its lack of statistical validity has been widely criticised. In addition, calibration methods alone cannot determine all parameter values when certain functional forms are employed. In these cases, econometric or semi-econometric methods are used in order to complete the set of initial parameters. This step also involves the choice of base-case values of exogenous variables.

The last step consists of writing and implementing a program within an appropriate software package and solving for the base-case or baseline equilibrium.
- Policy implementation in a CGE model

CGE models are specifically designed for policy analysis. In this case, one has to decide how to express the policy in terms of the variables, then run the simulation and compare the results to the base case. In order to evaluate the policy, the equivalent variation criterion is usually used as a measure of welfare.

- Recent developments of CGE models

The first generation of CGE models had two important shortcomings. On the one hand, delays and the adjustment costs associated with macro policies were not taken into account due to the static time dimension. In particular, this led to the non-differentiation of short-term and long-term effects of policies. The Walrasian set-up was also questioned. By contrast, the second generation of CGE models incorporates market imperfections due to either rigidities on prices, wages, or imperfect competition. In the first case, prices can be fixed by a mark-up rule and the adjustment operates by the quantities on good markets (Erlich et al., 1987; Mercenier and S. de Souza, 1994; Yeldan, 1997). These approaches have often been criticised for rigidities that are exogenous and postulated, rather than being derived from explicit micro-foundations (Schubert, 1993). Thus, a large number of studies have introduced either increasing returns to scale at the firm level or constant returns to scale production function associated with fixed costs. This allows the modelling of monopolistic or oligopolistic competition on the supply side, and to specify horizontal or vertical differentiation on the demand side. Harris (1984) was the first to apply this framework in a general equilibrium trade model. The result was that trade liberalisation was lead to a larger competition among producers, reduce mark-ups, augment the production volume and shows increase welfare. Among others, Burniaux and Waelbroeck (1992), Cox and Harris (1984), Devarajan and Rodrik (1991), Gasiorek et al. (1992), Mercenier (1994), Mercenier and Schmitt (1997) have developed general equilibrium trade models with imperfect competition. However, it has to be noted that the introduction of imperfect competition leads to additional technical difficulties. Mercenier (1995) shows that this static class of CGE models with imperfect competition leads to multiple equilibria. Both the first and second generations of CGE models produced comparative static analyses. In order to take into account adjustment delays or adjustment costs, one had to introduce a dynamic temporal dimension into CGE models. This is the third route pursued by CGE modellers. Two types of dynamics, which imply different equilibrium concepts and anticipations hypothesis, have been used in the literature. On the one hand, a first type of CGE models is characterised by a “sequential dynamics” – the economy is described as a sequence of temporary equilibria. The inter-periodic dynamics is then obtained by specifying either stock variations (capital stock or financial assets) or exogenous changes (Bourguignon et al., 1992; Burniaux et al., 1992). On the other hand, a recent type of CGE model introduces an intertemporal dynamics with forward-looking behaviour. Thus, consumers maximise their intertemporal utility function (over an infinite horizon) under budget constraints, whereas producers maximise their intertemporal profits under technical constraints. To capture the intertemporal dynamics on the supply side of the economy, two different behaviours have been taken into account. The first derives (or posits) an intertemporal investment equation (similar to the q-theory of Tobin, 1969) from the firm’s optimisation problem. These CGE models are dynamic neoclassical growth model in which technical progress is exogenous (McKibbin and Wilcoxen, 2000, Bismans and Bouzahzah, 2000; Berthélemy and Bourguignon, 1993; Diao and Somwaru, 2001; Devereux and Lee (1999), Wendner, 2001). The second relies on endogenous growth process, which is based on the recent new growth literature.
3. **AN EMPIRICAL AND THEORETICAL ASSESSMENT OF ENDOGENOUS GROWTH MODELS FOR CGE MODELS**

3.1. **The accumulation of capital stock**

3.1.1 **Linear growth models**

The linear growth or convex models of growth have been extensively studied by Jones and Manuelli (1990, 1997). We outline here the properties of this simplest class of endogenous growth models. As mentioned before, one controversial assumption of the Solow-Swan model is that of diminishing returns to capital, capital per worker is less and less productive, and the economy gradually converges to a steady state where saving provide investment only sufficient to maintain constant the capital per worker, without technological progress. Instead of postulating a decreasing marginal product of capital, these linear endogenous growth models try to find some plausible assumptions that would counteract diminishing returns to scale to the class of productive inputs that can be accumulated by some form of saving and investment.16 Often, they are based on straightforward generalisations of the Solow-Swan model and share the feature to introduce a basic aggregate technology as a simple convex specification.17 Therefore, this first generation of models only focuses on the production function (2) and ignores the explicit production of knowledge (1).18 For the sake of simplicity, this first class of models can be separated into two broad categories: (i) models where the marginal product of capital remains bounded and (ii) models that add mechanisms that will make the investment-output ratio constant. In others words, models can be distinguished according to their assumption on returns to capital and the dualism of accumulable and non-accumulable factors.

One strand of this literature, (e.g. Jones and Manuelli (1990)), allows for both accumulable and non-accumulable factors, and assumes a convex technology where the marginal product of capital is a decreasing function of its stock but does not vanish as the amount of capital per worker goes to infinity. Hence, returns to capital are bounded from below. Let us assume now that such a function exists and verifies the previous property.19 The growth process can be described by the following mechanism: as capital accumulates and the capital-labour ratio rises, the output-capital ratio (e.g. the average productivity)

---

16 As explained by Jones and Manuelli (1990, 1997), there is another way to understand the importance of the diminishing returns to scale of capital. Let us consider the standard Cass-Kosmams-Ramsey of optimal growth, that is, the saving decision of a representative agent with preference given by the discounted present value of instantaneous utility. Without technical change, the "Keynes-Ramsey" condition, involves $g, = \sigma(c)(r - \rho)$ where $\sigma(c), r, \rho$ denote the intertemporal elasticity of substitution, the real interest rate and the preference rate for present, or inverse of the discount factor, respectively. If the interest rate equals the inverse of the discount factor, the individual chooses a smooth pattern of consumption over time. Then aggregate consumption is stable over time and there is no growth. If the interest rate exceeds the discount rate ($r > \rho$), it will cause the time profile of consumption to be upward sloping. Thus a necessary condition for unbounded growth is to prevent the interest rate from falling to the level of the inverse of the discount factor. The standard arbitrage condition involves that the interest rate equals the marginal product of capital net of depreciation. This implies that the Solow-Swan model can not deliver long run growth because, as the stock of capital per worker increases, its marginal product is assumed to decrease without bound and, in turn, there exists an amount of the capital per worker which equals the inverse of the discount rate.


18 See Box 2.

19 The « Sobelow » production function, $F(K,L)=BK + g(K,L)$, where $g$ is a standard constant to scale Cobb-Douglas function, or a CES form when the elasticity of substitution between capital and labour is greater than one, are decreasing functions of capital where the marginal product is bounded from below.
will fall, approaching asymptotically a positive constant. Therefore, as agents continue to accumulate capital, growth continues over time. The immediate consequence is that now, asymptotically, the steady-state growth rate depends proportionally on the fraction of income saved and invested (e.g. is endogenous) and is unbounded, whereas it was independent of the investment rate in the Solow-Swan model. Anything that can affect the investment rate, (e.g. taxation), can permanently affect the rate of growth. In addition, both resources are allocated efficiently, the two fundamental welfare theorem hold. It is fair to say, however, that this “asymptotically endogenous” growth rate is less than an ideal solution. Most importantly this is because the property of endogenous growth results directly from the linear ad-hoc part of the production function.

A second strand of this literature adds mechanisms that will make investment proportional to output. If one supposes that the variation of the capital stock is simply proportional to the existing stock of capital and that production does not require labour, investment will be proportional to the stock of capital and the factor of proportionality is the output-capital ratio. In order to obtain a one-for-one relationship between output and capital, returns to capital have to be assumed constant. The simplest growth model using this type of production function is the contribution of Rebelo (1991), where the production function is given by \( Y = AK \). This is known as the “AK” model. Thus, replacing the hypothesis of diminishing returns to capital by constant returns to scale transforms what are "level" effects in the neoclassical growth model into "growth" effects. Shifts in the willingness to save or in the level of technology permanently affect the growth rate of the economy. On the other hand, a greater willingness to save or an improvement in the level of the technology only have a temporary effect and only alter the steady state levels of capital, output, consumption and investment per worker in the neoclassical model. This simple formulation of the production function also leads to the absence of transitional dynamics: the growth rate of the economy depends on initial conditions (hysteresis property). In other words, these models have the characteristic that consumption, capital and output are in their respective balanced growth paths. If the economy is perturbed from its balanced growth path, it will return instantaneously to its new balanced growth path. Consequently, this type of model suggests that the convergence does not take place: rich countries stay rich and poor countries stay poor. However, the existence of a transitional dynamics can be achieved in several ways: (i) by adding a stochastic process, (ii) by assigning a central role to a second state variable and (iii) by considering multi-sector models.\(^{20}\)

The “AK” specification has been criticised in several aspects. First, this model is based on a “knife-edge” restriction, that is, constant returns to scale in the unique accumulable factor. In effect, it is unlikely that the share of capital in the economy will be equal to one. Second, as pointed out by Solow (2000), the absence of labour as a factor of production is unrealistic. This defect can however be softened by taking a broader view of capital. This model can be extended along a variety of dimensions. In effect, even if the usual interpretation of “K” corresponds to physical capital, it is only one possible interpretation. All that is necessary is that there exist a fully reproducible productive input whose marginal productivity is bounded away as its level is increased without bound. As a first alternative, one may interpret “K” as the level of knowledge or the stock of human capital. Alternatively, “K” can be interpreted as being an aggregate of human and capital stock, where the ratio of human to physical capital is constant over time (Jones and Manuelli, 1990; Barro and Sala-I-Martin, 1995). “K” can also be interpreted as being the quality of the

\(^{20}\) To go beyond the aggregate character of the Solow-Swan model, one may allow for distinct sectors producing with their own technology, consumer goods, and investment goods. The initial endowments of the economy (capital, labour) are thus allocated between the sectors and a full set of prices is determined to equilibrate the markets. This literature begins with Uzawa (1961, 1963), Burmeister and Dobell (1970), and, recently, Jones and Manuelli (1991, 1994) and Rebelo (1991) propose n-sector growth models. The steady-state properties of the multi-sector models turn out to be very different from those of the one-sector model. In particular, assuming different factor intensities in each sector, there is convergence towards a steady-state under some conditions.
capital good used in production. This simple form is then sufficient to capture notions of quality improvements (Stockey, 1991; Grossman and Helpman, 1991a). However, in all these extensions, the part of knowledge controlled by firms is an ordinary input in production. Knowledge is non-rival and is not associated with increasing returns. Third, empirical evidence does not support the evidence of constant returns to scale to capital. Thus, both theoretical and empirical considerations seem to indicate the limited relevance of this class of models for CGE models.

3.1.2. Non intentional accumulation of knowledge: endogenous growth models with externalities

Within the “first generation” of endogenous growth models, recent papers emphasise the role of learning or knowledge as a positive externality of other economic activities. Romer (1986, 1987) follows Arrow’s formulation and assumes that new investments in physical capital lead to technology improvements in the form of “learning-by-doing”. These externalities arise because of spillovers of knowledge (Arrow, 1962). This is assumed to be external to the firm, so that the production function exhibits constant returns to scale at the firm level, but increasing returns to scale at the aggregate level – implying a decentralised equilibrium with externalities. Thus, the assumption of perfect competition can be retained in the presence of increasing returns to scale. This idea was first espoused by Marshall in his Principles of Economics: externalities contribute to reconciling increasing returns to scale and perfect competition among firms. Equilibrium is possible because the physical capital, K, and the physical labour, L, are the only factors that receive explicit compensation; knowledge is more like a public good. The different nature of returns to scale at the private and aggregate levels implies that the decentralised equilibrium is not Pareto optimal. The social planner will take into account the externality (and thus the social return to capital); individual agents only take care of the private return to capital. This, in turn, justifies the use of policies in order to restore optimality.

As argued by Arrow (1962), increasing returns arise because new knowledge is discovered as investment and production take place. The increasing returns are external to individual firms because most knowledge becomes publicly-known. However, Arrow (1962) restricts attention to the case where the aggregate elasticity of output with respect to capital and knowledge is less than one. This latter assumption implies the existence of a steady state of the economy, and population growth becomes the only source of growth in the model. In contrast, Romer (1986) develops a fully-specified model of long-run growth in which knowledge is assumed to be an input in production that has increasing marginal productivity. As explained before, the crucial point is the departure from the usual assumption of diminishing returns. In this model, knowledge is assumed to be the product of a research technology that exhibits diminishing returns. It should be noted that the research technology is not explicitly modelled, and knowledge is simply viewed as the cumulative sum of individual firm’s knowledge, either in capital per worker or the stock of capital. Therefore, the creation of new knowledge by one firm is assumed to have a positive external effect on the production possibilities of other firms, because knowledge becomes public and cannot be perfectly patented or kept secret. Most importantly, the production of the consumption good is a function of the stock of knowledge and other inputs, (e.g. private capital or raw labour), and exhibits constant returns to scale. More precisely, knowledge may have an increasing marginal product. (Knowledge will grow without bound).

In this respect, the resolution of the model leads to the following conclusions: (i) when social returns to capital are constant, the exogenous parameters of the economy. (e.g. the intertemporal discount rate) have an impact on the long run growth rate; (ii) given the difference between the social and marginal product, each firm would choose to acquire less than the socially optimal amount of knowledge, and the decentralised long run growth rate is below the centralised long run growth rate chosen by the social planer; (iii) the long run growth rate is entirely based on the presence of externalities, that is, the non-
remuneration of knowledge; and (iv) convergence does not take place because the beneficial external effects of capital accumulation outweigh the detrimental consequences of increasing capital per worker. In addition, the difference between the social and decentralised long run growth rate is one justification of an active role for macroeconomic policies.

This model can be criticised in several aspects. **First**, in effect, the never-ending growth process requires a never-ending accumulation of knowledge. As pointed out by Young (1993), the assumption that the potential productivity gains from learning are essentially unbounded is far from being unquestionable. When new technical progress first occurs, rapid learning takes place as the productive potential of that process is explored. But this learning will approach an inherent limit, and will eventually stop after some time. This bound on learning has frequently been introduced in partial equilibrium analysis (Spence, 1981; Fudenberg and Tirole, 1983 and Stokey, 1986). **Second**, the existence of balanced growth path also requires a “knife-edge” restriction, the sum of the elasticity of production to the externality and the elasticity of production to private capital must be equal to one (Box 2). If this is not the case, we have two possibilities: (i) an elasticity greater than one implies that the growth rate will still be endogenous but continually increase; (ii) an elasticity less than one implies that the growth rate will be exogenous. In the latter case, output is no longer proportional to the individual stock of capital, and the growth process will be constrained by decreasing factorial returns. Together, these conditions imply that the aggregate production function must be similar to the “AK” production function.

**Third**, it should be noted that the specific form of the externality has an important counterpart. For example, if one assumes that the externality is simply the sum of the individual stocks of capital (e.g. the total stock of capital), then the growth rate will depend on a scale effect (Box 2): An increase in the number of firms will generate more and more knowledge and will augment the rate of growth. Consequently, there is a positive correlation between the scale of the economy, measured by the total population, and the long run growth rate. Hence, these scale effects provide an important mechanism linking growth to trade. In models with learning-by-doing, opening to trade leads to an increase of the size the markets for producers, and thus to greater specialisation and a higher average scale of production. However, it follows that the growth rate of population cannot continually grow – the dynamics of the economy would explode because of the presence of the scale effect. At the opposite end, if one supposes that the externality is the average capital per worker in the economy, the scale effect does not matter. This scale effect is of particular concern when we study innovation-based endogenous growth models. **Fourth**, as explained by Dasgupta and Stiglitz (1984), the assumption of a competitive economy is only valid if firms can learn costlessly, completely, and instantaneously from the experience of others (i.e. if learning spillovers are complete). This extreme assumption is not supported by empirical evidence. These limits cast doubts on the use of such a framework in dynamic CGE models. **Fifth**, these models introduce a second knife-edge condition between elastic and inelastic labour supply. Typically, labor supply is assumed to be inelastic. It can be shown that the structure of the equilibrium changes when an elastic (i.e., endogenous labour supply is assumed). The reason is that the productivity of capital will positively depend upon the fraction of time devoted to labour. Furthermore, when workers are assumed to be endowed with a fixed unit of time, the steady-state allocation between labour and leisure must be constant. This arbitrage between leisure and

---

22 There is however a difference between the linear endogenous growth models and externality-based model in the sense that, in the former, we impose a linear specification of the production function or an asymptotic linear specification, whereas, in the latter, the linear specification is obtained as a reduced form after taking into account the existence of spillover effects.

23 This link between growth and trade is also present in other endogenous growth models, as for instance models with investment in human capital or with R&D.

24 Romer (1994) recognised himself that « when I look back on my work on growth, my greatest satisfaction comes from having rejected the first round of external effects models that I tried [...] ».
work has some important consequences when one studies a closed economy or a small open economy. When the labour supply is inelastic, both the marginal and the average productivity of capital are constant in a closed economy and the growth rate of consumption is equal to the growth rate of income or product. In a small open economy, equilibrium occurs where it can sustain differential growth rates of consumption and domestic output. In contrast, if the labour supply is elastic, the real variables will grow at a common rate, whether or not we consider either a closed economy or a small open economy. However, it should be noted that many CGE models do not incorporate the preference for leisure, whereas this preference has an important role in Real Business Cycles Theory (henceforth RBC). This may be justified by the fact that RBC models are much more concerned with short-run fluctuations.

Box 2. Production functions in models with accumulation of capital stock

We present here the different production functions in models with accumulation of capital stock. There is no knowledge function.

- The benchmark case: the Solow-Swan model

In order to compare the different models, we assume that the exogenous growth rate of the labour force and the technological progress are zero. The production function is a standard Cobb-Douglas production function:

$$Y_t = K_t^a L_t^{1-a}.$$  (1)

The production function exhibits constant returns to scale and diminishing factoral returns. In particular, the marginal productivity of the capital stock converges towards zero when the capital stock approaches infinity. In this respect, the growth rate of the economy is zero at the steady-state equilibrium.

- Linear endogenous growth models: Jones-Manuelli model

The production here is given by:

$$Y_t = B K_t + h(K_t, L_t)$$  (2)

where $h$ is a standard Cobb-Douglas production function.

The production function also exhibits constant returns to scale but the marginal productivity of capital approaches $B$ when the capital stock tends to infinity. In this case, the growth rate of the economy is asymptotically follows:

$$g_K = g_C = g_Y \rightarrow sB - \delta > 0 \quad \text{if and only if} \quad sB > \delta$$

where $s, \delta$ are the exogenous saving rate and the rate of depreciation of the capital stock, respectively.

The growth rate of the economy thus depends on structural parameters, e.g. endogenous. The Rebelo-model

The production here is given by:

$$Y_t = BK_t.$$  (3)

The production function also exhibits constant returns to scale and constant marginal productivity of capital, $B$. In this case, the growth rate of the economy is instantaneously given by:

25 In a closed economy, the growth rate is constrained by the domestic resources and thus all real variables would have to grow at the same rate.
The growth rate of the economy thus depends on structural parameters, e.g. endogenous.

- *Non-intentional accumulation of capital*

The Romer-model postulates the following production function:

\[ Y_t = K_t^\alpha L_t^{1-\alpha} B_t^\beta \]  

where \( B_t = K_t \) or \( \frac{K_t}{L_t} \) is an external effect.

At the aggregate level, the production function can be written as:

\[ Y_t = K_t^{\alpha + \beta} L_t^{1-\alpha} \]  
\[ Y_t = K_t^{\alpha + \beta} L_t^{1-\alpha - \beta} \text{ if } B_t = \frac{K_t}{L_t} \]

In the first case, the external effect is defined as the total stock of capital in the economy, whereas it is the average level of capital in the second case. To ensure the existence of a long-run growth path, one has to impose the knife-edge condition: \( \alpha + \beta = 1 \). Thus, the reduced form is an “AK” model (Rebelo, 1991) at the aggregate level, that is:

\[ Y_t = K_t L_t \]  
\[ Y_t = K_t \text{ if } B_t = \frac{K_t}{L_t} \]  

(5)

Thus, the growth rate of the economy is given by:

\[ g_K = g_L = g_Y \rightarrow sL_t^{1-\alpha} - \delta > 0 \text{ if } B_t = K_t \]

\[ g_K = g_L = g_Y \rightarrow s - \delta > 0 \text{ if } B_t = \frac{K_t}{L_t}. \]

The definition of the externality matters for the existence of a scale effect. In the first case, the larger the economy is, the greater is the growth rate of level variables. In the second case, the scale effect does not matter.

Finally, it has to be noted that the social and private marginal productivity are different. The former is derived from equation (5), while the latter comes from equation (4) (after imposing the knife edge condition). This causes the decentralised growth rate to be below the centralised growth rate (chosen by a social planner) and thus to use macro policies.

### 3.1.3 Empirical evidence

In order to test for empirical evidence for this class of endogenous growth models, one may compare the predictions of these models with the Solow-Swan model. In particular, empirical studies provide evidence about convergence, the elasticity of production to capital, the scale effect and the permanent or
transitory effect of policies. \(^{26}\) **First**, the literature concerning convergence tends to show that the neoclassical model generates better predictions than the “AK” model. \(^{27}\) However, this conclusion has been criticised in many aspects. **On the one hand**, the large empirical literature regarding convergence suggests that convergence rates are more variable and sensitive to time periods and to the set of countries than originally suggested by the works of Barro and Sala-I-Martin (1992a, 1995), Mankiw, Romer and Weil (1992), in which a benchmark estimate of the convergence rate is established at 2-3%. For example, Islam (1995) estimates a rate of convergence of about 4.7% (9.7%) for non-oil countries (OECD countries). By contrast, Temple (1998) estimates this rate to be between 0.3% and 6.7% (1.5% and 3.6%) for non-oil countries (OECD countries). Evans (1997) finds a convergence rate around 6% and Caselli et al. (1996) obtain a higher rate of convergence of around 10%. At the same time, Jones and Hall (1997), and Jones (1999) provide evidence about convergence in level income rather than the growth rate of income in a sample of OECD and developed countries. Finally, a number of studies point out that the convergence process may be due to other factors, as for instance technological transfers instead of the accumulation of capital stock. \(^{28}\) **On the other hand**, it is possible to build “AK” models with transitory dynamics. In this context, growth can become stochastic (Kelly, 1992, Kocherlakota and Yi, 1995). Kelly (1992) shows that evidence on convergence cannot be used as evidence against endogenous growth models, that is, convergence can be made consistent with some classes of endogenous growth models. In that paper, a stochastic specification of an endogenous growth model is used that is based on Long and Plosser (1983) and demonstrates economies converging towards each other even with stochastic constant returns to scale. Canova and Marcet (1995) and Den Haan (1995) investigate alternative specifications for the disturbance and point out the difficulty in interpreting the empirical findings from standard regression analysis in favour of the neoclassical model. Lueng and Quah (1996) generalise these results and show that some stochastic technological perturbations may engender a convergence process even if the production function exhibits stochastic constant returns to scale or stochastic increasing returns to scale. Finally, while the convergence hypothesis is often used to discriminate between endogenous and exogenous growth models, Durlauf and Johnson (1992), argue that the research cannot. In effect, the two models have a important different feature. In the case of the Solow-Swan model, the convergence process does not depend on initial conditions, that is, the economy moves towards the steady-state whatever the initial endowment of capital per worker (non hysteresis property). In the case of the “AK” model, the long-run growth rate depends on initial conditions: (hysteresis property). Hence, a direct comparison of both models needs to take this difference into account. In this respect, Bernard and Durlauf (1996) show that cross-section regressions can reject the null of no-convergence too often when the data are generated by a specific new growth model. **Second**, as mentioned before, most of the results of the neoclassical model depend on the capital share. This is also the case for the Romer-model. Early empirical evidence for the U.S. production function

\(^{26}\) The levels and growth accounting methods also provide some evidence. For a detailed description, see Greenwood and Jovanovic (1997), Barro (1998). The objective of this literature is to estimate the contributions of physical capital, labour, educational attainment, and technological progress in level or growth rates of output. In the former, Mankiw et al. (1992) show that cross-country variations in capital intensities (the capital stock to output ratio and the human capital to output ratio) explain a large fraction of the disparity in income while Klenow and Rodriguez-Clare (1997), and Hall and Jones (1997) find that these differences account for much less. In the latter, Young (1995), Hsieh (1997) provide evidence that growth performances of the Asian countries are due in large part to factor accumulation instead of the total factor productivity.

\(^{27}\) For example, Evans (1996) pointed out that if countries differ in their trend growth rates, the cross-country variance of log incomes will be stationary in first differences. In contrast, Evans (1996) shows that this variance has not trended upward in a sample of thirteen industrial countries and rather finds evidence of convergence. Evans concludes that “either endogenous growth models are fundamentally flawed, or else the effects they predict must be relatively unimportant for the countries considered here”.

\(^{28}\) For further discussion on the methodological problems of standard econometric approaches, see Temple (1998a, 1999).
(Romer, 1987) and U.S. manufacturing (Caballero and Lyons, 1992) supports the type of external effect embodied in the model of Romer but more recent work which uses better measures of capital services in estimating production functions (Burnside, 1996) finds no evidence of these externalities. In addition, a number of empirical results (Benhabib and Jovanovic, 1991; Benhabib and Spiegel, 1994) show that usual econometric methods tend to overestimate the elasticity of production to capital. Consequently, one of the strongest results in the investment-growth literature is that the returns to physical capital are almost certainly diminishing (King and Levine, 1992). However, this finding does not rule out the possibility of externality to investment. In particular, De Long and Summers (1981) studied the potential importance of investment to equipment. They show that the returns to equipment investment may be important. However, their results appear to be weak in OECD countries. Finally, some empirical studies confirm the existence of learning-by-doing or learning-by-using at the firm level. Intensities of these effects appear to be large but results are not robust. Third, the importance of the “scale effect” is controversial in the model of Romer (1986). If we assume that this scale effect is measured by the aggregate labour force of a country, the prediction will be that countries with more workers tend to grow faster – even in per capita terms. However, the empirical results (Jones, 1995a, 1995b) indicate that in a large number of countries the growth rate of per capita GDP bears only a weak positive relationship to the size of the working age population. Backus et al. (1992) look for the scale effect predicted by three stylised models of endogenous growth in which dynamic returns to scale arise from either learning-by-doing, investment in human capital, or development of new products. In the case of learning-by-doing, they assume that spillovers take place at the national level either within a national industry (e.g. spillovers are industry specific, the rate of increase in learning is proportional to total industry output) or across industries within a country. They find little evidence of a strong relationship between the growth rate of the income per capita and their measures of scale. They do find empirical support in the case of the manufacturing sector – thus, these findings do not reject a minor effect. It also possible that the scale variable for spillovers does not relate closely to aggregates measured at the country level. For example, the relevant scale can be larger than the size of the domestic economy if producers benefit from knowledge accumulated in other countries. Kremer (1993) argues that the correct scale variable might be the world population. Alternatively, if the externalities are limited to close neighbours (in terms of industry or of geographical position), the appropriate scale measure may be smaller than the total domestic population. In particular, Chua (1993) argues that countries can benefit from increased economic activity only in their close geographical neighbours. Finally, as we mentioned before, the scale effect of the Romer model can be eliminated by arguing that the externality relies on the economy’s average capital per worker. Lucas (1988) develops the same trick in a model with accumulation of human capital: each producer benefits from the average level of human capital in the economy, rather than from the aggregate of human capital. Fourth, Jones (1995a, 1995b) tests models of endogenous investment in new technology proposed by Romer (1986, 1987), and Rebelo (1991). The argument goes as follows. In these models, permanent changes in variables that are potentially affected by government policy lead to permanent changes in the growth rate. Jones shows that there are persistent changes in investment rates for advanced countries, while there are no persistent changes in growth rates. In this respect, Jones concludes that AK-style models are clearly rejected by the empirical evidence, and thus do not provide a good description of growth in advanced economies. In contrast, Bean (1990) and Ireland (1990) use Granger-causality implications of endogenous versus exogenous models and show that results are in favour of endogenous growth models. Lau and Shin (1994) apply cointegration techniques to discriminate between version of the Solow-Swan model and the model of Romer (1986). They find some evidence in support of the latter. Kocherlakota and Yi (1996) present evidence supporting endogenous growth models that emphasise public structural capital. Finally, McGrattan (1998) shows that the key predictions of AK-theory are consistent with the data, and that the patterns pointed out by Jones are short-lived. In particular, McGrattan considers a slightly more general formulation than the one tested by Jones (1995a, 1995b), by assuming that the capital-output ratio or the labour/leisure decisions are affected by

29 Jones also tests the Lucas model and the R&D-based endogenous growth model of Romer (1990a). See below.
government policies. In this case, the AK-type model does not necessarily predict that growth rates will change one-for-one with investment rates. In addition, McGrattan examines data over long time horizons in order to capture trends between investment rates and growth rates. Extending the sample considered by Jones, he finds empirical support that Jones’s deviations from investment and growth trends are relatively short-lived. Fifth, recent papers quantitatively explore the extent to which differences in economic policies account for differences in growth rates of incomes. In particular, Lucas (1990), Kim (1992), Jones et al. (1993), and Stokey and Rebelo (1995) have studied the effects of changes in tax rates on long-run growth rates. Stokey and Rebelo (1995) note that there was little or no change in the long-run US growth rate despite a large increase in the income tax rate. Their results suggest that standard AK models in which government policies (e.g., fiscal policies) affect the marginal productivity of capital (thereby altering the growth rate of the economy) predict implausibly large growth effects of fiscal policies. They show, in particular, that King and Rebelo (1990) and Jones et al. (1993) yield such results. In contrast, Stokey and Rebelo (1995) find that the two-sector endogenous growth model offers a good description of the data, if it is parameterised as in Lucas (1990) or Kim (1992). McGrattan and Schmitz (1999) find similar results, however, in some circumstances they conclude that the exogenous growth model does a better job for the dispersion of incomes, the lack of persistence in growth rates and the range of cross-country growth rates. This was the case when they looked at the implications of two parameterised models, an extended version of the standard exogenous growth model à la Ramsey and a detailed version of an AK endogenous growth model. In sum, there is a clear need for the quantitative approach of growth have to be extended, and results should be interpreted with caution at this stage.

Learning-by-doing behaviour has been introduced in CGE models. Especially in the case of energy/environmental modelling, Gritsevsky and Nakicenovic (1999) present numerical results of a model of global carbon dioxide emissions in which there is a wide range of energy technologies and resources, coupled with learning-by-doing. Learning spillover effects are introduced in technology clusters. Papathanasiou and Anderson (2000) also represent the costs of the renewable energy alternative by an elementary learning-curve. However, these models do not directly rely on the mechanisms of the Romer-model and the knife-edge condition. In general, technology-based models of learning-by-doing have a large impact when studying climate change. In particular, the results suggest that the long-run economics of climate change policy are significantly altered by the presence of learning-by-doing. However, evidence to the contrary can also be found. Goulder and Mathai (2000) develop an aggregated macro-economic model of learning-by-doing in carbon abatement, and find a more modest influence of learning-by-doing. The major difficulty is the degree of disaggregation found in the model when learning-by-doing is introduced. In fact, as pointed out by Grubb et al. (2000), if one assumes a low level of sectoral disaggregation, the marginal returns from learning-by-doing are likely to vary greatly between industries at

30 In effect, McGrattan assumes in his simplest formulation that capital stock is not interpreted broadly but rather includes human capital and capital stock separately. The production function exhibits constant returns to scale as the exponents on the two accumulable factors sum to one. However, the McGrattan-model relies on a crucial assumption: households choose investment so as to achieve a constant ratio of human to physical capital. This implies that the marginal productivity of the physical capital equals the marginal productivity of the human capital along the steady-state path. This assumption allows him to obtain a reduced AK form and the main implication of the AK model is not changed: investment is the engine of growth. It should be noted that McGrattan proposes different AK-based models. Specifically, in order to show that the findings of Jones are short-lived, McGrattan develops a model where the households has two types of capital (equipment and structures) and the leisure/labour decision is endogenous. Based on the crucial assumption that the after-tax marginal productivity of the capital stock and the human capital are the same, he shows that growth rates will not change one-for-one with investment rates. Greenwood et al. (1997) find similar results.

31 Cooley and Ohanian (1997) estimated the effects of British government policies over the postwar period with a two-sector version of the model used by McGrattan. They showed that their model fits the data on investment and growth well.
different stages of development. Many recent CGE models with endogenous technical change lack sectoral disaggregation. That is, they do not attempt to capture spillovers within an industry, spillovers across industries, or spillovers among trade partners. However, some CGE models with endogenous technical change do take into account international spillovers or human capital spillovers.

3.2. Accumulation of human capital stock

Human capital plays different roles in various theories of economic growth. In the Solow-Swan model, no special role is given to human capital in the production of output. Mankiw, Romer and Weil (1992) have included such a role in the production function: output depends on capital stock, raw labour and human capital. Their production function also exhibits constant returns to scale and diminishing returns to scale in each factor. However, to make the model one of endogenous growth, human capital is assigned a more central role. More concretely, Aghion and Howitt (1998) observe that the role of human capital can be divided into two broad categories. The first category simply broadens the concept of capital to include human capital. In this class of models, sustained growth is due to the accumulation of human capital over time (e.g. Uzawa, 1961; Takayama, 1963; Lucas, 1988; Bond et al., 1996). The second category of models attributes growth to a stock of human capital that generates innovations, (as in Romer 1990b), or improves a country’s ability to imitate or adapt new technologies (e.g. Rustichini and Schmitz, 1991; Acemoglu and Zilibotti, 2000). To emphasise the differences between the two categories of models, it is worth noting that, in the former, human capital is the skills embodied in a worker. It results from an arbitrage between present consumption and education. Thus, the use of these skills in one activity prevents its use in another activity. That is, human capital is a rival good. In the latter, however, ideas or knowledge may be non-rival and may in some circumstances be non-excludable. Note that this distinction means that theories of technological progress differ from models of accumulation of human capital. In particular, the growth of human capital in the model of Lucas affects output growth, while the stock of human capital in the Romer model has a scale effect (Krueger and Lindahl, 2001). We only present here the model of Lucas.

This model (Lucas, 1988) can be described as follows. The agents’ decisions at any instant of time are: (i) how much to consume; (ii) how to allocate physical and human capital across the two sectors, and (iii) at what rate to accumulate total physical and human capital over time. Lucas assumes that production uses physical capital and the skills of labour or human capital. Hence, he models a two-capital goods economy, where the human capital variable resembles physical capital. In contrast to raw labour, it is possible to increase human capital by investment, just as it is possible to increase the stock of physical capital. Thus, this kind of model may be viewed as allowing for physical labour as well as human capital, provided these two inputs are good substitutes. In particular, if these two inputs are perfect substitutes, human capital can simply be taken to be the sum of tangible and intangible human capital. In contrast to "A" of the “AK” model, if one wants to replicate a productive activity, it will incur a cost, (e.g. to train additional workers, in order to produce more human capital). The model is therefore a two-capital good

32 The second category of models is exposed in the next section.
33 The model of Romer is explained in section 3.3.
34 The second model of Lucas is close to a model developed by Krugman (1988). The key point is that increasing returns to scale have some positive and normative effects on trade.
35 In this respect, the model of Lucas can be interpreted within an international context in which physical capital is identified with traded capital and human capital with non-traded capital.
36 Mankiw, Romer and Weil (1992) also assume that there are accumulation of human capital and accumulation of capital stock. But they did not provide explicit micro-foundations to the accumulation of human capital and the production function does not exhibit increasing returns to scale due to the presence of externalities.
model. Consumption goods and physical capital are produced in the first sector.\textsuperscript{37} Hence, the production function exhibits constant returns to scale.\textsuperscript{38} The key sector for determining the growth rate of the economy is the sector for producing new human capital. Skills, knowledge, or human capital are generated in the “research and education” sector as follows: more education today lowers production today, but increases production tomorrow, because it raises the productivity of labour and capital. There is therefore an incentive to pursue learning activities and the incentive is clearly dependent on intertemporal preferences. The crucial assumption is that the accumulation of human capital is generated by a linear technology. According to the formulation (equation 1, section 2.3.), this means that $G$ is linear and therefore human capital grows at a rate which depends on the time devoted to education or creating human capital and a maximum achievable growth rate of human capital. Thus, if the accumulation of human capital is proportional to the product of the existing stock of human capital and the amount of raw labour devoted to training, the rate of growth of human capital is proportional to the amount of raw labour devoted to training. Anything that will affect this quantity will alter the growth rate of human capital, and thus the growth of output. Therefore, in steady state, output and human capital grow at the same rate, and depend on both the maximum achievable growth rate of human capital, as well as the determinants of the equilibrium value of the time devoted to training.

In this context, the model generates unbounded growth but does not rely on aggregate increasing returns to scale.\textsuperscript{39} Sustained growth arises because there are constant returns in the production of human capital, and all inputs can be accumulated – there are no essential fixed factors. If human capital grows without bound, its effect on the output equation is like the effect of exogenous technological change in the Solow-model or the exogenous population growth in the Arrow model. Consequently, never-ending growth requires never-ending increases in human capital. But, as pointed out by McCallum (1996), never-ending growth is implausible for such a variable because skills are possessed by individual human beings and are not automatically passed on to workers in succeeding generations. In this respect, human capital is different from the stock of knowledge, which is possessed by society in general, and is passed on from generation to generation. McCallum concludes that it is some form of knowledge, not human capital, that can plausibly provide the basis for never-ending growth. This criticism is levelled directly at the linear form of accumulation of human capital and is a crucial blow, given that Lucas emphasises that the endogenous character of the growth rate depends on this ad hoc assumption. If the technology of the human capital accumulation is less than linear (diminishing returns to scale in human capital), the growth rate of the economy is exogenous. Moreover, Solow (2000) pointed out that the Lucas model also relies on the specific form of the utility function. If the utility function depends on leisure, the growth rate will be exogenous. That is, a preference for leisure implies that supplied labour decreases with the growth of human capital. In this case, in order to obtain a balanced growth path, one has to assume that human capital is constant, (i.e. that endogenous growth no longer exists). In addition, it should be noted that the formulation of the externality matters for the existence of a scale effect. This externality can be written in different forms, which are not neutral in terms of the long-run properties of the model (Backus et al., 1992). In effect, by formulating spillovers as affecting the production rather than the accumulation function, Lucas generates growth without a scale effect. By contrast, if one assumes that the external effect in production depends on the total stock of human capital held by the labour force\textsuperscript{40}, there is no direct scale

\textsuperscript{37} In Uzawa-model, the production function exhibits constant returns to scale.

\textsuperscript{38} Lucas also suggests that there are increasing returns to scale, with positive external effects that are associated with human capital. While this externality arises in the model of Romer from an indirect link between physical capital and knowledge, Lucas emphasises direct interaction effects of human capital. This externality depends on the average level of human capital rather than on the total amount of human capital.

\textsuperscript{39} In Uzawa’s formulation, there is also unbounded growth with no increasing returns and no external effects.

\textsuperscript{40} In this case, the more educated workers are, the more new ideas to improve productivity efficiency there will be.
effect but there is a population growth effect. A scale effect arises if one assumes that human capital accumulation depends on the average fraction of time spent accumulating human capital and the size of the human capital sector. Finally, as noted by Solow (1999), investment in physical capital and in human capital are alternative uses of aggregate output in the Lucas model. In this case, the choice between them deserves to be modelled in some less mechanical way than fixed shares in each sector.

In order to test for empirical evidence of human capital-based endogenous growth models, one may exploit the results of both micro-economic and macro-economic studies. On the one hand, the micro-economic labour literature has provided several estimates of the rate of return to investment in education. Specifically, most micro-economic studies using a Mincerian human capital earning function 41 imply that a change of the average level of schooling provides a substantial change to income (Krueger and Lindahl, 2001; Card, 1999 and Psacharopoulos, 1994). However, it is unclear whether the social return to schooling exceeds the private return. On the other hand, the macro-economic growth literature has mainly investigated whether the initial level of education and/ or changes in education are related to the GDP growth rate. The results differ according to the data and the method used. Growth accounting exercises (Jorgensen et al., 1987; Young, 1994, 1995) find a significant impact of human capital accumulation in growth. But, as pointed out by Topel (1999), this result depends on the estimation approach used. In related work, Romer (1989, 1990) shows that there is a significant effect of the initial stock of human capital or literacy, but that implicit output elasticities are too high. At the same time, the variation of human capital appears not to be significant – this contradicts the results of the Lucas model in which a change of the human capital variable alters the long run balanced growth path. These results were also confirmed by Benhabib and Spiegel (1994) and Pritchett (1996). However, Krueger and Lindahl (2001) present evidence suggesting that these results are linked with the quality of education data. In particular, they show that estimates of the change in schooling are severely reduced by measurements errors in growth equations and the econometric method used. They suggest that both the change and the initial level of education are positively related with economic growth (Gemmel, 1996; Topel, 1999). Finally, even if these two lines of research are disparate, they provide evidence that human capital is an important source of growth, and has a positive effect on income. However, as mentioned before, these methods do not really test a structural endogenous growth model, but rather a reduced form. Jones (1995a, 1995b) tests the predictions of the Lucas model and shows that the empirical evidence does not support the results of Lucas (1988). In contrast, Bassanini and Scarpetta (2001) find a robust relationship between human capital and output growth as well as plausible output elasticities and some empirical support to an endogenous growth model à la Lucas.

The Lucas formulation has been implemented in CGE models. Ambler et al. (1999) analyse the quantitative link between export-promoting commercial policies and economic growth. They develop a dynamic CGE model in the case of a small open economy in which the economy’s equilibrium is suboptimal due to the presence of imperfect competition in the manufacturing sector and a human capital accumulation externality. The economy is characterised by two sectors: a capital-intensive manufacturing sector and labour-intensive non-manufacturing sector. In the former, monopolistically competitive firms produce differentiated intermediate goods with labour and capital. These intermediate goods are combined in order to produce a final good used for both consumption and investment. In the latter, firms are price takers and only produce a good for final consumption. Human capital accumulation is a by-product of learning-by-doing and is a function of the level of activity in the manufacturing sector (Box 3). Ambler et al. (1999) use the linear specification detailed below. This externality is justified by the effects of learning-by-doing and trade on growth rates (Lucas, 1988; Young, 1991). This asymmetry in human capital accumulation is explained by the following variables: the years of schooling, the experience and the experience squared. In this formulation, the time spent in school is the key determinant of earnings.

41 Mincer (1974) developed an equation in which the natural log of the wage for an individual is explained by the following variables: the years of schooling, the experience and the experience squared. In this formulation, the time spent in school is the key determinant of earnings.

42 See also Pritchett (1996).
accumulation externality between the two sectors is often found in the literature. In this respect, Ambler et al. (1999) show that promoting exports in the manufactured sector can lead to an increase in both welfare and growth rates for a wide range of parameters value.

Box 3. Production and knowledge functions in models with accumulation of human capital stock

We describe here the different formulations of the Lucas model. Then, we present the human capital accumulation equation in Ambler et al. (1999).

- Non-intentional human formation

Following our common representation of endogenous growth models, the production function is described as follows:

\[ Y_t = F(X_t, A_t) = (H_t)^\beta K_t^{\alpha} A_t^{\delta} \]  

where \( H_t \) is the human capital stock and \( \delta \) is an externality parameter.

The externality is defined by:

\[ H_t = H_{t-1} + \delta H_{t-1} \]  

The analytical framework is the same as Romer (1986).

- Intentional human formation

1. \[ Y_t = (N_t h_t)^{1-\alpha} K_t^\alpha \]  

where \( N_t \) is the total amount of time spent on human capital accumulation or the size of the human capital sector.

The accumulation function is defined as follows:

\[ h_{t+1} = h_t (1 + \beta n_{2t} B_t^\delta) \] and \( B_t = N_{2t} \)

where \( n_{2t} \) is the average fraction of time spent accumulating human capital and \( B_t \) measures the external effect of capital accumulation.

In this case, the growth rate includes a scale effect: countries with larger human capital sectors grow faster.

2. \[ Y_t = (N_t h_t)^{1-\alpha} K_t^\alpha B_t^\delta \]  

where \( N_t \) is the total amount of time spent on human capital accumulation or the size of the human capital sector and \( B_t \) measures the external level of human capital labour force.

The accumulation function is defined as follows:

\[ h_{t+1} = h_t (1 + \beta n_{1t} B_t^\delta) \]  

By formulating spillovers as affecting the production function rather than the accumulation function, the model generates growth without scale effects.

3. The external effect in the production function can depend on the total stock of human capital held by the labour force. Thus, equation (5) is unmodified and \( B_t \) is now equal to \( N_t h_t \). The model still generates growth without scale effect.


In their general equilibrium trade model, Ambler et al. suppose that human capital accumulation is a by-product of learning-by-doing and is a function of the level of activity in the manufacturing sector. The knowledge function is thus given by:

\[ H_{t+1} = (1 - \delta_H + \eta N_t) H_t \]

where \( \delta_H \) is the rate of depreciation of human capital, \( \eta \) captures the sensitivity of human capital accumulation to resource allocation and \( \lambda \) measures the diminishing returns to scale of resource allocation on human capital accumulation.

3.3. The accumulation of new ideas: endogenous growth with research and development

3.3.1. Standard R&D-based growth models

We just reviewed different ways of counteracting diminishing returns to scale in capital either by redefining the role of capital or introducing human capital. But none of these re-definitions provided us
with a theory of technological progress. These models do not allow us to think of firms as undertaking investments aimed at producing new products and production methods. The work of Romer (1990b), Grossman and Helpman (1991) and Aghion and Howitt (1992, 1998) provides explicit theoretical models of endogenous technological changes. The arguments go as follows. As mentioned before, many types of innovation involve substantial efforts in the development of blueprints so that, once they have been produced, they can be costlessly used to guide the manufacturing process for new goods. However, firms will not be willing to pay the fixed costs associated with the development of the product if the production of the new good will take place in a competitive environment where nothing is left to pay for the R&D activity. Thus, innovations have to be temporarily or perpetually protected (e.g. patents) and must provide a rent to innovators. In this respect, these models produce equilibria where the growth rate of technology is affected by economic incentives on capital accumulation or R&D or both. At the same time, they are explicitly based on micro-foundations of the activity of R&D. Thus, innovations induce inefficiencies due to imperfect competition and, at the same time, create further impediments for growth. These models have in common the structure presented in section 1.3 – they mainly differ in the story they tell regarding "A" of the “AK” model (see equation 2, &18). In these models, the economy has three sectors.

(i) Consumption/final goods

(ii) Intermediate goods

(iii) R&D

The first sector produces consumption/final goods and is in perfect competition. The second sector produces intermediate goods that are necessary inputs in the consumption/final good. Finally, the third sector undertakes R&D in the form of new varieties of intermediate goods or quality ladders. In the case of the Romer and Grossman-Helpman models, new intermediate goods or new quality ladders expand the possibilities of production, and do not perfectly substitute with existing goods. In contrast, the Aghion-Howitt model made precise the Schumpeter’s “creative destruction”. Finally, the models assume that only the R&D and the final good sectors employ the labour force.

Romer (1990b) proposes a theory of endogenous technical progress that uses the Dixit-Stiglitz model of monopolistic competition (Dixit and Stiglitz, 1977). The Romer model uses four basic inputs: capital, labour, human capital and an index of the level of technology. Capital is measured in units of consumption goods, and is assumed to be accumulated as forgone output, (capital goods are produced in a separate sector that has the same technology as the final output sector). In this respect, total capital is implicitly defined as being proportional to the sum of all different types of capital or intermediate goods. The population and the supply of physical labour are both constant. The concept of human capital is a

43 It should also be noted that these R&D based growth models reviewed here are relevant for economies on the frontier of technological development. These models can be extended to take into account technology adoption. In this case, the issue is not whether to devote resources to innovation but whether to adopt technologies that have been developed by others. There are important learning-by-doing dynamics in adopting new technologies. This learning-by-doing process has been emphasised by Lucas (1988), Stokey (1991) and Young (1991) in human capital accumulation-based models. For example, Greenwood and Yorukoglu (1997) develop a model in which a major technological revolution can lead to prolonged decline in productivity.

44 This formulation with a continuum of goods is also close to that used by Judd (1985). However, in the model of Romer, differentiated goods are capital goods and not consumption goods. It should be noted that the model can be only written in terms of physical labour devoted to manufacturing (final/consumption good) and on physical labour devoted to the R&D sector, that is, one may exclude the stock of human capital in order to simplify the analysis. However, we present here the original contribution of Romer, which includes both human capital and physical labour.
distinct measure of the cumulative effect of activities, as in the case of formal education and on-the-job training. It relies on the total years of schooling or training. The total stock of human capital in the population is fixed and can be decomposed into the fractions devoted to the manufacturing sector and the R&D sector. Finally, the model separates the rival component of knowledge (e.g. the stock of human capital), from the non-rival technological component (e.g. the number of new goods). Technological progress takes the form of an expansion in the number of differentiated intermediate goods (the index of the level of technology).

The final consumption good is produced by combining physical labour, human capital and differentiated intermediate goods, which embody the technological progress undertaken in the R&D sector. This final good is produced by a large number of competitive firms. The elasticity of production to each intermediate good is assumed to be the same and the marginal product of an intermediate good is independent of the quantity employed of other intermediate goods. The Cobb-Douglas production function, exhibits overall constant returns to scale, but diminishing returns in each factor of production.

In the intermediate sector, Romer chooses a simple linear technology, with units such that one unit of forgone consumption invested by consumers produces one unit of intermediate good. This simple formulation involves a trade-off between consumption today and knowledge that can be used tomorrow to produce more consumption. In particular, factor intensities are the same in the production of final output and in the production of intermediate capital goods. As intermediate goods are not perfect substitutes in the technology used to produce the final consumption good, a firm has to pay a fixed cost to produce but then receives a permanent monopoly rent in the intermediate good they create. Therefore, the intermediate

---

45 Hence the concept used here is more limited that the one presented in King and Rebelo (1986), Lucas (1988) and Becker et al. (1990).

46 Grossman and Helpman (1991) use the setting of a variety of consumer products (Spence, 1976) in order to study the links between technological change and economic growth. The Grossman-Helpman model assumes that consumer goods come in numerous varieties (instead of intermediate goods in the Romer model) and that households take care of an index of these varieties. In this case, a continuing expansion of the varieties of consumer products occurs only if the cost of R&D (in terms of goods) declines steadily. This assumption about the R&D technology seems implausible and thus the introduction of varieties of consumer goods adds little to the understanding of the growth process. It should be noted however that one can combine the models of expanding variety for producer and consumer goods.

47 In this model, human capital and physical labour can be interpreted as the number of skilled workers and the number of unskilled workers, respectively.

48 The production function can be written as follows:

$$ F(H_y, L, x) = g(H_y, L) \int_0^A x(i)^{\alpha} di $$

where $g$ is a Cobb-Douglas production function homogenous of degree $1 - \alpha$, $A$ is the number of goods’ varieties.

The specification of $g$ has a particular incidence on results (see further).

49 It follows that the number of firms can not be determined because of constant returns to scale. Thus, Romer assumes that there exists a large number of identical firms, which produce the final good, and that the final sector is characterised by perfect competition. Each firm maximises its profits and optimally chooses its demands of physical labour, human capital and intermediate goods.

50 Assuming that each unit of intermediate goods costs a constant unit of output, the solution of the monopolist’s problem is to charge a mark-up over marginal cost. If the marginal cost is the same for all producers, then price will also be the same. Thus in equilibrium all producers supply the same quantity, charge the same price and inverse demand curve are isoelastic. In addition, at equilibrium, the optimal monopoly level of intermediate goods and the optimal price are constant because the interest rate, the level of human capital in manufacturing and labour are constant.
goods sector consists of individual firms so that each firm produces only one variety of intermediate good. A potential entrant must weigh the fixed cost against the monopoly profits that result from the ability to charge a price above marginal cost throughout the infinite life-cycle of the product. This permanent monopoly rent relies on a patent, (i.e. a patent is viewed as a barrier to entry in the production of a new intermediate good). However, if it is assumed there is free entry into the production of new intermediate goods, it will follow that monopoly profits must be exactly outweighed by the cost of inventing a new product. Hence, the value of a patent will be defined as the present discounted value of monopoly profits that the holder of the patent can extract. The imperfect market structure of intermediate goods is of particular relevance because it determines the innovator’s remuneration and therefore the intensity of technological progress. On the one hand, an excess of competition reduces the monopolistic rent, and thus the rate of innovation and the technological frontier. On the other hand, an excess of market power leads to a higher rate of innovation. Finally, while an element of monopoly is necessary to generate rent to support the research activity, the existence of a patent system or some other structure that protects intellectual property rights provides the ex-ante economic incentives to allow ongoing generation of ideas and technological progress but also engenders ex-post inefficiencies that need to be corrected by economic policies.

In the R&D sector, Romer uses a linear specification where the number of new inventions (e.g. new goods) is determined by the part of human capital devoted to R&D, the total stock of ideas or innovations discovered in the past, and an exogenous parameter reflecting the effectiveness of the research. At any time, the allocation of human capital between the manufacturing (final good) and research sectors is determined by the requirement that the returns to human capital are the same in both sectors. Given the allocation of human capital, the production of knowledge exhibits constant returns to scale in the stock of human capital when the new designs are held constant, (devoting more human capital to research leads to a higher ratio of production of new intermediate goods). Hence, the rate of growth of new designs is unceasing in the amount of human capital employed, and the productivity of a unit of human capital is increasing in the total number of designs that currently exists. This positive dependence on the amount of human capital devoted to the R&D process is the crucial assumption in the model. The linearity in the variation of new designs is what makes unbounded growth possible. It follows that the amount of human capital determines the rate of growth of knowledge and thus the rate of growth of the economy. In this

---

51 The introduction of imperfect competition in a growth framework was first introduced by Nordhaus (1969), Judd (1985) and Romer (1987). Romer introduces many goods and maintains an explicit dynamic model of accumulation. His model does not rely on external effects to support a decentralised equilibrium, but rather uses an imperfect competition framework (monopolistic competition). In effect, as pointed out by Marshall (1861) and emphasised by Young (1920), there is a second source of external economies, trade among different firms offering unique specialised goods caused a form of increasing returns that is external to individual firms. Instead of describing specialisation in terms of a competitive equilibrium with externalities, Romer models the price setting behaviour of firms. This model contains most of the features of Romer (1990). But it does not introduce an R&D sector. In the equilibrium, growth takes place at a constant exponential rate. This rate of growth is too low relative to the rate that a social planner could achieve, and it increases with any intervention that increases savings. The main value of this model is that it does not rely on growth with spillovers or with a broad view of capital. In effect, the problem of diminishing returns to scale is counteracted by the continual introduction of new goods.

52 Following Rivera-Batiz and Romer (1991), another possibility is to assume that to invent new goods the economy must invest a fixed amount of output into the R&D process. The underlying assumption is that the R&D sector has the same factor intensity as the technologies that generates consumable and intermediate goods. Rivera-Batiz and Romer describe their specification as the “lab-equipment” model of R&D. In this respect, the specific assumption used in Romer (1990a) must be regarded as a tractable approximation.

53 See below.
sense, unbounded growth will be more like an assumption than a result of the model. This aspect of the model has been criticised by other researchers (Jones, 1995, 1998, 2001). This specification also induces that knowledge enters into production in two distinct ways. On the one hand, a new idea or design allows the production of a new intermediate good that can be immediately used in the final sector. A new idea also increases the total stock of knowledge and the marginal productivity of the stock of human capital devoted to the R&D process. On the other hand, innovators can use the available stock of knowledge (due to its public good nature) and build a new design. Thus, researchers achieve their success at least partly by “standing upon the shoulders of giants”. The stock of knowledge accumulates from the spillovers of previous ideas, (i.e. this stock acts as a positive externality). These public good aspects of knowledge create economy-wide increasing returns.

Based on these formulations, the reduced form of the model can be described by a set of two equations. In the first, the production function is linear in the number of products available in the economy, and exhibits increasing returns to scale. In the second, the growth rate of new goods linearly depends on the stock of human capital devoted to the R&D sector. Thus, the assumptions of constant human capital stock and physical labour involve that, on a balanced growth path, the rate of growth of new goods is equal to the rate of growth of output. In other words, the productivity of the R&D sector directly influences the growth rate of the economy. It turns out that this endogenous rate of growth depends on a number of factors: (1) the household’s preference parameter; (2) the intertemporal elasticity of substitution; (3) the efficiency term of the R&D process; and (4) a scale effect. Regarding factor (1), a weaker willingness to save lowers the growth rate of the economy. This occurs because the higher is the degree of impatience, the less resources agents will devote to the future. That they will reduce savings and thus the accumulation of physical capital. But the capital stock also allows them to produce intermediate goods, which in turn allows production of the final good. If the capital stock falls, it follows that the production costs will increase. This reduces production and, in turn, the value of patents – which depends indirectly on the production of the final consumption good. Consequently, the incentives to research will gradually be weakened and thereby lower the number of innovations and the growth rate of the economy. For factor (2), a shift of the intertemporal elasticity of substitution has the same effect : the higher the elasticity of substitution is, the more agents will substitute future consumption into present consumption and thus reduce savings. Regarding factor (3), the rate of growth positively depends on the efficiency term of the R&D process. An augmentation of this term leads to an increase of the remuneration of researchers and thereby of the amount of resources devoted to the R&D sector. For factor (4), a larger endowment of human capital in the R&D process raises the rate of growth. This scale effect is similar to those that take place in the model of learning-by-doing with spillovers (Romer, 1986; Lucas, 1988) and with the Barro model of public expenditures and goods (Barro, 1990). However, the mechanism that generates the scale effect is different here. It is the fundamental link between ideas and returns to scale that gives rise to this effect: a new product, which involves a cost of inventing, can be used in a non-rival manner across the economy. This result stems from the assumption that there is a fixed cost associated with inventing a new good that is unrelated with the number of units of the final good manufactured in the future. In a technical sense, this effect appears as a by-product of the linearity assumption of the knowledge-based technology. For the sake of simplicity, if we assume that the factors of production are only capital, ideas, and the labour force ;and that a constant fraction of the fixed population is devoted to R&D, it will immediately follow that the number of goods invented depends on the size of the market measured by the total labour force. In effect, an increase of the size of the population, (other things being equal), raises the number of researchers and leads to an increase in the growth rate of per capita income. Hence, a larger economy will grow faster.

\[ \text{Romer (1990c) shows that an increase in the fixed supply of labour has two opposite effects in his original formulation (Romer, 1990a). On the one hand, it directly leads to an increase of the marginal product of human capital used in manufacturing. On the other hand, it increases the marginal product of each of the intermediate goods causing the monopoly rent to go up. This causes the price of new design to increase, which in turn raises the returns to human capital in research. These two effects exactly cancel with the Cobb-Douglas specification retained by Romer (see footnote 35). Consequently, an increase of the labour}\]
while, in a small economy, the fixed cost paid to develop many goods is not compensated by the size of the market because the cost of invention per unit of the labour force is too large. This argument has been mainly used in models of endogenous growth with trade liberalisation (Grossman and Helpman, 1991). We further discuss the empirical evidence of such scale effects.

As we mentioned before, the imperfectly competitive structure of the market for capital variety and the non-rival and only partially excludable nature of the stock of knowledge involve ex-post inefficiencies. These two major market imperfections suggest that the market-driven rate of economic growth is likely to be less than the growth rate in the planned economy (chosen by the social planner). This lower growth rate results from a gap between the private return and the social rate of return from various activities. Such effects were obtained in the models of learning-by-doing with spillovers (Romer, 1986), in the models of human capital accumulation (Lucas, 1988, 1993), and in the models of public goods (Barro, 1990; Saint-Paul, 1993) with positive or negative externalities. It also affects the sharing of the value added between consumers and the patent holders. This effect is sometimes called the “surplus effect of consumers”. An innovator obtains a monopoly profit but the social benefits of his idea correspond to the consumer’s surplus. Consequently, the incentive to innovate (private gains) is less than the social gains. In other words, the socially optimal level of R&D is larger than the private level of R&D. Finally, due to the public good characteristic of knowledge, the market does not compensate researchers for their contributions to the existing stock of ideas, or to improving of the productivity of future innovators (“knowledge externality effect”). In practise, these distortions may be important and may be partially corrected by economic policies, as for instance subsidies to purchases of intermediate goods, for final output, or for research (Barro and Sala-I-Martin, 1995). As pointed out by Jones (1995, 1998), this also gives rise to the question of the optimal level of investment in the R&D process.

Before examining the main theoretical criticisms of the Romer-model in the next section, we briefly present the contributions of Aghion and Howitt (1992, 1998) and Grossman and Helpman (1991, 1994). These models use the same structure of the Romer model, but differ in two key assumptions. They are, however, subject to the same drawbacks. In particular, the knowledge technology is a linear form and the rate of growth and features the same scale effect of population encountered in the variety extension model. Aghion and Howitt (1992, 1998) and Grossman and Helpman develop models that incorporate the life-cycle aspect of innovation. In contrast to the Romer model, they do not assume that new varieties of intermediate goods coexist with old varieties and innovation does not involve risk-taking, (i.e. that the model employs a stochastic R&D technology). Their “quality ladders” models embody two key concepts: (i) profit-seeking firms try to achieve market power by producing a better good than their competitors; and (ii) over time, new goods displace old ones, firms earn profits for some period of time, and the new goods are displaced in turn. Potential entrants must weight the cost of improving an existing good against the benefits of monopoly power associated with successful innovation. But the monopoly power is only

---

55 Negative externalities or spillovers correspond to congestion or pollution effects. Such effects may also be present in models with R&D (Jones, 1995).

56 In the same area of research, Caballero and Jaffee (1993) develop a model that incorporates both the quality ladders and the expanding variety models. Acemoglu (1997) uses a version of the quality ladder model to study how technological progress can respond to changes in factors of production. In this respect, Acemoglu shows that the scale effect can help us understand the evolution of skill premium in the United States. Finally, a number of extensions of the Aghion-Howitt model are presented in Aghion and Howitt (1998).
temporary. The process of R&D advances the technological frontier by increasing the index of quality. Due to the proximity of both models, we present here a simplified version of the Aghion-Howitt model.\footnote{For a complete description of the Aghion-Howitt model, see Aghion and Howitt (1998). The general representation of the Grossman-Helpman model can be summarised as follows. Total factor productivity depends on the available assortment of intermediate goods. These intermediates can be either horizontally differentiated (different types of inputs) or vertically differentiated on quality ladders. New intermediate inputs are created through R&D undertaken by profit-seeking firms. Hence the total number of available intermediate goods is a function of past R&D investments. This leads to the conclusion that the more intermediates are used in production, the higher is total factor productivity.}

In contrast to the Romer model, Aghion and Howitt assume that the R&D activity is a random process. In fact, the emergence of a new innovation depends on instantaneous probability driven by a stochastic Poisson’s process. The new intermediate good displaces the old intermediate goods. Hence the final production sector only uses the last intermediate good. This is the major difference with the Romer model. The producer of the final good demands the intermediate good relative to its price, which constitutes the monopoly demand curve of the unique intermediate good. The monopoly profit determines the remuneration of the innovator. But the flow of expected value created by a patent is now not equal to the instantaneous monopoly profit: a new patent may appear at any time and thus may reduce this expected value. This is the “creative destruction” process identified by Schumpeter. Workers are assumed to be identical and are indifferent between manufacturing and R&D sectors. Consequently, the wage must be equal to the expected value of the remuneration of a researcher. This arbitrage condition and the equilibrium condition on the labour market determine the equilibrium of the economy. The effects of various parameters on the (average) growth rate of the economy are closely similar to those in Romer. However, in this case the probability of discovering a new design alters the rate of growth. An increase of this probability has two opposite effects. On the one hand, it causes the supply of new designs to increase. On the other hand, this also causes a reduction in the remuneration of an innovator. Overall, the first effect is larger than the second one, i.e. per capita income grows. The use of a random process in order to create a new intermediate good substantially complicates the model and poses several questions of calibration in a computable general dynamic equilibrium model. For example, Caballero and Jaffe (1993) develop a model in the spirit of Aghion and Howitt and Grossman and Helpman (1992) that gives a simple relationship for the effect of new products on the value of existing ones. In addition, they estimate some equations by using market value and patents on 567 large U.S. firms. However, they point out that the overall system of equations was not estimated because “the theoretical and empirical compromises that are necessary to find empirical counterparts to the model constructs cannot really be applied consistently across the different parts of the models” (Caballero and Jaffee, 1993).

### 3.3.2. Growth with or without scale effects

The models of Romer (1990a, 1990c), Aghion and Howitt (1992, 1998) and Grossman and Helpman (1992, 1994) are summarised in Box 4. At a theoretical level, these models use the same technical tricks for generating sustained growth. Here, sustained growth arises because producing new products or products with better quality involves factors of production that can be accumulated, so innovation will never reach an upper bound that would slow down the rate of expansion. In addition, these models involve a knife-edge condition in the accumulation function. The production of new ideas is homogenous of degree one, or exhibits constant returns to scale in the present stock of ideas when the stock of human capital is held constant. If one assumes that the technology function has diminishing (increasing) factorial returns to the stock of ideas, it follows that the rate of growth is exogenous unceasing). This condition is analogous to those described in physical or human capital-based models of endogenous growth (sections 3.1. and 3.2.), where the sum of the elasticity of production to the externality and the elasticity of production to the accumulable must be strictly equal to one. Finally, as explained before, the endogenous growth rate is more an assumption than a result of these models.
All these models share two common features. First, these R&D-based growth models are often characterised by having scale effects, (i.e. variations in the size or scale of the economy affect the size of the long-run growth rate). Second, these models require the production function to exhibit constant returns to scale in the factors that are being accumulated endogenously. This second limitation is linked with the knife-edge condition of the accumulation function (the linear specification of the creation of new ideas), but it also includes additional conditions regarding the production function. Both conditions have been extensively discussed and criticised in the theoretical and empirical literature (Solow, 1994). These considerations have led to the emergence of non-scale growth models or “semi-endogenous” growth models (Jones, 1995a, 1995b, 2001; Segestrom, 1998 and Young, 1998).

The basic formulation of non-scale growth models has two distinct properties. First, balanced growth paths exist under quite general production structures. Indeed, the knife-edge condition, which imposes a scale effect in the standard R&D-based growth model, is obtained as a special case by imposing a restriction on an exogenous parameter. Hence, non-scale equilibria are the rule rather than the exception. Second, long-run growth is proportional to the population growth rate by a factor that reflects the productivity of labour and capital in the aggregate production function. Therefore, it is independent of macro policy instruments. In what follows, we describe a general representation of the semi-endogenous approach based on the works of Jones (1995a, 1998), Kortum (1997) and Segestrom (1998).

The semi-endogenous approach aims at incorporating the decreasing R&D productivity of innovations, as illustrated by Jones (1995a, 1995b), as well as the empirical and theoretical works of knowledge spillovers from R&D. The seminal works of Jones (1995, 1998) incorporates the general settings of R&D-based endogenous growth models. But they mainly differ in two aspects: (i) standard R&D-based growth models appear as special cases for parameter values; and (ii) the accumulation function is reconsidered. In the simplest formulation of R&D-based growth models, the change in knowledge is equal to the number of people attempting to discover new ideas multiplied by the stock of knowledge discovered in the past. In addition, the accumulation function is linear. In contrast, Jones (1995, 1998) supposes that the creation of new ideas is less than linear in the existing stock of knowledge. If the externality parameter (e.g. the knowledge externality effect) is positive, the formulation of Jones allows for increasing returns to scale in the production of new ideas, corresponding to the case in which previous discoveries increase the productivity of current R&D. If the externality parameter is negative, there are decreasing returns to scale in the production of new ideas. This corresponds to the case referred to in the productivity literature as “fishing-out effect”, in which the rate of innovation decreases the R&D productivity. The standard linear accumulation function is thus obtained as a special case, when the degree of externality is exactly equal to one. Finally, Jones introduces the possibility that, at a point in time, the duplication and overlap of research can reduce the total number of innovations developed by the units of labour. Using the decentralised set-up of Romer (1990), together with the assumption that the labour force grows at an exogenous, constant rate, Jones finds that the growth rate of the economy depends only on the population growth rate and some exogenous parameters, which determine the external returns as well as the returns to scale in the R&D sector. In this case, if we impose the restriction that the accumulation function is linear in the existing stock of knowledge, no balanced growth path exists in this economy because of the growing rate of the labour force. Thus, once one relaxes the assumption of a unitary coefficient of the knowledge externality in favour of a coefficient less than one, the scale effects of the Romer/Grossman and Helpman/Aghion and Howitt models are replaced by a dependence on the growth rate of the labour force. In this respect, changes in the intensity of research no longer affect the long-run balanced growth path but rather the level of per capita variables. This implies that a change in the intensity of research has only a transitory effect. Indeed, the long-run growth rate is determined very much as in the Solow-Swan model. In addition, the steady-state growth is invariant to government polices. This

---

58 Segestrom (1998) assumes that in each industry the most obvious ideas are discovered first, making it harder to find new ideas subsequently.
result contrasts with those of Romer/Grossman and Helpman/Aghion and Howitt models in which the steady-state growth rate depends on policy variables such as subsidies to R&D.

In response to Jones’ criticisms, Aghion and Howitt (1998), Young (1998), Dinopoulos and Thompson (1998a, 1998b), Howitt (1999) neutralise the scale effect in standard R&D based growth models. In these models, the variety of consumption goods is proportional to the population of the economy. Thus, the growth rate of new ideas now depends on research effort per variety, which means that the scale effect on growth is eliminated. Their results are based on the assumption that an increase in population results in a proportionate increase in the number of sectors in the economy. This means that the size of each sector and the number of researchers in each sector does not change in response to a rise in population. However, these models still impose a knife-edge condition, e.g. the variety of consumption goods is exactly proportional to the total population in the economy. Young (1998) relaxes this assumption of proportionality and shows that the model can generate either positive or negative scale effects on growth. Jones (1999) reinterprets Young’s results and characterises the asymptotic growth rate in a general model including the standard R&D growth models, the Jones/Kortum/Segestrom models and Aghion-Howitt/Young/Thompson-Dinopoulos models. In this respect, Jones concludes that all of the models reviewed here exhibit scale effects: the size of the economy affects either the long-run growth rate in standard R&D-based growth models or the level of per capita income in Jones/Kortum/Segestrom models. The key point is the implication in terms of government policies. In the former, macro instruments have an effect on the long-run equilibrium growth path. In the latter, macro policies have large effects on long-run levels, but only a transitory effect on the growth rate of the economy.

---

59 We consider here a common representation of these models, as exposed in Jones (1999). For further details, see Aghion and Howitt (1998), Dinopoulos and Thompson (1998a, 1998b), Young (1998).
Box 4. **Production and knowledge functions in models with R&D**

We present here a general representation of scale and non-scale R&D-driven growth models. Following of Jones (1999), we only describe the reduced form of the production and accumulation functions. For important insights concerning the micro-foundations of growth readers are referred to the original source.

- **Scale growth models**

  **The Romer/Grossman-Helpman/Aghion and Howitt models**

  The aggregate production function is described as follows:

  \[ Y_t = A^\alpha L_t \]  

  where \( L_t \), \( A_t \) denote the labour force used in the final production sector and the stock of ideas, respectively.

  In this economy, the growth rate of the population is assumed to be zero. The production function exhibits constant returns to scale in the primary rivalrous inputs (the labour force here) and increasing returns to scale to combined labour and ideas. In this respect, the parameter \( \alpha \) measures the extent to which increasing returns to scale are important.

  The accumulation function is described as follows:

  \[ A_t = L_t A_{t-1} \xi \]  

  where \( L_t \), \( \xi \) denote the labour force devoted to the R&D activity and the effectiveness of research, respectively.

  The accumulation or knowledge function is linear in each of its arguments. New ideas (the variation of \( A_t \)) are produced using the existing stock of knowledge and the labour force devoted to the R&D sector. Thus, this function exhibits constant returns to scale in each factor.

  Following Jones (1999), we assume that a constant fraction of labour is devoted to each sector, that is:

  \[ (1 - \beta)Y_t = L_t \] and \[ \beta L_t = A_t \]  

  with \( 0 < \beta < 1 \).

  Finally, the growth rate of output per worker is given by:

  \[ g_y = \frac{Y_{t+1}}{Y_t} - \alpha \frac{A_{t+1}}{A_t} = \alpha \xi L_t = \alpha \xi b L. \]  

  Thus, there is a scale effect: the larger the economy is, the higher the growth rate is.

- **Non-scale growth models**

  **The Jones/Kortum/Segestrom model**

  By contrast, the accumulation function is given by:

  \[ A_t = \xi L_t A_t^\phi \]  

  where \( \phi < 1 \) is imposed.

  With \( \phi > 0 \) the accumulation function allows for increasing returns to scale in the production of new ideas (knowledge spillover effect), whereas, with \( \phi < 0 \), the accumulation function exhibits decreasing returns to scale ("stepping on toes" effect). The Romer/Grossman-Helpman/Aghion-Howitt models are obtained as a special case when this parameter is set to one.

  If we assume that the growth rate of the labour force is non zero, the growth rate is as follows:

  \[ g_A = \frac{n}{1-\phi} \] and \[ g_y = \alpha g_A = \alpha \frac{n}{1-\phi} \]  

  The growth rate does not depend on the scale effect, but rather on the population growth rate, \( n \), and technological parameters.
3.3.3. Empirical evidence

A large number of studies have been undertaken at the level of individual firms, industries and countries in order to analyse the empirical relevance of R&D activities. As pointed out by Nadiri (1993), a first strand of the literature studies three questions pertaining to R&D investment: (i) Are there diminishing returns to innovation activities? (ii) What is the relationship between R&D spending and growth of output or total factor productivity, and what are the magnitudes in relation to a firm or industry’s R&D investment? (iii) What are the magnitudes of the benefits from R&D, undertaken by other firms and industries? The evidence on the first issue is still controversial (Griliches, 1990, 1992 and Nadiri, 1993) because of the lack of a precise measure of the output of R&D and an adequate measurement of the inputs to the R&D process. The empirical literature considers, however, that innovative activities exhibit increasing returns to scale if proper measures are used and spillover effects of R&D activities are taken into account. The evidence on the second issue suggests pervasive and strong relationships between R&D expenditures and the growth of output or total factor productivity. The estimated magnitudes of the elasticities and rates of return to R&D investment vary considerably depending on the type of data (cross-sections or time-series), the unit of analysis (firms, industries or countries), and the econometric method. At the firm level, elasticity of R&D tends to be around 10% to 30%, and rates of return around 20% to 30%. At the industry level, these values range around 8% and 30%, and between 20% to 40%, respectively (Nadiri, 1993, Table 1; Cameron, 1998, Table 1). In addition, almost all available studies suggest that: (i) the returns to process R&D are different from the returns to product R&D, with process R&D being higher; (ii) rates of return on privately financed R&D are much higher than those on publicly financed R&D; (iii) privately financed R&D and publicly financed R&D are either complementary or substitutes; (iv) the returns to basic research are different from the returns to applied R&D, with the former usually yielding higher returns; and (v) the returns to R&D vary significantly between industries, with R&D in research-intensive sector having higher returns. On the third question, the measurement of R&D spillovers has proved to be quite difficult. Available studies differ in the extent to which they attempt to model knowledge spillovers. The benefits of R&D spillovers are widespread so that each firm will benefit from its own effort as well as the research results of other firms. Patents, scientific literature, technology licences and technology embodied in imported or exported goods provide different ways to diffuse technology throughout the domestic and world economy. In this respect, it becomes difficult to measure all spillover effects associated with R&D. For example, inter-industry and inter-firm spillover effects are difficult to capture and to incorporate into total factor productivity analysis (growth accounting exercises). In this context, in order to measure such spillover effects, the literature uses some proxy for the flow of spillovers based on either input-output tables, patent concordances, innovation concordances or proximity (Nadiri, 1993). In spite of these difficulties, Griliches (1992) argues that the overall impression remains that R&D spillovers are both prevalent and important and social rates of return remain significantly above private rates. This conclusion is also supported by Nadiri (1993), whose survey of empirical evidence points to the social rates of return to R&D close to an average of 50%. Finally, as noted earlier, there are four mechanisms by which R&D causes externalities. These are through the effects of “surplus appropriability”,

---

60 For recent surveys of the evaluation of R&D activities, see Mohnen (1990), Griliches (1992), Nadiri (1993) and Cameron (1998).

61 The empirical literature usually measures the input and output of R&D by expenditures and patents, respectively.

62 A survey of David et al. (1999) argues that the available econometric evidence gives rise to conflicting answers about the link between public and private R&D spending. On the one hand, public R&D spending may be complementary and thus additional to private R&D expenditures. On the other hand, public R&D spending can substitute for and tend to crowd-out private R&D. In this respect, Guelllec and van P. de la Potterie (2002) conclude that public and private R&D are complementary in a sample of OECD countries.

63 For a review of the evidence, see Griliches (1992), Klette, Moeer and Griliches (1999).
“standing on shoulders”, “creative destruction”, and “stepping on toes”. Jones and Williams (1997) derive an endogenous growth model in the spirit of Romer (1990a), which incorporates all four of these externalities. By comparing the estimate of the social return to R&D with the true social return from their model, they provide evidence that empirical estimates surprisingly represent a lower bound on the true social rate of return. Using an estimate of the rate of return to R&D of about 30%, optimal R&D investment is at least four times larger than actual investment in the United States.  

A second strand of the empirical literature studies the impact of technological diffusion across countries, (e.g. international spillovers). The standard literature on convergence assumes that the technology is the same among countries and does not capture the mechanisms of technological diffusion or the different levels of technology. In contrast, in the literature on international spillovers, convergence across countries is assumed to depend on the degree of international technology diffusion. Strong diffusion equates differences in technology across countries and leads to convergence. Conversely, the absence of international technology diffusion leads to divergence across countries. Two basic mechanisms for international economic activities to lead to technology diffusion have been emphasised: (i) direct learning about foreign technological knowledge; and (ii) employing specialised and advanced intermediate products that have been invented abroad. In this respect, two major determinants of successful technology diffusion are: (i) human capital (Nelson and Phelps, 1966); and (ii) R&D activities.

Different approaches are used to test the validity of international R&D spillovers. One approach is found in the empirical literature which studies productivity effects of foreign R&D (with other factors) on domestic productivity with the so-called “international R&D spillover regression”. A production function is used in order to relate the total factor productivity to measures of domestic and foreign R&D activities. The latter is usually defined as a weighted average of other countries’ R&D. By comparing total factor productivity (TFP) elasticities of domestic and foreign R&D, Coe and Helpman (1995) provide evidence in a sample of 22 industrialised countries that these elasticities are significant and important. In particular, they estimate domestic (foreign) R&D elasticity of about 8% (12%) for the 15 small countries. In the case

---

64 A large empirical literature reports estimates of the rate of return to R&D ranging from 30% to 100%. Therefore, Jones’s conclusion is based on a lower bound of the estimated rate of return to R&D in the literature.

65 Alternative views on technological knowledge and diffusion have been exposed by Mankiw (1995), Parente and Prescott (2000) and Quah (2001a, 2001b). In the former, the differences in individual incomes lie in complementary factor accumulation of physical and human capital. However, empirical evidence does not support this idea (Klenow and Rodriguez-Clare, 1997; Hall and Jones, 1999). In the latter, productivity differences are not explained in the availability of rival factors but rather in the actually employed technological knowledge. Finally, Quah (2001a, 2001b) explains international differences by the fact that technological knowledge is disembodied, codified and global whereas human capital is embodied, tacit and local knowledge.

66 Other growth mechanisms that are not related to international technology diffusion can be advanced: (i) trade and domestic monopoly, (ii) trade and knowledge accumulation through learning-by-doing and (iii) trade and endogenous change. In the first case, liberalisation of international trade may reduce monopoly power of domestic firms, which affects the pricing behaviour and thus the R&D sector. In the second case, trade liberalisation affects the domestic resource allocation. For example, Young (1991) develops a two-country model in which the prevailing level of learning-by-doing determines each country’s growth rate. This learning-by-doing is defined as a cost-lowering effect from cumulative production and is high for more recently invented products than for older products. In this case, trade liberalisation may have a negative effect on the growth rate of the poor country. In the third case, Grossman and Helpman (1991) extend their model in an international context and extensively discuss the links between trade and endogenous change.

67 For example, Mohnen (1992) and Coe and Helpman (1992) used import shares as weights.
of the G7 countries, they also find a domestic R&D elasticity of about 23% and a foreign R&D elasticity of about 6%, (i.e. a share of about 20% for foreign R&D). In contrast, Park (1995) estimates that foreign R&D accounts for about two-thirds of the total effect. However, this method gives rise to some important methodological issues (Keller, 2001).

A second approach uses general equilibrium models in which productivity is related to increases in the quality of intermediate goods. Eaton and Kortum (1994, 1997) calibrate a model of international technology diffusion and show that domestic productivity growth depends on foreign technology improvements. Eaton and Kortum (1999) point out that the part of the productivity that is due to domestic (as opposed to foreign) R&D is between 11% and 16% in Germany, France and the UK. This part is around 35% in Japan and 60% in the United States. (However, these models rely on strong assumptions).

A third approach studies the channel of technology diffusion. Three channels have been mainly identified: trade, foreign direct investment and communication patterns (Coe and Helpman, 1995). Coe and Helpman (1995) provide evidence on the importance of trade for international technology diffusion. Using their “international R&D spillover regression”, Coe and Helpman (1995, 1999) show that a country’s productivity is increasing in the extent to which the country imports from high knowledge countries (“import composition effect”) and that international spillovers are related to the overall level of imports. Coe et al. (1997) present analogous results in their analysis of foreign technology differences between highly and less developed countries. By contrast, Eaton and Kortum (1996) find that bilateral imports do not help to predict bilateral patenting activity. Keller (1998) states that the import composition of a country does not have a strong influence in his regression results. In the context of inter-industry technology flows, Keller (1997) also fails to find an import composition effect. However, Xu and Wang (1999) show that the import composition effect is robust when one considers capital goods instead of all manufacturing goods. In the case of foreign direct investment, Lichtenberg and van P. de la Potterie (1996) argue that a country’s outward and foreign direct investment gives access to foreign technology, the reverse (inward investment technology) is not supported by empirical evidence. In contrast, Baldwin et al. (1999) provide evidence of positive inward spillover effects at the industry level. Overall, however, the results are mixed.

A fourth approach concentrates on the geographical localisation of international technology diffusion. In this respect, income convergence depends on whether technology spillovers are local or global. Global technology spillovers are in favour of convergence across countries, whereas geographically localised diffusion makes divergence more likely. Empirical evidence shows that within-country spillovers are much stronger than between-country spillovers (Jaffe et al., 1993; Jaffe and Trajtenberg, 1998; Eaton and Kortum, 1999) and geographical localisation of international technology diffusion appears more often (Keller, 2001a, 2001b). By contrast, Irwin and Klenow (1994) study learning-by-doing spillovers in the semi-conductor industry and find the opposite. However, their results are more related to human capital models than technology-based models.

Some researchers have criticised the research-based growth models on empirical grounds. In an influential paper, Jones (1995a) shows that growth rates in OECD countries have not shown a persistent upward trend in spite of policy changes, as for instance trade liberalisation, increase in investment of human capital, and R&D. In particular, the productivity of research, (at least as measured by realised patents), shows long declines from the mid-1950s until the mid-1980s. Jones points out that this may indicate some systematic relationship between the knowledge stock and research benefits or it may represent an exogenous fall in research productivity. In this respect, Jones (1995a, 1998) criticises the linearity assumption in the knowledge function or knife-edge condition (Box 4: Production and knowledge functions in models with R&D) underlying such R&D-driven growth models, (i.e. the benefits of R&D are

---

68 For further details, see Keller (2002), Eaton and Kortum (2001a, 2001b).
related to the existing stock of knowledge). This in turn leads to the conclusion that R&D-based growth models are rejected by the empirical evidence. However, this critique depends crucially on measurements of input and output of the R&D process. In effect, in order to test the plausibility of the linear accumulation function in standard R&D-based growth models, Jones takes the number of engineers engaged in R&D as a proxy for the level of resources devoted to innovation process and total factor productivity as a proxy for the output of innovations. He then asserts (Jones, 1995a, 1995b) that he can reject the assumption of a unitary learning elasticity since “total factor productivity exhibits little or no persistent increase, and even has a negative trend for some countries, while the measures of $L_A$ (Box 4), exhibit strong exponential growth”. His finding is however based on a poor proxy for the rate of innovation, as measured by the total factor productivity (Nelson and Romer, 1996). In particular, total factor productivity data depend critically on aggregate price indices and these systematically underestimate the impact of quality and variety. Hence, while the contribution of Jones (1995a, 1995b) is important, the linear specification of the accumulation function cannot be ruled out without more detailed empirical work. Backus et al. (1992) also look for the scale effects predicted by endogenous growth models in which dynamic returns to scale arise from the development of new products. They find little empirical evidence between the growth rate of income per capita and the measures of scale used at the national level, but some empirical support in the case of manufacturing sector.

The empirical evidence supports the stylised fact that R&D is an important source of growth. However, due to the complexity of the concepts used and the measurement problems, empirical studies differ in the estimated magnitudes of R&D at the firm level, industry or country level. In addition, the debate on whether policy affects the long-run balanced growth rate of just the steady-state level of income is almost impossible to solve. Given the presence of large scale effects, Temple (1999) point out that distinguishing between endogenous and semi-endogenous R&D models is not as pressing as it might seem. The important point is that macro policies can have a major impact on a country’s level welfare.

3.3.4. R&D-based CGE models

A number of studies have begun to introduce elements of the R&D based growth models into fiscal-oriented policies, trade, and environment/energy CGE models. Macroeconometric models have also attempted to look at the consequences of international spillovers. The recent evolution of computable equilibrium trade models is helping to elucidate the effects of endogenous technical change in the context of macro policies. This evolution can be summarised in four steps. The first generation of CGE models uses a static framework with perfect competition and thus only studies the impact of trade policies using comparative static analysis. The effect of trade liberalisation here was found to be limited. In contrast, the second generation allows for imperfect competition, and scale economies (Harris and Cox, 1982). In this case, trade liberalisation has a larger economic effect. By allowing for trade policy to affect factor supplies, especially the physical capital stock (Baldwin, 1989, 1992), the third generation finds even larger effects of trade liberalisation. Finally, the fourth generation introduces endogenous R&D technical change based on standard R&D growth models. In this case, trade policy affects long-run growth.

---

69 Bayoumi et al. (1996, 1998) extend the IMF’s multicountry econometric model (MULTIMOD) by allowing for externalities in the form of international spillovers of R&D expenditures on total factor productivity. In doing so, total factor productivity becomes endogenous as suggested by the New Growth Theory. The theoretical basis is provided by Grossman and Helpman (1991). However, the extension of Bayoumi et al. does not incorporate a reduced “AK” form, that is, their model incorporates diminishing returns to scale to the reproducible factor of aggregate production, as for instance physical and R&D capital. This implies that a transitory or permanent shift in R&D investment will only have a level effect on output. However, the impact on economic growth is long-lived as it take a long time to converge towards the new steady-state.
Of course the differences between the last two approaches most concern the subject of this paper. In the former, modellers concentrate on induced capital formation, (i.e. “trade-induced investment-led growth”). In the latter, applied researchers focus on the accumulation of knowledge capital, that is, “trade-induced productivity-led growth”. This difference has important effects in terms of trade policies. The same trend also occurs in computable equilibrium environment/energy models (Grubb, Khoehler and Anderson, 2001). The first generation (GREEN, G-cubed, DICE) considers exogenous technical change while the second generation (WARM, GEM-E3) introduces some elements of the R&D-based growth models described in previous sections. The two varieties of models also have different policy implications. In the former, long-run growth was not affected by the introduction of exhaustible resources and pollution. In the latter, the interactions between endogenous growth and environmental issues requires thinking about the various links between the technology “block”, the preference “block”, and the environment “block”. In particular, one has to consider that the relation between economic growth and the environment is not homogenous. There is a rather large amount of heterogeneity depending on the specific pollutant, the natural resources used, and the reversibility of damages done to the environment. These models are more designed to answer the following questions: Is there a trade-off between economic growth and environmental preservation? Is sustainable growth optimal? What is the effect of R&D policy on environment and the effect of environmental policy on R&D and economic growth? In this respect, for example, a simple comparison of simulations about abatement costs for Europe based on exogenous and endogenous growth formulations show important differences in the case of the Kyoto Protocol (Fougeyrollas et al., 2001). We present here some recent R&D-based CGE models in the fields of trade and environment/energy policies. In the light of the previous sections, we discuss the structure, the functional forms used and the results of these models.
Box 5. A common representation of a dynamic CGE model

We present, formally, the general features of a Dynamic Computable General Equilibrium Model as a precursor to discussing their adoption of endogenous growth (Box 6). We consider the case of a closed economy without government and then extend our simple model to introduce international trade and government. There are two sectors in the multi-good economy: the final production/consumption goods sector and the intermediate goods sector. This general representation has been extensively used in the dynamic CGE literature.

**A closed economy without government**

In the first case, the model is described as a set of three groups of equations. The first group determines households’ maximisation program, the second, the program of producers, and the last the market-clearing conditions. These equations are either static or dynamic.

1. The representative household’s program

We consider a representative household/final producer who offers inelastically their labour services and are endowed with one unit of time. His preferences are represented by the following standard intertemporal utility function:

\[ U(C) = \sum_{t=0}^{\infty} \beta^t u(C_t) \quad (1) \]

where \( u(C_t) = \frac{C_t^{1-\sigma}}{1-\sigma} \) is the instantaneous utility function which depends on CES basket \( C_t \) of the different goods \( i : C_t = \sum_{i=1}^{I} \frac{1}{\gamma_i} I_{it}^{1+\gamma_i} \).

We assume that the price of the basket of consumption goods is normalised to one.

The intertemporal budget constraint is:

\[ t \left( t Y_{it} \right) + t K_t + t P_{it} = C_t \quad (2) \]

where \( K_t \) is the only available asset in the economy, \( Y_{it} \) is the total output in the final goods sector, \( Y_{it} \) is the total output of the intermediate sector and \( P_{it} \) is the relative price of intermediate goods.

Thus, the representative agent maximises intertemporal utility (1) given the intertemporal budget constraint (2). In doing so, one obtains static and dynamic first order conditions.

The first set of conditions are static and intra-temporal:

\[ \frac{C_i}{C_j} \frac{\partial u}{\partial P_{ij}} = \frac{P_{ij}}{P_{ij}} \quad \forall i, j = 1, \ldots, I \text{ and } i \neq j \quad (3) \]

These conditions imply that the marginal substitution rate between two goods equals the ratio of their prices.

We define the following price index of consumption goods, which is normalised to 1, in order to verify the aggregation condition \( P_{it} C_t = \sum_{i=1}^{I} p_{it} c_{it} = 1 \):

\[ P_t = \left( \sum_{i=1}^{I} \frac{\partial z}{\partial P_{it}} \right)^{-\gamma} \quad (4) \]

The second optimal condition is dynamic, it is the standard Keynes-Ramsey condition on the basket of consumption goods between \( t \) and \( t + 1 \):
Thus, the forward change of consumption between two adjacent periods can derived as a function of the real interest rate. A high \( r_{r+1} \) implies large future returns from saving/investment so consumption increases because of high returns.

2. The programs of the producers

There are two sectors of production: a final production/consumption sector, indexed by “f”, which is competitive and an intermediate goods sector, indexed by “int”, which is characterised by an imperfect competition environment.

(i) Sector of final consumption goods:

This sector employs three inputs: labour, capital and intermediate goods. The number of intermediate goods is set to \( A \). In the case of R&D-based growth models, this number is not exogenous: the number of new intermediate goods is governed by a knowledge function (Box 6).

The technology is described by a Cobb-Douglas function for each consumption good i: \( Y_i = L_i^\alpha K_i^\beta \sum_{a=1}^{A} X_{i,a}^\delta \). Each production function exhibits constant returns to scale, that is, \( \alpha + \delta = 1 \). In addition, the elasticities of intermediate goods to final consumption goods are the same.

Without adjustment cost of capital the program of the firm is static, the firm maximises its instantaneous profit under technological constraint. The first order optimality conditions give the factor demands which depend on the factor prices, the wages \( w_i \), the interest rate (or rental price) \( r \) and the prices of intermediates goods \( w_{a,i} \).

\[
\begin{align*}
(\text{labour}) & \quad \alpha_i \frac{Y_{i,j}}{L_{i,j}} = \frac{w_i}{p_{i,j}} \\
(\text{capital}) & \quad \alpha_k \frac{Y_{i,j}}{K_{i,j}} = r_i + \delta \\
(\text{intermediate}) & \quad \alpha_{\alpha} \frac{Y_{i,j}}{\sum_{a=1}^{A} X_{i,a}^{\alpha_{a,i,j}-1}} = \frac{w_{a,i}}{p_{i,j}}
\end{align*}
\]

Equations (6) and (7) are familiar but but equation (8) implies that the marginal product of each intermediate good is equal to its price.

(ii) Sector of intermediate goods:

To analyse the behaviour of the producers in the intermediate goods sector, we use the set-up of an imperfect competition à la Dixit-Stiglitz. This assumption has been extensively used in endogenous growth and CGE models. Thus, the competition is monopolistic: intermediate goods are imperfect substitutes for the firms in the final sector. We assume that a firm must pay an ex-ante fixed sunk cost (patent’s cost) in order to produce an intermediate. Following Romer (1990), the patent represent a barrier to entry: a new intermediate can only be produced by one firm. The firms here only employ a capital good with a one-to-one technology. The demand curve can be expressed as \( w_{a,i} = X_{a,i,j}^{\alpha_{a,i,j}-1}N_i \) where \( X_{a,i,j} = \sum_{i=1}^{A_i} X_{i,a,i,j} \) and \( N_i \) is the number of firms. Each firm acts as a monopolist so that the price \( w_{a,i} \) is the product of the mark up (>1) and the marginal cost. The latter is equal to the rental price of capital, \( r \), due to the one-to-one technology used here. Thus,
3. Market-clearing conditions

The equilibrium is defined by sequences of $c_{t,i}$, $K_t$ for each date $t$.

(i) Intermediate goods markets:

The right hand side of the equation (9) does not depend on $a$, so that, the Nash equilibrium is symmetric: $P_{it} = w_{it}$.

(ii) Capital market equilibrium:

The capital demand of firms, $K^d = \sum_{i=1}^{i=I} K_{it} + \sum_{a=1}^{a=A} K_{at}$, is equal to the capital supply by the household $K_t$. This determines the rental price of capital and the marginal productivity of capital is obviously the same across sector.

(iii) Labour market equilibrium:

The demand of labour, $L^d = \sum_{i=1}^{i=I} L_{it}$, is equal to the labour supply $L_t$. This determines the wage, which is the same across sectors.

(iv) Goods markets:

The demand for each good $i$ is equal to each good supply. This determines the price of each goods.

4. The foreign sector

To capture the role of imports in the economy, we introduce the standard Armington’s specification (1969). Imported goods are considered to be imperfect substitutes in consumption for the domestically produced goods. Thus, the firms can use intermediate goods which are produced by domestic and foreign (imported goods) producers. The representative household can consume goods which are distinct by their country origin. The home country’s demanders choose domestic and foreign goods into a composite for each product at a minimum cost according to a constant elasticity substitution function (a CES function):

$$C_i = \Lambda \left[ \nu M_i^\varepsilon + (1-\nu) D_i^\varepsilon \right]^{\frac{1}{\varepsilon-1}} (10)$$

where $C_i$ is now the "composite good", $M_i$ is the imported good, $D_i$ is a domestic good, produced and consumed domestically, $\Lambda$ is a parameter, $\nu$ is the share of foreign goods and $\varepsilon$ is the elasticity of substitution.

The sub-program of the consumer consists in minimising $P_{it}D_i + (1+t_{im})P_{mt}M_i$ subject to equation (10), where $P_{it}$ is the price of the good $D_i$, $P_{mt}$ is the price of the good $M_i$ and $t_{im}$ is a tariff.

The demand from foreign countries for each good produced in the home country is aggregated and derived from a constant elasticity of transformation function (a CET function):

$$Y_i = \Gamma \left[ \mu EX_i^{\eta} + (1-\mu) D_i^{\eta} \right]^{\frac{1}{\eta}} (11)$$
where \( Y_i \) is the final production, \( EX_i \) is the good exported, \( D_i \) is the good produced and consumed domestically, \( \Gamma \) is a parameter, \( \mu \) is the share of the foreign good and \( \eta \) is the elasticity of substitution.

So there is a second program for the firms \( i \) in the final consumption good sector \( i \) which consists in maximizing \( P_D D_i + P_{EX} EX_i \) subject to equation (10), where \( P_{EX} \) is the price of the good \( EX_i \).

5. Government

The introduction of government in many CGE models is simplistic. The government’s expenditures often are not arguments in the utility function or productive factors for firms. In models where they do appear we have the possibility of studying fiscal policy by introducing taxes: consumption taxes, labour taxes, capital taxes, incomes taxes, tariffs. Revenues from these taxes can be devoted to government expenditures or redistributed to households as a lump sum. If, we assume that the government provides public goods and services at an exogenous level \( \bar{G} \) (indefinitely), and we assume that \( \bar{G} \) enters the utility function in an additively separable manner, there is no need to model the impact of public provisions on consumer well-being (its impact is isolated). However, we still have to account for the cost of providing public sector services through the introduction of taxes. When tariffs, consumption and output taxes are the only sources of revenue, the public sector constraint can be written as:

\[
P_{G_i} := \sum_{i} m_{i,t} M_i + t_{C,t} C_i + t_{Y,t} Y_{i,t} \quad (12)
\]

The left hand side represents the cost of public expenditures and the right hand side represents tax revenue from tariffs, consumption taxes and final output taxes, respectively.

A few dynamic models impose inter-temporal closure by imposing inter-temporal budget constraints on agents in the model. The more sophisticated models impose such constraints on consumers, firms, governments and even countries as a whole. When combined with forward-looking behaviour, such models are relatively complex because financial assets must be consistent with future changes imposed both by policy as well as the closure conditions.

**Steady-state reference**

The calibration of a CGE model begins with the assumption that the data are obtained from an economy in some type of “equilibrium”. In the static version of the model, this is usually an equilibrium point. In our simple version of a dynamic model, one needs to assume that the entire path of the reference case represents an equilibrium or steady state of the economy. Hence, parameters are calibrated for this reference case. This ensures that the model will generate an equilibrium solution that is compatible with the benchmark data of the studied economy. In most applications, the balanced growth rate is assumed to be zero. Specifying a positive growth rate requires elimination (or detrending) the exogenous, balanced-growth trends in the reference run. Finally, numerical models can only be solved for a finite number of periods; hence, some adjustments are required in order to approximate an infinite horizon problem with a finite horizon solution. A standard approach consists of imposing steady-state conditions at some future terminal period. On the one hand, as long as the transversality condition is satisfied, (i.e. the discount rate and the rate of time preference are positive and greater than the balanced growth rate by the terminal condition), the demand side does not require further conditions. On the other hand, we need to impose at least one condition on the supply side. One natural condition is to suppose that the investment-capital ratio satisfies a certain relation, which is given by the resolution of the model, at the terminal periods for all simulations. This in turn requires the choice of the terminal date and to test the robustness of the numerical model when specifying different terminal dates. In the case of exogenous growth models, this approximation does not matter, and different methods are now available in the literature (Lau et al., 2000). In the case of endogenous growth models, the methodology is the same, but also depends on the determination of the balanced endogenous growth trend.
The seminal contributions involving R&D-based CGE models are Diao and Roe (1997), Diao et al. (1999), Rutherford and Tarr (1998, 1999), Baldwin and Forslid (1999). These models aim at studying the effects of trade policies. Due to the large number of equations, we only present here the general representations of these models and highlight the major differences. First, these models differ in terms of the degree of disaggregation and the economy studied. The Diao-Roe model is calibrated to US level data. Their model contains four final sectors: agriculture and food processing, mineral and materials, manufacturing, and services. Each sector is assumed to produce a single output using inputs of two non-augmented factors – differentiated capital, (which accumulates), and a set of other intermediate goods. In contrast, the Diao-Roe-Yeldan model is specified and calibrated to data from Japan. Trade partners are aggregated into three countries/regions: the United States, Europe and the Rest of the World. Their model contains seven final output sectors and an R&D sector (intermediate goods). Each sector produces a single output using units of labour, engineers, a non-human resource and a set of differentiated capital goods and other intermediates. The Rutherford-Tarr model is a two-sector model. The final goods sector uses three inputs: labour, capital and an intermediate good. The intermediate good is produced by both foreign and domestic firms in a second sector. Finally, the Baldwin-Forslid model is a multi-sector, multi-good and multi-factor version of an endogenous growth model. There are four regions, three standard factors of production (skilled labour, unskilled labour and capital) and two types of goods (differentiated products and perfect competition goods). Each region produces a non-traded perfect competitive good and several varieties of differentiated goods. Each region has a R&D sector. The model is however calibrated to a notional data set.

Second, while these models have in common the introduction of a general equilibrium with R&D driven endogenous growth, they differ in terms of the formulation of the externalities and the specification of the accumulation function. The Diao-Roe model assumes that the accumulation of new designs is proportional to accumulated knowledge (Romer-model). Thus, the R&D production function (accumulation function) exhibits constant returns to scale to the two primary factors (labour and capital), chosen by the firm engaged in the R&D process. The firm takes the stock of knowledge as given, whereas at the aggregate level, the existing stock of knowledge depends on the levels of past additions to knowledge. Thus, the production function exhibits overall increasing returns to scale. The linear specification of knowledge in the accumulation function gives rise to the standard spillover effect à la Romer. Finally, in order to link the R&D sector to producers of intermediate goods, Diao and Roe (1997) use the Romer-scheme, that is, each new design is associated with only one firm in the sector of intermediate goods – there exists an ex-ante fixed cost associated with one type of differentiated capital variety and the number of firms equals the number of designs. By contrast, the Diao-Roe-Yeldan model accounts explicitly for domestic and cross-borders technological spillovers. In the R&D sector, each variety of heterogeneous capital is developed and produced by an individual firm. Firms also have to pay an initial up-front cost. To conduct R&D, the firm employs three primary factors of production: labour, engineers and a non-human resources. The R&D accumulation function exhibits constant returns to scale to all primary factors and is linear in two additional factors: a productivity coefficient and the existing stock of knowledge. These terms capture international and domestic spillovers, respectively. Based on the works of Coe et al. (1997) and Wang and Xu (1997), Diao et al. (1999) assume that cross-border spillovers are generated through the imports of investment goods. Thus, they suppose that the productivity coefficient is a function of an international spillover coefficient and the initial level of technological coefficient in the R&D production. The international spillover coefficient is defined as a function of the

---

70 For a general representation of dynamic CGE models, see Box 5. 
71 The model is specified and calibrated for Tunisia. In fact, the intermediate good sector is built as an aggregate of intermediate sectors in national accounts. 
72 The two primary factors of production are mobile among sectors but immobile internationally. 
73 For a discussion of empirical results, see section 3.3.
import-weighted pool of foreign technological knowledge. A linear specification in this coefficient is used in order to ensure convergence to the steady-state. Thus, the R&D accumulation function allows links foreign countries’ technological knowledge to domestic R&D activities. The **Rutherford-Tarr model** also uses the Romer-model and Grossman-Helpman model. Using the work of Keller (1995), they assume that an increase in the number of foreign varieties has the spillover of reducing the costs of blueprints for domestic firms. They also make a special assumption about the life-cycle of products and firms: the number of vintage firms that produce at late dates declines exponentially at a rate equal to the depreciation rate of physical capital. In the **Baldwin-Forslid model**, the growth rate of the knowledge capital is assumed to depend on the labour force devoted to each R&D sector (scale effect) and an international externality. The extent to which externalities occur internationally is regulated by an exogenous parameter.
Box 6. Examples of Dynamic Computable General Equilibrium Model with R&D-driven growth

We present here different ways to model the R&D sector in a CGE model. We use the general framework presented in Box 6.

A simple version

In the simple model presented in Box 5, we have assumed a constant number of patents which is exogenous. Here we introduce a Research and Development sector which creates new blueprints. The number of researchers is equal to $L_2$. This sector is competitive and the existing stock of blueprints give rise to externalities. The knowledge function summarises both scale and non-scale R&D-based growth models.

The number of new blueprint is given by:

$$A_{t+1} - A_t = eL_{2,t}A_t^\phi$$  \hspace{1cm} (1)

where $e$ represents the effectiveness of research.

If $\phi=1$, the growth rate will be endogenous, whereas, if $\phi<1$, the growth rate will be semi-endogenous. In the latter, the total population must grow in order to obtain a positive growth rate.

The non-profit condition in this sector determines the instantaneous wage. So the necessary equality between the wages of the final consumption sector and the research sector gives the optimal allocation of workers between each sector. In addition, $L = \sum_{i=1}^n L_{1,i} + L_{2,i}$.

A simplified version of Baldwin and Forslid (1999)

The Baldwin-Forslid model assumes that the production of new knowledge is as follows:

$$1^{/2} + (1 - \lambda) tA_t + \lambda - g = \frac{A_{t+1}}{A_t} - 1 = L_2 (1 + \lambda)$$  \hspace{1cm} (2)

where $\lambda$ measures the extent to which externalities occur internationally.

In particular, $\lambda = 1$ or $0$ indicates perfect or zero international spillovers, respectively.

A simplified version of Dia and Roe (1997)

The Diao-Roe model uses the following representation of the R&D sector:

$$A_{t+1} - A_t = g(L_{1,t}, B_t)A_t$$  \hspace{1cm} (3)

where $L$ and $B$ are the two primary factors of the R&D sector and $g$ is a constant returns to scale Cobb-Douglas function.

Firms pay a fixed cost of new design before production takes place. The total number of firms equals the number of designs A. There is monopolistic competition in factor market for differentiated capital.

A simplified version of Diao, Roe and Yeldan (1999)

The Diao-Roe-Yeldan model accounts explicitly for domestic and cross-border technological spillovers. The knowledge function is described as follows:

$$A_{t+1} - A_t = e, g(L_{1,t}, G_t, B_t)A_t$$  \hspace{1cm} (4)

where $L, G$ and $B$ are the three primary factors of the R&D sector, $g$ is a constant returns to scale Cobb-Douglas function and $e$ is a productivity coefficient.

It has to be noted that this coefficient is constant in equation (1). A simple manner to introduce additional spillovers into the R&D sector is to endogenise the efficiency parameter $e$ of this sector (see equation 1). In fact, this parameter could depend on cross border spillovers in an open economy due to “the import–weighted pool of foreign technological knowledge”:

$$e_t = \left(1 + \xi_t \right)\bar{e}$$  \hspace{1cm} (5)

where $\bar{e}$ is the initial level of technological coefficient in the R&D production function and $\xi(t)$ is the international spillover coefficient:

$$\xi_t = \sum_r \omega_r A_{rt}$$  \hspace{1cm} (6)

where $\omega$ is the spillover elasticity in domestic R&D production function, $\omega_r$ is the share of investment good imports from country $r$ and $A_{rt}$ is the stock of R&D of the country $r$ at date $t$. 

49
Third, these CGE models differ in terms of the restrictions imposed on coefficients so as to ensure a balanced growth path\textsuperscript{74} and the methodology used to determine the coefficients. Several approaches have been used for parameter estimation: calibration and econometric evaluation. The calibration approach focuses on the selection of a particular functional form and its associated parameter. The parameters to be calibrated are determined by forcing the model to replicate a chosen benchmark year. While this is a standard approach in CGE models, it gives rise to additional problems in the case of endogenous technical change. Moreover, calibration alone can not determine all parameter values when certain functional structures are employed, as for instance substitution parameters in the CES structure. In addition, there is a risk that multiple equilibria may exist, giving the result that different values of the same parameters lead to different long-run equilibrium growth paths. The econometric approach can be used to alleviate this problem by empirically deriving the value of certain parameters. (Subject to the caveat that, as was pointed out before, the empirical evidence on endogenous growth models is mixed.) For example, Diao et al. use the Coe and Helpman regressions in order to calibrate the R&D productivity elasticities with respect to foreign and domestic R&D stock. These estimated TFP elasticities vary in the empirical literature. For different spillover elasticities, the simulation outcomes may be different.\textsuperscript{75} Finally, numerical models can only solve for a finite number of periods, hence some adjustments are required in these models to produce a good finite horizon approximation of the infinite horizon model. For example, Baldwin and Forslid (1999) use the q-approach to characterise the steady-state growth rate, whereas Rutherford and Tarr use a method developed in Lau et al. (2000).\textsuperscript{76}

Other R&D-based CGE models have been developed in the environment/energy literature. Den Butter and Wollmer (1996) use a nested constant elasticity of substitution production function incorporating capital with spillovers into human capital. Van Bergeijk et al. (1997) consider nested CES production functions in which both investment and R&D expenditures have spillovers into human capital. Nordhaus (1999) introduces induced innovation in an updated version of the DICE model. Nevertheless, the major conclusion of Nordhaus is that endogenous technical change is likely to be a less powerful factor influencing climate change policy than the substitution of energy by capital and labour. Wing (2001) also develops a detailed sectoral model for the United States based on induced innovation. Technological knowledge accumulates in response of the arbitrage of households between quasi-rents from physical and R&D investments. Carraro (1997) and Carraro and Galeotti (1997) assume that changes in the polluting and non-polluting capital stock are dependent on technical change. But only two forms of capital are considered. In fact, only the recent work of Fougeryrollas et al. (2001) have only introduced an operational framework for innovation in GEM-E3 model. They take into account the important effects of inter-sectoral and international diffusion of knowledge to sectoral technology generation. In a given sector for a particular country, external knowledge comes from four different sources: intra-sectoral and inter-sectoral, of other national or international origin.

Thus, the introduction of an R&D sector in dynamic models gives rise to a number of issues. First, modellers must be confident that the economic structures assumed actually reflect the reality. However, as is explained before, the empirical evidence is not clear and available econometric evidence tends to show mixed results about the assumptions and predictions of the R&D-driven CGE models. In particular, these models use the same technical tricks and functional forms of the new growth literature. In the light of both

\textsuperscript{74}. For example, Diao et al. (1999) assume that the share parameters for differentiated capital in the value-added functions have to be the same for all sectors. However, this assumption lacks empirical evidence.

\textsuperscript{75} In this respect, Diao et al. use two initial values of the elasticity parameter and compare the outcomes.

\textsuperscript{76} Baldwin and Forslid use the Tobin’s q-theory of investment, which focuses on the ratio of a firm’s stock market value to the replacement cost of capital. Thus, they determine the present market value of the firm. In contrast, Rutherford and Tarr (1998) use some specific terminal conditions and approximate the infinite horizon problem into a finite one.
empirical and theoretical problems, it is therefore not possible to obtain a clear conclusion about the impact of macro policies on the growth rate of the economy. **Second**, existing CGE models typically use a low level of disaggregation. In particular, the models considered here specify only one R&D sector. One may reasonably think that each type of industry or sector has its own R&D sector. In addition, the nature of this R&D activity may differ across sectors. For example, some R&D sectors may use non-renewable resources. **Third**, the calibration procedure and defining forward-looking algorithms are often difficult. For example, as pointed out by Baldwin and Forslid (1999), the trade-growth literature finds that productivity in OECD countries proceeds at a steady rate in the post-war period, whereas global openness is trended upward. This means that the degree of openness and the growth rate can not be monotonically linked in CGE models. However, this assumption is usually made in the R&D-driven CGE models described in the previous sections. **Fourth**, the economy may be characterised by multiple stable equilibria. This is troubling because comparisons of macro policies will appear to be arbitrarily determined in such circumstances – unless the boundaries between the equilibria can be rigorously established. In some cases, the indeterminacy of the equilibrium can be solved locally when the equilibrium satisfies a local "saddle-point" property. Nonetheless, for many solution algorithms this requires linearisation of the dynamics of the economy around a locally-determined steady-state, and imposition of restrictions on eigenvalues of the Jacobian matrix. **Fifth**, the models reviewed here integrate one among other factors of growth. To our knowledge, a model that incorporates both human capital and R&D processes, has not yet been introduced in the CGE literature. However, it may appear reasonable to consider both human capital accumulation and knowledge accumulation when one models the R&D sector. Finally, when modellers are interested in the welfare impact of macro policy changes as well as in its impact on the growth rate, calculations need to be made of the transition path and the steady-state, neither of which are commonplace.
4. CONCLUDING REMARKS

The new growth literature provides new insights into the sources of growth and explains how to endogenise different factors of growth (e.g., human capital and R&D). In particular, the R&D-driven growth models are of particular interest when one wants to take care of environmental issues. However, both theoretical and empirical considerations lead to mixed results. On the one hand, endogenous growth models use either some specific linear forms or some knife-edge conditions. This gives rise to the question of the robustness of results. On the other hand, the empirical literature has not yet been able to clearly distinguish between competing models and magnitudes of the growth sources vary between studies. Finally, the question of macro policy effect on either the steady-state balanced growth rate or the steady-state values seems to be impossible to solve. However, policymakers may be much more concerned by medium-term analysis. In this respect, the endogenous growth framework might provide some useful and interesting results.

At this stage, the direct application of the endogenous growth theory in CGE models depends on policies variables that are difficult to measure and on models that have predictions relying on controversial parameterisations. Further progress can be made to model and quantify growth in policy models. This is not to say that economists should lessen their efforts to model endogenous growth in CGE models. First, previous attempts to incorporate R&D-based endogenous growth models in dynamic CGE models provide some interesting results, especially in the case of environmental issues. Second, standard accounting exercises show that over the three majors contributors of economic growth – labour, capital and total factor productivity – the latter is non-negligible. Thus, we need to provide policy insights into this economic force affecting growth. Third, progress can be made with better measures of factor inputs and policies variables and greater synthesis of theory and data. In particular, the quantitative theory approach can be used instead of the econometric approach, in order to discriminate among competing models and their predictions.
REFERENCES


Baldwin R. and R. Forslid, “Putting growth effects in computable equilibrium trade models”, in ...


Solow R.M. (1990), Sienna Lectures on growth theory.


