VARIETY, SPILLOVERS AND MARKET STRUCTURE IN A MODEL OF ENDOGENOUS TECHNOLOGICAL CHANGE*

Pietro F. Peretto

Department of Economics
Duke University

Revised: September 1995

ABSTRACT

I study an economy where oligopolistic firms establish in-house R&D programs to produce a continuous flow of cost-reducing (incremental) innovations. The scale of firms' R&D operations determines the rate of productivity growth. I first study the role of concentration, firm size, and demand, appropriability, and opportunity conditions, when the number of firms is exogenous. I find hump-shaped relationships of knowledge appropriability and the number of firms with steady-state growth that give a first intuition about the forces at work in the model. In symmetric equilibrium, static economies of scale due to variety of input supply and knowledge spillovers yield increasing returns to the number of firms. A larger number of firms leads to higher output, higher (aggregate) R&D, and faster growth. Offsetting this force, market fragmentation leads to smaller firms, smaller R&D programs, and slower growth. Next, I let the number of firms be endogenous and study the balanced-growth path of the economy in two dimensions: productivity growth and the number (variety) of goods. The price, investment, entry, and exit decisions are interdependent. In particular, R&D is a fixed (sunk) cost and in zero-profit equilibrium is negatively related to the number of firms. This additional feed-back reinforces the market fragmentation effect. More importantly, many parameters have no longer the effects predicted by the model with an exogenous number of firms (and the standard models in the literature). For example, the scale effect of population size may be negative. Large markets have many firms, high output, and high aggregate R&D. Market fragmentation may offset this force and result in small firms and slower growth.

Keywords: Growth, Market Structure, Product Variety, Spillovers, Endogenous Technological Change.

JEL Classification Numbers: E10, L16, O31, O40

---

* Address: Department of Economics, Duke University, Durham, NC 27708. Phone: (919) 6601807. Fax: (919) 6848974. E-mail: Peretto@lewis.econ.duke.edu. I thank Richard Levin and Xavier Sala-i-Martin for helpful discussions. This is a revised version of the paper presented at the International Conference on Increasing Returns and Economic Theory in honor of Kenneth J. Arrow held at Monash University, Clayton, Victoria, Australia. I thank discussant Alfred Guender for useful comments. All errors are my responsibility.

1. Introduction

Growth theorists have produced a number of interesting models investigating the idea that technological progress, the engine of growth in income per capita, is endogenous to the economic system and driven by market forces. These models are radically different from the traditional theory of economic growth based on capital accumulation and emphasize the incentives for profit-seeking agents to undertake R&D aimed at developing new products and processes, or incrementally improving old ones.

The key problem that these models must tackle is the appropriability problem. That is, they must specify an economic or institutional device that allows innovators to appropriate the returns from their R&D activity. Such a device must be at the heart of any theory of endogenous innovation because without the expectation that the (ex post) economic returns from innovation are appropriable, profit-seeking agents would not incur the (ex ante) sunk costs that R&D requires.

Models of endogenous innovation characterize R&D as the development of blueprints for a new (consumption or intermediate) good. Anyone with access to the blueprint can produce the new good. Innovations are essentially new pieces of knowledge and knowledge, by its very nature, is non-rival and can be reproduced at will and at zero cost by anyone with access to it. To induce profit-seeking agents to undertake R&D the information contained in the blueprints has to be protected. Accordingly, the institutional device that generates the incentives to innovate is the patent law that restricts the right to use the product-specific knowledge to the patent holder.¹

These models have two important features. First, infinitely lived and perfectly enforceable patents generate local monopolies by assigning to the innovator the exclusive right to manufacture and sell the new good. In addition, production firms conducting in-house R&D to develop their own innovations have no

---

¹ The variety-expansion model (Romer 1990, Grossman and Helpman 1991, Ch. 3) formalizes innovation as the expansion of the range of intermediate (consumption) goods that enter symmetrically the production of final output (consumers' utility). R&D is deterministic and the continuous introduction of horizontally differentiated goods, each one produced by a new local monopolist, drives growth. The quality-ladder model (Segerstrom, Anant and Dinopoulos 1990, Grossman and Helpman 1991, Ch. 4, Aghion and Howitt 1992) formalizes innovation as the introduction of higher-quality versions of existing intermediate (consumption) goods. R&D is stochastic and the arrival at (random) discrete intervals in time of vertically differentiated goods drives growth. Goods of different quality are perfectly substitutable and product obsolescence from innovation generates firm turnover over time. Therefore, each generation of the good is produced by a local monopolist whose tenure is temporary.

specific advantage over independent research firms that sell in the patent market the rights to use their innovations. The emphasis on the non-rivalry of knowledge and the patent law makes clear that imperfect markets and monopoly power are necessary for profit-seeking agents to undertake R&D and innovation. However, it focuses too much on monopolistic competition and neglects that most innovations are carried out by established producers.² In so doing, the models neglect the central role of the modern oligopolistic corporation and its in-house integration of manufacturing and R&D. As a consequence, the theory does not address those components of the market structure, like concentration and firm size, emphasized by Schumpeter (1942) as key determinants of the R&D activity of profit-seeking firms.

In this paper, I extend the theory beyond the case of local, temporary or permanent, monopolies and consider oligopolies with an endogenous number of firms undertaking in-house R&D. Business and economic historians stress the central role of large corporations with integrated manufacturing and R&D activities in generating innovation and growth (see, e.g., Freeman 1982, Mowery and Rosenberg 1989, Chandler 1990). To begin with, R&D generates, in addition to specific innovations, knowledge that is not innovation-specific and that can be used in subsequent R&D activity. Moreover, the patent market for the exchange of innovation-specific knowledge contained in blueprints is far from perfect and suffers from informational problems that do not allow to write perfect contracts between users and developers of innovations. Firms, therefore, have strong incentives to internalize the R&D function and to keep knowledge secret.³ Evidence in IO shows that firms build up their competitive position by consistently investing in R&D over time and accumulating in-house knowledge protected by secrecy, patents and other appropriability devices (see, e.g., Levin et al. 1987 and Mowery and Rosenberg 1989).

---

² Segerstrom (1991) developed a quality-ladder model with innovation and imitation that relaxes the emphasis on monopolistic competition by allowing innovators and imitators to coexist as collusive duopolists. This model, however, still neglects the role of in-house R&D and does not allow to study oligopolistic markets with more than two firms. Barro and Sala-i-Martin (1995, Ch. 7) developed a version of the quality-ladder model where innovation is carried out by the leader-incumbent and there is no firm turnover. This model takes into account the role of in-house R&D by established producers but like many others it considers monopolistic competition only.

³ In contrast, models based on (intertemporal) knowledge spillovers (see footnote 1) posit that the knowledge generated as a by-product of R&D is public and can be used by any subsequent innovator for free. Since developing an innovation does not give any advantage beyond the patent relative to the new product, there is no incentive to integrate in-house manufacturing and R&D activities.

The Industrial Organization (IO) literature, theoretical and empirical, has typically assumed that R&D is undertaken by established firms and focused on those features of the business environment, the number and size of firms, barriers to entry, diversification, that determine market rivalry and, therefore, that determine the opportunities and constraints that firms face when planning R&D activities.⁴ The IO literature has also identified a number of additional features of the industry environment that are at least as important as the concentration rate and the average size of firms. These features have proven so important empirically that in recent work the term market structure has come to denote the entire vector of industry characteristics that are relevant for the R&D decisions of firms: concentration, firm size, characteristics of demand (e.g., volume, growth, degree of product substitution), and appropriability and opportunity conditions (see, e.g., the discussion in Cohen and Levin 1989).

In the general equilibrium (GE) model that follows, I study an economy where oligopolistic firms establish in-house R&D facilities to produce over time a continuous flow of cost-reducing (incremental) innovations. In GE, firms compete not only for sales but also for resources. In symmetric equilibrium, the number of firms determines concentration and firm size. These determine the scale and the efficiency of firms' R&D operations and, therefore, the rate of innovation. If the number of firms is exogenous, one can study the relationship between number of firms, elasticity of product substitution, appropriability and opportunity conditions, and the rate of growth of the economy. The number of firms, however, must be endogenous and determined, jointly with the rate of growth of the economy, by the zero-profit condition. This generates some new interesting effects. To isolate these effects from those discussed above, I study the implications of the model under two assumptions: exogenous and endogenous number of firms.

In section 2, I describe the structure of the model. In section 3, I derive the price and R&D strategies of the typical oligopolist and the properties of the symmetric Nash Equilibrium (NE). In section 4, I construct the GE of the economy and argue that, when the number of firms is exogenous, the model

---

⁴ See, e.g. the surveys by Kamien and Schwartz (1982), Baldwin and Scott (1987), Dosi (1988), Tirole (1988, Ch. 10), Cohen and Levin (1989), and Scherer and Ross (1990, Ch. 17). The emphasis on rivalry is important. Scherer and Ross (1990, Ch. 1), in particular, define market rivalry as "a striving for potentially incompatible positions ..... combined with a clear awareness by the parties involved that the positions they seek to attain may be incompatible" (p.16). Monopolistic competition models, with their emphasis on atomistic firms, do not capture very well the notion of rivalry.

does not have transitional dynamics and the economy is in steady state at all times. I characterize this steady state and discuss how opportunity, appropriability, and demand conditions, and the number of firms shape the balanced-growth path (BGP) of the economy. Not surprisingly, most results at this stage resemble those of standard models of endogenous innovation. Interesting, however, are the effects of the number of firms and the degree of appropriability on steady-state growth. These depend on the elasticity of product and knowledge substitution. When the elasticity of knowledge substitution is high, growth is highest if appropriability is perfect (spillovers are zero). When the elasticity of knowledge substitution is low, a hump-shaped relationship between appropriability and growth arises. Similarly, depending on the values of the elasticity of product and knowledge substitution, the relationship between number of firms and growth can be either monotonically increasing or hump-shaped.

These results give a first intuition about the economic forces at work in this class of models. There are macroeconomic complementarities arising from two sources: static economies of scale due to variety of input supply and knowledge spillovers yield increasing returns to the number of firms. A larger number of firms leads to higher aggregate output, higher R&D in GE, and faster growth. Offsetting this force, the fragmentation of the market due to a larger number of firms leads to smaller firms and slower growth. Market fragmentation and macroeconomic complementarities work in opposite direction since in this model the size of the firm determines the scale of the firm's R&D program and this determines the firm's rate of innovation. To see in detail how these forces interact, in section 7, I study the endogenous determination of the number of firms in zero-profit equilibrium. There are no exogenous fixed costs and the only factor that checks the number of firms is R&D expenditures. These, however, are endogenous to the market equilibrium and are not necessarily positive. Provided appropriability is above a threshold level, when the economy is large enough and/or the productivity of R&D is high enough, a zero-profit equilibrium with positive growth and a finite number of firms exists. Indeed, two such zero-profit equilibria exist and I discuss some alternative selection mechanisms. The most interesting implication of the endogeneity of the market structure is that the model now represents the BGP of the economy in two dimensions, the rate of growth of productivity in the intermediate sector and the number of differentiated inputs it supplies. In addition, the endogeneity of the number of firms leads to interdependent price,

investment, entry, and exit decisions. The effects of many parameters may be reversed with respect to standard models and the case where the number of firms is exogenous. For example, the scale effect, the effect on growth of an increase in the size of population, is no longer unambiguously positive and in some cases it is actually negative.

2. Structure of the Model

I consider an economy with a two-tiered manufacturing sector. One competitive firm produces output with a constant returns to scale (CRS) technology defined over a continuum of intermediate goods. Oligopolistic firms produce intermediate goods with a technology exhibiting increasing returns to scale (IRS) to labor and (labor-augmenting) knowledge. Intermediate oligopolists accumulate knowledge to increase factor productivity, reduce production costs, and win market share by offering lower prices.⁵ Manufacturing and research have the same factor intensity and output is either consumed or invested to accumulate knowledge capital. This section describes in detail the production side of the economy.

2.1. Final Producer

One representative firm produces and sells a homogenous final good, Y , in a competitive market. The production technology is defined over a continuum of differentiated, non-durable intermediate goods,

$$(2.1) \quad Y = \left[\int_0^N X_i^{(\epsilon-1)/\epsilon} di \right]^{\epsilon/(\epsilon-1)},$$

where $\epsilon > 1$ is the elasticity of substitution between intermediate goods, X_i is the final producer's use of each intermediate good, and N is the number of intermediate goods supplied at a moment in time. The production technology is defined over the interval $[0, \infty)$ because the list of intermediate goods is infinitely long, all of these goods are known, and intermediate firms have the ability to produce them.⁶ However, at

---

⁵ Quality-improving and cost-reducing technological progress are formally equivalent when the quality of the good determines the cost at which the good delivers its services (see, e.g., Spence 1984 and Tirole 1988). I focus on cost-reducing technological progress only, but the model can easily be rewritten in terms of quality-improving technological progress with the same qualitative results.

⁶ Thus, there is no research involved in expanding the range of intermediate goods used in final production. All of these goods and the technology to produce and use them are known. The number of available intermediates, however, is an endogenous variable because of the entry and exit decisions of the intermediate producers (see below). The focus here is not on how new intermediates are introduced into the economy. Rich models addressing this issue already exist (see the papers cited in footnote 1). The focus is on how, at any moment in time, rivalry in the intermediate sector determines the variety of input supply

a moment in time only N of such goods are supplied. The final producer, therefore, faces the production technology (2.1) defined over the interval $[0, N]$. Facing N available inputs, he uses all of them because each intermediate's marginal product approaches infinity as its use approaches zero.

The production technology (2.1) exhibits economies of specialization. Suppose, for example, that the final producer uses all the available intermediate goods in equal quantity. One can write

$$(2.2) \quad Y = N^{\epsilon/(\epsilon-1)} \cdot X,$$

where X is the quantity of each intermediate input. Holding constant the use of each intermediate good, output increases by increasing the number of intermediate goods. These are not dynamic gains, like those emphasized by Romer (1990). Since all possible intermediate goods are already known, these are static economies of specialization, similar to those discussed by Adam Smith in his famous argument that the size of the market determines the extent of the division and specialization of labor.

To close this description of the final output sector, I now solve for the optimal behavior of the final producer. Let the price of the final good be normalized to one, $P_Y \equiv 1$. The final producer acts as a price-taker in the input and output markets and maximizes profits

$$(2.3) \quad \Pi_F = Y - \int_0^N P_i X_i di$$

subject to the production technology (2.1). The final producer allocates total expenditure on intermediates according to (see, e.g., Dixit and Stiglitz 1977)

$$(2.4) \quad X_i = Y \cdot \left[P_i^{-\epsilon} / \int_0^N P_i^{1-\epsilon} di \right],$$

where total expenditure on intermediate goods is equal to total revenues that accrue to the final producer because in a competitive market he just breaks even. Equation (2.4) is the instantaneous demand schedule faced by each intermediate producer.

2.2. Intermediate Oligopolists

Each intermediate producer faces the production technology

$$(2.5) \quad X_i = A_i \cdot L_i,$$

where

---

and, therefore, specialization in the final sector. I augment the model and introduce research costs for the introduction of new intermediates in Peretto (1995b).

$$(2.6) \quad A_i \equiv \left[\alpha Z_i^{(\eta-1)/\eta} + (1-\alpha) \int_0^N Z_j^{(\eta-1)/\eta} dj \right]^{\eta/(\eta-1)},$$

is a labor productivity index determined by the firm's (labor-augmenting) knowledge capital (hereinafter knowledge), Z_i , and by spillovers of knowledge from the other firms, Z_j for all $j \neq i$, X_i is output, and L_i is labor employment.⁷ Knowledge is non-rival within one firm and is partially excludable between firms. The parameter $\eta > 1$ measures the elasticity of knowledge substitution. The parameter $0 \leq \alpha \leq 1$ measures the appropriability conditions.⁸ Knowledge is a public good when $\alpha = 0$, while it is a private good when $\alpha = 1$. Since knowledge is non-rival, the production technology exhibits CRS to labor alone, the rival input, and IRS to labor and knowledge. I assume CRS to knowledge to ensure that constant endogenous growth be feasible in steady state.

Equation (2.6) captures two different effects of knowledge accumulation. The first is the direct effect of increasing the firm's labor productivity for given external opportunities, as measured by the knowledge of the rivals. The second is the indirect effect of increasing the marginal contribution to the firm's productivity of each unit of outside knowledge. Knowledge accumulation, thus, enhances the "absorptive capacity" of the firm, the firm's capacity to learn from the surrounding environment and apply to its production process ideas and methods developed elsewhere (Cohen and Levinthal 1989). Equation (2.6), moreover, exhibits economies of specialization arising from the variety and substitutability of knowledge across firms. Consider a symmetric equilibrium where firms have the same knowledge stocks,

$$(2.7) \quad A = Z \cdot [1 + (1-\alpha)(N-1)]^{\eta/(\eta-1)}.$$

Average productivity, A , is proportional to average knowledge, Z . This reflects that one firm's knowledge contributes to the pool of knowledge spillovers and improves the technological opportunity conditions that all firms face. The rate of proportionality, is decreasing in the elasticity of knowledge substitution, η ,

---

⁷ I abstract from physical capital for simplicity. Physical capital has very interesting implications for the transitional dynamics of the model but does not alter significantly the properties of the balanced-growth path that I discuss in this paper. See Peretto (1995a) for an example.

⁸ The IO literature has developed this concept to denote technological and institutional factors like the protection granted to intellectual property rights, the effectiveness of industrial secrecy, or some measure of the "natural" transferability of technology across firms (e.g., tacit vs. formal knowledge). Levin et al. (1987) show that there is a variety of means of excluding the other firms from benefiting from one firm's technological advance. Thus, the parameter α is a summary measure of the effectiveness of all these means (see, e.g., Cohen and Levinthal 1989).

because the more substitutable are the different knowledge sources, the lower is the marginal contribution to the productivity of the individual firm of the pool of knowledge spillovers. The rate of proportionality is decreasing in the appropriability parameter, α , and increasing in the number of firms, N , because the less knowledge is a private good (the lower is α) and the more knowledge sources there are (the larger is N), the larger is the pool of knowledge spillovers and the higher is each firm's productivity.

Knowledge is accumulated according to the simple R&D technology

$$(2.8) \quad \dot{Z}_i = \theta I_i,$$

where $\dot{Z}_i$ is the increment in knowledge obtained by firm i from investing I_{Zi} units of final output for an interval of time dt and $\theta > 0$ is a parameter of technological opportunity in research, the extent to which institutional and other environmental factors determine R&D productivity.⁹

Equations (2.5) through (2.8) emphasize local, incremental technological progress and formalize it as accumulation of labor-augmenting knowledge. The focus is on those costly learning activities, not necessarily formal R&D, that firms undertake to improve the productivity of the labor force (see, e.g., Malerba 1992).¹⁰ Differently from most models in the literature, in equations (2.5) through (2.8) IRS to knowledge accumulation apply at the firm level, not at the industry level. This assumption is important.

By its very nature, disembodied knowledge is non-rival across firms as well as within the firm. It is certainly excludable, however, often to a substantial degree (see, e.g., Levin et al. 1987), due to patent laws, protection of intellectual property rights, and secrecy or tacitness of knowledge itself. In addition, a relevant component of knowledge is embodied in the firm and its members as part of an organization and is rival across firms (see, e.g., the discussion in Nelson and Winter 1982 and Dosi 1988). Thus, knowledge

---

⁹ In IO literature this concept denotes, among others, factors like the contributions of university and government research to industrial R&D. See, e.g., Levin et al. (1987), Cohen and Levin (1989) and Klevorick et al. (1993) for a detailed discussion.

¹⁰ In all recent models that rely on aggregate intertemporal spillovers to sustain growth in the long run (e.g., Romer 1990, Grossman and Helpman 1991, Ch. 3), knowledge accumulation is formally equivalent to labor-augmenting, disembodied technological progress. Recently, Lucas (1993) has stressed that, as an explanation of aggregate growth, there is only a negligible difference of interpretation between models of human capital accumulation and models of technical change. In his view, the crucial question is whether to emphasize human capital accumulation at school or by firms. Models of technological change emphasize accumulation of human capital by firms. In particular, they emphasize accumulation of human capital through research rather than as a by-product of experience in production (learning-by-doing). With the differences discussed in the text, I follow this tradition.

does not diffuse immediately and at no cost across firms both because it is partially excludable, in its disembodied and embodied parts, and because it is partially embodied in the firm as an individual entity. It is important to emphasize that excludability and rivalry are related but not the same. If knowledge were totally embodied in the members of the organization, it would be totally rival across firms and therefore excludable. If knowledge were totally disembodied, it would be totally non-rival across firms but still partially excludable. In both cases, however, knowledge would be non-rival at the firm level. When the firm undertakes an R&D project that develops a new piece of knowledge, all members of the organization benefit from the project and learn the piece of knowledge.

This view of technological change is different from the one formalized by Romer (1990) and Grossman and Helpman (1991, Ch. 3-4). It is indeed more similar to the one that Romer (1986) proposed in the seminal paper that started the theory of endogenous growth.¹¹ Technology does not consist of blueprints that are taken off the shelf and put to use in production. New technology and its commercial application must be learned at a substantial cost by all those involved in its use. An R&D project does not consist of the research division of the firm, or an outside contractor, developing a new blueprint that then is passed along to the production division and put to commercial use. Rather, it consists of a set of integrated activities, formal research in a lab being only one of these, that accumulate useful knowledge as human capital embodied in the skills of the labor force. When new products and processes are developed, or old ones incrementally improved, a blueprint is only one step toward commercial implementation. The labor force has to be trained and develop new skills, old facilities have to be upgraded or new ones set up, managers have to learn to run a different organization of labor and handle different problems. These activities develop the body of knowledge that ultimately determines what and how efficiently the firm can do. This knowledge for the most part cannot be formalized and articulated in blueprints and technical manuals and is embodied in the human capital of the firm and its people.

---

¹¹ There are two key differences. Romer assumes that knowledge accumulates as a by-product of investment in physical capital (learning-by-doing or learning-by-investing), while I assume that knowledge accumulates through specific R&D activity. In addition, Romer assumes that knowledge is a public good, while I assume that it is partially appropriable.

3. Knowledge Accumulation in Oligopoly Equilibrium

Intermediate oligopolists behave non-cooperatively and maximize their value. In this section, I define the equilibrium with free entry and free exit for the economy's intermediate sector and study its properties. All firms face the same production and R&D technologies and demand schedules. I restrict the analysis to symmetric equilibria to simplify the model and focus on the key properties of the BGP.

3.1. Definition of Equilibrium

The typical firm chooses price and investment policies that maximize the present discounted value of net cash flow,

$$(3.1) \quad V_i(t) = \int_t^{\infty} R(\tau) [P_i(\tau)X_i(\tau) - wL_i(\tau) - I_i(\tau)] d\tau,$$

where wL_i is total production cost and I_i is R&D expenditure. With perfect foresight, V_i is the stock market value of the firm, the price of the ownership share of an equity holder.

I consider a symmetric Nash Equilibrium (NE) in open loop strategies. Let $a_i = [P_i(\tau), I_i(\tau)]$ for $\tau \geq t$ be firm i 's strategy. This strategy is a time-path of price and investment. To simplify the analysis, I assume that entry and exit involve zero costs and firms do not have any scrapping value. The number of firms is free to jump to its equilibrium level. I construct an equilibrium where at time t firms commit to time-paths of price and investment, while simultaneously free entry and exit determine the number of firms. The price and investment strategy induces time-paths of production, sales, and knowledge accumulation. Following Dasgupta and Stiglitz (1980), I define the equilibrium as follows. Let A_N be the strategies of the N firms. At time t , $[N, A_N]$ is an instantaneous equilibrium with free entry and free exit if for all i

$$(3.2) \quad V_i[N, A_N] \geq V_i[N, A'_N] \geq 0$$

and for all $N' > N$

$$(3.3) \quad V_i[N', A_{N'}] \leq 0,$$

where $[A'_N]$ denotes the strategies when firm i deviates from its optimal time-path of price and investment while the other firms do not deviate. Condition (3.2) requires that the active firm maximize the present value of net cash flow, taking as given the competitors' strategies, and that this maximized value be non-negative. The latter inequality is the free-exit condition since the scrapping value of an active firm, the

opportunity cost of incumbency, is zero. Condition (3.3) is the standard free-entry condition that the value of the marginal firm be non-positive.

I study the properties of this NE in two steps. I first study the properties of the GE of the economy treating the number of firms as a parameter, and then study the endogenous determination of the number of firms and its interaction with the rate of growth. The first step of this procedure determines the value of incumbency for a given number of firms. To determine the NE with free entry and free exit, one compares this value to the opportunity cost of incumbency and the opportunity cost of entry. These are both zero.

3.2. Knowledge Accumulation

Deriving the cost function from the production technology (2.5), the typical firm solves

$$(3.4) \quad \max_{\{P_i, I_i\}} V_i = \int_t^\infty R(\tau) [P_i - w/A_i] \cdot X_i - I_i d\tau,$$

subject to the research technology (2.8), the demand schedule (2.4), $Z_i(t) = Z > 0$ (initial knowledge is given), $Z_j(\tau)$ for all $\tau \geq t$ given for all $j \neq i$ (the firm takes as given the R&D paths of its rivals), and $\dot{Z}_i(\tau) \geq 0$ for all $\tau \geq t$ (knowledge accumulation is an irreversible investment).

There are different approaches to this problem. One is to observe that the objective function (3.4) is weakly separable in the price and investment decision and two-stage budgeting is feasible for the firm. Consider stage-one profit maximization. Given the firms' knowledge stocks, the optimal Bertrand-Nash price strategy is

$$(3.5) \quad P_i = \xi_i w / A_i (\xi_i - 1),$$

where $\xi_i = \varepsilon - (\varepsilon - 1) \cdot S_i > 1$ is the price elasticity of demand that firm i faces and $S_i \equiv P_i^{1-\varepsilon} / \int_0^N P_i^{1-\varepsilon}$. If firms follow the optimal strategy in price and quantity, S_i is the firm's market share, defined as the output of the firm divided by the total output of the industry. A Bertrand-Nash equilibrium with positive output exists if the price elasticity of demand faced by each firm is larger than one. This requires $0 \leq S_i < 1$. In symmetric equilibrium, this condition is met if $N > 1$.

In stage-two profit maximization, firms hold constant the price elasticity of demand they face, ξ_i . Substitution of the price policy into the demand schedule (2.4) determines the demand for each variety as a function of the N labor productivity indexes. This yields the instantaneous profit function,

$$(3.6) \quad \pi_i = \frac{Y}{\xi_i} \cdot \frac{[A_i(\xi_i - 1)/\xi_i]^{\varepsilon-1}}{\int_0^N [A_i(\xi_i - 1)/\xi_i]^{\varepsilon-1} di},$$

expressed as the firm's instantaneous revenues gross of R&D costs.¹² Maximizing equation (3.5) subject to equation (2.8) requires maximizing the Current Value Hamiltonian

$$(3.7) \quad CVH_i = \pi_i - I_i + q_i \theta I_i,$$

where the costate variable, q_i , is the shadow value of knowledge, the firm's knowledge, Z_i , is the state variable and R&D investment, I_i , is the control variable.

The Hamiltonian is linear in R&D investment, thus giving the optimal policy $I_i = \infty$ when $\partial CVH_i / \partial I_i > 0$, and $I_i = 0$ when $\partial CVH_i / \partial I_i < 0$. The former case violates the GE conditions and is ruled out. The latter case corresponds to an instantaneous equilibrium with zero R&D. The first order conditions when $\partial CVH_i / \partial I_i = 0$ and $I_i > 0$ are given by the equality between the marginal revenue from one unit of R&D and its marginal cost, $1/\theta = q_i$, the transversality condition that at the end of the planning horizon the firm's knowledge has no value, $\lim_{t \rightarrow \infty} R(t) \cdot q_i(t) \cdot Z_i(t) = 0$, the constraint on the state variable given by the R&D equation (2.8), and a differential equation in the costate variable

$$(3.8) \quad r = \frac{\dot{q}_i}{q_i} + \frac{\partial \pi_i}{\partial Z_i} \cdot \frac{1}{q_i}.$$

Since the shadow value of knowledge is the inverse of R&D productivity, the transversality condition is satisfied as long as the discount factor is positive. In the steady state with constant endogenous growth, this condition requires that knowledge does not grow at a rate faster than the interest rate (see below). The first order conditions reduce to

$$(3.9) \quad r = \theta \cdot \frac{\partial \pi_i}{\partial Z_i}.$$

Equation (3.9) is a perfect-foresight, no-arbitrage condition which requires that there are no unexploited profit opportunities in R&D: the marginal profit from one R&D project must be equal to the cost of the

---

¹² This procedure is equivalent to the following reasoning. First, firms choose to charge a mark-up over marginal cost. They treat this mark-up as an unknown to be determined by the solution of the system of optimality conditions and think of prices as proportional to marginal costs. Given this pricing rule, they realize that their knowledge affects their competitors' prices via the spillovers. The optimal knowledge accumulation decision must take into account this effect. Differentiating profits with respect to the firm's knowledge requires differentiating the demand schedule with respect to the competitors' prices as well. To compute the optimal mark-up, firms take the usual first order condition with respect to their price (see, e.g., Beath and Katsoulacos 1991).

R&D project financed by borrowing at the rate r (direct cost of R&D), and this must be equal to the return from a riskless loan, at the rate r , of the resources required for the R&D project (opportunity cost of R&D). The two interpretations of the cost of R&D are equivalent under perfect foresight and imply that the firm can finance its own R&D either through debt or through equity.

The NE of the R&D game is given by the set of first order conditions and dynamic constraints for all the active firms. Since I have assumed that all firms in the industry start out with the same amount of knowledge, I have indeed assumed that the model is symmetric not only in the fundamentals that firms face, but also in the starting values of the state variables. Since the dynamic behavior of each firm is dictated by the same equations and by the same initial conditions, firms accumulate knowledge at the same rate and the NE is symmetric at all times.¹³

3.3. The Returns to R&D in Bertrand-Nash Equilibrium

Theory and evidence in IO support the idea that firms account for the partial appropriability of knowledge when planning their investment (see, e.g., Spence 1984, Levin et al. 1987, Cohen and Levinthal 1989, Cohen and Levin 1989). Firms realize that knowledge spillovers accruing to competitors do not contribute to improving their market position because they do not give any advantage over the competitors. Thus, what matters for market competition is the knowledge that can be appropriated.

Consider the marginal gross revenue from knowledge accumulation

$$(3.10) \quad \frac{\partial \pi_i}{\partial Z_i} = \int_0^N \frac{\partial \pi_i}{\partial A_j} \cdot \frac{\partial A_j}{\partial Z_i} dj.$$

The returns to R&D are determined by the effect of labor productivity on profit and by the effect of knowledge on labor productivity. The first effect is captured by the derivatives of the instantaneous profit

---

¹³ An additional argument is needed here. The solution discussed in the text follows from the fact that the Hamiltonian is linear in investment. The corresponding bang-bang strategies yield immediate convergence to the firm's steady state only if there are no constraints to investment. Although I did not specify it formally, in GE firms are constrained by the saving behavior of households. Thus, an additional assumption is needed to simplify the analysis. Specifically, the NE of the R&D game is given by the set of first order conditions from the intertemporal maximization problem. If investment is unconstrained all firms converge instantaneously to their symmetric steady state. To avoid problems deriving from the constraints induced by the GE structure of this model, I assume that the N firms start out with the same level of the knowledge stock. For an introductory discussion of the properties of this type of dynamic (capital accumulation) games see Fudenberg and Tirole (1986) and Tirole (1988, Ch. 8). Fudenberg and Tirole (1991, Ch. 13) provide a formal textbook treatment.

function with respect to the firms' productivity indexes. The second effect captures the role of the partial excludability and partial substitutability of knowledge. Taking derivatives and rearranging for symmetry equation (3.10) yields, after substitution into equation (3.9),

$$(3.11) \quad r = \frac{Y}{\varepsilon(N-1)+1} \cdot \frac{\theta\alpha(\varepsilon-1)(N-1)[1+(1-\alpha)(N-1)]^{1/(\eta-1)}}{N} \cdot \frac{1}{A}.$$

The equation captures a number of effects that derive from the interaction between firms. The terms $Y/[\varepsilon(N-1)+1]$ and $[(\varepsilon-1)(N-1)/N] \cdot \alpha[1+(1-\alpha)(N-1)]^{1/(\eta-1)}$ measure, respectively, the gross-profit and the business-stealing effects. Given size of the market Y (the volume of demand), the NE rate of return to R&D is given by the gross profits earned for a given market share and by the increase in the market share achieved by the R&D project. The gross-profit effect is decreasing in the elasticity of product substitution, ε , because the oligopoly mark-up is lower the higher is ε . The business-stealing effect, on the other hand, is increasing in the elasticity of product substitution, ε , because the effectiveness of cost-reducing R&D in stealing customers from other firms is higher the higher is ε .

The business-stealing effect is a hump-shaped function of the appropriability parameter, α . Spillovers have two opposite effects on the incentives to undertake R&D (Spence 1984). If appropriability is zero, if knowledge is a public good, the marginal contribution of knowledge to the rivals' productivity is equal to its marginal contribution to the firm's productivity. The firm does not gain any marginal advantage over the competitors and the R&D project is worth zero. This is the incentive effect of spillovers and reduces the firms' R&D effort for given market and technology conditions. However, if appropriability is low, large spillovers from the rivals' knowledge increase the marginal contribution of the firm's own knowledge to productivity and increase the rate of return to R&D. This efficiency effect of spillovers increases the productivity of knowledge for given market and technology conditions.

The gross-profit effect is decreasing in the number of firms, N , because in the symmetric NE both the market share and the oligopoly mark-up are lower the larger is N . The business-stealing effect has two components. The first, the market interaction between firms, captured by the term $(\varepsilon-1)(N-1)/N$, is increasing in the number of firms, N , because it captures the firm's potential gain of market share,

measured by the rivals' total market share.¹⁴ The second, the technological interaction between firms, is captured by the term $\alpha[1+(1-\alpha)(N-1)]^{1/(\eta-1)}$ and is increasing in the number of firms, N , because the marginal value of knowledge increases with the pool of spillovers.

4. Balanced-Growth with an Exogenous Number of Firms

To construct the GE of the model, I impose three conditions: labor and output market clearing, and equality between the rate of return to investment and the rate of return to saving. I assume that there exists a fixed population, L , of identical individuals endowed with L units of labor that they supply inelastically in a competitive labor market. Individuals consume and save. The capital market is perfect.

4.1. Labor and Output Market Equilibrium

Substituting the price strategy (3.5) into the conditional labor demand obtained from the production function (2.5), and using the firm's demand schedule (2.4), yields $L_i = Y(\xi_i - 1)S_i/w\xi_i$. Integration over firms yields aggregate employment, and since labor is in fixed supply equal to L , the labor market clearing condition, $\int_0^N L_i di = L$, yields the equilibrium wage rate, $w = (\xi - 1)Y/L\xi$. The final producer breaks even and $\Pi_F = 0$, where profits for the final producer have been defined in equation (2.3).

In the symmetric NE the production technologies (2.1) and (2.5) and $L_i = L/N$ yield

$$(4.1) \quad \frac{Y}{A} = LN^{1/(\epsilon-1)},$$

that can be plugged into (3.11) to peg the rate of return to R&D,

$$(4.2) \quad r = \frac{L\alpha\theta(\epsilon-1)(N-1)N^{1/(\epsilon-1)}[1+(1-\alpha)(N-1)]^{1/(\eta-1)}}{[\epsilon(N-1)+1]N}.$$

With an exogenous number of firms, there are no transitional dynamics and the economy is always in steady state (see below). Output market clearing yields the economy's resource constraint

$$(4.3) \quad Y = LE + N\dot{Z}.$$

Using equations (2.7) and (4.1), this can be written

$$(4.4) \quad c = N^{1/(\epsilon-1)}[1+(1-\alpha)(N-1)]^{\eta/(\eta-1)} - \frac{N}{L\theta} \cdot g,$$

---

¹⁴ Borrowing terminology from Baldwin and Scott (1987), I call this the "cannibalization" effect since it captures the idea that cost-reducing investment transfers surplus across firms. This effect is closely related to the notion of "creative destruction" proposed by Schumpeter (1942, Ch. 7).

where $g \equiv \dot{Z}/Z$ is the average rate of growth of knowledge and $c \equiv E/Z$ is consumption per effective worker. Taking logs and time derivatives of equations (4.1) and (2.7) yields $\dot{Y}/Y = \dot{A}/A = \dot{Z}/Z$ since the number of firms, N , is constant.

4.2. Capital Market Equilibrium and Dynamics

The typical household maximizes lifetime utility

$$(4.5) \quad U(t) = \int_t^\infty e^{-\rho(\tau-t)} \frac{E(\tau)^{1-\sigma} - 1}{1-\sigma} d\tau,$$

subject to the intertemporal budget constraint that the present discounted value of expenditure cannot be greater than the present discounted value of wages and dividends plus initial wealth,

$$(4.6) \quad \int_t^\infty R(\tau)E(\tau)d\tau \leq \int_t^\infty R(\tau)[w(\tau) + d(\tau)]d\tau + B(t),$$

where $\sigma > 0$ is the inverse of the constant elasticity of intertemporal substitution, $\rho > 0$ is the individual discount rate, $R(\tau) \equiv e^{-\int_t^\tau r(s)ds}$ is the cumulative discount factor from time t to time τ , B is the individual's assets holding, w is the wage rate, and d is dividends.¹⁵ The household chooses the optimal plan

$$(4.7) \quad \frac{\dot{E}}{E} = \frac{r - \rho}{\sigma}.$$

Taking logs and time derivatives of consumption per effective worker, and using the expression

for the rate of return to R&D in equation (4.2), one obtains

$$(4.8) \quad \frac{\dot{c}}{c} = \frac{1}{\sigma} \left[\frac{L\alpha\theta(\varepsilon - 1)(N - 1)N^{1/(\varepsilon-1)}[1 + (1 - \alpha)(N - 1)]^{1/(\eta-1)}}{[\varepsilon(N - 1) + 1]N} - \rho \right] - g.$$

Figure 1 shows the dynamics of the GE of the economy in (g, c) space. Paths that eventually yield negative consumption per effective worker are not equilibria since they yield negative utility. Paths that eventually yield infinite consumption per effective worker violate the GE conditions since aggregate consumption, LE , will exceed output, Y , in finite time after productivity growth drops to zero. The unique GE is the saddle point where per capita consumption, E , and average productivity, A , grow at the same constant rate

$$(4.9) \quad g = \frac{1}{\sigma} \left[\frac{L\alpha\theta(\varepsilon - 1)(N - 1)N^{1/(\varepsilon-1)}[1 + (1 - \alpha)(N - 1)]^{1/(\eta-1)}}{[\varepsilon(N - 1) + 1]N} - \rho \right].$$

---

¹⁵ Profits in the intermediate sector are distributed as dividends in equal shares to households.

The economy grows if the rate of return to investment and saving is greater than the intertemporal discount rate. Since R&D expenditures must be non-negative, the rate of growth is zero whenever this condition is not satisfied.¹⁶

4.3. Comparative Statics

In the rest of this section, I discuss how opportunity, appropriability and demand conditions, and the exogenous number of firms shape the BGP of the economy.

Population Size and the Scale Effect

The size of the population, L , affects positively the rate of growth, g . R&D investment is a sunk cost producing a non-rival good. The value of knowledge depends on the scale of production to which it is applied. This intuition is captured by the gross-profit effect (see section 3). The larger demand, the higher the rate of return to R&D, and the higher the rate of growth of the economy.¹⁷ This effect is related to the notion of "demand pull" (see, e.g., Schmookler 1966 and Dosi 1988) and to the scale effect present in most endogenous growth models. With an exogenous number of firms, therefore, the role of population size is the same as in standard models of endogenous innovation.

R&D Productivity

The effect of R&D productivity, θ , on the rate of growth, g , is positive because the marginal product of R&D is higher the higher is θ . This effect is related to the notion of "technology push" (see, e.g., Rosenberg 1976 and 1982, and Dosi 1988).

The Intertemporal Discount Rate and Elasticity of Substitution

The intertemporal discount rate, ρ , and the inverse of the intertemporal elasticity of substitution, σ , have the standard negative effects on g . The more impatient society or the lower the propensity to smooth consumption over time, the lower the saving rate and the lower the rate of growth.

The Elasticity of Product Substitution

---

¹⁶ The condition $(1 - \sigma)g < \rho$ ensures that the consumer's lifetime utility is bounded.

¹⁷ See Backus, Kehoe, and Kehoe (1992) and Barro and Sala-i-Martin (1995) for a discussion of scale effects in endogenous growth models. The scale effect in this model can also be interpreted as a relationship between growth and the size of the labor force because labor market clearing yields that the labor force determines the scale of activity in the intermediate goods sector.

The effect of the elasticity of product substitution, ϵ , has two distinctive aspects which are captured by the gross-profit and business-stealing effects (see section 3). A higher elasticity of product substitution implies that the scope for competition based on productivity improvement is larger, but it also implies that the mark-up over marginal cost is lower. The balance between these two effects is captured by the term $(\epsilon - 1)/[\epsilon(N - 1) + 1]$ in equation (4.2).

In GE there is an additional effect. Equation (4.1) states that there is a constant ratio between final output, Y , and average manufacturing productivity, A . Two ideas are captured by the term $N^{1/(\epsilon-1)}$. On one hand, the larger is the variety of intermediate goods, the higher is productivity in the final output sector, and the larger is total demand for intermediates. This effect increases the returns to each intermediate good. On the other hand, the higher is the elasticity of product substitution between intermediate goods, the less demand for intermediate goods concentrates on each intermediate, and the lower are the returns to each intermediate producer.

To sign the net effect of the elasticity of product substitution, ϵ , on steady-state growth, g , I need to sign the term $(\epsilon - 1)N - [\epsilon(N - 1) + 1]\ln N$. Therefore, I have a U-shaped relation with a minimum at $\epsilon^* \equiv (N + \ln N)/(N + \ln N - N \ln N)$. For intermediate values of ϵ , growth may be zero.

The Elasticity of Knowledge Substitution

An increase in the elasticity of knowledge substitution, η , decreases steady-state growth, g . The more substitutable (homogeneous) is knowledge across firms, the lower is the marginal contribution to the firm's productivity of the pool of knowledge spillovers (see section 2). The efficiency effect of spillovers falls, and the business-stealing effect falls (see section 3). The rate of return to R&D in equation (4.2) falls, and the rate of growth falls.

Appropriability

There is a lower bound α_0 on the degree of appropriability below which steady-state growth, g , is zero. Firms invest in R&D only if they can appropriate a large enough fraction of the benefits from innovation. When $\alpha = 0$ and knowledge is a public good, the incentives to undertake R&D disappear, firms do not invest, and the economy does not grow. In addition, there is a value of the appropriability parameter, α , that maximizes steady-state growth. Let this value be $\alpha^* \equiv \min\{1, (\eta - 1)N/\eta(N - 1)\}$.

This yields $\alpha^* = 1$ when $\eta \geq N$, and $\alpha^* = (\eta - 1)N / (\eta(N - 1)) < 1$ when $1 < \eta < N$. The former case arises when knowledge substitution is high and/or there are few firms. The latter case arises when knowledge substitution is low and/or there are many firms.

To interpret these results, consider the condition $1 < \eta < N$. The first element of the inequality measures the substitutability of knowledge and the second element measures two related but opposite effects of the number of firms: the variety of knowledge sources and the fragmentation of the market. Each firm is both a source of technological opportunity and a competitor in the product market. As a source of learning, each firm improves the efficiency of the other firms. As a competitor in the product market, each firm reduces the market shares of the other firms. The condition $1 < \eta < N$ compares the substitutability of knowledge with the number of sources. This suggests, that when knowledge becomes more homogeneous across firms, the benefits from spillovers decrease and rival firms are more competitors in the product market than sources of productivity advance in the production process. Thus, when the substitutability of knowledge is lower than the number of firms, the trade-off in the appropriability level occurs. Too high appropriability prevents the exploitation of knowledge spillovers more than it induces higher R&D investment by excluding the rivals from benefiting from one firm's knowledge accumulation. Conversely, when the elasticity of knowledge substitution is larger than the number of firms, $\eta \geq N$, the degree of appropriability that maximizes steady-state growth is the one corresponding to knowledge being a private good. The trade-off in the appropriability level now does not occur because spillovers allow too many rivals to benefit from one firm's knowledge accumulation, while the benefits that the firm derives from spillovers from the competitors are not significant since knowledge is highly substitutable.¹⁸

These results emphasize the trade-off between the role of appropriability as a private incentive to invest and the role of spillovers in creating externalities (Dosi 1988, Levin et al. 1987). It is interesting that while for $\alpha = 0$ the argument inside equation (4.6) is always negative, for $\alpha = 1$ it can be either positive or negative. In the latter case there exists an upper bound on appropriability above which the economy

---

¹⁸ This intuition is further confirmed by the result discussed above that the higher is knowledge substitutability, the lower is steady-state growth. The important effect, in fact, is that the more homogenous is knowledge across firms, the less firms learn from each other, and the less spillovers contribute to higher productivity.

exhibits zero growth. The absence of spillovers from other firms makes R&D and the creation of new knowledge too costly an enterprise for the individual firm.

The Number of Firms

There exists a value of the number of firms, N_0 , below which the economy does not grow. Too few firms are in the market and the scope for innovation and the advantage from interaction through technological spillovers are too low to induce the firms to invest in R&D. When there are too few firms, although average profits are high, the business-stealing effect is so small that the rate of return to R&D becomes very small and drives to zero the incentives to save and invest. Firms do not pursue market expansion through R&D but profit by charging high mark-ups on marginal costs.

Recall from the previous discussion that each firm is both a source of technological opportunity and a competitor in the product market. As a source of learning, each firm improves the efficiency of the other firms. As a competitor in the product market, each firm reduces the market shares of the other firms. Two parameters drive the balance between these effects: the elasticity of knowledge substitution and the elasticity of product substitution. The substitutability of knowledge determines the effect of the rivals' knowledge on the firm's marginal return to R&D. The elasticity of product substitution has three main effects. A higher elasticity of product substitution implies a lower oligopolistic mark-up and lower profits for a given market share (gross-profit effect), but it also increases the scope for R&D competition because it increases the potential of each price reduction to steal customers from the competitors (business-stealing effect). Moreover, in GE the higher the elasticity of product substitution, the lower the ratio between final output production and average manufacturing productivity in the intermediate sector (specialization effect), and the lower the rate of return to R&D.

Consider now an increase in the number of firms, N . The industry pool of knowledge spillovers becomes larger and the rate of return to R&D rises. The scope for R&D competition increases because the market share of the competitors increases and the potential for stealing business from the competitors through price reductions is larger. In addition, an increase in the number of firms increases the ratio between total demand for intermediate goods and average productivity in the intermediate sector, and increases sales per unit of knowledge and the returns to R&D. On the other hand, an increase in the number

of firms reduces the market share of the firm, the mark-up it charges on marginal cost, and reduces its profits and its incentives to undertake R&D.

The overall effect of the number of firms, N , on the rate of growth, g , depends on the balance between the role of market fragmentation and knowledge and product variety. For given elasticity of knowledge and product substitution, η and ϵ , an increase in the number of firms, N , increases market fragmentation, thus depressing the returns to R&D, increases product variety, thus improving the returns to R&D (since $\epsilon > 1$), and increases the sources of knowledge, thus improving the returns to R&D (since $\eta > 1$). The net effect of an increase in the number of firms on steady-state growth depends on the term $[1/(\epsilon - 1) + 1/(\eta - 1)] - 1$. When it is positive, knowledge and product variety dominate the fragmentation of the market and steady-state growth increases monotonically with the number of firms. When the term is zero, a balance is reached where asymptotically these forces offset each other and the rate of growth converges to a constant.¹⁹ When the term is negative, market fragmentation dominates knowledge and product variety and as the number of firms goes to infinity steady-state growth goes to zero.

5. Balanced-Growth in Zero-Profit Equilibrium

The analysis of the previous section has discussed how opportunity, appropriability and demand conditions shape the BGP of the economy when the number of firms in the market is exogenous. I now characterize the BGP when the zero-profit condition holds and the number of firms is endogenous.

5.1. Constructing the Zero-Profit Equilibrium

In zero-profit equilibrium, firms spend all their cash flow on R&D. Recall that the NE in price and investment strategies yields the required rate of return to R&D, not how much firms spend on R&D. In an instantaneous equilibrium with free entry and exit, this is determined by the zero-profit condition. Equation (3.6) yields

---

¹⁹ It is always reasonable to assume that this constant, $L\alpha\theta(1 - 1/\epsilon)(1 - \alpha)^{1/(\eta-1)} - \rho$, is positive since the population, L , is large.

$$(5.1) \quad \frac{I}{N} = \frac{Y}{\varepsilon(N-1)+1}.$$

This equation expresses average R&D expenditure as a function of the size of the market for intermediate goods and the number of intermediate producers. R&D spending is decreasing in the number of firms for two reasons. First, the larger is the number of firms, the lower, in symmetric equilibrium, is each firm's market share. In addition, the mark-up charged on marginal cost is lower. There is another important aspect of equation (5.1). With no exogenous fixed costs, the only factor that limits the number of firms in the industry is the endogenous sunk costs due to R&D spending. Since these R&D costs are endogenous, a zero-profit equilibrium where the number of firms is finite does not necessarily exist.

Equation (5.1) can be rewritten, by using equations (2.5), (2.7), (2.8), and (4.1), as

$$(5.2) \quad g = \frac{LN^{1/(\varepsilon-1)}\theta[1+(1-\alpha)(N-1)]^{\eta/(\eta-1)}}{\varepsilon(N-1)+1}.$$

In (N, g) space this identifies the locus of the pairs of rate of growth and number of firms such that the typical oligopolist makes zero profit. I call this locus the ZZ locus. Equation (4.9), identifies the locus of the pairs of rate of growth and number of firms such that the intermediate sector is in NE and the economy is in GE. I described this locus as the balanced-growth path of the economy and referred to it as the BGP locus. The crossing of the two loci identifies the zero-profit, aggregate equilibrium.

Equation (5.2) gives the instantaneous relation between productivity growth and the number of firms. Productivity growth is decreasing in the number of firms because of the effects on margins and market shares discussed above. Not surprisingly, it is increasing in population size, L , and the R&D productivity, θ . There are two additional effects. One is the role of the static economies of scale from variety of input supply, captured by the term $N^{1/(\varepsilon-1)}$. The other is the role of technological spillovers, captured by the term $[1+(1-\alpha)(N-1)]^{\eta/(\eta-1)}$. These effects result from macroeconomic complementarities across firms and yield that aggregate output is increasing in the number of firms. As aggregate output increases, more resources can be devoted to R&D and growth rises.

5.2. Dynamics and Equilibrium Selection

When the BGP and ZZ loci cross, two zero-profit equilibria with a finite number of firms exist. Since I have set up the model with zero costs of entry and exit, both equilibria are relevant as the number

of firms, N , is a jumping variable and the economy is free to select either one at any time. In static, partial market models, one usually selects the stable equilibrium by imagining that there exists some long-run adjustment process such that the industry converges over time to the zero-profit equilibrium. Thus, although the dynamic process is not made explicit, one invokes a heuristic stability argument to rule out the equilibrium that is locally unstable. One imagines that at any moment in time the economy is always on the BGP locus. When the BGP locus is above the ZZ locus, firms are spending too much on R&D and profits are negative. Exit moves the economy to the left along to BGP locus. When the BGP locus is below the ZZ locus, firms are spending too little on R&D and profits are positive. Entry moves the economy to the right along the BGP locus. Local stability requires that the BGP locus be steeper than the ZZ locus at the equilibrium point. Figure 2 illustrates the case where the ZZ locus is everywhere decreasing and the BGP locus hump-shaped (the other cases are similar). The zero-profit equilibrium in N_1 is locally stable, while the one in N_2 is locally unstable. Initial conditions pin down whether the economy converges to the stable steady state or not. If the economy starts out with a number of firms larger than N_2 , the number of firms explodes to infinity and equations (5.2) and (4.9) fail to hold together. This is intuitive, since R&D expenditures are a fixed cost that bounds from above the number of active firms. An infinite number of firms can be supported in equilibrium only if R&D spending is zero. With zero fixed costs, the market structure should immediately degenerate to one of static monopolistic competition in the sense of Chamberlin (see, e.g., Beath and Katsoulacos 1991).²⁰

These arguments, however, are not really rigorous since one should take into account the explicit dynamics of the model. To see this, plug equation (5.2) into equation (4.8) to obtain

$$(5.3) \quad \frac{\dot{c}}{c} = \frac{LN^{1/(\varepsilon-1)}\theta[1+(1-\alpha)(N-1)]^{\eta/(\eta-1)}}{\varepsilon(N-1)+1} \cdot \left[\frac{\alpha(\varepsilon-1)(N-1)}{\sigma N[1+(1-\alpha)(N-1)]} - 1 \right] - \frac{\rho}{\sigma}.$$

Next, plug equation (5.2) into equation (4.4) to obtain

---

²⁰ A well-defined equilibrium of this type exists only if output does not explode to infinity as the number of intermediate firms becomes infinite. This argument points out that when the zero-profit equilibria with positive R&D do not exist because the NN and ZZ loci do not cross, the model may not have a long-run equilibrium. It is possible to modify the model in a way that this does not happen and a well-defined GE always exists (see, e.g., Peretto 1995a and 1995b). I omit this complication since I am not interested in the global dynamics of the system but in the properties of the BGP.

$$(5.4) \quad c = \frac{(\varepsilon - 1)(N - 1)N^{1/(\varepsilon - 1)}[1 + (1 - \alpha)(N - 1)]^{\eta/(\eta - 1)}}{\varepsilon(N - 1) + 1}.$$

One now has two choices: use equations (5.3) and (5.4) to discuss the GE dynamics in (N, c) space, or take logs and time-derivatives of equation (5.4) and reduce the system to a differential equation in the number of firms, N . The second way is more intuitive and yields

$$(5.5) \quad \frac{\dot{N}}{N} = \frac{\frac{LN^{1/(\varepsilon - 1)}\theta[1 + (1 - \alpha)(N - 1)]^{\eta/(\eta - 1)}}{\varepsilon(N - 1) + 1} \cdot \left[\frac{\alpha(\varepsilon - 1)(N - 1)}{\sigma N[1 + (1 - \alpha)(N - 1)]} - 1 \right] - \frac{\rho}{\sigma}}{\frac{\varepsilon N(N - 1) + (\varepsilon - 1)}{(\varepsilon - 1)(N - 1)[\varepsilon(N - 1) + 1]} - \frac{\eta(1 - \alpha)N}{(\eta - 1)[1 + (1 - \alpha)(N - 1)]}}$$

Setting $\dot{c} = 0$ in equation (5.3) gives exactly the two solutions that obtain from setting equation (4.9) equal to equation (5.2). It is not possible to derive exact conditions that ensure that the two loci cross. Inspection suggests that, provided the term in brackets in equation (5.3) is positive over some range of the number of firms, if population, L , is too small and/or if R&D productivity, θ , is too low, the zero-profit equilibria do not exist. One can show that the necessary (not sufficient) condition that the term in brackets be positive imposes a lower bound on appropriability which depends on the elasticity of product substitution, ε , and on the inverse of the elasticity of intertemporal substitution, σ .

Equation (5.5) describe the dynamics in $(N, \dot{N}/N)$ space. When the ZZ locus is monotonically increasing, equation (5.5) is continuous. Alternatively, when the ZZ locus is hump-shaped, equation (5.5) is discontinuous with a vertical asymptote at the point, $\tilde{N}$, where the ZZ locus has its maximum. In Figure 3, panel (a) corresponds to the case when the ZZ locus has a maximum to the right of N_2 . N_1 is a saddle point while N_2 is locally stable. The stable manifold is the interval $[N_1, \tilde{N})$. (The stable manifold is $[N_1, \infty)$ when the ZZ locus is monotonic.) There is global indeterminacy and local indeterminacy around N_2 . Panel (b) corresponds to the case when the maximum of the ZZ locus is between N_1 and N_2 . Both steady states are saddle points and the stable manifold is the pair $\{N_1, N_2\}$. Local indeterminacy does not arise but there is again global indeterminacy. In panel (c), corresponding to the case when $\tilde{N}$ is to the left of N_1 , there is local indeterminacy around N_1 while N_2 is a saddle point. The stable manifold is the interval $(\tilde{N}, N_2]$. In all cases, paths that start out outside the stable manifold converge to $(1, 0)$, to $\tilde{N}$, or

explode to $N \rightarrow \infty$. I have already argued that $(1,0)$ is not an equilibrium (see section 3). In addition, $\tilde{N}$ and $N \rightarrow \infty$ are not equilibria since equations (5.2) and (4.9) fail to hold together.

This discussion is clearly not exhaustive and much more work is needed to understand the dynamics of this class of models. I address this issue elsewhere (Peretto 1995a and 1995b).²¹ One point, however, is clear. Differently from the heuristic argument that concludes that only the steady state with few firms is stable, the explicit dynamics of the model suggest that both steady states are equally likely.

5.3. Comparative Statics

Once the steady states N_1 and N_2 are found, one can determine the comparative statics effects of the parameters on steady-state growth by looking at equation (5.2). Each parameter has a first order effect, the partial derivative of equation (5.2) with respect to the parameter evaluated at the steady state point, and a second order effect, the change in growth in response to the endogenous change in the number of firms evaluated at the steady state point. Formally,

$$(5.5) \quad \frac{dg}{d\varpi} = \frac{\partial g(N, \Omega)}{\partial \varpi} + \frac{\partial g(N, \Omega)}{\partial N} \cdot \frac{\partial N}{\partial \varpi},$$

where $\varpi \in \Omega \equiv \{L, \theta, \varepsilon, \alpha, \eta, \rho, \sigma\}$ and $g(N, \Omega)$ is equation (5.2), the ZZ locus. Consider, for example the scale effect, an increase in the size of population that leads to an increase in the rate of growth. Equation (5.5) shows that there is a positive first order effect. The second order effect, however, can take any sign. In particular, a negative scale effect is possible whenever the second order effect is negative and larger in magnitude than the first. Two sets of conditions may yield this result. First, the ZZ locus is locally downward sloping and an increase in population size, L , leads to an increase in the number of firms. Equation (5.3) shows that this is always true in the case of the steady state in N_2 . Second, the ZZ locus is locally upward sloping and an increase in population leads to a fall in the number of firms. This is always the case in the N_1 steady state.

An alternative way of checking the comparative statics is to plug equation (5.2) into (4.9) to get

---

²¹ These papers, in particular, show that the property of indeterminacy is robust to modifications of the model that introduce a pre-determined state variable. In Peretto (1995a), I introduce physical capital in the model with zero entry costs and spillovers and study the dynamics in the space of capital and consumption per effective worker. In Peretto (1995b), I introduce entry costs in the model with no physical capital and spillovers and study the dynamics in the space of the number of firms and consumption per effective worker. In both cases, I find multiple equilibria and indeterminacy.

$$(5.6) \quad g = \frac{\rho N[1 + (1 - \alpha)(N - 1)]}{\alpha(\epsilon - 1)(N - 1) - \sigma N[1 + (1 - \alpha)(N - 1)]}.$$

By construction, this locus includes the zero-profit condition in the expression for the rate of growth of the economy along its BGP. I refer to this locus as the NN locus. Figure 4 gives one example of possible equilibria given by the crossing of the ZZ and NN loci.

The sign of the comparative statics effects depend on the local slopes of the ZZ and NN loci. The slope of the ZZ locus depends on how the interaction between product and knowledge substitution and the fragmentation of the market affects the incentives to entry, given the incumbents' NE. When the locus is downward sloping, entry reduces equilibrium R&D because market fragmentation leads to lower margins and profits more than product and knowledge variety lead to higher productivity. Accommodating entry requires a reduction in fixed sunk costs and this leads to slower growth. When the locus is upward sloping, the reverse is true and entry leads to faster growth. This does not mean that accommodating entry requires higher R&D spending. Rather, it means that entry improves the productivity of knowledge, via its effects on productivity in the final sector and spillovers in the intermediate sector, more than it reduces margins and profits. Additional resources are made available by the boost to aggregate productivity, and growth can increase together with the number of firms.

The slope of the NN locus depends on how the interaction between product and knowledge substitution and the fragmentation of the market affect the incentives to engage in R&D, given that the number of firms in the market be determined by the condition that incumbents just break even. When the locus is downward sloping, market fragmentation leads to lower margins and profits more than product and knowledge variety lead to higher returns to R&D and higher aggregate productivity. Entry, then, reduces the incentives to engage in R&D and leads to lower growth. When the locus is upward sloping, the reverse is true: entry contributes to R&D efficiency, through knowledge spillovers in the intermediate sector, and higher productivity in the final sector, through a larger variety of intermediate goods, more than it induces lower R&D spending. The rate of growth increases with entry.

Consider, for example, the equilibrium in N_1 , where the NN locus is downward sloping. As the size of the market, L , increases, the number of firms falls and the rate of growth rises. This reproduces

Sutton's (1991) result that when endogenous sunk costs matter, concentration may actually increase as the market gets larger because firms escalate R&D expenditures. Note that the scale effect is always positive in this steady state. The equilibrium in N_2 shows that as the market gets larger, it does not only allow higher growth but it also accommodates a larger number of firms. This reflects the belief widely held before Sutton's work that as the market gets larger concentration falls. Nothing in the model, however, prevents the NN and ZZ loci to cross when they are both downward sloping in N_2 . This gives rise to a truly novel result. While in the previous two cases the scale effect on growth is positive, now a larger market gives rise to lower concentration and this determines too thin a dispersion of R&D resources across firms. Knowledge spillovers and economies of specialization do not compensate the fall in average R&D and the rate of growth of the economy falls.²²

8. Conclusion

I have explored the role of oligopolistic markets in a model of endogenous market structure and endogenous growth. I studied the properties of the economy's steady-state growth path focusing in particular on the role of product and knowledge variety, technological spillovers and the number of firms in determining the rate of growth of the economy. To highlight the importance of this approach, I studied the relationship between spillovers, number of firms, and growth in two cases: when the number of firms is exogenous and when the number of firms is endogenous. In the former case, things are pretty straightforward. Including the number of firms as a parameter highlights some interesting aspects of the oligopolistic interaction between firms when various types of pecuniary and technological externalities are present. Quite interesting, for example, is the (possible) hump-shaped relationship between growth and the number of firms or the degree of appropriability of knowledge.

When the number of firms is endogenous, however, things are quite different. On one hand, there are many interacting, and often offsetting, forces that make difficult to predict unambiguously the effects of changes in the parameters. I do not consider this as a shortcoming of the model. On the contrary, it is a

---

²² The scale effect is always negative, for example, if appropriability is perfect and spillovers are zero. With a little more algebra, one can show that this is also the case when the elasticity of knowledge substitution, η , is one and the productivity index (2.6) becomes Cobb-Douglas.

realistic feature that is quite consistent with the body of empirical evidence developed in Industrial Organization (see, e.g., Cohen and Levin 1989). The analysis of endogenous technological change in imperfect markets is difficult precisely because of the variety and complexity of interactions involved. In addition, the macroeconomic complementarities between firms that arise in General Equilibrium make difficult even to predict unambiguously which steady-state equilibrium the economy converges to. Global and local indeterminacy calls to attention the emphasis that Schumpeter (1934, 1942) and Rosenberg (1976) put on entrepreneurs' expectations as a driving force of the process of technological change. The assessment of this issue, however, requires much more work on the dynamics of the interaction between growth and market structure. The agenda for future research is a challenging but promising one.

References

- Aghion P. and Howitt P., 1992, A Model of Growth Through Creative Destruction, *Econometrica*, 60:323-351.
- Backus D., Kehoe P., and Kehoe T., 1992, In search of Scale Effects in Trade and Growth, *Journal of Economic Theory*, 57:377-409.
- Baldwin W. and Scott J., 1987, *Market Structure and Technological Change*, New York, Harwood Academy.
- Barro R. and Sala-i-Martin X., 1995, *Economic Growth*, New York, McGraw Hill.
- Beath J. and Katsoulacos Y., 1991, *The Economic Theory of Product Differentiation*, Cambridge, Cambridge University Press.
- Chandler A., 1990, *Scale and Scope: the Dynamics of Industrial Capitalism*, Cambridge, Harvard University Press.
- Cohen W. and Levin R., 1989, Empirical Studies of Market Structure and Innovation, in *Handbook of Industrial Organization*, edited by Schmalensee R. and Willig R., Amsterdam, North Holland.
- Cohen W. and Levinthal D., 1989, Innovation and Learning: the Two Faces of R&D, *Economic Journal*, 99:569-596.
- Dasgupta P. and Stiglitz J., 1980, Industrial Structure and the Nature of Innovative Activity, *Economic Journal*, 90:266-293.
- Dixit A. and Stiglitz J., 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.
- Freeman C., 1982, *The Economics of Industrial Innovation*, Cambridge, Cambridge University Press.
- Fudenberg D. and Tirole J., 1986, *Dynamic Models of Oligopoly*, New York, Harwood Academy.
- Fudenberg D. and Tirole J., 1991, *Game Theory*, Cambridge, MIT University Press.
- Grossman G. and Helpman E., 1991, *Innovation and Growth in the Global Economy*, Cambridge, MIT University Press.
- Kamien M. and Schwartz A., 1982, *Market Structure and Innovation*, Cambridge, Cambridge University Press.
- Klevorick A., Levin R., Nelson R., and Winter S., 1993, On the Sources and Significance of Interindustry Differences in Technological Opportunity, *Cowles Foundation Discussion Paper n.1052*, New Haven, Cowles Foundation for Research in Economics at Yale University.
- Levin R., Klevorick A., Nelson R., and Winter S., 1987, Appropriating the Returns from Industrial R&D, *Brookings Papers on Economic Activity*, 783-820.

- Lucas R., 1993, Making a Miracle, *Econometrica*, 61:251-272.
- Malerba F., 1992, Learning by Firms and Incremental Technical Change, *Economic Journal*, 102:845-859.
- Mowery D. and Rosenberg N., 1989, *Technology and the Pursuit of Economic Growth*, Cambridge, Cambridge University Press.
- Nelson R. and Winter S., 1982, *An Evolutionary Theory of Economic Change*, Cambridge, The Belknap Press of Harvard University Press.
- Peretto P., 1995a, Investment in Knowledge and Capital in Oligopoly: Balanced Growth, Poverty Traps and Indeterminacy, mimeo, Duke University.
- Peretto P., 1995b, Sunk Costs, Entry, and Endogenous Technological Change, mimeo, Duke University.
- Romer P., 1986, Increasing Returns and Long-Run Growth, *Journal of Political Economy*, 94:1002-1036.
- Romer P., 1990, Endogenous Technological Change, *Journal of Political Economy*, 98:S71-S102.
- Rosenberg N., 1976, *Perspectives on Technology*, Cambridge, Cambridge University Press.
- Rosenberg N., 1982, *Inside the Black Box: Technology and Economics*, Cambridge, Cambridge University Press.
- Scherer F. and Ross D., 1990, *Industrial Market Structure and Economic Performance*, Boston, Houghton Mifflin Co.
- Schmookler J., 1966, *Invention and Economic Growth*, Cambridge, Harvard University Press.
- Schumpeter J., 1934, *The Theory of Economic Development*, New York, Oxford University Press.
- Schumpeter J., 1942, *Capitalism, Socialism, and Democracy*, New York, Harper and Row.
- Segerstrom P., Anant T., and Dinopoulos E., 1990, A Schumpeterian Model of the Product Life Cycle, *American Economic Review*, 80:1077-1092.
- Segerstrom P., 1991, Innovation, Imitation, and Economic Growth, *Journal of Political Economy*, 99:807-827.
- Spence M., 1984, Cost Reduction, Competition, and Industry Performance, *Econometrica*, 52:101-121.
- Sutton J., 1991, *Sunk Costs and Market Structure*, Cambridge, MIT University Press.
- Tirole J., 1988, *The Theory of Industrial Organization*, Cambridge, MIT University Press.