Specialization, Knowledge Dilution, and Scale Effects in an IO-based Growth Model

by

Pietro Peretto
Department of Economics
Duke University

and

Sjak Smulders
Department of Economics and CentER
Tilburg University

Abstract
We present a model where accumulation of non-rival knowledge drives growth but where the scale effect, which may be positive or negative, vanishes asymptotically. This result stems from the interaction between technological differentiation and market structure dynamics. Firms are linked to each other in networks of spillovers determined by the technological proximity of their activities. These spillovers-networks span only a fraction of the total economy and the average technological distance between firms increases with the size of the economy. When the economy expands, less related activities become profitable and specialization increases. As a result, the networks expand at a slower pace than the overall economy. In the limit, the networks cease to grow with the size of the economy. A larger economy, therefore, accumulates a larger total stock of knowledge but not necessarily a larger effective stock of knowledge that is useful to the individual firm. The reason is that the latter expands with the size of the network to which the firm belongs. The scale effect vanishes asymptotically because the effective stock of knowledge that each firm exploits is unrelated to the size of the economy when this is very large.

This version: January 7, 1998

JEL code: O40

Key words: growth, specialization, scale effects, spillovers, innovation, entry.

SRN fields: macroeconomics, industrial organization and regulation

Correspondence: j.a.smulders@kub.nl or peretto@econ.duke.edu

We would like to thank Jan Boone, Theo van de Klundert, and Richard Nahuis for useful comments on an earlier version. Smulders' research is financed by the Dutch Science Foundation (NWO).

1. Introduction

One of the puzzles in the theory and empirics of endogenous technological change is the so-called *scale effect*, a positive relation between the size of an economy and its rate of growth. Models in which growth is driven by the accumulation of non-rival knowledge predict that larger economies grow faster because (a) they have more resources to devote to knowledge creation and (b) the larger scale on which non-rival ideas can be applied raises the returns to innovation. This prediction is difficult to reconcile with empirical evidence. In their study of the cross-sectional evidence, Backus, Kehoe and Kehoe (1992) find that GDP growth is not related to the scale of the economy, although TFP growth in manufacturing is positively related to the scale of the manufacturing sector. In his study of the time-series evidence, on the other hand, Jones (1995) finds that the behavior of TFP growth and R&D investment in the manufacturing sectors of OECD countries is inconsistent with models that exhibit scale effects.

We present a model in which non-rival knowledge drives growth without necessarily inducing the scale effect. In particular, the scale effect may be positive or negative but always vanishes asymptotically. What distinguishes our solution from several others, discussed below, is that our model is consistent with the microeconomic evidence on (a) the characteristics of R&D processes and knowledge spillovers within firms and industries and (b) the role of market structure in shaping these processes.

Central to our solution is the interaction between two elements: (a) technological differentiation and (b) market structure dynamics. In our model, technology is firm-specific in the sense that firms have unique and privileged knowledge of products and processes that they accumulate over time through systematic and continuous R&D activity. In other words, firms design products, run production processes, and undertake R&D in-house in order to improve product quality and reduce production cost. In this environment, spillovers across firms are limited because firms accumulate differentiated knowledge. Although knowledge is non-rival, the usefulness of spillovers decreases with the *technological distance* between the creator of knowledge and the receiver of spillovers (Jaffe, 1986). Firms are linked to each other in networks of spillovers determined by the technological relatedness, or proximity, of their activities. These *spillovers-networks* span only a fraction of the total economy. We posit, in particular, that average technological distance increases with the size of the economy. This is where the second element in our model, market structure dynamics, comes into play. Existing knowledge can be used to improve existing technologies or to create new ones. When the economy expands, spillovers-networks expand as well but, because less related activities become profitable and specialization increases, they expand at a slower pace than the overall economy. In the limit, the size of the networks ceases to grow with the size of the economy. This is the mechanism that drives our model. A larger economy accumulates a larger total stock of knowledge but the effective stock of knowledge that is useful to an individual firm expands with the size of the network to which the firm belongs. The scale effect vanishes, therefore, because the effective stock of knowledge that each firm exploits is unrelated to the size of the economy when this is very large.

It is useful to discuss this mechanism in some detail. An increase in the number of technologies affects both productivity and levels of R&D efforts. For a given level of aggregate R&D expenditures, an increase in the number of firms reduces the returns to R&D since the existing pool of knowledge is now exploited for more diverse purposes. Following Adams and Jaffe (1996), we label this effect the *dilution* of knowledge. Moreover, entry of new firms generates *dispersion* of total R&D activity over a larger number of R&D projects. A large economy supports high aggregate R&D activity and a large knowledge stock but it also supports a large number of firms with differentiated technologies and innovation paths. Dilution of knowledge and dispersion of R&D resources reduce the productivity and intensity of R&D at the firm level and may offset the positive effect that the large aggregate volume of R&D acitivity should have on growth. A corollary to this result is that low entry costs (relative to firm-specific R&D-costs) reinforce the basic mechanism and mitigate scale effects. Indeed, very favorable entry conditions may result in negative scale effects.

Microeconomic research on innovation, spillovers, and market structure provides strong evidence for the building blocks of our model. In Section 2 we relate the ingredients of our model to the stylized facts that arise from the empirical evidence. Section 3 confronts our model with existing growth models. We sort out the crucial assumptions that generate scale effects and show how they can be relaxed. In Section 4, we present an overview of the model that we specify in section 5. We characterize the steady state in Section 6. We discuss our results on the scale effect in Section 7.

2. In-house R&D, Entry, and Scale: Microeconomic Evidence

The concept of technological distance is widely applied in microeconomic empirical studies on R&D and knowledge spillovers. These studies, summarized in Griliches (1992), robustly find an important role for spillovers in the R&D process. Total R&D expenditures (within the industry or agggregated over related industries) affect the productivity of a firm's own R&D expenditure. However, the relevant spillover variable is a *weighted* sum of R&D expenditures (or cumulated stocks, as a proxy for knowledge). Firms borrow different amounts of knowledge from different sources according to technological and economic distance from them. The weights used to construct the spillover variable measure the effective fraction of knowledge that is borrowed. Jaffe (1986), for example, uses the distribution of firms' patents over patent classes to characterize their relative positions and construct measures of technological distance. Griliches (1990, p. 1698) summarizes the importance of technological distance as follows:

> To the extent that an invention either reduces the cost of production or develops entirely new products, it has an aspect of increasing returns to it. The same invention could produce the same proportional effect, in different size markets or economies. The public good nature of most inventions and the "multiplicative" aspect of their impact do not require their number to grow just to sustain a positive rate of productivity growth. On the other hand, economies do not grow just by

replication and expansion; they also get more complex, proliferate different products and activities, and develop in different geographical and economic environments. To that extent, the "reach" of any particular invention does not expand at the same rate as the growth of the overall economy, but only at the rate of its "own" market.

Empirical evidence of dilution of R&D is documented in Adams and Jaffe (1996). They match firm-level R&D expenditures to plant-level total factor productivity. They find that total firm-level R&D expenditure affects plant-level productivity, but the number of plants negatively affect these intra-firm spillovers. Also, spillovers from technologically related firms depend on R&D expenditure per plant rather than total expenditures. Adams and Jaffe consider the total number of plants in the industry as an exogenous variable. We argue that further insights can be gained by endogenizing market structure. In our model, we incorporate costly entry that allows us to investigate to what extent dilution and dispersion of R&D endogenously offset the scale effect.[1]

As mentioned in the Introduction, innovation is a firm-specific activity.[2] As a consequence, firm-level characteristics determine growth. We are interested in reconciling the empirics of economic growth with micro-level evidence on the determinants of innovation. While, as noted above, positive scale effects on growth are strongly disputed, empirical studies unambiguously find positive scale effects at the firm level. Cohen and Klepper (1996) review the literature of the last 30 year on R&D and firm size. They conclude that (a) the likelihood of a firm reporting positive R&D rises with firm size, and (b) R&D rises monotonically with firm size among performers of R&D. Larger firms benefit from larger output levels over which the fruits of R&D can be spread. This mechanism of *cost-spreading* determines the positive relation between firm size and returns to R&D.

In our model, we distinguish between innovation and entry. We thus allow market structure dynamics to play a more crucial role (viz. that of dilution and dispersion explained above) than in most models of innovation-driven growth where innovation is necessarily associated with the formation of new firms. The empirical evidence summarized by Dosi (1988) reveals that technology often improves in an incremental and gradual way within existing firms that build on their own history. The introduction of new technologies is not generally associated with entry of new firms, like in variety-expansion models (e.g., Romer 1990; Grossman and Helpman 1991, Ch. 3), or with turnover of firms, like in quality-ladder models (e.g., Aghion and Howitt; Grossman

---

[1] Alternatively, existing firms may diversify into related product lines which generate spillovers to the core activities. Diversification driven by internalization of knowledge spillovers among different product lines is studied by Jovanovic (1993) and, in growth models, by Smulders and van de Klundert (1995) and Tse (1997).

[2] In our model, there is no trade in patents. Levin *et al*. (1987) document that the role of patents is indeed modest. Secrecy and tacitness of knowledge are more effective ways to appropriate the returns to innovation. In addition, absorption of knowledge spillovers requires own R&D effort.

and Helpman 1991, Ch. 4).[3] While many innovations occur without entry, entry often requires innovation. If entry is costly, an entrant has to introduce a technology (or marketable product) that is different from that of incumbents, to be able to recoup the entry cost. Innovations introduced by entrants have more radical effects than incremental inhouse R&D: they affect the technology structure of the economy and the pattern of specialization. Moreover, entry affects competition and market structure which affect incentives for innovation. This suggests that entry and innovation should be regarded as distinct, but related, phenomena.

Geroski (1994, Ch. 5) documents the relation between innovation and entry. After controlling for fixed industry effects, he finds that in the time-series dimension entry has a negative effect on innovation while innovation has no effect on entry. Hence, innovation does not open opportunities for entry. This supports our view that explicit consideration of market structure requires to abandon the view that there is a one-to-one correspondence between the formation of new firms and innovation. Geroski also finds that across industries high innovation is correlated with low barriers to entry. Our model predicts the same. We assume that entrants build on the same knowledge pool as incumbents. Industries with large spillovers provide both high technological opportunity for incumbents and low barriers to entry for entrants.

3. Related Growth Models

R&D-based endogenous growth models that feature scale effects (e.g., Romer 1990; Grossman and Helpman 1991, Chs. 3-4; Aghion and Howitt 1992) have two elements in common. First, R&D generates new, unique ideas. In particular, no systematic duplication occurs and diffusion does not require resources: general-purpose knowledge diffuses costlessly while blueprints are traded in the patent market at no transaction cost. Second, the economy's total stock of knowledge determines R&D productivity in a linear fashion.

There are several growth models that do not exhibit scale effects. In most of them the scale effect is removed by departing from the two assumptions mentioned above. First, the scale of the economy is related to the number of agents, either consumers/producers or firms, that incur costs to accumulate their own agent-specific knowledge. Second, productivity of agent-specific learning depends on some measure of accumulated knowledge that is independent of the scale of the economy. This independence may stem from the assumption that (a) spillovers among agents are absent, or that (b) spillovers depend on average knowledge rather than total knowledge, or that (c) spillovers occur only among a given number of agents.

[3] Malerba and Orsenigo (1995) and Malerba, Orsenigo and Peretto (1997) show that industries in which persistent innovation by existing firms is the dominant mode of technological advance are much more numerous than industries characterized by creative destruction. Thompson and Waldo (1994) build on this evidence and construct a growth model of "trustified capitalism" in which neither entry nor creative destruction occur and in which a given number of firms undertake R&D in order to compete for market shares. In our model, the number of firms plays a crucial role and is endogenous.

Following the terminology recently introduced by Jovanovic (1995), we can say that models with scale effects focus on innovation while models without scale effects focus on adoption. Jovanovic argues that the bulk of expenditures on knowledge acquisition consists of adoption costs rather than the costs of inventing new ideas. If spillovers in adoption are small, this is sufficient to dwarf scale effects. Jovanovic, in particular, shows that Lucas' (1988) model of growth driven by human capital accumulation can be interpreted as a model in which invention and adoption are inseparable. No scale effect arises because each agent has to accumulate his own human capital and the productivity of human capital accumulation depends on the average human capital in the economy. In other words, each agent duplicates all research.

Young (1995, Section 2) focusses on knowledge accumulation by firms rather than workers. The productivity of knowledge accumulation is independent of the scale of the economy as it depends on the productivity of the most productive firm only. A larger economy consists of more firms replicating the behavior of firms in a small economy and, more importantly, replicating research results. Young labels this the principle of "equivalent innovation", following Gilfillan who argued that "inventions are not only duplicated about the same time by identical solutions ... but are also paralleled by equivalent but unlike means for reaching the same goal around the same time" (Young, 1995, p. 5).

Yang and Borland (1991) develop a model of the division of labor where knowledge is person-specific, as in Lucas (1988). There are no knowledge spillovers. Person-specific learning-by-doing and increasing returns drive growth. Each person consumes a fixed number of goods and has to decide how many goods to produce and how many to buy from others. In steady state, each producer/consumer engages in a trade-network that is independent of the scale of the economy. A larger economy has more trade-networks but each network is of the same size. There is no incentive to increase the trade-network and reap more benefits from increasing returns or learning by doing because trade involves a transaction cost that does not depend on economy-wide variables. Absence of spillovers and the presence of agent-specific transaction costs imply that agents look only at their local circumstances, no matter how large is the economy in which they operate.

In Rustichini and Schmitz (1991) knowledge is person-specific. Spillovers occur among a given number of agents, i.e., the size of spillovers-networks is fixed. A larger number of symmetric agents does not increase the knowledge pool and leaves per capita knowledge creation unaffected.

Summarizing, these models without scale effects have the property that a large economy replicates a small economy. They thus ignore that a larger economy allows a larger degree of specialization and complexity, as emphasized by Griliches. The reason is that non-rivalry of knowledge no longer applies. Although these models allow for spillovers, they assume that agents, be they firms or workers, do not learn more from a larger population. The implicit assumption is that all knowledge is duplicated. In our opinion, it is preferable to allow for non-rivalry of knowledge in Romer's (1990) sense that the aggregate stock of knowledge, as opposed to the average stock of knowledge, determines the productivity of R&D.

Jones (1995) follows Romer and assumes that R&D builds on the aggregate stock of knowledge but imposes diminishing returns with respect to this stock. Increases in the scale of the economy raise research activity but diminishing returns eventually set in. The scale of the economy, therefore, affects per capita productivity levels, not long-run growth rates. The drawback of this model is, as noted by Young (1995) and many others, that growth is not sustained unless an exogenous force offsets diminishing returns (either population growth or exogenous technological improvements).

Groth (1997) develops a model in which growth requires both invention (following Romer, 1990) and human capital formation (following Lucas, 1988). The result is a combination of Lucas' and Romer's results: scale effects are smaller than in the original Romer model but still positive.

Xie (1997) assumes that R&D is undertaken in two sectors producing final goods. Only R&D in the first sector contributes to the economy's total stock of knowledge on which R&D in both sectors builds. The second sector features increasing returns to scale in production. If the economy is larger, production costs in the second sector fall, labor is reallocated to this sector, and incentives to undertake R&D in the first sector fall. Since growth is driven by R&D in the first sector, it may be lower in a larger economy. This model acknowledges that a larger economy differs in structure from a smaller one, thus partially capturing the spirit of Griliches' argument. However, the scale effect is offset only if the engine of growth is located in sectors where increasing returns in production are relatively weak. Whether this assumption is empirically valid is an open question. Moreover, Xie's model does not explain why in time-series data there is no proportionality between TFP growth and the aggregate amount of resources devoted to R&D, as documented by Jones (1995).

Our model generates growth without scale effects and does not have the unappealing features of replication, duplication, or "equivalent innovation". The degree of specialization of the economy depends on the scale of the economy. We introduce firm-specific knowledge so that costs of adoption (following Jovanovic 1995) and dilution of R&D (following Adams and Jaffe 1996) become important, but we maintain Romer's (1990) idea of non-rivalry. The introduction of market structure dynamics (entry of new firms) allow us to merge these ideas without generating scale effects. The resulting model is very tractable and is directly related to empirical evidence from the Industrial Organisation (IO) literature and, therefore, is readily applicable to explaining industry-level and aggregate patterns of innovation and growth. Moreover, it avoids unrealistic features like the "hit-and-run" nature of competition in Young (1995),[4] or the absence of firms in Yang and Borland (1991). Finally, it does not rely on exogenous sources of growth, like in Jones (1995), nor on restrictive assumptions on sectoral structure, like in Xie (1997).

We build on our previous work in which we analyzed several IO-based growth models (Peretto 1996a,b,c; Smulders and van de Klundert 1995; van de Klundert and Smulders 1997). In earlier papers we have already shown (in passing) that it is possible to eliminate scale effects by assuming that spillovers among firms depend on average knowledge per firm (see Smulders and

[4] In Young (1995, Section 2), the entire stock of knowledge in the economy is fully public so that entrants and incumbents face identical initial conditions. The absence of firm-specific knowledge makes entry and innovation a "hit-and-run" affair where in each period all firms are replaced by new firms. In contrast, we model incremental, path-dependent innovation by incumbents and path-breaking innovation by entrants.

van de Klundert 1995 p. 152; Peretto 1997a,b,c). However, this assumption is restrictive and not very appealing, as argued above. In the present paper we introduce the concepts of technological distance and knowledge dilution, and we model the process of entry in a more general and fundamental way. In particular, models where spillovers depend on average knowledge are the asymptotic limit for a very large number of firms of the model presented here. Hence, this paper provides a more robust microfoundation for our approach to modeling growth and market structure dynamics.

4. Overview of the Model

Knowledge creation is driven by costly in-house R&D. Firms devote resources to improve the production process and the quality of their product.[5] In so doing, they accumulate knowledge that is firm-specific. Other firms that produce different products cannot use this knowledge for their production activity. Furthermore, knowledge is (at least partly) tacit so that even if there were well established and protected property rights, trading technology in a patent market is not feasible. Firms benefit from knowledge developed by other firms only indirectly. First, outside knowledge provides ideas that can be used by incumbents in their own R&D activity, i.e., there are spillovers among firms in the domain of general-purpose knowledge. The larger the pool of knowledge on which a firm can build, the more productive its R&D effort. Second, outside knowledge provides ideas for establishing new firms. Hence, incumbents and entrants alike exploit the stock.

Figure 1 illustrates the process of knowledge creation and knowledge exploitation. N incumbents are in the market. They contribute to the public knowledge stock S, as depicted by the solid arrows, and they exploit the stock S in their R&D activity, as depicted by the broken arrows. Because knowledge is non-rival, the stock of public knowledge does not deplete as it is exploited. In addition to public knowledge, incumbents use in their R&D activity their own firm-specific knowledge, as depicted by the backward bending broken arrows. Firm-specific knowledge provides the basis to absorb outside knowledge and apply it to the firm's production process. In other words, it determines the firm's *absorptive capacity* (Cohen and Levinthal, 1989), its capacity to adapt ideas and methods developed elsewhere to its own production and research activity. An entrepreneur wishing to start up a new firm must create its own initial stock of firm-specific knowledge. He can do so only by exploiting the existing pool of public knowledge S. He has to sort out relevant information out of the mass of available knowledge and adapt it to create a niche of its own. This is a costly process. Once the entrepreneur has sunk the set-up costs, the new firm also contributes to the stock of public knowledge.

[5] Since cost-reducing and quality-improving innovation are formally similar, we focus on productivity growth with the understanding that our results apply to quality improvement as well.

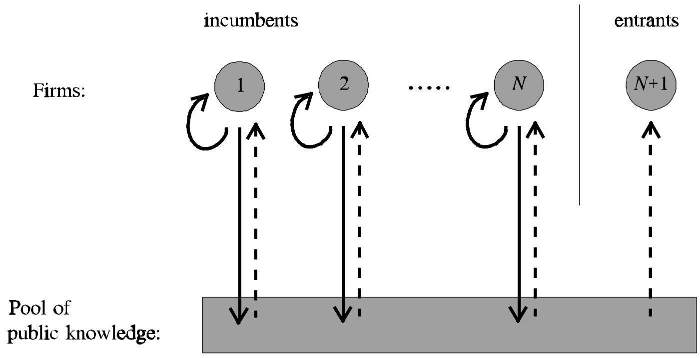

Figure 1 Exploitation and creation of knowledge

To keep the analysis tractable, we make assumptions that allow us to impose symmetry across firms.[6] In particular, entrants enter with a level of productivity that is equal to the average productivity of incumbents. Figure 2 illustrates this assumption. Incumbents operate at productivity level $Z(t)$ at date t. Over time, productivity increases as a result of in-house R&D and incumbents climb their productivity ladders along the vertical dimension. We obtain symmetry at all dates by assuming that a new product line that is establlished at date t starts at productivity level $Z(t)$. Hence, firms that start at later dates start at higher productivity levels. It is important to note that this does not provide an incentive for incumbents to stop doing R&D, wait a while, and then enter at the higher productivity level that prevails in the economy at the later date. First, entry is costly so that an incumbent should compare the cost of in-house R&D to the entry cost. Achieving a given productivity increase dZ over a period of time dt may be cheaper when done incrementally through in-house R&D than when done discretely through this "wait and reenter" policy. Second, and more fundamentally, entry is the creation of a new product line that is qualitatively different from existing ones. That is, the only way to climb an existing productivity ladder is to do it incrementally through in-house R&D. Along the vertical dimension in Figure 2, productivity growth is cumulative, idiosyncratic, and path dependent. In this sense, technological advance along the vertical dimension is firm-specific and is qualitatively different from establishing a new product line. The former corresponds to the cumulative solutions of problems that are specific to a particular product line. The latter corresponds to the creation of the initial condition on which the process of in-house R&D builds.

[6] This is a strong assumption that we make in order to focus on the macroeconomic aspects of the model. Within our conceptual framework, where knowledge is firm-specific, there are many forces that work against symmetry. However, our symmetric model gains in tractability what is lost in realism. In the Appendix, we argue that because of knowledge spillovers and decreasing returns to knowledge at the firm level, in an asymmetric situation small firms grow faster than large firms. The resulting convergence supports a symmetric steady state for the process of industrial dynamics underlying our model. To avoid having to study a system with as many state variables as firms, we impose symmetry at all times.

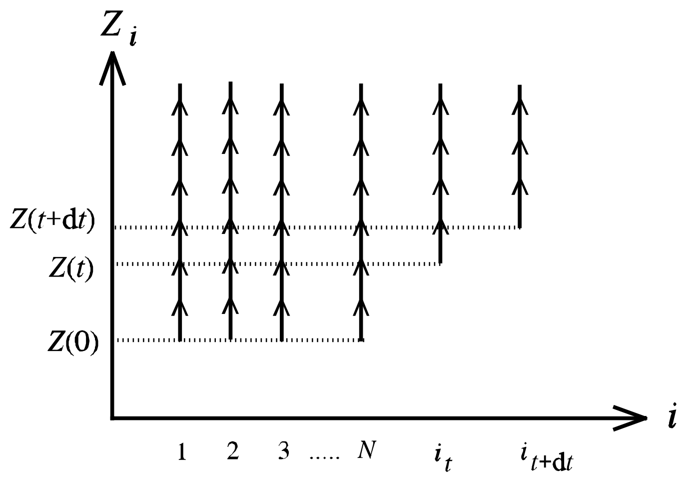

Figure 2 Incumbents versus entrants

The final element in our model is the notion of technological distance. Firms are qualitatively different in terms of the knowledge they use and the product they produce. Even if bits of knowledge could be perfectly transferred to other firms (i.e., if we ignore tacitness), the knowledge of one firm is less useful to another firm since the latter operates along a different technological trajectory as a result of its firm-specific innovation history. We assume that spillovers from technologically distant firms do not contribute much to one firm's R&D. In contrast, spillovers from technologically close firms can be more easily adapted to the firm's own production line. To capture this idea, we assume that the *effective* pool of spillovers S_i accruing to an individual firm consists of the knowledge stocks of other firms weighted by technological distance. The main implication of this assumption is that effective spillovers do not depend on the aggregate knowledge stock, but vary less than proportionally with the number of firms. As more firms are active in the market, the economy becomes more specialized in the sense that the technological distance between firms becomes larger on average. Each new firm that enters the market is on average less related to existing firms and therefore contributes less to each firm's effective pool of spillovers.

5. Description of the Model

5.1. Consumers

The preference side of the economy is modeled as simply as possible. Consumers derive utility from a CES index C of differentiated products, with elasticity of substitution $\varepsilon > 1$. Utility is logarithmic in the consumption bundle C and the intertemporal discount rate ρ is constant:

$$W = \int_0^\infty \ln C \cdot e^{-\rho t} dt; \tag{1}$$

$$C = \left(\int_0^{\infty} C_i^{\frac{\varepsilon - 1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon - 1}}. \tag{2}$$

The solution to the household maximization problem are the well known Dixit/Stiglitz demand schedule and the Keynes/Ramsey rule:

$$X_i = LE \left(\int_0^N P_j^{1-\varepsilon} dj \right)^{-1} P_i^{-\varepsilon}; \tag{3}$$

$$\hat{E} = r - \rho, \tag{4}$$

where E is household expenditure, L is the number of households (and our measure for the scale of the economy), $X_i=LC_i$ is aggregate demand for variety i, N is the number of varieties in the market (the number of firms), and r is the market rate of interest. Hats denote growth rates.

5.2. Incumbents

At any moment in time, N firms are active. Each incumbent i controls his own stock of firm-specific knowledge Z_i that allows him to produce quantity X_i of his own variety and sell it in a monopolistic market against price P_i. Labor is the only factor of production. L_{xi} units of labor are allocated to production with productivity depending on technology Z_i according to the production function

$$X_i = Z_i^{\theta} L_{xi}. \tag{5}$$

In-house R&D employs L_{Zi} units of labor to expand the stock of firm-specific knowledge according to the knowledge production function

$$\dot{Z}_i = Z_i^{1-\psi} S_i^{\psi} L_{Zi}. \tag{6}$$

Productivity of labor in R&D depends on own knowledge Z_i and spillovers S_i. This captures two important characteristics of firm-specific knowledge. First, at the applied end of the spectrum, technology is path-dependent and outside firms with no practical experience of producing good i cannot improve productivity of good i,s production process. In other words, firms with zero stock Z_i cannot develop product- or process-specific knowledge that applies to this activity. This is the rationale for assuming that R&D is undertaken in-house. Second, opportunities to exploit spillovers depend on the firm’s own knowledge stock.

Firms maximize the net present value of profits subject to demand (3) and technology (5) and (6). Taking the wage rate as the numeraire, profit flows are

$$\Pi_i = P_i X_i - L_{xi} - L_{Zi}. \tag{7}$$

The first order conditions read (see Appendix):

$$P_i = \frac{\varepsilon}{\varepsilon - 1} Z_i^{-\theta}, \tag{8}$$

$$r_{RD} = \theta (S_i/Z_i)^{\psi} L_{xi} - \psi \hat{S}_i, \tag{9}$$

$$L_{Zi} > 0 \quad \textit{only if} \quad r = r_{RD}. \tag{10}$$

(8) is the standard mark-up pricing rule. (9) is the no-arbitrage condition for R&D investment. The firm's rate of return to R&D is denoted by r_{RD}. Three determinants of this return are worth noting. First, r_{RD} increases with the size of the firm L_{xi}. This is due to the nonrivalry of knowledge: a larger firm can apply a single new idea to a larger volume of production. Second, r_{RD} decreases with Z_i because there are diminishing returns to firm-specific knowledge in R&D (ψ<1). Third, r_{RD} increases with the stock of public knowledge S_i (contemporaneous spillover effect) but decreases with its rate of change $\hat{S}_i$ (intertemporal spillover effect). A relatively large stock of public knowledge raises the returns to R&D because it raises the firm's R&D productivity, as captured by the term $(S_i/Z_i)^{\psi}$; however, firms do not internalize the contribution of knowledge spillovers to reducing future R&D costs of other firms, as captured by the term $-\psi\hat{S}_i$. (10) is a no-arbitrage condition stating that R&D must yield the required rate of return dictated by the capital market r. Profits follow from (5) and (8):

$$\Pi_i = \frac{1}{\varepsilon - 1} L_{xi} - L_{Zi} \tag{11}$$

This is also the profit flow that accrues to entrants once they become incumbents.

5.3. Entrants

By sorting out and adapting existing knowledge, entrepreneurs create new product lines. Enterpreneurs consider entry as long as the sunk cost is lower than the value of the firm. Hence, in equilibrium

$$V_j \leq \frac{1}{\beta}\left(\frac{Z_j}{S_j}\right)^{\gamma} \quad \wedge \quad \dot{N} \geq 0, \text{ with at least one equality} \tag{12}$$

where V is the (post-entry) value of a firm and the right-hand-side in the first (in)equality is the entry cost in labor units. For simplicity, we treat the number of firms N as a continuous variable. The entry cost increases with the productivity level Z at which a new firm is established, and decreases with spillovers S available to the entrepreneur. β and γ are positive coefficients, indicating the level and the steepness of the entry cost function.

5.4. Spillovers

We assume that each firm contributes to every other firm's knowledge pool but that the contribution decreases with its technolgical distance from the recipient firm. The effective pool of spillovers for firm i is thus the weighted sum of the knowledge stocks of all other firms. Technological distance determines the weights: if firm j is at distance δ_{ij} from firm i, the latter receives $exp(-\delta_{ij})Z_j$ effective spillovers from the former. To fix ideas, assume that technological distance can be measured in a one-dimensional way so that firms can be ordered in terms of technological congruence. The distance between firm i and firm j is $\delta_{ij}=\delta\cdot|i-j|$, where δ is a positive constant. If δ is small, technological differences among firms are small. A large δ indicates a high degree of technological diversity. The total range of technologies employed in the economy is N. We call this the *technology space* of the economy. The spillovers pool availabe to firm i is thus

$$S_i = \int_0^N Z_j e^{-\delta|i-j|} dj .$$

Without loss of generality, we can set $i=0$ and calculate

$$S_0 = \int_0^N Z_j e^{-\delta j} dj = Z\left(\frac{1-e^{-\delta N}}{\delta}\right). \qquad (13)$$

where our symmetry assumption implies $S_i=S_0$ for all i.

The spillover pool S is linear in the average knowledge stock Z and less than linear in the number of firms N. Furthermore, the elasticity of S with respect to N, which equals $\delta e^{-\delta N}/(1-e^{-\delta N})$, becomes zero if N grows very large. These three features drive our results. Other specifications that are more general than (13) yield the same properties. In particular, it seems appropriate to relax the assumption of fixed distance between firms. Hence, we shall use a more general function where S/Z depends positively on N but subject to diminishing returns:[7]

[7] Note that we generalize (13) by assuming that the lower bound of S/Z may be strictly positive. This simplifies the general equilibrium analysis a lot, see proposition 1 below.

$$S = Z \cdot d(N), \tag{14a}$$

$$0 \leq d(0) \leq d(N) \leq \lim_{N\to\infty} d(N) < \infty, \tag{14b}$$

$$n(N) \equiv \frac{\partial(S/Z)}{\partial N}\frac{N}{S/Z} = \frac{d'(N)\,N}{d(N)}, \qquad 0 < n(N) < 1, \qquad \lim_{N\to\infty} n(N) = 0. \tag{14c}$$

When a new firm enters, it contributes to the pool of public knowledge but at the same time it enlarges the technology space of the economy. Entry, in other words, brings into existence new firm-specific technological trajectories. As a consequnce, the typical firm becomes more specialized and, on average, knowledge developed by other firms is less useful (the average spillover $S/NZ=d/N$ is decreasing in N as captured by n<1). As N becomes very large, firms become so specialized that entry of a new firm, establishing a new technological trajectory, does not contribute to the effective spillover pool of existing firms (as captured by n=0 for $N\rightarrow\infty$).

5.5. General equilibrium

Goods market equilibrium implies that total spending equals the value of total supply. Taking into account (5) and (8) we find

$$LE = P_i N X_i = \frac{\varepsilon}{\varepsilon - 1} N L_{xi} \tag{15}$$

Labor devoted to entry is

$$L_N = \frac{1}{\beta}\left(\frac{Z}{S}\right)^{\gamma}\dot{N}. \tag{16}$$

Total supply of labor is fixed at L. Labor market equilibrium requires

$$N L_{xi} + N L_{Zi} + \frac{1}{\beta \cdot d(N)^{\gamma}}\dot{N} = L. \tag{17}$$

Capital market equilibrium is determined by the interaction among households, incumbents and entrants. Households are willing to supply funds by postponing consumption as long as the rate of return is large enough. In particular, the required rate of return on savings follows from (4) and (15),

$$r = \rho + \hat{L}_{xi} + \hat{N}. \tag{18}$$

Competition between incumbents and entrants (equivalently, arbitrage by investors between productivity improvement and creation of new product lines) ensures that either both groups attract funds against the rate of return that households require, or that only the group that offers the higher rate of return obtains funds.

To find the rate of return that entrants offer, suppose that entrepreneurs sell stocks to finance the sunk cost of entry. The return on stocks equals

$$r = \frac{\Pi}{V} + \frac{\dot{V}}{V} \tag{19}$$

If entrants are active (i.e., if $\dot{N}>0$), the value of a firm V equals the sunk cost of entry. Substituting (12), (11), (13) and (14) into (19), we find the rate of return to entry,

$$r_N = \frac{\beta}{\varepsilon - 1} d^{\gamma} L_{xi} - \beta d^{\gamma} L_{Zi} - \gamma n \hat{N}. \tag{20}$$

The rate of return that incumbents offer is given by (9). Using (13) and (14), we can express this rate of return as

$$r_{RD} = \theta d^{\psi} L_{xi} - \psi d^{\psi} L_{Zi} - \psi n \hat{N}. \tag{21}$$

We are now ready to analyze the interaction between entrants and incumbents. To fix terminology, we shall refer to the creation of new product lines as “entry” and to in-house R&D undertaken by incumbents as “R&D”. Four regimes are possible in this model, depending on whether entrants and/or incumbents or able to raise funds. A *free-entry equilibrium* is an equilibrium with interior solution $r=r_N=r_{RD}$ and is the one we focus on. A *blocked- entry equilibrium*, is an equilibrium where no entry takes place because $r=r_{RD}>r_N$. A *blocked-R&D equilibrium* is an equilibrium where no in-house R&D is undertaken because $r=r_N>r_{RD}$. Finally, a *no-growth equilibrium* is an equilibrium where both entry and in-house R&D yield too low a rate of return to attract funds.

6. Steady state

We focus attention on steady states where R&D drives growth while entry peters out. With a constant allocation of labor, unbounded productivity growth can be sustained because accumulation of firm-specific knowledge is linear in Z (combine (6) and (13) to see this). However, the returns to entry fall as entry takes place. Combining (16) and (13), we find

$$\hat{N} = \beta L_N d(N)^{\gamma} / N. \tag{22}$$

As long as $n\gamma<1$, the rate of entry is decreasing in N. This condition is satisfied because n, the elasticity of average tehnological distance d with respect to the number of firms N, approaches zero as N grows large; see (14). The steady state is therefore characterized by a constant number of firms.

With a constant number of firms, R&D is the only source of growth. Steady state economic growth is captured by the rate of productivity growth, denoted by g, where,

$$g = \theta \dot{Z}/Z = \theta(S/Z)^{\psi} L_{Zi} = \theta d^{\psi} L_{Zi}. \tag{23}$$

In steady state, capital market equilibrium implies $r=\rho$; see (18). Moreover, a steady state is characterized by time-invariant values for L_{xi}, L_{Zi}, N, and g.

Whenever growth is positive, the rate of return to R&D should equal the steady-state required rate of return ρ. After substitution of $r=r_{RD}$, $\hat{L}_{xi}=0$, and $\dot{N}=0$ into (17), (18), (21), and (23), this condition can be written as

$$g = \frac{\theta}{\theta+\psi}\left[\frac{\theta d(N)^{\psi} L}{N} - \rho\right] \equiv g^{SS}(N). \tag{24}$$

In free-entry equilibrium, entry and innovation yield the same rate of return. Substituting $r_{RD}=r_N$ and $\dot{N}=0$ into (20), (21), and (17), solving for L_{Zi}, and substituting the result into (23), we find the relation between growth and the number of firms in a free-entry, steady-state equilibrium:

$$g = \frac{\theta d^{\psi} L}{N}\left(\frac{\theta(\varepsilon-1)d^{\psi} - \beta d^{\gamma}}{\theta(\varepsilon-1)d^{\psi} - \beta d^{\gamma} + (\psi d^{\psi} - \beta d^{\gamma})(\varepsilon-1)}\right)$$

Using (24) to eliminate $\theta d^{\psi} L/N$, we find

$$g = \frac{\theta\rho}{\beta[\psi-\theta(\varepsilon-1)]}\left[\theta(\varepsilon-1)\cdot d(N)^{\psi-\gamma} - \beta\right] \equiv g^{FE}(N). \tag{25}$$

Since we want to study the impact of entry on the growth rate, we focus on the situation in which conditions of entry, as captured by (25), as well as opportunities for innovation (as captured by (24), determine the steady state equilibrium. In other words, we rule out that one type of investment (entry or R&D) dominates the other investment activity for all states of the economy. The following proposition defines sufficient conditions for which this situation arises.

Proposition 1 *Define* $d_0=\lim_{N\to 0}d(N)$ *and* $d_\infty=\lim_{N\to\infty}d(N)$. *Assume (a)* $0<d_0<\infty$ *and (b)* $\psi>\theta(\varepsilon-1)>\max\{\beta d_0^{\gamma-\psi},\beta d_\infty^{\gamma-\psi}\}$. *Then,* *i.* *there is a unique free-entry steady state with positive growth and a positive number of firms, $N=N^*>0$ and $g=g^{SS}(N^*)=g^{FE}(N^*)>0$;*

ii. *there is a unique perfect foresight dynamic general equilibrium:*

- *if $N<N^*$, the economy jumps on the saddle path and converges over time to the free-entry steady state;*
- *if $N>N^*$, the economy enters immediately a steady state with no entry.*

Proof. Appendix.

The restriction $\beta d^{\gamma-\psi}<\theta(\varepsilon-1)$ ensures that entry is less profitable than R&D when no R&D is undertaken, thus providing the incentive to invest in R&D rather than entry. The restriction $\psi>\theta(\varepsilon-1)$ ensures that the returns to R&D fall quick enough with research efforts (large ψ), so that for large enough research efforts entry and R&D become as profitable. Note that under these assumptions the FE curve slopes upward if $\psi>\gamma$ and downward if $\psi<\gamma$.

Figure 3 characterizes growth in steady state. Equations (24) and (25) are depicted as the FE and SS curves in the *(g,N)* plane under the assumption that $\beta d^{\gamma-\psi} < \theta(\varepsilon\text{-}1) < \psi$ holds for all *N*. The intersection of the FE-curve and SS-curve represents a free-entry equilibrium in which the number of firms is endogenously determined (it arises after a transition period with entry). It is, however, not the only long-run equilibrium. All bold segments in Figure 3 represent steady-state equilibria. These segments correspond to the blocked-entry and no-growth regimes. Intuitively, if *N* is larger than in free-entry equilibrium, the returns to entry fall below the returns to R&D and entry is no longer considered. As a result, the long-run equilibrium is only determined by R&D opportunities, which are described by the g^{SS} curve. If *N* is very large, however, R&D stops as well because firm size is very small and the fruits of R&D cannot be spread over a large volume of production and the returns to R&D fall below the required rate of return ρ.

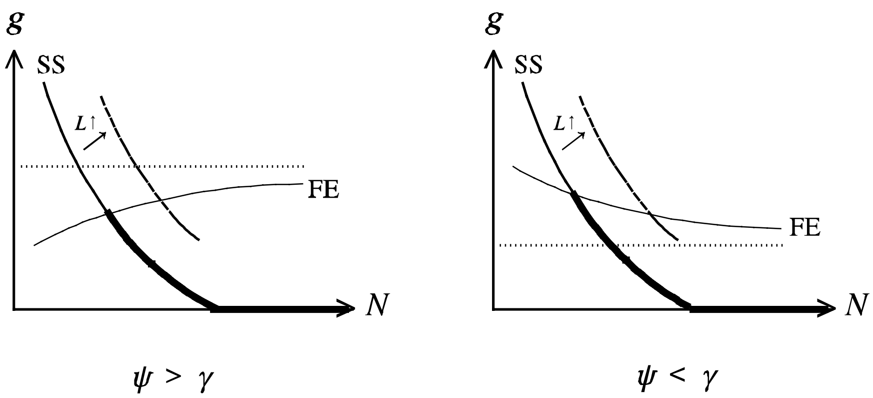

Figure 3 Steady-state equilibria and the scale effect

7. The scale effect

In steady state, the scale effect on growth can be either negative or positive but it is always bounded. If the labor force increases, the long-run growth rate is less and less affected by the scale of the economy and asymptotically approaches a constant. This is illustrated in Figure 4. An increase in the scale of the economy, L, affects only (24), the SS curve, that shifts up as L increases. The free-entry growth rate is determined by the intersection with the FE curve, given by (25). The free-entry steady-state growth rate for a larger labor force is given by a point on the FE curve further to the right. Note that the FE curve is upward or downward sloping, depending on whether ψ is larger or smaller than γ, but approaches a horizontal asymptote, which implies that growth is bounded and that the scale effect vanishes in the limit.

An increase in the size of the economy increases the returns to investment in knowledge because knowledge is non-rival. This means that the cost of producing an idea is independent of the scale of production to which it is applied (R&D is a fixed cost). This basic scale effect in the creation of a non-rival good is captured by the upward shift of the SS-curve: other things being equal (constant N), growth increases with the scale of the economy. We call this the *direct or pre-entry scale effect*. However, a larger market attracts entry, if the sunk cost of entry is not too large, because the value of the firm increases.[8] Entry of new firms offsets the direct scale effect in three ways: (a) it generates *dispersion* of resources over a larger number of firms; (b) it generates an *expansion* of the public knowledge stock; (c) and it generates *dilution* of public knowledge due to specialization. The balance of these forces ultimately determines whether the scale effect is positive or negative in this economy. First, a larger number of firms implies, other things equal, a smaller size of the average firm: labor is dispersed over more business units. Hence, while the scale of the total economy increases, individual firms may become smaller. Since knowledge is firm-specific, it is the firm size that determines the return to R&D, rather than the economy size. Hence, dispersion mitigates the scale effect. Second, entry of firms expands the public stock of knowledge, since entrants create new product lines that may inspire other firms. This increases the returns to R&D and reinforces the scale effect. Finally, start ups of new technologies increase the economy's technology space and reduce average spillovers among firms.

Dispersion of resources over a larger number of firms is never offset completely by increased knowledge spillovers. The reason is that entrants develop specialized knowledge that is less useful to other firms. A one percent increase in the number of firms reduces average firm size by one percent, but increases the effective knowledge pool by less than one percent since only a fraction of the entrants' knowledge is useful to other firms. In other words, if the economy directs R&D efforts to more product lines, it dilutes the aggregate knowledge stock. Hence our model captures Adams and Jaffe's (1996) empirical findings.

Summarizing, entry of specialized firms mitigates or reverses the pre-entry scale effect on

---

[8] If the economy starts in a blocked-entry equilibrium, the rise in firm value may not be enough to make entry profitable. In Figure 3, if initial N is to the right of the intersection between FE and SS, a small upward shift of the SS curve not necessarily shifts the intersection beyond the initial N so that the economy remains in the blocked-entry equilibrium. See next section.

the growth rate. Obviously, in models in which the rate of entry is proportional to the rate of innovation and growth, this mechanism cannot apply. There, positive growth requires positive entry rates by construction so that on a balanced growth path the number of firms should *not* affect the growth rate. In R&D growth models driven by variety expansion (Grossman and Helpman 1991 Ch. 3, Romer 1990, Barro and Sala-i-Martin 1995 Ch. 6), entry does not affect the growth rate since the dispersion effect is exactly offset by the knowledge expansion effect. In these models R&D entails the development of blueprints for new intermediate goods. A larger number of intermediate goods implies less resources available to produce each good, which reduces revenues per blueprint, but the larger number also increases the stock of public knowledge, which reduces costs.

If technological distance is very large, either because firms are very different (δ large) or because N is very large, the dilution effect is strong and offsets the knowledge expansion effect of entry (the FE curve becomes horizontal for large N or large δ since in both cases d is insensitive to N). Effective spillovers are hardly affected by entry (d approaches a constant and n approaches zero) because the entrant shows up with knowledge that is too distant to be useful. In this case, there is no scale effect because the dispersion effect exactly offsets the pre-entry scale effect. The reason is that rates of return now depend on firm-level variables, see (20) and (21), and no longer on the number of firms through knowledge spillovers. An increase in L raises firm size L_{xi} which triggers higher R&D investment and entry of new firms. Entry causes rates of return to fall gradually via the dispersion effect. The steady-state rate of return, $r_N = r_{RD} = \rho$, can only be realized with L_{xi} and L_{Zi} at their pre-shock level. Hence, entry stops when firms are of the same average size as before the expansion of the market.

Dilution of knowledge in our model is related to the degree of specialization in the economy. The extent of the market determines the degree of specialization, as Adam Smith argued a long time ago. If L increases, a larger variety of firms emerge, each following its firm-specific technological trajectory. Consumers have love-of-variety preferences and benefit from entry.[9] Hence, specialization has a positive effect on utility *levels*. However, the effects of specialization on *growth* are ambiguous. Specialization means that technological distance between firms increases on average, so that firms learn less from each other. Since spillovers are smaller, larger firm-specific R&D investments are required to sustain the same rate of growth. If the increase in the size of the market triggers a large increase in specialization, this may generate a negative scale effect.[10]

Whether the scale effect is on balance positive or negative depends on the relative

---

[9] We could interpret C in (1) as consumption of a homogenous final good and (2) as the production function of the final good by means of intermediate inputs. In this case, an increase in specialization, i.e., an increase in N, implies higher totap factor productivity levels rather than utility levels. For examples, see Peretto (1996b, 1996c).

[10] In all variety-expanding models (e.g., Romer 1990) and also in the more recent work by Kelly (1997) growth is driven by an ongoing process of specialization so that level effects cannot be separated from growth effects.

magnitudes of parameters ψ and γ (recall that the slope of the FE curve in Figure 3 is determined by the sign of γ-ψ). If γ exceeds ψ, knowledge spillovers benefit entry more than they benefit R&D, see (22) and (23). As a result, when the economy expands, investors shift their portfolios from shares in established firms to shares in new firms because these benefit more from the knowledge expansion effect of the larger economy. New firms enter, rapidly reducing average firm size. In the new steady state, firms are smaller and employ less labor in R&D. In contrast, if creation of new firms relies on the ingenuity of the entrepreneur rather than on public knowledge, while established firms easily absorb public knowledge ($\gamma<\psi$), incumbents can easily finance higher R&D while only few new firms start up in response to an expansion of the labor force. Hence, the additional amount of labor available in the economy is mainly allocated to existing firms. In the steady state, average firm size is larger and supports higher R&D spending per firm and faster growth.

8. The role of entry and exit

Under our assumptions (14) and (26), the steady state is characterized by zero entry. Moreover, since exit never occurs, both net entry and gross entry are zero. This is in contrast with some existing growth models, as noted in the introduction, and deserves a closer inspection.

Let us first consider under what conditions steady state growth can be driven by continuous entry. Equation (22), reveals that -- with an upper limit on labour input in entry activities ($L_N \leq L$) -- the rate of entry is only prevented to fall to zero in the long run if d^γ/N is non-decreasing, that is if the elasticity of d with repect to N is large enough for any level of N ($n(N) \geq 1/\gamma$). This cannot be reconciled with the notion of technological distance and its consequences for specialization and knowledge dilution. It requires that any specialized firm should contribute enough to the public knowledge stock to prevent the returns to further entry to fall. In other words, it requires a bound to the dilution effect, which violates our hypothesis that in an economy with very many specialized firms, an additional firm has a negligible effect on public knowledge.

It can be considered as support for our assumptions that empirical studies show that in most industries net entry is very small. Cable and Schwalbach (1991) compile the results of more or less comparable empirical studies on entry for 8 countries. They conclude that annual average net entry is less than 0.5% in terms of market share.

However, gross entry might significantly contribute to productivity growth. Baldwin and Geroski (1991) find that while net entry is almost zero, gross entry and exit are significant. Since entrants are on average more productive than exiters, replacement of firms is associated with productivity growth. In its present formulation, our model cannot account for this fact since it does not feature exit of firms. In contrast, quality-ladder models that exhibit creative destruction (e.g., Grossman and Helpman 1991, Ch.4; Aghion-Howitt 1992) take the other extreme and explain all innovation as the replacement of incumbent firms by new, more productive firms (while net entry is zero). In reality, only a small fraction of total productivity growth stems from entry (cf. Baldwin and Geroski's results for Canada). Moreover, most entry and exit concerns small firms: entrants have little effect on incumbents and replace other small firms that exit. Turnover of small firms on the fringe of a particular industry is different from the type of entry that we model in this paper, but is definitely a research topic worth pursuing in the future.

A recurring theme in the IO literature is that of barriers to entry. In our model, the cost of introducing a new product acts as a barrier to entry. The situations discussed above of free entry and blocked entry, therefore, should be distinguished. It is natural to assume that the economy starts with a small number of firms and converges to the free-entry steady state. Once in the free-entry steady state, the economy may be stuck in the blocked-entry region when shocks that require a reduction in the number of firms hit. Because firms enter whenever profitable but never exit, the model exhibit hysteresis. Starting from a free-entry steady state, an increase in scale induces entry and a small (positive, negative, or zero) effect on growth by shifting outward the SS curve. In contrast, a reduction in scale causes growth to fall sharply while the number of firms remains unchanged. The *positive* direct (or pre-entry) growth effect of an *in*crease in scale is always offset by entry but the direct *negative* growth effect of a *de*crease in scale is not offset by exit.

If we allowed for exit, the hysteresis effects would be smaller. Firms that have entered because of favorable shocks in the past might exit under adverse circumstances. The simplest way to introduce exit in is to assume that each firm has to incur a fixed overhead cost of f units of labor. This fixed cost must be subtracted from the right-hand-side of (7) and (11) while the total amount of labor in overhead activities Nf must be added to the left-hand-side of the labor market clearing condition (18). Recalculating the SS and FE curves yields:

$$g = \frac{\theta}{\theta+\psi}\left[\theta d(N)^{\psi}\left(\frac{L}{N} - f\right) - \rho\right] \equiv g^{SS}(N). \qquad (24')$$

$$g = \left(\frac{\theta}{\psi-\theta(\varepsilon-1)}\right)\left[\theta(\varepsilon-1)d(N)^{\psi}\left(\frac{\rho}{\beta d(N)^{\gamma}} + f\right) - \rho\right] \equiv g^{FE}(N). \qquad (25')$$

From the new expression for profits, the new labor market clearing condition and the definition of g in (23), we find a relationship between the growth rate and the number of firms for which profits are zero:

$$\Pi_i \geq 0 \quad \Leftrightarrow \quad g \leq \theta d^{\psi}\left(\frac{L}{\varepsilon N} - f\right) \equiv g^{ZP}(N). \qquad (26)$$

These three relationships can be depicted in the N,g plane as the SS, FE and ZP curve respectively. Steady state equilibria are located on the part of the SS curve that is below the FE curve and below the ZP curve, see the thick segment in figure 4. To the left of this segment, new firms enter. To the right of this segment firms incur losses and therefore exit. As expected, the larger the fixed cost f, the smaller the maximal number of firms that can be supported in the economy. Without fixed cost, firms always realize positive post-entry profits and there is no upper bound on N (formally, if f=0, $g^{SS}(N)<g^{ZP}(N)$ for each N).

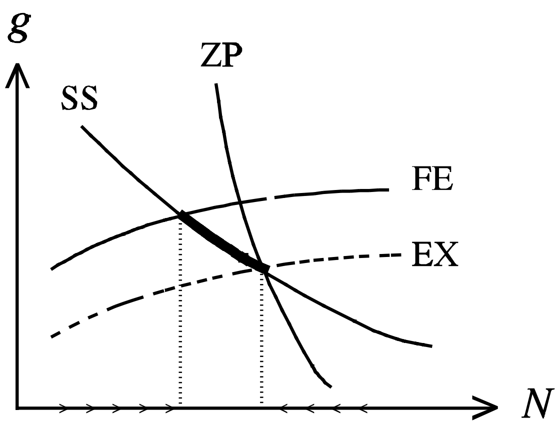

Figure 4 Steady state equilibria with fixed overhead costs

The range of possible equilibria becomes smaller if the entry cost becomes smaller. To see this, substitute (24') into (26) to eliminate the scale variable:

$$g = \left(\frac{\theta}{\psi - \theta(\varepsilon - 1)}\right)\left[\ \theta(\varepsilon-1)d(N)^{\psi} f \ - \ \rho\ \right] \ \equiv \ g^{EX}(N). \qquad (27)$$

By construction, this equation defines a line in the N,g plane that intersects the SS and ZP curve in their point of intersection, which marks the right endpoint of the range of possible equilibria. This line has the same shape as the FE curve, as can be seen by subtracting (27) from (25'):

$$g^{FE} - g^{EX} = \left(\frac{\theta^2(\varepsilon-1)d(N)^{\psi}}{\psi - \theta(\varepsilon-1)}\right)\frac{\rho}{\beta d(N)^{\gamma}}. \qquad (28)$$

Note that $\rho/\beta d^{\gamma}$ is the annualized value of the entry cost for an entrepreneur that considers entry if the economy is in the steady state (the "marginal entrant"). This value determines the range of steady state equilibria. If the entry cost is negligible ($\beta \rightarrow \infty$), FE, SS and ZP intersect in one point and there is no hysteresis (the FE curve converges to the EX curve).

9. Conclusion

Dilution of knowledge is the main reason in our model why the scale effect on growth is at least mitigated. In a larger economy, the stock of public knowledge is applied by a larger and more diverse set of firms so that on average public knowledge is less productive: its content is diluted.

Technological distance between firms explains in turn why dilution occurs. New firms introduce new technologies, so that the unweighted stock of public knowledge rises. But at the same time, if this knowledge is highly specialized, it will not raise the effective knowledge stock that is applicable by the average firm in the economy.

Dilution only occurs if new technologies are introduced. This is where the interaction between innovation and market structure dynamics becomes important. Favorable conditions of entry, which are well-defined and closely related to innovation opportunities in our model, imply a significant change in the structure of the market and technology. As a result, dilution effects on the rate of innovation are strongly negative and the overall scale effect may become negative. In contrast, with only small changes in market structure, mainly existing firms benefit from an expansion of the market and growth is boosted by the scale effect.

Our model covers, of course, only a few stylized facts and insights from the industrial organization (IO) literature. Nonetheless, its results show that the puzzle of the scale effect in growth theory can be solved by seriously building growth models on IO foundations.

Appendix

I. Firm's maximization problem

The Hamiltonian implied by (5), (6) and (7) reads:

$$H = P_i X_i - X_i Z_i^{-\theta} - L_{Zi} + q_i Z_i^{1-\psi} S_i^{\psi} L_{Zi} + \lambda_i L_{Zi}, \tag{I.1}$$

where q_i is the co-state variable, Z_i is the state variable, and X_i and L_{Zi} are the control variables. Firms take S_i as given. The first order conditions are:

$$(\partial P_i / \partial X_i) X_i + P_i - Z_i^{-\theta} = 0, \tag{I.2}$$

$$L_{Zi} = \begin{cases} 0 & for \quad 1 > q_i Z_i^{1-\psi} S_i^{\psi} \\ L_{Zi} & for \quad 1 = q_i Z_i^{1-\psi} S_i^{\psi}, \\ \infty & for \quad 1 < q_i Z_i^{1-\psi} S_i^{\psi} \end{cases} \tag{I.3}$$

$$X_i \theta Z_i^{-\theta-1} + (1-\psi) q_i \hat{Z}_i + \dot{q}_i = r q_i. \tag{I.4}$$

The assumption that firms are atomistic allows us to approximate the price elasticity of demand by ε (the impact of P_i on the expression in brackets in equation (3) is negligible, so that $(\partial P_i/\partial X_i)X_i/P_i = -\varepsilon$). Substitution of this result into (I.2) yields (8).
The linearity of the Hamiltonian in L_{Zi} implies that R&D is either at a corner or indeterminate at the firm-level, see (I.3). General equilibrium conditions determine L_{Zi}. The corner solution $L_{Zi}=\infty$ can be ruled out because it violates the labor market constraint. Hence, we may write:

$$L_{Zi}\left[1 - q_i Z_i^{1-\psi} S_i^{\psi}\right] = 0 \tag{I.3'}$$

In an interior solution ($L_{Zi}>0$), we have $q_i=1/(Z_i^{1-\psi}S_i^{\psi})$ from (I.4). Using this result to eliminate q_i and $\dot{q}_i$ from (I.4) and noting that $L_{xi}=X_iZ_i^{-\theta}$ from (5), we find (9) and (10). Whenever $r \neq r_{RD}$, the conditions for an interior solution are violated, and we must have $L_{Zi}=0$, which explains why $L_{Zi}=0$ in a capital market equilibrium with $r<r_{RD}$.

It is important to note that since firm i takes the spillover pool S_i and its production scale L_{xi} as given, diminishing returns to firm-specific knowledge ($0<\psi<1$) imply that large firms face lower returns to innovation. In particular, the bang-bang R&D policy (A.1) implies that only the smallest firm undertakes R&D since this is the firm that offers the highest rate of return. This force implies that if firms start out with asymmetric market positions, they converge to symmetry over time. We can thus simplify the analysis by assuming that entrants join the industry at the average level of productivity so that the industry is in symmetric equilibrium at all times.

II. Characterization of the steady state (proposition 1).

Proposition 1 in the main text characterizes equilibrium under certain pararmeter restrictions. This appendix characterizes steady state equilibria under less restrictive cases.

Proposition A1. Three growth regimes arise in the steady state, according to the following conditions:

i. A steady state Nash equilibrium with $r_{RD}=r_N=\rho$ requires $\psi>\theta(\varepsilon-1)>\beta d^{\gamma-\psi}$;

ii. No R&D takes place in a steady state equilibrium for $\beta d^{\gamma-\psi}>\theta(\varepsilon-1)$;

iii. No entry takes place any temporary equilibrium for $\theta(\varepsilon-1)>max\{\psi,\beta d^{\gamma-\psi}\}$.

Proof Consider Figure 5, where we depict the rates of return to entry and R&D given by (20) and (21) in *(r,g)* space. Firms decide how much to invest in R&D taking as given L_{xi} and N. This follows from (3), (5), and (8) which show that the size of the firm is determined by the size of the market *LE* and the number of firms *N*, variables that the firm does not control. We can represent this decision as incumbents choosing $g=\theta d^{\psi}L_{Zi}$. We evaluate equilibrium assuming no entry (i.e., $\dot{N}=0$) since we characterize steady states. The panels represent different candidates for a steady-state equilibrium. Proofs of the individual statements are as follows:

i. Point S in panel (a) corresponds to the free-entry equilibrium ($r_N=r_{RD}=\rho$). Firms have no incentive to deviate from the equilibrium value of *g* and this is a stable Nash equilibrium. Stability obtains if the r_{RD} line is steeper than the r_N line at point S. An intersection exists if $\psi>\theta(\varepsilon-1)>\beta d^{\gamma-\psi}$. This condition implies that the slope condition $\psi>\beta d^{\gamma-\psi}$ is satisfied.

ii. $\beta d^{\gamma-\psi}>\theta(\varepsilon-1)$ implies that the following two situations are possible: (a) at the intersection point the r_{RD} line is flatter than the r_N line and the Nash equilibrium is unstable because reducing growth raises the rate of return to entry more than it raises the rate of return to R&D; (b) the r_N line is everywhere above the r_{RD} line. In both cases entry dominates R&D and incumbents cannot raise funds on the capital market. These situations, therefore, identify the free-entry steady states with zero R&D.

iii. $\theta(\varepsilon-1)>max\{\psi,\beta d^{\gamma-\psi}\}$ implies that the r_N line is everywhere below the r_{RD} line. Hence, R&D dominates entry and only established firms can finance investment. This situation, therefore, identifies the blocked-entry steady states. The analysis of this regime is discussed below. ■

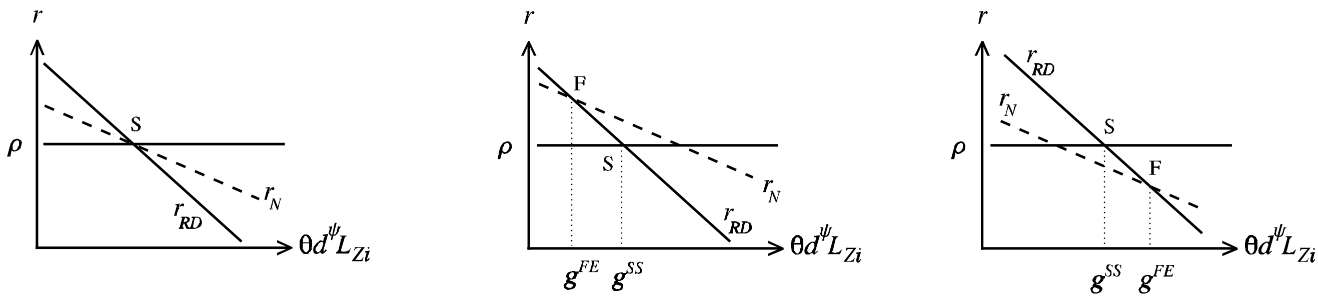

Figure 5 Steady-state Nash equilibria

Key: The downward sloping lines represent equations (21) and (20), respectively, drawn for given L_{xi}, N, and for $L_N=0$, under the assumption $\beta d^{\gamma-\psi}<\theta(\varepsilon-1)<\psi$.

The following proposition characterizes the set of steady state equilibria in which conditions of entry as well as opportunities for innovation determine the steady state equilibrium, i.e. on the situation described in part i of proposition A1.

Proposition A2. Assume $\theta(\varepsilon-1)<\psi$ and define Q as the range (set) of values of N for which $\beta d(N)^{\gamma-\psi}<\theta(\varepsilon-1)$.

i. Steady-state growth is zero for all $N\notin Q$.

ii. Steady-state growth is zero for all N such that $N\cdot d(N)^{-\psi}>\theta L/\rho$.

iii. No $N\in Q$ such that $g^{FE}(N)<g^{SS}(N)$ can be sustained in a steady states.

iv. All pairs (N,g) such that $N\in Q$ and $0<g=g^{SS}(N)<g^{FE}(N)$ can be sustained in a steady state.

Proof

i. See Proposition A1, Part ii.

ii. A zero growth steady state is characterized by $L_{Zi}=0$, $L_{xi}=XZ^{-\theta}=L/N$, and $r=\rho$. Substituting these results into (I.4) and solving for q yields $q=\theta L/N\rho Z$. Hence, $1-qZ(S/Z)^{\psi} = 1-\theta Ld^{\psi}/\rho N > 0$ so that $L_{Zi}=0$ is indeed the optimal Nash strategy for firms, see (I.3).

iii. Consider Figure 5: $N\in Q$ implies $\beta d^{\gamma-\psi}<\theta(\varepsilon-1)<\psi$ so that $r_N=r_{RD}$ is a feasible and stable equilibrium, see proposition A1, part i. $g^{FE}(N)<g^{SS}(N)$ implies that this equilibrium yields a rate of return higher than ρ. Hence, it cannot be a steady state. Figure 5, panel (b) illustrates. The intersection of the r_N and r_{RD} lines determines g^{FE} (point F). $r_{RD}=\rho$ determines g^{SS}. Given the slopes and intercepts of the two lines, $g^{FE}<g^{SS}$ implies $r>\rho$ when the returns to R&D and entry are equalized (point F).

iv. $g=g^{SS}$ implies $r_{RD}=\rho$ by construction. Also by construction, $r_{RD}=r_N$ only if $g=g^{FE}$. Hence, $g^{FE}>g^{SS}$ implies that we need a higher growth rate to equate r_{RD} to r_N than we need in a steady state with only R&D. Since both rates of return are decreasing in g, $r_N<r_{RD}=\rho$ if $g^{SS}<g^{FE}$, which implies that entry is blocked. Figure 5, panel (c) illustrates. Point S represents the steady state equilibrium (it corresponds to the *SS* curve). There is no incentive to enter since $r_N<r_{RD}$. Entry and R&D yield the same return only if g is larger, for example at point F. However, this cannot be a steady-state because the rate of return is smaller that the required rate of return ρ. ■

Proof of proposition 1

Proposition 1 in the main text makes the assumption that the necessary condition for a free entry steady state Nash equilibrium mentioned in prop. A1, part i is satisfied for all positive values of N (i.e. that $Q=\mathbb{R}^+$).

i. Assumption (a) implies $\lim_{N\to\infty}g^{SS}(N)<0$ and $\lim_{N\to 0}g^{SS}(N)=\infty$. Assumption (b) implies that $0<g^{FE}(N)<\infty$. Hence, there is a unique positive value for N, say N^*, for which $g^{SS}(N)=g^{FE}(N)>0$. Assumption (b) ensures that $\beta d^{\gamma-\psi}<\theta(\varepsilon-1)$ for all $N>0$ so that, by prop. 1 part i, the equilibrium is a Nash equilibrium.

ii. $N>N^*$ implies $g^{SS}<g^{FE}$. This is a steady state by proposition A2, part iv.
$N<N^*$ implies $g^{SS}>g^{FE}$. This is no steady state by proposition A3, part iii. We proof that $\dot{N}>0$ during some transition period by showing that $\dot{N}=0$ violates the equilibrium conditions, so that we must have $\dot{N}>0$. Note that we now have to worry about capital market equilibrium outside the steady state.
   (a) consider a capital market equilibrium with $r_N<r=\rho$, $r_{RD}<r=\rho$, $\dot{N}=L_{Zi}=0$. This is impossible since with $N<N^*$, $g^{SS}(N)>0$ which implies by construction that $r_{RD}=\rho$ has to hold in an equilibrium where $\dot{N}=0$.
   (b) consider a capital market equilibrium with $r_N<r_{RD}=r$ and $\dot{N}=0$, $L_{Zi}>0$. Using (17) to eliminate L_{Zi} in (21), we see that r depends positively on L_{xi}. Hence, L_{xi} has to jump immediately to the value for which $r_{RD}=\rho$ in order to satisfy (18) at all moments in time. Such a steady state is impossible, see proposition A3, part iii.
   (c) consider a capital market equilibrium with $r_N=r_{RD}=r$ and $\dot{N}=0$, $L_{Zi}>0$. Since $N<N^*$, $g^{FE}<g^{SS}$ and $r>\rho$. The Ramsey rule (18) can only be satisfied for $\dot{N}=0$ if $\hat{L}_{xi}>0$. However, equality of interest rates $r_N=r_{RD}$ and labor market equilibrium (17) can hold for constant N only for a unique level of L_{xi}. Hence, $\dot{N}=0$ is impossible.

Hence, $N<N^*$ implies an equilibrium with $r_N=r>\rho$ and $\dot{N}>0$. ■

Alternatively, part (ii) and (iii) can be proofed by constructing a Phase diagram in the (N,NL_{xi}) plane and showing that a unique saddle path exists. This space-consuming proof is available upon request.

References

Adams, J.D., and A.B. Jaffe, 1996, "Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data", *Rand Journal of Economics*, 27, 700-721.

Aghion, P., and P. Howitt, 1992, "A Model of Growth Through Creative Destruction," *Econometrica*, 60, 323-351.

Baldwin J. and P. Geroski, 1991, Entry, Exit and Productivity Growth, in: Geroski and Schwalbach (eds) 1991, 244-256.

Cable J. and J. Schwalbach, 1991, International Comparisons of Entry and Exit, in: Geroski and Schwalbach (eds) 1991, 257-280.

Cohen, W., and R. Levin, 1989, "Empirical Studies of Market Structure and Innovation", in R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization*, Amsterdam, North Holland.

Cohen, W. and D. Levinthal, 1989, 'Innovation and Learning: the two faces of R&D', *Economic Journal* 99, 569-596.

Cohen, W., and S. Klepper, 1996, "A Reprise of Size and R&D", *Economic Journal*, 106, 925-951.

Dixit, A., and J. Stiglitz, 1977, "Monopolistic Competition and Optimal Product Diversity", *American Economic Review*, 67, 297-308.

Dosi, G., 1988, "Sources, Procedures, and Microeconomic Effects of Innovation", *Journal of Economic Literature*, 36, 1120-1171.

Geroski, P., 1994, "Entry and the Rate of Innovation", in *Market Structure, Corporate Performance and Innovative Activity*, Oxford, Clarendon Press. (First published in *Economics of Innovation and New Technology*, 1, 203-214, 1991.)

Geroski, P. and J. Schwalbach (eds), 1991, *Entry and Market Contestability, an international comparison*, Blackwell, Oxford.

Griliches, Z., 1990, "Patent Statistics as Economic Indicators: A Survey", *Journal of Economic Literature*, 28, 1661-1707.

Griliches, Z., 1992, "The Search for R&D Spillovers", *Scandinavian Journal of Economics*, 94 (supplement), S29-S47.

Grossman, G., and E. Helpman, 1991, *Innovation and Growth in the Global Economy*, Cambridge, MIT University Press.

Groth, C., 1997, "R&D, Human Capital, and Economic Growth", mimeo, University of Copenhagen.

Jaffe, A.B., 1986, "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value", *American Economic Review*, 76, 984-1001.

Jones, C.I. 1995a, 'Time Series Test of Endogenous Growth Models', *Quarterly Journal of Economics* 110, 495-525.

Jones, C.I., 1995b, "R&D based Models of Economic Growth", *Journal of Political Economy*, 103, 759-784.

Jovanovic, B., 1993, "The Diversification of Production", *Brookings Papers on Economic Activity* Microeconomics, 197-247.

Jovanovic, B., 1995, "Research, Schooling, Training, and Learning by Doing in the Theory of Growth", mimeo, University of Pensylvania.

van de Klundert, T., and S Smulders, 1997, "Growth, Competition and Welfare", *Scandinavian Journal of Economics*, 99, 99-118.

Levin, R., A. Klevorick, R. Nelson, and S. Winter, 1987, "Appropriating the Returns from Industrial R&D", *Brookings Papers on Economic Activity*, 3, 783-820.

Lucas, R.E. 1988, 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, 22, pp. 3-42.

Malerba, F., and L. Orsenigo, 1995, "Schumpeterian Patterns of Innovation", *Cambridge Journal of Economics*, 19, 47-65.

Malerba, F., L. Orsenigo, and P. Peretto, 1997, "Persistence of Innovative Activities, Sectoral Patterns of Innovation, and International Technological Specialization", *International Journal of Industrial Organization*, 15, 801-826.

Peretto, P,.1996a, "Sunk Costs, Market Structure and Growth", *International Economic Review*, 37, 895-923.

Peretto P., 1996b, "Firm Size, Rivalry and the Extent of the Market in Endogenous Technological Change", Duke University, Department of Economics, Working Paper No. 96/07.

Peretto, P. 1996c, "Technological Interdependence, Industrial Development, and Balanced Growth," Duke University, Department of Economics, Working Paper No. 97/10.

Peretto, P., 1997a, "Cost Reduction, Entry, and the Interdependence of Market Structure and Economic Growth", mimeo, Duke University, Department of Economics. (Revised version of Working Paper No. 95/48.)

Peretto, P., 1997b, "Technological Change, Market Rivalry, and the Evolution of the Capitalist Engine of Growth", *Journal of Economic Growth*, forthcoming.

Peretto P., 1997c, "Technological Change and Population Growth", Duke University, Department of Economics, Working Paper No. 96/28.

Romer, P., 1990, "Endogenous Technological Change", *Journal of Political Economy*, 98, S71-S102.

Rustichini, A. and J.A. Schmitz, 1991, "Research and Imitation in Long-run Growth", *Journal of*

Monetary Economics, 27, 271-292.

Smulders, S., and T. van de Klundert, 1995, "Imperfect Competition, Concentration and Growth with Firm-Specific R&D", *European Economic Review*, 39, 139-160.

Thompson, P., and D. Waldo, 1994, "Growth and Trustified Capitalism", *Journal of Monetary Economics*, 34, 445-462.

Tse, C. Y., 1997, "Diffusion Lag of Knowledge, Diversification of Firms and Growth", mimeo University of Hong Kong.

Xie, M.X., 1997, "Economic Integration and Economic Growth with Science-pushed Innovation", *Review of International Economics*, forthcoming.

Yang, X. and J. Borland, 1991, "A Microeconomic Mechanism for Economic Growth", *Journal of Political Economy*, 99, 460-482.

Young, A., 1995, "Growth without Scale Effects", NBER working paper No. 5211.