Innovations, Patent Races and Endogenous Growth

Joseph Zeira

Hebrew University of Jerusalem and CEPR

June 2004*

Abstract

This paper presents a model of innovations and economic growth, in which patent races emerge as a result of two assumptions: R&D is directed toward specific innovations, and the cost of innovation increases with distance from the technology frontier. The paper then examines the effects of these assumptions on growth, welfare, and the market structure of R&D. There are three main results to the paper. First, the effect of scale on growth is significantly reduced. Second, R&D is Pareto-inefficient, as there is too much search for easy innovations, and too little search for difficult ones. Third, risk aversion leads to concentration of R&D in few firms, which reduces growth and increases duplication.

JEL Classification: O31, O40.
Keywords: Endogenous Growth, Innovations, Patent Races, R&D Sector.

Address:
Department of Economics
The Hebrew University of Jerusalem
Mt. Scopus
Jerusalem 91905
Israel
E-mail: mszeira@mscc.huji.ac.il

* This paper is dedicated to my uncle, the inventor Jack Drillick (1925-1991). In working on this paper I have benefited significantly from discussions with Federico Etro. I am very grateful to him. I am also indebted to Daron Acemoglu, Gadi Barlevy, Jacques Crémer, Dani Givon, Chad Jones, Shaul Lach, Albert Marcet, Motty Perry and Avi Simhon for valuable comments. Remaining errors are all mine.

Innovations, Patent Races and Endogenous Growth

1. Introduction

This paper presents a model of innovations and economic growth, which examines how patent races emerge, what is their market structure, and what is their effect on economic growth. A patent race is defined in this paper as a race between many research teams, all trying to find the same specific innovation, and the first to find it gets the patent rights. The paper adds to the standard R&D based growth models three assumptions. First, research is directed, or innovation-specific. Second, the cost of innovation increases with distance from the technological frontier. Third, there is a minimum time required to find an innovation. The paper then shows that under these plausible assumptions patent races are likely to emerge, the effect of scale on growth is significantly reduced, and competitive R&D is Pareto-inefficient, due to duplication of innovative activity.

Before describing the model we expand on the scale effect and its importance. The R&D-based growth models of Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992) have been very successful in utilizing the increasing returns to scale of innovations to explain persistent global growth of output per capita over the recent two centuries.[1] But they also faced criticism, since the scale effect they predict, namely that the rate of growth increases with the scale of the economy, is not supported by the data, as shown by Jones (1995a), Segerstrom (1998) and others.

Several attempts were made to reduce the scale effect in such endogenous growth models. The main one, by Jones (1985b), Kortum (1997), and Segerstrom (1998),

[1] In the years 1820-1992 world GDP increased 40 fold, and GDP per capita increased 8 fold. US GDP per capita increased 17 fold. See Maddison (1995).

assumes that the difficulty of innovation rises over time. This assumption reduces the scale effect, but also leads to the problematic result that economic growth is not sustainable without population growth.[2] This result also implies that technical progress is positively related to the rate of population growth, which contrasts with the experience of the last two hundred years that have seen dramatic demographic changes, while the rate of TFP growth has been quite stable. This paper introduces a small change to the assumptions of Jones and Segerstrom, namely that the cost of innovation is not increasing absolutely, but only within each period, as it becomes distant from the frontier. This change in assumption leads to very different results with respect to the dynamics of innovation, to the market structure of R&D, and to economic efficiency.

The model presents an economy, which grows through innovations that increase productivity of workers. Innovations differ by cost: the amount of researchers needed to find a specific innovation is increasing with distance from the frontier. Hence, the cost of innovations is rising within each period, but waiting till the frontier advances can reduce it. This can be justified by accumulation of knowledge over time. Thus, for example, the internal combustion engine could not be invented at the same time as the steam engine, and airbags could be invented shortly after cars, but only after the necessary experience has been accumulated. The rest of the model is similar to the original R&D-based models, where individuals choose between production and R&D, thus equalizing expected utilities across the two sectors.

As the scale of the economy – the size of population – increases, the gains from each invention increase, which induces entry to the R&D sector. Innovators can either

[2] Another solution to the problem is suggested by Young (1998) and Howitt (1999), who assume that scale cannot increase sectors, but add sectors. Hence, scale has no effect over growth rates.

begin research on a new innovation, which is more costly, as it is farther from the frontier, or join an existing patent race on a less costly innovation. This significantly reduces the effect of scale on innovation and growth. The mechanism this paper describes is that increased R&D activity leads to larger patent races and thus to more duplication, with only a small or no increase in new innovations.

Another result is that equilibrium is Paretto-inefficient, since there are too many innovators searching for the low-cost innovations. It is shown that this holds even if there are multiple research strategies for each innovation, so that researchers can diversify. In that case, they tend to crowd the more promising research strategies, which have higher probability of success. While there is too much R&D in the lucrative strategies, there is too little R&D in the less promising and more difficult lines of research.

The paper then examines when patent races do not emerge. One case is where the required time to find an innovation is not constant but flexible. In that case research teams can reduce the time of innovation by hiring more researchers and thus win the race. That eliminates patent races in equilibrium. But the paper shows that even without patent races, the two main results of the paper remain intact. Namely, the scale effect is reduced as the research team becomes too large, and it also leads to inefficiency. Another case where patent races might not be observed is when innovators are risk averse. Since participating in a race is very risky, there is incentive to share the gains from successful innovation. This can happen when a single firm employs many teams (or even all) that search for an innovation, thus internalizing the patent race. Such a firm hires many teams, even if one team alone is sufficient to find the innovation, in order to increase the probability of being first and to deter potential competitors. Hence, although patent races are not observed, the main results of the paper remain intact. Furthermore, such concentration leads to more R&D than in competitive patent races and to less growth.

In addition to the literature on R&D based growth, this paper is also related to the literature on duplication in innovation, like Dasgupta and Stiglitz (1980 a, b). In a way it combines these two literatures together. While original endogenous growth models assume that all potential innovations are equally costly, so innovators can always turn to a new innovation instead of duplicating, here duplication is restored by assuming that the cost of innovation is not uniform. The connection between duplication and patent races has already been acknowledged by Stokey (1995) and Jones and Williams (1998, 2000), but without explicitly modeling it. Segerstrom, Anant and Dinopoulos (1990) and Etro (2002), explicitly describe patent races, but examine very different issues.

The paper is organized as follows. Section 2 presents the benchmark model and Section 3 analyzes the equilibrium. Section 4 examines the effect of scale and Section 5 the inefficiency of equilibrium. Section 6 extends the analysis to multiple research strategies, while Section 7 examines what happens if the time length of innovation is variable. Section 8 shows how risk aversion leads to concentration of R&D by large firms and Section 9 concludes.

2. The Model

Consider an economy in a discrete time framework. There is a single final good in the economy, which is produced by labor. Output of the final good in period t, Y_t, depends on the labor input in the production sector L_t through the following production function:

(1) $$Y_t = a_t L_t .$$

The productivity of labor a_t reflects all available technologies at time t. This productivity rises from one period to the next through new innovations, which are therefore called in the literature "process innovations."[3] Innovations are infinitesimal. Each innovation increases productivity of each worker by an amount, which is proportional to last period productivity.[4] Innovation j increases productivity by ba_{t-1}, where $b > 0$. Thus, if I_t innovations are found in period t, the change in productivity over time is given by:

(2) $$a_t = a_{t-1}(1 + I_t b).$$

We next describe R&D activity. Innovations are searched and found by research teams. Each team searches for a specific potential innovation. Hence, unlike many endogenous growth models, research is directed and innovation-specific. Potential innovations are ordered on the real line [0, ∞). Each potential innovation j requires a different size of a team, which depends on how far the innovation is from the technological frontier. The required size of the team is therefore $s(j - f_{t-1})$, where f_{t-1} is the last innovation found in $t - 1$, and the function s is increasing. The size of a team can be greater than 1, if more than 1 innovator is needed, or smaller than 1, if an innovator can find j within less than one period of time. The function s is shown in Figure 1.

[Insert Figure 1 here]

The assumption that the cost of innovation is rising with distance from the frontier is central to the model. It means that finding innovations requires both researchers and waiting time. The marginal productivity of innovators is diminishing within each period, but waiting increases this productivity. We can find an innovation either by hiring a

[3] See Dasgupta and Stiglitz (1980 a, b). Other models, like variety, or quality ladders, have similar results.
[4] This proportionality assumption is common to all endogenous growth models. It reflects what is sometimes called the "spillover effect" of innovations.

larger team now, or by waiting till more knowledge is accumulated, and then use a smaller team next period. The production function of innovations by researchers and by waiting time displays constant returns to scale.[5]

Assume that a single research team is sufficient to find a potential innovation.[6] It is possible though that the number of teams that search for the innovation may not be equal to 1. First, it is possible that a team of the required size searches for the innovation only part time $q,\ q \leq 1$. In this case it has probability q of finding the innovation.[7] More importantly, it is possible that more than one team searches for the same innovation, but only one team finds it first. This first team gets the patent rights and sells the use of the innovation to producers of the final good. Patent rights hold for one period only and from the next period on, innovations become public knowledge. It is assumed that the probability of being first is equal to all, conditional on whether the team searches full or part time. The first team is randomly realized after the innovation effort has been invested by all the teams in the patent race. The size of the patent race, the number of teams that search for innovation j in period t, is denoted $n_{j,t}$. Then, the probability of success is:

$$(3) \qquad P_{j,t} = \begin{cases} \dfrac{1}{n_{j,t}} & \text{if team searches full time,} \\[2ex] \dfrac{q}{n_{j,t}} & \text{if team searches } q \text{ time and } n_{j,t} \geq 1 \end{cases}$$

We next describe individuals in the model. The population consists of non-overlapping generations, where each individual lives one period. There is no population

---

[5] Doubling the amount of researchers by spreading them equally in two consecutive periods of time doubles the amount of innovations.
[6] We therefore implicitly assume that there is only one way to search for each innovation. In Section 6 we replace this assumption with a more realistic one of multiple search strategies.
[7] Alternatively we can assume that an innovation is found only if team works full time. In that case patent races have only integer size and not continuous size as in this paper, but the results are the same.

growth and the size of each generation is *N*. As we see below, *N* is the variable that determines the scale of the economy. Individuals are assumed to be risk neutral.[8] The utility they derive from consumption *c* is described by:

$$u = c . \tag{4}$$

There is free entry to work either in production or in research. A producer derives income from producing the final good but if she uses a new innovation in production, she has to pay patent fees to the patent holder, who has won the patent race in the same period.[9] Innovators, who belong to a successful team, receive income from patent fees. Unsuccessful innovators do not earn income. Individuals decide whether to work in innovation or in production before knowing which team is first in the patent race.

3. Equilibrium

3.1 Markets for Innovations

Consider an innovation *j* found in period *t*, which is sold to producers in the same period *t*. The inventing team with the patent rights has monopoly over the innovation. Denote the patent fee paid in period *t* for innovation *j* by $z_{j,t}$. The production workers are willing to purchase the innovation as long as their net income from using it is greater or equal to their income without using the innovation:

$$ba_{t-1} - z_{j,t} \geq 0 . \tag{5}$$

Hence, the demand for the innovation is described by the following step function:

$$Q_{j,t}(z_{j,t}) = \begin{cases} L_t & \text{if } \; z_{j,t} \leq ba_{t-1} \\ 0 & \text{if not.} \end{cases} \tag{6}$$

[8] This assumption is changed in Section 7, where the effect of risk aversion is analyzed.
[9] A version of the model where innovations are used one period later yields similar results.

The monopoly patent holder maximizes profits by setting the fee at the maximum price and all workers purchase the innovation and use it. The patent fee is therefore equal to

(7) $$z_{j,t} = ba_{t-1} .$$

Hence, patent fees are equal for all innovations in period t.

3.2. Income and Employment

Due to free entry, expected income of workers must be equal for production workers and for innovators alike. Denote the income of production workers in period t by w_t. Since their gains from new technologies are equal to what they pay as patent fees we get:

(8) $$w_t = a_{t-1} .$$

The income of a successful innovator, whose team wins the patent race for innovation j in period t, is equal to:

(9) $$\frac{z_{j,t} L_t}{s(j)} = \frac{ba_{t-1} L_t}{s(j)} .$$

The income of an unsuccessful innovator is equal to 0.

The size of employment in the production sector L_t is related to the size of the R&D sector, R_t, through the equilibrium condition in the labor market:

(10) $$L_t = N - R_t .$$

3.3. The Size and Number of Patent Races

We next turn to determine the amount of innovations in period t, I_t, and the size of each patent race $n_{j,t}$. The expected income of an innovator, whose team searches for innovation j an amount of time q is:

(11) $$\frac{qz_{j,t} L_t}{s(j) n_{j,t}} = bqa_{t-1} \frac{N - R_t}{n_{j,t} s(j)} .$$

Innovators compare this expected income to the alternative income from production, qa_{t-1}. Hence, a research team has an incentive to raise q as much as possible at j, if $n_{j,t}s(j)$ is low, until $q = 1$, or until $n_{j,t}s(j) = b(N - R_t)$, namely until the patent race is full. As long as expected income of innovators is higher, more teams join and $n_{j,t}$ increases, until expected income in the two sectors is equal. Hence, the equilibrium number of teams working on innovation j in period t is:

$$n_{j,t} = \frac{b(N - R_t)}{s(j)}. \qquad (12)$$

The size of the patent race for innovation j is negatively related to the required size of the team $s(j)$ since it is positively related to the returns from innovation. The more lucrative innovations attract more research teams.

We next turn to determine how many innovations are found in each period. As the size of research team increases, the size of the patent rate declines, until it reaches 1. There cannot be a smaller race, since then $b(N - R_t) < s(j)$, and as a result the expected income of innovators in such a part time team is lower than income from production, as $bqa_{t-1}(N - R_t)/s(j) < qa_{t-1}$. Hence, no one joins this race. As a result, research stops exactly at the patent race of one team only, which determines the amount of innovations in period t:

$$s(I_t) = b(N - R_t). \qquad (13)$$

The number of innovations in period t depends positively on the size of the production sector $N - R_t$. This is the scale effect.

Hence, innovations depend negatively on the size of the R&D sector, as shown by curve IN in Figure 2. As the size of the research sector increases the production sector decreases and with it the incentive for innovation. When the R&D sector reaches the size $\overline{R} = N - s(0)/b$, the production sector becomes so small that no innovation is profitable and $I = 0$.

[Insert Figure 2 here]

3.4. Equilibrium R&D

As shown above, the size of each patent race depends on the overall size of the R&D sector R_t. At the same time the size of the R&D sector itself depends on the individual patent races, since:

(14) $$R_t = \int_0^{I_t} n_{j,t} s(j) dj \, .$$

Using (12) and (13) we can rewrite the size of the patent race as a function of the amount of innovations I_t only:

(15) $$n_{j,t} = \frac{s(I_t)}{s(j)} \, .$$

Substitute (15) in (14) and get:

(16) $$R_t = I_t s(I_t) \, .$$

This function, which describes how the size of the R&D sector depends positively on the amount of innovations, is described by the curve *RD* in Figure 2.

The intersection of the two curves, *RD* and *IN*, determines a unique equilibrium, which describes both the rate of innovation and the size of the R&D sector. The equilibrium is a steady state, and from here on we delete time subscripts. The equilibrium amount of innovation can be described by the following equilibrium condition, which is derived from (13) and (16):

(17) $$s(I)(1+bI)=bN.$$

The amount of innovations determines also the equilibrium rate of growth, *g*:

(18) $$g=\frac{a_t-a_{t-1}}{a_{t-1}}=bI_t=bI.$$

We can now calculate the size of the R&D sector. From (13), (16) and (18) we get: $R=Ib(N-R)=g(N-R)$ and hence:

(19) $$\frac{R}{N}=\frac{g}{1+g}.$$

The growth rate is related to the relative and not to the absolute size of the R&D sector.

In order to better understand the equilibrium in this model and how it is reached, it is useful to focus on the equilibrating mechanism in comparison with other endogenous growth models. Consider a situation where the expected income of innovators in equation (11) is high, raising the incentive to become innovators. In the original endogenous growth models equilibrium is reached by reducing the production sector, as more people become researchers, until the scale of production is small enough. In Jones (1995b) and Segerstrom (1998) equilibrium is reached by reduction of expected returns through searching for more costly innovations, namely increasing the denominator in (11). This paper presents an additional mechanism, which further increases the denominator in (11), namely increasing the size of the patent race, thus reducing the probability of success.

3.5. Patent Races

As shown above the equilibrium determines not only the size of the research sector and the amount of new innovations per period, but also the maximum size of patent races *n*, which is equal to:

(20) $$n = \frac{b(N-R)}{s(0)} = \frac{s(I)}{s(0)}.$$

Hence, patent races emerge if the difficulty of innovation is increasing. Intuitively, as some innovations create higher income than others, they attract more research teams, so that patent races emerge. Note, that patent races can become large only if difficulty of innovation increases enough, so that $s(I)$ is much larger than $s(0)$.

4. The Effect of Scale

In this section we study the effect of changes in scale, namely in size of population N, on the equilibrium. We use this analysis in order to show that our basic assumptions reduce significantly the strong scale effect in such a R&D based endogenous growth models. An analysis of the effect of scale is best done by use of Figure 2. As N increases, the IN curve shifts to the right. That increases both the size of the R&D sector R and the amount of innovations I. The size of patent races also increases with scale, as $s(I)$ increases.

Note first, that innovation and growth begin only when the economy passes some threshold scale or size. As long as $N < N_0 = s(0)/b$, the IN curve in Figure 2 is zero everywhere and there are no innovations and no economic growth. The reason is that the scale is so small that the incentives for innovation are too low. Only when the size of population exceeds the threshold N_0, innovations and growth begin.

Also, the positive effect of scale on innovations is diminishing with scale. One reason is that additional innovations require more researchers, since s is increasing. The other reason is that as scale increases, some of the additional innovators do not search for new innovations but join existing patent races. They increase these races and do not add to innovation and growth. To see it formally use the equilibrium equation (17), and get:

(21) $$\frac{\partial I}{\partial N} = \frac{1}{s'(I)(I+b^{-1})+s(I)} < \frac{1}{s(I)}.$$

Hence, the effect of scale on innovations is positive but diminishing, as $s(I)$ increases. The effect of scale on the rate of economic growth is similar, since $g = bI$. Furthermore, if the function s is not only increasing but unbounded, namely if $s(j) \xrightarrow[j\to\infty]{} \infty$, then the scale effect is not only diminishing but it converges to zero as scale increases.[10]

A comparison of these results with the original endogenous growth models reveals many similarities. First, an economy with a fixed population can enjoy sustained economic growth through innovations. Another similar result is that scale is critical to the take-off of economic growth, as there is a threshold size necessary for innovation and growth. Another similarity is that scale increases the size of the R&D sector. But here the models depart significantly. While in the original endogenous growth models all new researchers contribute to finding innovation, in this model only some find new innovations, while others join existing patent races. Hence, the scale effect is diminishing and can even approach zero.

The reduction of the scale effect enables us to better fit the model with the empirical evidence. While scale of production increased significantly in the world in the recent two centuries, TFP growth has been fairly stable, at a rate of around 1% annually.[11] We can therefore specify the function s to better fit these stylized facts of global growth. This is achieved if we assume that the function s rises to infinity at some finite amount of innovations. Formally, if we assume that there exists a level I^* such that:

(22) $$s(j) \xrightarrow[j\to I^*]{} \infty.$$

[10] The assumption that s is unbounded is equivalent to assume an Inada condition in the production of innovations, namely that the marginal productivity of inventors is diminishing to zero.

This restriction on s yields the following results. First, since the curve *RD* in Figure 2 is now bounded by I^*, the equilibrium amount of innovations is bounded by I^* as well. Hence, the rate of growth is bounded as well: $g \leq bI^*$. This specification therefore fits the empirical findings of Jones (1995a) and others.

5. Patent Races and Pareto Inefficiency

If many innovation teams search for the same innovation, while one team alone can find it, there is inefficiency due to misallocation of resources. To see this formally, consider a central planner, who reallocates individuals between production and R&D. This planner can assign only one innovation team for each innovation, and assign all others to work in production. This reduces the size of the R&D sector from $\int_0^I n_j s(j)dj$ to $\int_0^I s(j)dj$, as $n_j \geq 1$ for all j, and increases the size of the production sector. The rate of growth remains the same, while the level of output in each period increases. That means the equilibrium is not Pareto-Optimal.

This inefficiency is due to duplication of innovation activity in patent races. Such duplication inefficiency has been already observed in microeconomic studies of innovation, such as Dasgupta and Stiglitz (1980 a, b). Most endogenous growth models do not have duplication, since R&D is not direct. Stokey (1995) and Jones and Williams (1998, 2000) acknowledge the possibility of duplication and its effect on endogenous growth, which Jones and Williams call the “stepping on toes effect,” but do not model it explicitly. This paper fills this gap by modeling endogenous patent races.[12]

---

[11] See Maddison (1995), Table 2.6.

[12] Segerstrom, Anant and Dinopoulos (1990) have patent races without duplication. Etro (2002) has patent races and duplication, but focuses on different issues than this paper.

The inefficiency of equilibrium raises the question whether there is too much or too little R&D, which has been recently discussed in Stokey (1995), Jones and Williams (1998, 2000), and Li (2001). This model gives a different answer than these papers, as R&D is not general but innovation-specific. Hence, there is both too much and too little R&D. Due to duplication there is too much R&D on innovations, which attract a patent race of more than 1 team. We next show that there is not enough R&D on the innovations beyond *I*, which are delayed to future periods, for a number of reasons. First, because patent life is less than infinity and second, because of the usual external spill-over effect, as innovations benefit not only direct users, but also future innovators. Finally, our model introduces an additional external effect to innovations. Innovating today pushes the technology frontier forward and thus reduces the cost of finding future innovations. Namely, innovating today produces knowledge that makes future innovations less costly.

We next show these effects formally. Assume that the central planner maximizes the discounted sum of present and future incomes, since individuals are risk-neutral. The subjective discount rate is an intergenerational rate *r*. Increasing innovation only in period 0 increases present and future discounted output by:

$$a_{-1} L \frac{1+r}{r-g} b \, .$$

Present production falls by $a_{-1}s(I)$. Since according to equation (13), the equilibrium rate of innovation satisfies $s(I) = bL$, the benefit from increasing innovation is greater than the cost. Therefore, equilibrium innovation is below the social optimum. Hence, while R&D should not be supported for innovations that attract large patent races, it should be supported for marginal innovations that at *I* and beyond.

6. Multiple Research Strategies

As shown in Section 5, the equilibrium in the benchmark model leads to Pareto-inefficiency, since too many innovators are trying to find the same innovation. It seems that this result depends on an implicit assumption, that all teams search for the innovation in the same way, except that only one of them finds it first. Hence, the inefficiency might disappear if innovators could search in many different ways, when new teams can diversify the search for the innovation and increase the probability of finding it. In this section we examine this possibility of multiple research strategies for each innovation. Interestingly, having multiple research strategies does not remove the inefficiency and patent races are still overcrowded. The research strategies, which are mostly crowded, happen to be those with the highest success probabilities.

For the sake of simplicity consider the case, where the size of research teams is fixed and equal to 1, namely one innovator is required for each team, for innovations $0 \leq j \leq I*$, while s is infinite for innovations beyond I*. Assume that for each innovation $0 \leq j \leq I*$ there are infinitely many different ways to search for it. We call them 'research strategies' and number them by m, $1 \leq m < \infty$. The probability of finding the innovation by using strategy m is p_m. The strategies are ordered by decreasing probability of success: $p_1 > p_2 > p_3 > \ldots$. The strategies are disjoint, so that the probability of finding the innovation is $\sum_{m=1}^{M} p_m$ if strategies 1, …, M are followed. If all strategies are followed the probability of success is 1. In other words, one and only one of the research strategies can lead innovators to find the innovation.

The probabilities of the research strategies can in principle vary over innovations and over time, but for tractability, assume that they are the same for all innovations and for all times. Hence, we must further specify these probabilities to be: $p_m = p(1-p)^{m-1}$ for all *m*, where $0 < p < 1$. Under this specification, if an innovation is not found in period *t*, after using methods 1,…, *M*, innovators can use the remaining untried research strategies from next period on, and by renumbering the strategies, *M*+1 becoming 1 etc., and the conditional probabilities of success become the original probabilities. Hence, the probabilities of the various research strategies are the same, whether research on the innovation has just begun, or whether it has been going on for some time.

We assume, as in the benchmark model, that a research strategy can be used by a number of teams. Conditional on the success of this strategy, if it leads to finding the innovation, each of the teams who follow it has equal probability of reaching the innovation first, if working full time. Hence, if the number of teams who try to find innovation *j* using research strategy *m* in time *t*, is $n_{m,j,t}$, the success probability of each is:

$$P_{m,j,t} = \begin{cases} \dfrac{p_m}{n_{m,j,t}} & \text{if the team works full time,} \\ \dfrac{qp_m}{n_{m,j,t}} & \text{if the team works } q \text{ time and } n_{m,j,t} \geq 1, \\ qp_m & \text{if the team works } q \text{ time and it is the only team.} \end{cases}$$

Note that most of the analysis of the benchmark case carries through to this case of multiple research strategies. The market for each innovation looks the same, patent fees are the same and so are the wage rate and the employment levels. The expected returns for an innovator, who uses research strategy *m* to find innovation *j* are:

$$\frac{qp_m b a_{t-1}(N - R_t)}{n_{m,j,t}}.$$

Hence, innovators look for the strategy with the highest $p_m / n_{m,j,t}$ and then use it for the maximum amount of time. If the patent race of the *m* strategy is too large they work part time on *m* and part time on *m*+1. This equates the expected returns from all strategies. In equilibrium expected income from innovation and from production is equal, and we get:

(23) $$n_{m,j,t} = p_m b(N - R_t) .$$

Hence, patent races emerge in this case as well. The strategies with the highest success probability have the largest races. As *m* increases and the success probability diminishes, the size of the patent race declines as well. The research strategies that are followed are $m = 1,...,M$, where *M* is the highest that satisfies[13]

(24) $$p_M \geq \frac{1}{b(N - R)} .$$

This condition determines the strategies that are followed and also the probability *P* of finding the innovation in each period of time: $P = p_1 + p_2 + ... + p_M$.

In order to close the equilibrium we calculate the size of the R&D sector:

$$R = I * \sum_{m=1}^{M} n_m = I * b(N - R) \sum_{m=1}^{M} p_m = I * b(N - R)P .$$

A simple calculation yields:

(25) $$\frac{R}{N} = \frac{I * bP}{1 + I * bP} .$$

Hence, the share of the R&D sector in population increases with the probability of success *P*. Note that *I*bP* is the rate of economic growth *g*. Equation (25), which reflects the amount of researchers needed to achieve this success level in equilibrium, is presented by the curve *RN* in Figure 3.

The probability of success depends on the size of the R&D sector also through the scale effect, since:

(26) $$P = \sum_{m=1}^{M} p_m = \sum \left\{ p_m \middle| p_m \geq \frac{1}{b(N-R)} \right\} = \sum \left\{ p_m \middle| \left(1 - \frac{R}{N}\right) p_m \geq \frac{1}{bN} \right\}.$$

This step function, which describes how the probability of success depends on scale, is presented by the curve *PR* in Figure 3. The intersection of the two curves determines the equilibrium in the economy.

[Insert Figure 3 here]

The equilibrium determines the number of research methods adopted each period *M*, the probability of finding innovations *P*, the rate of growth $g = I * bP$, and the relative size of the R&D sector $g/(1+g)$. The effect of scale on innovation and growth is similar to the benchmark model. As *N* increases, innovations become more profitable, the *PR* curve shifts to the right, more research strategies are followed, and more innovations are found. But this positive effect of scale on growth is diminishing, as patent races increase as well.

The main result of this extension of the model is that despite the many research strategies for each innovation, we still have duplication of innovative activity through patent races. Actually, there are too many innovators using the more promising strategies, with low *m*, while the economy can benefit from putting innovators to work on the marginal strategies, like *M*+1, as that increases the chance of finding the innovation. Hence, the equilibrium is not Pareto-efficient. A Pareto-improving policy should aim at reducing the number of innovators working on the more promising research strategies 1,

[13] We suppress the time index from here on, as it is clear that the equilibrium is a steady state.

…, M, while promoting the more risky research beyond M. The problem is that such a policy is very hard to implement, due to moral hazard and other practical difficulties.

7. Rushing to Be First

This paper presents a theory of patent races based on three assumptions: R&D is innovation-specific, the cost of innovation is increasing with distance from the frontier, and also there is a fixed time required to find an innovation, which is 1 period of time. In this section we relax this third assumption and examine how it affects patent races, innovations, growth and efficiency. We assume that by increasing the size of the research team we can reduce the time required to find the innovation. Clearly, this eliminates all patent races. If there is a patent race of size n, each team knows that if it increases its size a bit it will reach the innovation before the others and win the race. Hence each team has an incentive to reduce the time of finding the innovations as long as $n > 1$. Hence, in equilibrium innovations will be found sufficiently fast, so that patent races are of size 1. We next show that despite the elimination of patent races in this case, the main results of the paper are kept intact under very plausible assumptions. Namely, we show that the scale effect is significantly reduced and the equilibrium is inefficient.

In order to make timing endogenous we introduce a few changes to the benchmark model. First, we assume that time is continuous. Second, we assume that individuals are infinitely lived, with discount rate r, but the size of population is N as above. Third, we assume that size of the team required for finding innovation j depends both on its distance from the frontier and on the time it takes to find the innovation, T. The size of the team is described by:

(27) $$s(T, j - f_t) = \begin{cases} h(T) & \text{if } j - f_t \le I^* \\ \infty & \text{if } j - f_t > I^*. \end{cases}$$

Namely, it is assumed for simplicity that the number of innovations is bounded by *I**, as in section 6. The function *h* is decreasing in *T*, but it is assumed that *T* is bounded from below by some minimum time to innovate *T**. further assumed that its elasticity is larger than 1, namely if we want to reduce the time of finding an innovation by half, we have to more than double the number of researchers. This is a very plausible assumption that reflects the additional coordination efforts of using a larger team in order to find an innovation earlier.

We next solve for the equilibrium in the economy, assuming for simplicity that all *I** innovations are searched simultaneously. Since timing of innovation is set such that only one team searches for each innovation, net profits for an innovator in time 0 are:

$$\frac{a_0 b(N-R)}{h(T)} \int_T^\infty e^{-rt} dt - a_0 \int_0^T e^{-rt} dt.$$

In equilibrium these net profits are equal to 0 and thus we derive the following equilibrium condition:

$$b(N-R) = h(T)(e^{rT} - 1).$$

To this condition we add another equilibrium condition on the size of the R&D sector:

$$R = I^* h(T).$$

Remembering that the rate of growth is: $e^{gT} = 1 + bI^*$ and combining the equilibrium conditions we get the following general equilibrium condition:

(28) $$bN = h(T)(e^{rT} - 1 + bI^*) = h(T)T\left(\frac{e^{rT} - 1}{T} + \frac{bI^*}{T}\right) \approx h(T)T(r+g).$$

As scale N increases, T is reduced, as can be seen from (28). As a result growth rises, since innovations are produced more rapidly and the size of the R&D sector rises as well. Note though, that g rises by less than the increase in N, since $h(T)T$ rises as well. Hence, the scale effect is reduced in this extension of the model as well. The reason is that the single research team for each innovation becomes very large to enable it to reach the patent rapidly. Furthermore, if h is more elastic, the effect of scale on growth becomes smaller.

We next show that in this case the equilibrium may be inefficient as well, when teams become too large when they rush to reach the patent first. The present value of output in the economy is described by:

$$V = (N - R)\sum_{n=0}^{\infty} a_0 e^{(g-r)Tn} \int_0^T e^{-rt} dt.$$

In order to insure convergence of welfare we assume that the rate of growth g is lower than r, which imposes a lower bound on T. This means that the function h goes to infinity as T approaches some minimum level T^*. After some manipulation of the present value of output we get:

$$V = \frac{a_0}{r} \frac{e^{rT} - 1}{e^{rT} - 1 - bI^*} \left[N - I^* h(T) \right].$$

Clearly, reducing the time of research T increases the present discounted value of output of each production worker, but reduces the number of these workers. To find the optimum we calculate the derivative of V with respect to T and get:

(29) $$\frac{bN - bh(T)I^*}{e^{rT} - 1} = h'(T) \frac{e^{rT} - 1 - bI^*}{re^{rT}}.$$

Note that the equilibrium condition (28) can be rewritten as:

(30) $$\frac{bN - bh(T)I^*}{e^{rT} - 1} = h(T).$$

Since $h(T)/h'(T)$ goes to zero as T diminishes, the RHS of (29) exceeds the RHS of (30) from some point on. Since the LHS in both equations is identical it means that optimal T is larger than competitive T. Hence, competitive innovation is more rapid than optimal innovation.

Hence, removing the assumption of a minimum required time for innovation, leads to equilibrium without patent races, but with the main results of the paper, namely a reduced effect of scale on innovation and growth, and inefficiency, as there are too many researchers. These results therefore do not depend on patent races, but rather on the other two basic assumptions in this paper, namely that R&D is innovation-specific and that the cost of innovation is increasing with distance from the frontier.

8. Risk Aversion and Concentration of R&D

This section presents another case in which we may observe much less patent races than predicted by the benchmark model. This section shows that when individuals are risk averse, the search for innovations tends to be concentrated in a small number of large R&D firms and in many cases even carried by a single monopolistic R&D firm. Hence, our theory can also account for the well known empirical phenomenon of concentration of much of R&D in large firms. But this section shows, that despite the concentration of search within few firms, the main results of the paper remain intact, namely the scale effect is significantly reduced and much of R&D is inefficient.

Next we return to the benchmark model, but replace the original assumption of risk neutrality with risk aversion. Note that although the gains from innovation are very high, the probability of success is low, if the patent race is large. This creates a strong incentive for innovators to share this risk. This can be done in two possible ways. One is by insurance across innovations. The other is by forming a coalition of teams who search for the same innovation, where they share the gains from innovation if one of them finds it first. This way they have the same expected income, but less risk. The first type of insurance is usually represented in reality by venture capital. The second type of insurance, of forming a coalition of teams, is usually hard to observe in reality, due to problems of free riding and contract enforcement. These problems can be solved by large research firms, which hire many research teams to search for an innovation. We show later that both types of insurance lead to the same aggregate results. This section focuses on the second type of insurance, created by large firms that internalize the patent races within them.

Interestingly, such firms do not eliminate the inefficiency created by patent races, despite their incentive to reduce the number of innovators, in order to increase the returns from successful innovation per innovator. The reason is that such firms have an opposite incentive to increase the number of innovators in order to increase the chance of getting first, and also to deter potential entrants. Hence, they still hire too many researchers.

To formalize the analysis, we change the benchmark model and assume that consumers are risk averse and utility is described by:

(31) $$u = \log c .$$

Let us also assume that innovators do not work full time in innovation, but work a share x of their time as production workers, $x < 1$.[14] Hence, even in case of failure in innovation

[14] Otherwise no individual enters a patent race with risk of zero income, and the model becomes trivial.

they have some positive income: $xw_t = xa_{t-1}$. We further assume for simplicity that each innovator can work at one innovation at most, namely that $s(0) > 1$.[15] Let us denote the size of fees per innovator in a winning team by f:

$$f = \frac{b(N-R)}{s(j)}.$$

For the sake of simplicity assume, as in the benchmark model, that there is only one research strategy for each innovation.

We first solve the model under perfect competition, where research monopolies are not allowed, or if the cost of forming them is very high. A team enters the patent race if its expected utility is greater or equal to that of working in production. Hence, if the number of the other teams is n, an additional team enters until:

$$\frac{1}{n+1}\log(f+x) + \left(1-\frac{1}{n+1}\right)\log x = 0. \tag{32}$$

This condition describes how the sizes of patent races depend on the fee per innovator f, which depends on the size of the research sector R. The equilibrium is determined by this condition together with a condition that relates the size of the R&D sector R to the sizes of the patent races.

We next turn to the case of concentrated research within firms, which enable an increase of expected utility through risk sharing. For the sake of simplicity we overlook the problems of contract enforcement, and view such firms as groups of researchers, who divide equally the returns from a successful innovation. We further assume that firms participate in a Cournot competition. Let the number of teams in such a firm be k. This

[15] The ability to develop more than one innovation already supplies some insurance and reduces the need to create mechanisms for risk sharing.

number is chosen to maximize the expected utility of each innovator, taking the number of teams in the other firms, *n*, as given. Hence, *k* is determined by maximization of:

$$\left(\frac{k}{n+k}\right)\log\left(\frac{f}{k}+x\right)+\left(1-\frac{k}{n+k}\right)\log x=\log x+\frac{k}{n+k}\log\left(\frac{f}{kx}+1\right). \tag{33}$$

Note that this maximization presents the opposing interests of the firm. On the one hand it has an incentive to reduce *k* in order to increase the return to innovator in case of success, *f*/*k*. On the other hand it has incentive to increase *k* in order to increase the probability of success *k*/(*n*+*k*).

The overall amount of research activity for this innovation is determined by the entry condition. Firms enter the market until the expected income of innovators equals that in the production sector, namely until (33) is equal to zero:

$$\log x+\frac{k}{n+k}\log\left(\frac{f}{kx}+1\right)=0. \tag{34}$$

The maximization of expected utility and the entry condition together determine the equilibrium *k* and *n*. The number of research firms is (*n* + *k*)/*k*.

We can now outline the first result of this section. Since (33) is larger than (32), as *k* is not restricted to be 1, the number of teams in case of cooperation, *n* + *k*, is larger. In other words, allowing firms to internalize the patent race and to share risk increases the number of teams searching for the same innovation. The intuition for this result is straightforward: the ability to share risk induces more people to become innovators and to join the patent race. Hence, allowing firms to concentrate research for innovation leads to more R&D, and to more duplication.

A second result is obtained when *f* declines, as *s* increases and innovations become more difficult. It can be shown that as *f* declines, there are less firms and less teams as well. Thus, as the size of innovation teams increases, the patent race becomes more concentrated. For the most difficult innovations we have a research monopoly, which completely internalizes the patent race. Such a monopoly completely eliminates the risk of participation in the patent race and hence innovators enter as long as expected income from innovation exceeds income from production. Hence, the number of teams in such a monopoly k is determined by:

$$\frac{f}{k} = 1 - x .$$

As f further declines innovation stops when $k = 1$. In this case the research monopoly consists of one big team only, when $f = 1 - x$. This marginal innovation is the same as in the case of competition. But under cooperation $N - R$ is lower. The reason is that the size of the R&D sector is larger, due to larger number of teams for lower j. Hence, in order for f to be equal to $1 - x$ we need to have lower $s(I)$ and hence lower I. Thus, allowing concentration increases the R&D sector, but reduces the rate of innovation and the rate of growth in the economy.

This result is even exacerbated if instead of large research firms insurance is achieved by venture capitalists. If risk is completely diversified across all innovations, it further increases the incentive by innovators to enter, as it is as if the large firms above are all monopolistic. Hence, R increases by even more. Hence, there is even more duplication and there are less innovations in equilibrium.

9. Summary and Conclusions

This paper analyzes the role of patent races in an endogenous growth model. It introduces to the R&D-based growth literature two new elements. The first is that research is no longer general but directed, namely innovation-specific. The second element is the importance of time in the process of innovation. The time it takes to find an innovation, and how it is related to the size of the innovation team, and the time it takes to wait till an innovation becomes ripe and easier to find. These two elements enable us to examine when patent races emerge. Furthermore, directed research and cost reducing waiting time significantly reduce the positive effect of scale on growth. A larger economy increases the incentive to innovate, but most new innovators crowd existing patent races, while only few search for additional innovations.

But the paper's contribution goes beyond the effect of scale on growth, to the welfare analysis of R&D. It directs our focus to areas with too much R&D and to areas with too little. The paper shows that researchers tend to crowd the more promising research strategies, which have higher probabilities of success, while there is too little R&D at the low probabilities of success. It stops short of offering policy recommendations, as the practical problems of formulating an optimal policy are huge. An ideal policy should support only those innovators who travel the less frequented ways, namely those who try strategies with lower probabilities of success. Hence, this model suggests that research incentives should be given to those who deviate from the crowd and who are doing less standard and more risky research. It is hard to find practical policies that identify such researchers, but it is definitely worth looking for.

References

Aghion, Philippe, and Howitt, Peter. "A Model of Growth Through Creative Destruction," *Econometrica*, Vol. 60 (2), March 1992, p. 323-351.

Aghion, Philippe, and Howitt, Peter. *Endogenous Growth Theory*. Cambridge, MA: MIT Press, 1998.

Barro, Robert J., and Sala-i-Martin, Xavier. *Economic Growth*. New York: McGraw Hill, 1995.

Dasgupta, Partha and Stiglitz, Joseph. "Uncertainty, Industrial Structure, and the Speed of R&D," *Bell Journal of Economics*, Vol. 11 (1), Spring 1980 (a), p. 1-28.

Dasgupta, Partha and Stiglitz, Joseph. "Industrial Structure and the Nature of Innovative Activity," *Economic Journal*, Vol. 90 (358), June 1980 (b), p. 266-293.

Etro, Federico. "Patent Races and Schumpeterian Growth," mimeo, February 2002.

Grossman, Gene M. and Helpman, Elhanan. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press, 1991.

Howitt, Peter. "Steady Endogenous Growth with Population and R & D Inputs Growing," *Journal of Political Economy*, Vol. 107 (4), August 1999, p. 715-730.

Jones, Charles I. "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics*, Vol. 110 (2), May 1995 (a), p. 495-525.

Jones, Charles I. "R & D-Based Models of Economic Growth," *Journal of Political Economy*, Vol. 103, August 1995 (b), p. 759-784.

Jones, Charles I. And Williams, John C. "Measuring the Social Return to R&D," *Quarterly Journal of Economics*, Vol. 113, November 1998, p. 1119-1135.

Jones, Charles I. And Williams, John C. "Too Much of a Good Thing? The Economics of Investment in R&D," *Journal of Economic Growth*, Vol. 5, March 2000, p. 65-85.

Kortum, Samuel. "Research, Patenting, and Technological Change," *Econometrica*, Vol. 65 (6), November 1997, p. 1389-1419.

Li, Chol-Won. "On the Policy Implications of Endogenous Technological Progress," *The Economic Journal*, Vol. 111, May 2001, p. C164-C179.

Maddison, Angus. *Monitoring the World Economy 1820-1992*. Paris, France: Organization for Economic Co-operation and Development, 1995.

Panagopoulos, Andreas. "When Does Patent Protection Stimulate Innovation?" mimeo, 2003.

Romer, Paul M. “Endogenous Technical Change,” *Journal of Political Economy*, Vol. 98 (5, pt.2), October 1990, p. S71-S102.

Segerstrom, Paul S. “Endogenous Growth without Scale Effects,” *The American Economic Review*, Vol. 88 (5), December 1998, p. 1290-1310.

Segestrom, Paul S., Anant T.C.A. and Dinopoulos, Elias. “A Schumpeterian Model of the Product Life Cycle,” *The American Economic Review*, Vol. 80 (5), December 1990, p. 1077-1091.

Stokey, Nancy L. “R&D and Economic Growth,” *Review of Economic Studies*, Vol. 62 (3), July 1995, p. 469-489.

Young, Alwyn. “Growth without Scale Effects,” *Journal of Political Economy*, Vol. 106 (2), February 1998, p. 41-63.

Figures

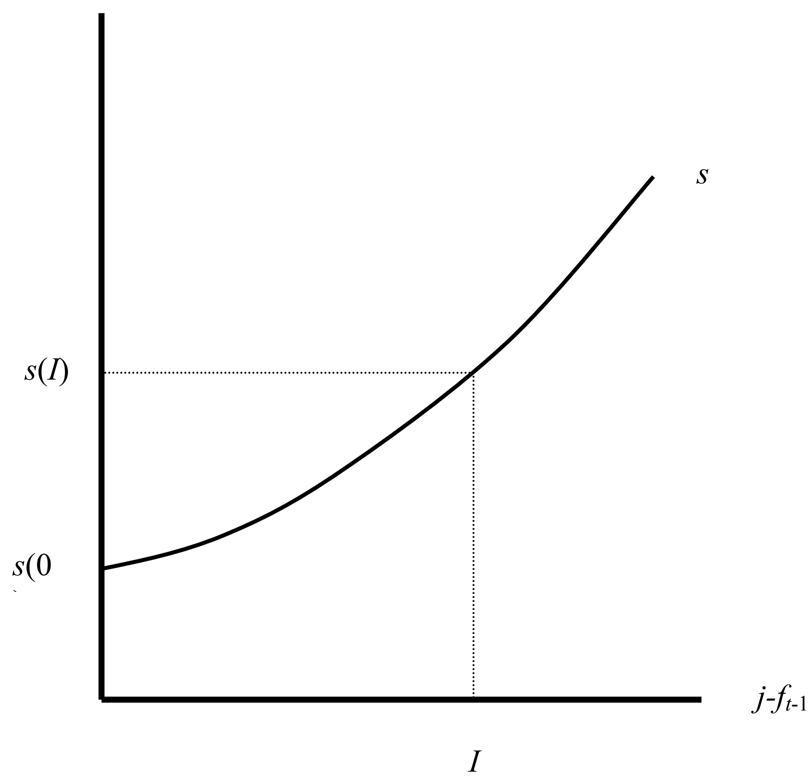

Figure 1

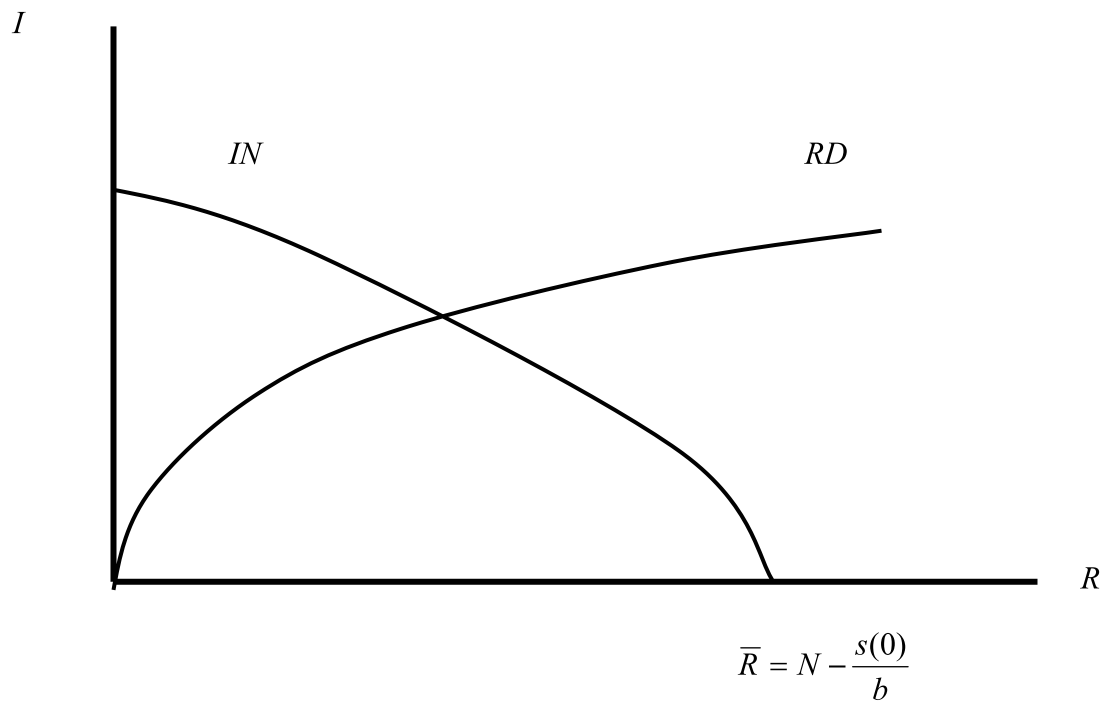

Figure 2

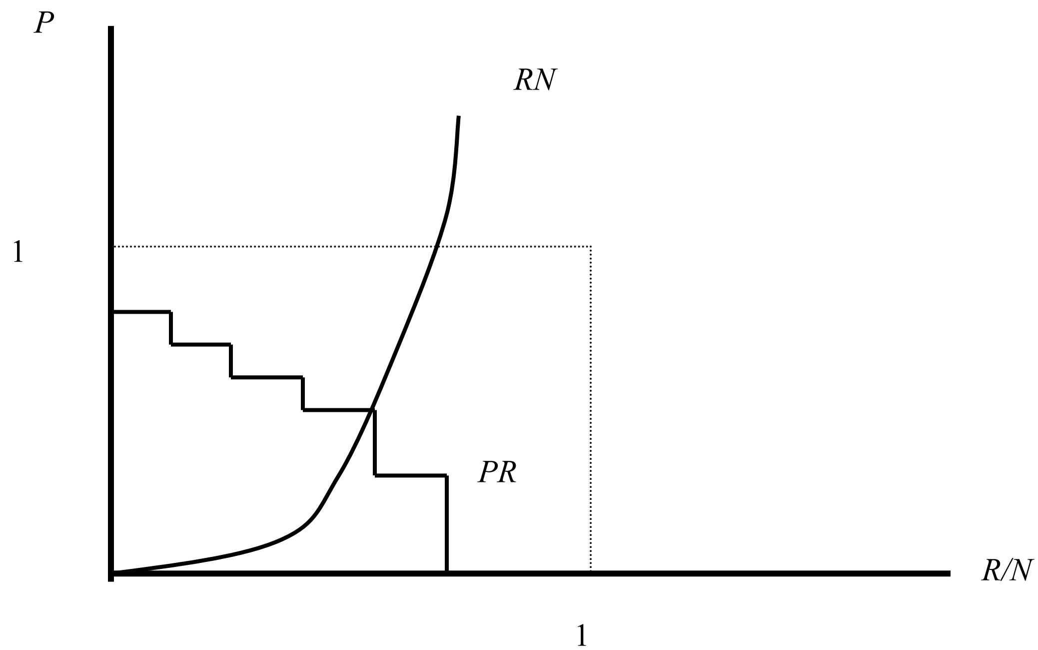

Figure 3