

Innovations, Patent Races and Endogenous Growth

Joseph Zeira

Hebrew University of Jerusalem, Harvard University and CEPR

February 2003*

Abstract

This paper presents a model of innovations and economic growth, which departs from earlier endogenous growth models by assuming that innovations require not only labor of innovators, but also time. Each innovation has to wait till its time comes, and speeding it up is costly. This change in assumption yields a number of important results. First, the model gives rise to patent races, where many researchers search for the same innovation. Second, the rate of growth does not rise too much with the scale of the economy. Third, R&D is Pareto-inefficient, as too many researchers look for the same breakthrough. Fourth, concentration of R&D by monopolistic firms is explained in this model by risk aversion, where patent races are internalized within one firm in order to share risk.

JEL Classification: O31, O40.

Keywords: Endogenous Growth, Innovations, Patent Races, R&D Sector.

Address:

Kennedy School of Government

Harvard University

69 JFK Street

Cambridge, MA 02138

E-mail: joseph_zeira@ksg.harvard.edu

* 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 Dani Givon, Chad Jones, Shaul Lach, 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 departs from the original R&D-based endogenous growth models in one single assumption.¹ While these models assume that it takes only innovators, or R&D workers, to find new inventions, this paper assumes that it takes both innovators and time. Namely, it is not enough to put many people to work on a project in order to develop it, if its time has not come yet. The need for time is inherent to the process of innovation, which is gradual as new innovations build on previous ones. Thus, for example, finding the proof for the last theorem of Fermat took many Mathematicians, but it also took some time as well. Similarly, the internal combustion engine could not be invented at the same time as the steam engine, but only years later, after the necessary experience and knowledge were accumulated. Sometimes even the need for an innovation emerges only after some time and experience, like air bags for cars.

Hence, this paper views the process of innovation as an intrinsically gradual process, which needs two inputs: innovators and time. This change in assumption leads to significantly different results from the original endogenous growth model. First, this model generates patent races, where many researchers, or research teams, search for the same potential innovation. Second, in this model the effect of scale on the rate of growth is diminishing, so the model does not have the strong scale effect, which original endogenous growth models have. Third, R&D is Pareto-inefficient, as there are too many

¹ The main R&D based endogenous growth models, to which I refer in the paper, are Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992).

research teams searching for the same innovation. This problem increases with the scale of the economy. Finally, the paper shows that concentration of much R&D by monopolistic research firms can be a result of risk aversion.

The paper develops an aggregate model of economic growth through innovations that increase the productivity of workers.² Innovation activity requires both inventors and time, so the number of innovators required for each innovation within a period of time rises with innovations. In other words, the cost of innovation is increasing within each period. Individuals can choose between production and innovation. If many innovators are searching for an innovation, only the one who finds it first, gets the patent rights and sells its use to production workers for patent fees. The sizes of the production sector and of the R&D sectors are determined by equalization of expected utilities across 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 begin research on a new innovation, which is more costly, or join the patent race in one of the less costly innovations. This explains two of the main results of the paper. First, patent races emerge, in which many research teams search for the same innovation. Second, an increase in scale leads only few researchers to new innovations, while many join existing patent races and do not raise the rate of innovation. This explain the smaller effect of scale in this model. The reduction of the scale effect is important, since the strong scale effect of the original endogenous growth models is in odds with the empirical evidence. Jones (1995a) shows that increases in R&D in OECD countries did not lead to higher growth rates. The explanation this paper offers to Jones' findings is that

² This paper therefore models technologies as 'process innovation,' as in Dasgupta and Stiglitz (1980 a, b). Other modeling devices, such as variety of intermediates, or quality ladders, lead to similar results.

increased R&D activity leads to larger patent races and thus to more duplication, with only a small or no increase in innovations.

The third result, that equilibrium is inefficient, follows from the above. There are too many innovators, who search for the same innovation. The paper shows that this waste due to duplication holds even if there are multiple research strategies for each innovation, so that innovators can join the search for the same innovation but use different methods. The paper shows that few innovators use new research strategies, while many join existing strategies, especially the most promising ones. Hence, there could be too much R&D activity in this case as well.

Finally, the paper discusses the effect of risk aversion of patent races. Risk aversion can lead to cooperation between participants of a patent race, since by sharing the gains from innovation they keep the same expected income but reduce risk. Such cooperation is hard to achieve voluntarily, but it can emerge when a single firm employs all the teams that search for an innovation, thus internalizing the patent race. Note that such a firm would hire many teams, even if one is sufficient to find the innovation, in order to deter potential competition. We show that such R&D monopolies have more teams than competitive patent races, so they are even less efficient.

This paper is related to the literature of endogenous growth, which began with Romer (1986) and Lucas (1988), and had contributed significantly to understanding global economic growth. Within this literature we focus on a series of papers that modeled the creation of technical progress within a macroeconomic framework of sustained growth, mainly Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992). These models have been very successful in using the increasing returns to

scale of innovations to explain the high rates of global economic growth over the recent two centuries.³ But these models also faced criticism, because their scale effect on growth appears to be too strong, and the prediction that the rate of growth increases unboundedly with scale contradicts the data, as shown by Jones (1995a), Segerstrom (1998) and others.

As a result of this critique, some attempts were made to reduce this strong scale effect. Some papers, like Jones (1985b) and Segerstrom (1998) assume that the difficulty of innovation rises over time. This assumption eliminates the scale effect, but these models have a problematic result, that growth stops without population growth, namely increases in scale become necessary to technical progress. Other attempts to eliminate the strong scale effect, like Young (1988) and Howitt (1999), assume that the scale of production is bounded within each sector, and if the overall scale of the economy increases, it translates into more sectors. In this way the scale has no effect over growth within each industry.

This paper follows a different route, and instead of bounding innovation over time, or bounding scale, it bounds innovations within each period, due to the gradual nature of innovations, which cannot be invented all at the same time, but follow previous innovations, one at a time. This view enables the paper to develop a micro-oriented analysis of the innovation process, of how innovators compete with one another, on how they might use similar or different strategies in their research. Hence, this paper has wider results than just elimination of scale effects. It explains patent races, their market structure, and how they affect efficiency.

³ 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).

The paper is also related to the literature on duplication in innovation activity, as described for a single innovation by Dasgupta and Stiglitz (1980 a, b). This model extends this literature to a general equilibrium framework with many innovations. Note, that the way one moves from a single innovation to many innovations is crucial. The original endogenous growth models assumed that all potential innovations are equally costly. Hence, innovators can always turn to a new innovation instead of duplicating another one, so that duplication and its inefficiency disappeared from these models. This model restores duplication by assuming that the cost of innovation is increasing.

The paper is organized as follows. Section 2 presents the model. Section 3 analyzes the equilibrium in the case of a single research strategy, while Section 4 examines the effects of scale. Section 5 discusses the inefficiency of equilibrium and Section 6 extends the analysis to multiple research strategies. Section 7 examines how risk aversion leads to concentration of R&D by large monopolies that internalize patent races. Section 8 concludes.

2. The Model

Consider an economy with a single final good, which is produced by labor. Time is assumed to be discrete. Labor is continuous. Productivity of labor, using all the available technologies at time t , is a_t . Namely, each infinitesimal worker i can produce an amount $a_t di$ of the final good. This productivity rises from one period to the other through innovations, which are therefore what is called in the literature “process innovations.” Innovations are infinitesimal as well. Each innovation increases productivity of each

worker by an amount, which is proportional to last period productivity.⁴ Innovation j increases productivity of worker i by $ba_{t-1}didj$, where $b > 0$. Thus, if last period productivity has been a_{t-1} and if I_t is the amount of innovations found in period t , this period's productivity is:

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

In each period innovations are found by research teams, or by innovation teams. The size of such a team is s , where s can be greater than 1, if more than 1 innovator need to work for a whole period of time, or can be smaller than 1, if an innovator can find the innovation within less than 1 period of time. The difficulty of finding innovations is increasing. Ordering potential innovations in period t on the real line $[0, \infty)$, then the size of necessary team to find innovation j , $s(j)$, is increasing. After finding innovations $[0, I_t)$ in period t , next period innovations move to the origin and the labor cost of finding them decreases, from $s(I_t)$ to $s(0)$, etc. This reflects our main assumption that in order to find innovations we need both researchers and time. In some sense we can say that the production function of innovations by researchers and time is of constant returns to scale.⁵ While it is sufficient to have a single research team in order to find a potential innovation, it is possible that more than one team might search for the same innovation. In that case only one team finds the innovation first with equal probability to all searching teams.⁶ This team gets the patent rights and sells the use of the innovation to

⁴ This proportionality assumption is common to all endogenous growth models. It reflects what is sometimes called the “spillover effect” of innovations: they increase productivity of future innovators.

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.

producers of the final good. If the number of teams that search for innovation j in period t is $n_{j,t}$, the probability of success for each team is:

$$(2) \quad P_{j,t} = \begin{cases} 0 & \text{if } n_{j,t} = 0 \\ \frac{1}{n_{j,t}} & \text{if } n_{j,t} \geq 1. \end{cases}$$

We next describe individuals in the economy. Assume a non-overlapping generations economy, where individuals live one period each. There is no population growth and the size of each generation is a continuum of size N . Individuals are assumed to be risk neutral, so that utility from consumption c is:⁷

$$(3) \quad u = c.$$

There is free entry to labor to either production or research. A worker, who produces the final good in period t , sells it in the market to earn income. If she uses a new innovation in production of the intermediate good, she has to pay patent fees to the patent holder, who has innovated it in the same period t .⁸ A successful innovator receives income from patent fees. We assume that patent rights hold one period only and from the next period on all innovations become public knowledge. We further assume that individuals decide whether to search for an innovation or work in production before they know whether they succeed or not in finding the innovation first.

3. Equilibrium

3.1 The Markets for Innovations

Consider an innovation j found in period t , which is sold to production workers in the same period t . The inventing team that has 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 after using it is greater or equal to their income without using the innovation:

$$(4) \quad ba_{t-1} - z_{j,t} \geq 0.$$

Hence, the demand for the innovation is a step function that depends on the amount of production labor in period t , L_t :

$$(5) \quad q_{j,t}(z_{j,t}) = \begin{cases} L_t & \text{if } z_{j,t} \leq ba_{t-1} \\ 0 & \text{if not.} \end{cases}$$

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

$$(6) \quad z_{j,t} = ba_{t-1}.$$

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

3.2. Income and Employment

Due to free entry the expected income of workers must be equal for production workers and for innovators as well, before knowing whether they have succeeded or not in their patent race. 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:

$$(7) \quad w_t = a_{t-1}.$$

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

⁷ This assumption is changed in Section 7, where the effect of risk aversion is analyzed.

⁸ A version of the model where innovations are used one period later yields similar results.

$$(8) \quad \frac{z_{j,t}L_t}{s(j)} = \frac{ba_{t-1}L_t}{s(j)}.$$

The income of an unsuccessful innovator, whose team does not find the innovation first, 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:

$$(9) \quad L_t = N - R_t.$$

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 $n_{j,t}$, namely how many innovators participate in the search for innovation j , $0 \leq j \leq I_t$. The expected income of an innovator who is searching for innovation j is:

$$(10) \quad \frac{z_{j,t}L_t}{s(j)n_{j,t}} = ba_{t-1} \frac{N - R_t}{n_{j,t}s(j)}.$$

When individuals decide whether to become production workers or innovators they compare this expected income with the alternative income of working in production, which is $w_t = a_{t-1}$. As long as the expected income of innovators who search for j is higher than a_{t-1} , more innovators join and $n_{j,t}$ increases. If the expected income of innovators is lower than a_{t-1} , innovators leave the race and $n_{j,t}$ is reduced. Hence, the number of innovators working on innovation j in period t is the largest integer that satisfies:

$$(11) \quad n_{j,t} \leq \frac{b(N - R_t)}{s(j)}.$$

As a result, if the size of the production sector is large enough, innovations, which require a small number of researchers to find them, namely a small s , tend to have a large patent

race, since the income per innovator in case of success is high. Hence, this model leads to emergence of patent races. Formally, the condition for patent races of more than 1 research team is: $b(N - R_t) > 2s(0)$.

We next turn to determination of the amount of innovations found in each period. As equation (11) shows, the size of patent races declines with s , as difficulty of research increases. Hence, the patent races for the most difficult innovations consist of one team only. Finally, we reach a point that the next innovation is not searched at all, when the size of the patent race falls below 1, namely when:

$$(12) \quad s(I_t) = b(N - R_t).$$

This condition determines the amount of innovations found in period t , and it clearly depends positively on the size of the production sector $N - R_t$. This is the scale effect.

3.4. Equilibrium R&D

As shown above, the sizes of patent races depend on the overall size of the R&D sector R_t . But the size of this sector itself depends on the sizes of individual patent races, since:

$$(13) \quad R_t = \int_0^{I_t} n_{j,t} s(j) dj.$$

Note that due to (11) and (12) we can redefine the size of the patent race as a function of the amount of innovations I_t only. Namely, the size $n_{j,t}$ is the largest integer that satisfies:

$$(14) \quad n_{j,t} \leq \frac{s(I_t)}{s(j)}.$$

Equations (13) and (14) together determine a functional relation between the amount of innovation and the size of the R&D sector:

$$(15) \quad R_t = RD(I_t).$$

This function is described by the curve RD in Figure 1. This is an increasing function with discontinuities, or jumps, which occur when the size of the largest patent race increases by 1, as the number of innovations increases. It can be shown that the maximum size of the patent race jumps from $n - 1$ to n at innovation $I(n)$, which is defined by:

$$(16) \quad s[I(n)] = ns(0).$$

[Insert Figure 1 here]

The second curve in Figure 1, IN , shows how the amount of innovations depends on the size of the R&D sector, as described by equation (12). Note, that 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 $\bar{R} = N - s(0)/b$, the production sector becomes so small that no innovation is profitable and $I = 0$. The intersection of the two curves, RD and IN , determines the equilibrium rate of innovation and the size of the R&D sector. The equilibrium is time independent, namely it is a steady state.⁹

Knowing the amount of innovations in equilibrium enables us to calculate the rate of growth of the economy, g , which depends on the equilibrium amount of innovations and the productivity of each innovation:

$$(17) \quad g = \frac{a_t - a_{t-1}}{a_{t-1}} = bI_t = bI.$$

The equilibrium also determines the size of the R&D sector R . A direct calculation of this size is quite complicated due to jumps, as the size of patent races changes, but we can

⁹ As a result we delete from here on the time subscripts.

calculate some bounds. It can be shown that the amount of researchers working on innovation j , $n_j s(j)$, always satisfies: $s(I)/2 \leq n_j s(j) \leq s(I)$. Hence the size of the R&D sector satisfies the following constraints:

$$(17) \quad \frac{Is(I)}{2} \leq RD(I) \leq Is(I).$$

Some calculation yields:

$$(18) \quad \frac{g/2}{1+g/2} \leq \frac{R}{N} \leq \frac{g}{1+g}.$$

This is an interesting result. Unlike the original endogenous growth models, which predict that the rate of growth g rises with the absolute size of the R&D sector R , here it is related to its relative size. This result fits better the empirical findings of Jones (1995a), who shows that recent decades have experienced a rather stable rate of TFP growth, together with a sharp increase in the R&D sector, though its relative size has been more stable.

In order to better understand the equilibrium in this section and how it is reached, it is useful to compare it with the equilibrating mechanism in the original endogenous growth models. If the scale of production is large, innovations are very lucrative and the incentive to become innovator is high. In this model the equilibrium is reached by reduction of expected returns, either by moving to more costly innovations or by increasing the size of the patent race and thus reducing the probability of success. In the original endogenous growth models the equilibrium is reached by reducing the production sector, as more people become researchers, until the scale of production becomes small again. This mechanism is clearly less realistic.

3.5. Emergence of Patent Races

As shown above the equilibrium determines the size of the research sector, the amount of new innovations per period and also the maximum size of patent races. This is determined in Figure 1 by the number of the steps in the RD curve, below the intersection with IN . Formally, it is the highest n , which satisfies: $I(n) \leq I$. Clearly, the main result of this model is the formation of patent races, with size greater than 1. This result holds in the model only if the IN curve intersects the RD curve above $I(2)$. It can be shown that a sufficient and necessary condition for that is the existence of $I(2)$, namely if s is sufficiently increasing so that for some I , $s(I) > 2s(0)$. Hence, the condition for emergence of patent races is that the difficulty of innovation is sufficiently increasing. Intuitively, as some innovations create higher income than others, they attract more research teams, so that patent races emerge.

Note, that in the original endogenous growth models the function s is constant, and hence no $I(2)$ exists and there are no patent races. Note further, that in order to have patent races, it is not enough to assume that s is increasing, but we have to assume that it is increasing enough to allow for patent races to form. Hence, from here on we further assume that the difficulty of innovation is not only increasing but it is unbounded, so:

$$(19) \quad s(j) \xrightarrow{j \rightarrow \infty} \infty .$$

This assumption can be justified in the following way. While assuming that s is increasing is equivalent to assuming that marginal productivity of innovators is diminishing, assuming that s is unbounded is equivalent to assuming that marginal productivity satisfies the Inada conditions, namely it goes to zero as the number of innovators increases. Assuming that s is unbounded means that patent races emerge, and

that they even become larger and larger with the scale of the economy, and their size is unbounded as well.

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 conduct a comparison between this model and the original R&D based endogenous growth models. We then contrast the results of the model with the historical experience of economic growth in the last two centuries in order to add some restrictions on the function s , which describes the increasing difficulty of innovation within the same period of time.

An analysis of the effect of scale is best done by use of Figure 1. As N increases, the IN curve shifts parallel to the right. That increases the size of the R&D sector R . As for the number of innovations I , it either increases or remains equal, if the two curves intersect in regions where the RD curve is horizontal. In such regions an increase in scale does not increase the number of innovations, but increases patent races for already researched innovations. Figure 2 describes how the equilibrium amount of innovations I depends on the scale N . This is a non-decreasing function with horizontal steps, which we denote by E . Note, that the effect of scale on the rate of economic growth is similar, since $g = bI$.

[Insert Figure 2 here]

If the economy has a small scale, such that $N < N_0 = s(0)/b$, the IN curve in Figure 1 is zero everywhere and there are no innovations and no economic growth in equilibrium. The reason is that the scale of the economy is too small and the incentives

for innovation are too low. When the size of population exceeds the threshold level N_0 , we begin to have innovations and growth. First there is only one team for each innovation and no patent races. As N increases, and the equilibrium amount of innovation exceeds $I(2)$, patent races begin to form. As N further increases, the maximum size of patent races increases as well. At each horizontal segment of the innovations curve in Figure 2 this maximum size increases by 1. This increase in size of patent races also explains why the curve E in Figure 2 is concave. As the size of the research sector R increases with scale, only some of the new researchers go to find new innovations. The others join existing patent races.

A direct analysis of the function $I = E(N)$ is complicated due to the integer sizes of patent races. But we can use equations (12) and (17) to calculate some bounds to this function. An upper bound \bar{I} is defined by:

$$(20) \quad N = \bar{I}s(\bar{I}) + \frac{s(\bar{I})}{b}.$$

A lower bound \underline{I} is defined by:

$$(21) \quad N = \frac{1}{2}\underline{I}s(\underline{I}) + \frac{s(\underline{I})}{b}.$$

Using these bounds it is easy to see that if, as assumed at the end of Section 3, s is unbounded, then the slope of the innovation curve in Figure 2 tapers off to zero:

$$(22) \quad \frac{\partial I}{\partial N} \xrightarrow{N \rightarrow \infty} 0.$$

Hence, the effect of scale on the rate of economic growth is positive but this effect is decreasing, due to increasing patent races, and it even goes to zero when scale becomes very large.

We can now turn to compare the results of this model to the original endogenous growth models. This model keeps the main result of this literature, namely that an economy with a fixed population can grow indefinitely at a fixed rate, through innovations. In both models scale is critical to the take-off of economic growth. Namely, there is some threshold size of population below it no innovations are found and there is no growth. When scale increases beyond this threshold, innovation and growth begin. In both models 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 find new innovations and contribute to growth, in this model only some find new innovations, while others join existing patent races. Hence, the scale effect on economic growth, which is so strong in the original endogenous growth models, where the slope of the E curve is fixed, is diminishing in this model and is even approaching zero.

The reduction of the scale effect is important, since it were found to be in strong contrast with the empirical evidence of the last two centuries. While scale of production increased significantly in the world, TFP growth has been fairly stable, at a rate of around 1% annually.¹⁰ In this model the scale effect is much weaker, but is not bounded, as it appears to be from the data. Hence, in order to better fit the main stylized facts of global growth in the last two centuries, namely increasing scale and bounded growth rates, we further specify the function s . From here on we assume that the function s not only rises unboundedly, but it does so at some finite amount of innovations. Formally, we assume that there exists a level I^* such that:¹¹

$$(23) \quad s(j) \xrightarrow{j \rightarrow I^*} \infty .$$

¹⁰ See Maddison (1995), Table 2.6.

This additional restriction on s yields the following results. The curve RD in Figure 1 is now bounded by I^* , and hence the equilibrium amount of innovations is bounded by I^* as well. As a result, the rate of growth is bounded too: $g \leq bI^*$. Hence, the effect of scale on growth is not only diminishing but bounded as well. Under this assumption, therefore, the model fits the basic stylized facts of global growth fairly well.

We next introduce a simple example for such a bounded innovation function. In this example the difficulty of innovation in each period is constant up to some level I^* , above which the difficulty becomes infinite, namely:

$$(24) \quad s(j) = \begin{cases} s^* & \text{if } j \leq I^* \\ \infty & \text{if } j > I^*. \end{cases}$$

It is easy to show that under this specific function the economy finds all potential I^* innovation each period, if the size of population exceeds $s^*(1/b + I^*)$. Then the rate of growth is $g = bI^*$, and the size of each patent race is the largest integer, which is smaller than $Ng/[s^*(1 + g)]$. Hence, in this example scale increases only patent races, while the rates of innovation and growth remain constant. We call this specific example the “constant growth case.”

5. Patent Races and Pareto Inefficiency

If many innovation teams are searching for the same innovation, while one team alone can find it, there is inefficiency due to misallocation of resources. To see this more formally consider a central planner, who reallocates individuals between production and R&D, in this economy. This planner can assign only one innovation team for each

¹¹This specification means that the production function of innovators within one period is bounded.

innovation, and assign all others to work in production. The size of the R&D sector will fall from $\int_0^I n_j s(j) dj$ to $\int_0^I s(j) dj$, and the size of the production sector will increase. Hence, the rate of growth will be the same, while the level of output in each period will increase. That means the equilibrium is not Pareto-Optimal.

The inefficiency is due to duplication of innovation activity in patent races. Inefficiency due to duplication of research has been already observed by economists, mostly in microeconomic studies of innovation, such as Dasgupta and Stiglitz (1980 a, b). But these studies discuss a single potential innovation, where innovators do not have the option to begin a new innovation project. When macroeconomic models of innovation, mostly endogenous growth models, have introduced many potential innovations within each period, duplication disappeared. Instead of doing the same as other innovators do and earning less, innovators could start a new project. What this paper claims is that this result was not only due to moving from a single innovation to many innovations, but also due to the assumption that all innovations have equal difficulty.

Note that in this result, the model drastically differs from former endogenous growth models, in most of which there is too little R&D. The reason is the spillover effect of innovations, namely that they increase productivity of future innovators, who do not pay for it. As a result of this externality, optimal R&D should exceed equilibrium R&D. While this argument holds in this model as well, there is also too much R&D due to duplication in patent races. We next try to examine what dominates, the waste in patent races, or the external effect of innovation? In other words, we ask whether there is too much or too little R&D in our model. To answer this question we examine maximization of the discounted sum of all present and future incomes, where the use of income justified

by risk-neutrality. The subjective discount rate is the intergenerational rate ρ . Hence, for each level of innovation I , the discounted sum of income is:

$$(25) \quad [N - RD(I)] \sum_{t=0}^{\infty} \frac{(1+g)^t}{(1+\rho)^t} = [N - RD(I)] \sum_{t=0}^{\infty} \frac{(1+bI)^t}{(1+\rho)^t} = (1+\rho) \frac{N - RD(I)}{\rho - bI}.$$

If we assume that $\rho > bI^*$, namely that the discount rate exceeds the maximum rate of growth, which is reasonable empirically, we get that the function (25) is concave and we can search for a maximum. The derivative of (25), where RD is continuous, at the equilibrium, where $s(I) = N - RD(I)$, has the same sign as $1 - \rho + bI$. Since time periods in this model are quite long, an annual discount rate of .05 becomes $\rho = 1.65$ for 20 years, and an annual growth rate of TFP of .01 becomes $bI = .22$ over 20 years. Hence $1 - \rho + bI$ is negative. Hence (25) is decreasing at the equilibrium level, either when the function is continuous or discontinuous since the jumps have a negative sign. That means that the R&D sector should be smaller even at the expense of a lower growth rate, for reasonable discount rates. Hence, the inefficiency due to duplication in patent races is quite significant, so that it overcomes the gains from increasing innovations and growth.

This result becomes even more apparent in the case of constant growth, which is presented in Section 4 by equation (24). In that case, if population is large enough so that the rate of innovation is I^* , increasing the R&D sector has no effect at all on the rate of innovation. It only increases the size of patent races, and thus increases the inefficiency in the economy. In this case, therefore, reducing the R&D sector is clearly a Pareto-improving policy, as long as patent races do not fall to size smaller than 1, namely as long as all the I^* available innovations are discovered each period.

6. Equilibrium with 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. Note, that this result seems to depend on an implicit assumption, namely that all innovation 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 for the same innovation in many different ways, since innovators can explore these new ways and increase the probability of finding the innovation. In this section we examine this possibility, by exploring the equilibrium if there are multiple research strategies for each innovation. Interestingly, having multiple research strategies does not remove the inefficiency in R&D and patent races are still overcrowded. The research strategies, which are mostly crowded, happen to be those with the highest success probabilities.

Let us consider for the sake of simplicity the constant growth case, where the size of research teams is fixed, s^* for innovations $0 \leq j \leq I^*$, while it is infinite for innovations beyond I^* . To simplify the analysis even further assume that $s^* = 1$, namely that there is one innovator in each research team. 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 while using strategy m is p_m . We order the strategies by decreasing probability of success: $p_1 > p_2 > p_3 > \dots$. The strategies are independent of one another, so that the probability of finding the innovation is $\sum_{m=1}^M p_m$ if strategies $1, \dots, M$ 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 enable innovators to find the innovation.¹² These probabilities can in principle vary over innovations and over time, but in order to simplify the presentation, we assume that these probabilities 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 $0 < p < 1$.¹³ Under this specification, if an innovation is not found in period t , after using methods $1, \dots, M$, innovators can use the remaining untried research strategies from next period on, and by renumbering the strategies, $M+1$ becoming 1 etc., 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 further assume, as in the benchmark model, that each research strategy can be used by a number of innovators. Conditional on the success of this strategy, namely that it leads to finding the innovation, each of the innovators who follow it has an equal probability of reaching the innovation first. Hence, if the number of innovators who try to find innovation j by using research strategy m in time t , is $n_{m,j,t}$, the success probability of each innovator is:

$$(26) \quad P_{m,j,t} = \begin{cases} 0 & \text{if } n_{m,j,t} = 0 \\ \frac{p_m}{n_{s,j,t}} & \text{if } n_{s,j,t} \geq 1. \end{cases}$$

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 as well and so are the wage rate and the employment levels. The

¹² One way to think about it is as if it is search in a large area. It is somewhere there, but each team can search only in part of the area.

¹³ Namely, the probability of success of each strategy, conditional on failure of prior strategies, is p .

expected returns for an innovator, though, depend on the research strategy chosen by the innovator. The expected returns from innovation j for an innovator using strategy m are:

$$(27) \quad \frac{p_m b a_{t-1} (N - R_t)}{n_{m,j,t}}.$$

Hence, the size of the patent race for innovation j of those who use strategy m should be the largest integer satisfying:

$$(28) \quad n_{m,j,t} \leq p_m b (N - R_t).$$

Hence, patent races emerge in this case as well. They are largest for the strategies with the highest success probability. As m increases and the success probability diminishes, so do patent races. The research strategies that are followed are $m = 1, \dots, M$, where M is the highest number that satisfies:¹⁴

$$(29) \quad p_M \geq \frac{1}{b(N - R)}.$$

Formally, the sizes of patent races are either n_m or $n_m - 1$, where n_m is the highest integer that satisfies: $n_m \leq p_m b (N - R)$.

Knowing the sizes of the patent races for each potential innovation and for each research strategy we can sum them up to find the size of the R&D sector:

$$(30) \quad R \leq I * \sum_{m=1}^M n_m \leq I * b (N - R) \sum_{m=1}^M p_m = I * b (N - R) P.$$

Note that P is the probability that an innovation is found, for each innovation and at each period of time. Note also that the difference between the LHS and RHS in the inequality

¹⁴ We suppress the time index from here on, as it is clear that the equilibrium is time invariant except for the level of productivity.

(30) cannot be very big and is bounded by I^*M . Hence, if we divide both sides of (30) by N , the inequality can be approximated by equality and we get:

$$(31) \quad \frac{R}{N} = \frac{I^*bP}{1+I^*bP}.$$

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 (31), 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 through another channel, which is the scale effect, since:

$$(32) \quad P = \sum_{m=1}^M p_m = \sum \left\{ p_m \mid p_m \geq \frac{1}{b(N-R)} \right\}.$$

We can rewrite this scale effect in terms of the relative size of the R&D sector and get:

$$(33) \quad P = \sum \left\{ p_m \mid \left(1 - \frac{R}{N}\right) p_m \geq \frac{1}{bN} \right\}.$$

This step function, which describes how the probability of success depends on scale, is presented by the curve PR in Figure 4. The intersection of the two curves determines the equilibrium in the economy.

[Insert Figure 4 here]

The equilibrium sets the number of research methods that are adopted each period M , the probability of finding innovations P , the expected rate of growth of productivity $g = I^*bP$, 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. If N increases, innovations become more profitable, which shifts the PR curve to the right. As a result more research

strategies are followed, more innovators are operating and more innovations are found. The economy grows at a higher rate, but this gain in growth rate is diminishing with scale, since the rate of growth is bounded by $I*b$.¹⁵

The main result of this version 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 1, ..., M , while promoting the more risky and marginal research beyond M . Such a policy is clearly very hard to implement, due to moral hazard and other practical difficulties.

Obviously, a policy that increases overall R&D activity, like a subsidy, might have adverse effects, especially if scale is large. The marginal benefit from increasing M is diminishing as p_M is diminishing, and hence it is diminishing with scale, as M increases with scale. Furthermore, patent races become larger with scale. Hence, the efficiency loss from more R&D increases with scale. Hence, the net marginal benefit of increasing R&D is diminishing with scale, and becomes negative at some level.

7. Risk-Aversion and Concentration of R&D

Our paper shows that if the global economy is large enough, there will be patent races for all innovations and many innovators will participate in these races. But in reality we

¹⁵ In a similar way to the benchmark model, a small enough scale leads to zero growth, since the PR curve becomes horizontal at 0.

observe many cases in which innovations are searched by a small number of large R&D firms and in many cases R&D is concentrated within a single monopoly. Interestingly, our model can account for this phenomenon as well, by attributing it to risk aversion. To see this let us replace our original assumption of risk neutrality by risk aversion. Note that although the gains from innovation are very high, the probability of success is low, as there are many innovators in the patent race. This creates a strong incentive for these innovators to form a coalition, where they will share the gains from innovation if one of them finds it first. This way they can retain expected income, but with smaller risk. Since such arrangements are hard to create cooperatively, due to problems of free riding and contract enforcement, these problems are usually solved by a research monopoly, which hires many research teams to search for an innovation.

Interestingly, such a monopoly that takes over all the research for a potential innovation does not eliminate the inefficiency created by patent races. On the one hand such a monopoly has an incentive to have as few as possible innovators, in order to increase the share of the returns from innovation to each worker. But on the other hand the monopoly needs to have a sufficient number of innovators within it to deter potential entrants to the race. Hence, even in this case, there are still too many researchers, and the equilibrium is inefficient.

We formalize these ideas by introducing a small change to the benchmark model. Assume that consumers are risk averse and that utility is described by:

$$(34) \quad u = \log c .$$

Let us also assume for simplicity that this is the constant growth case, where the size of innovation team is s^* , for $0 \leq j \leq I^*$, and infinite beyond I^* . We further assume that

$s^* < 1$, namely that an innovator works less than a life-time career on one innovation.¹⁶ For the sake of simplicity we assume, as in the benchmark model, that there is only one research strategy for each innovation.

First we solve the model under perfect competition, when research monopolies are not allowed, or when the cost of forming them is very high. Then, the number of innovators in each patent race n is determined by the following condition, which equates expected income when working part time on an innovation and when working full time in production:¹⁷

$$(34) \quad \frac{1}{n} \log[b(N - R) + 1 - s^*] + \left(1 - \frac{1}{n}\right) \log(1 - s^*) = 0.$$

This condition describes a negative relationship between R and n . The size of the R&D sector satisfies $R = I^* s^* n$, and this is a positive relation between the two variables. Together the two conditions determine the equilibrium levels of research activity.

We next turn to describe a research monopoly, in an economy where they are allowed to form. For the sake of simplicity we overlook the problems of contract enforcement, and view such monopolies as groups of researchers, who divide equally the returns from a successful innovation. Consider a monopolistic coalition of n researchers. In order to make sure they gain all the returns from the innovation they need to deter all potential competitors. A competitor is a research coalition itself, of size k . A competing firm decides on its optimal size k so as to maximize the expected utility:

$$(35) \quad \left(\frac{k}{n+k}\right) \log\left[\frac{b(N-R)}{k} + 1 - s^*\right] + \left(1 - \frac{k}{n+k}\right) \log(1 - s^*).$$

¹⁶ This assumption is added for technical reasons mostly, to overcome a situation that innovators are taking a large enough risk, working all their life on a project, which can fail.

¹⁷ For simplicity we ignore integer problems in this section and treat n as continuous.

The leading firm then chooses n such that potential competing firms decide to stay out of the patent race, namely that the expected utility they can generate is smaller or equal to that of a production worker:

$$(36) \quad \frac{k}{n+k} \log \left[\frac{b(N-R)}{k} + 1 - s^* \right] + \left(1 - \frac{k}{n+k} \right) \log(1 - s^*) = 0 .$$

Under this condition the leading firm remains the single firm in the market and it can give its researchers high certain income. The general equilibrium is reached by combining (36) with $I = s^* I^* n$.

Note that the size of the patent race, which is internalized by a monopolistic firm, is larger than in perfect competition. To see this formally, notice that the expected utility of (35), which is maximized over k , is greater or equal than the expected utility in competition (34), which is limited to $k = 1$. Hence, the number of innovators necessary to deter single competitors is smaller than the number of innovators necessary to deter coalitions.¹⁸ The intuitive reason is clear as well. Once coalitions are formed, innovators can join the race more easily, as they can diversify risk, as (35) shows. Hence, deterring competitors requires a larger number of innovators. Therefore, a research monopoly, which internalizes the patent race within itself, has more innovators than a competitive patent race. In other words, R&D monopolies are even larger and more inefficient than competitive patent races.

We end this section with the following observation. Creation of research monopolies, which internalize the patent race, is a result of risk sharing between innovators. That depends among other things on the divisibility of innovations.

¹⁸ Since (36) exceeds (34) and since both are decreasing functions of n , the value of n that solves (36) is larger than that that solves (34). Hence, its intersection with $I = s^* I^* n$ is at a higher n .

Intuitively, if it takes a whole career to work on an innovation or even more, individuals will not be able to bear the risk of doing it by themselves and research monopolies will emerge. If innovations require only part of a career, like a year or two, individuals can undertake them and competitive patent races emerge. Thus, size of minimum time needed to work on innovation can explain differences across sectors, where in some innovation activity is concentrated, while in others it is fairly competitive.

8. Summary and Conclusions

This paper departs from the initial R&D based endogenous growth models in one assumption. In this paper innovations require not only the labor of innovators, but also enough time to ripen up. That time is necessary both to accumulate more knowledge, and to develop the need for specific innovations as well. The paper then shows that changing this assumption affects the results of the analysis quite significantly. One result is that this model fits better the historical data on global growth rates and on the size of R&D activity. It shows that growth rates do not rise unboundedly with the size of the economy, and with the size of the R&D sector.

But the main deviation of the paper from the initial R&D based endogenous growth literature is that it leads to patent races, either between or within innovating firms. The increasing difficulty of innovation leads innovators to join others, who already search for the easier innovations. Thus patent races emerge. Patent races are wasteful due to duplication of innovative activity. The paper shows that this is true even if there are many different strategies to search for the same innovation. Then researchers tend to crowd the more promising strategies, which offer higher probabilities of success.

We stop short of offering policy recommendations, because any R&D policy, like subsidy, has mixed effects, according to this model. On the one hand a subsidy increases the number of innovations found and that increases the rate of growth. On the other hand, it increases incentives to innovators to join patent races, which already have too many participants. The ideal policy could be to support 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 do it, but it is worth searching for.

References

- Aghion, Philippe and Howitt, Peter. "A Model of Growth Through Creative Destruction," *Econometrica*, Vol. 60 (2), March 1992, p. 323-351.
- . *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.
- 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.
- Kortum, Samuel. "Research, Patenting, and Technological Change," *Econometrica*, Vol. 65 (6), November 1997, p. 450-457.
- Lucas, R.E. Jr. "On the Mechanics of Economic Development," *Journal of Monetary Economics*, Vol. 22, July 1988, p. 3-42.
- Maddison, Angus. *Monitoring the World Economy 1820-1992*. Paris, France: Organization for Economic Co-operation and Development, 1995.
- Romer, P.M. "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, Vol. 94, 1986, p. 1002-1037.
- . "Endogenous Technical Change," *Journal of Political Economy*, Vol. 98 (5, pt.2), October 1990, p. S71-S102.
- Seegerstrom, Paul S. "Endogenous Growth without Scale Effects," *The American Economic Review*, Vol. 88 (2), December 1998, p. 1290-1310.

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

Figures

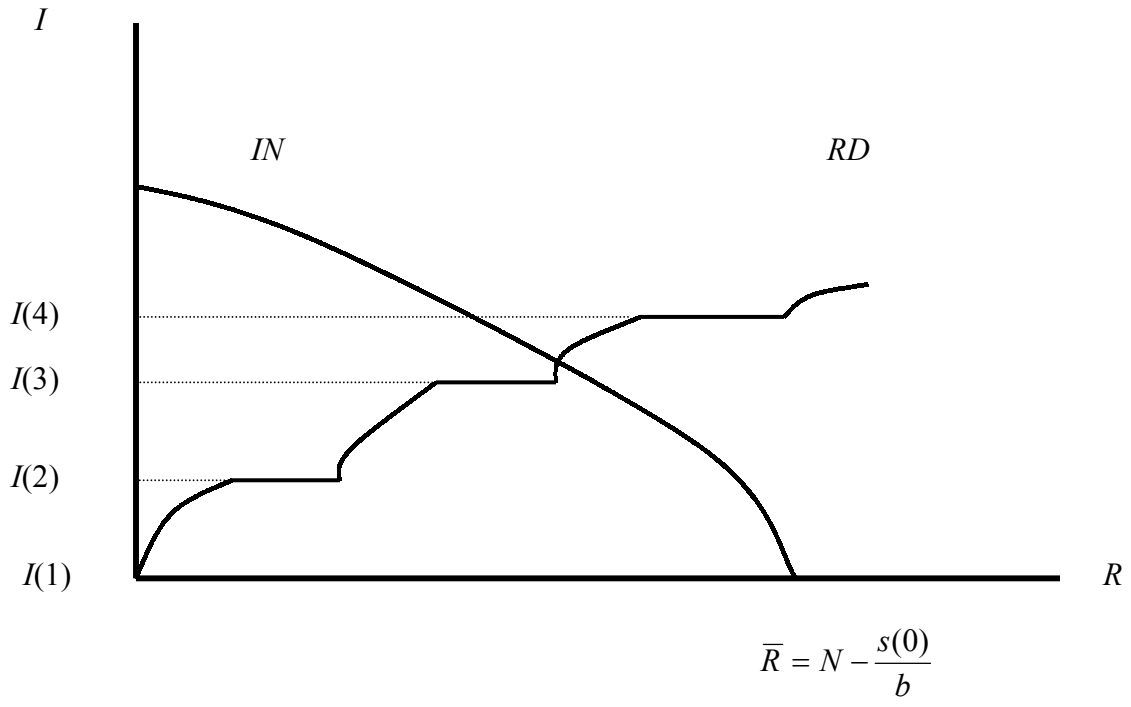


Figure 1

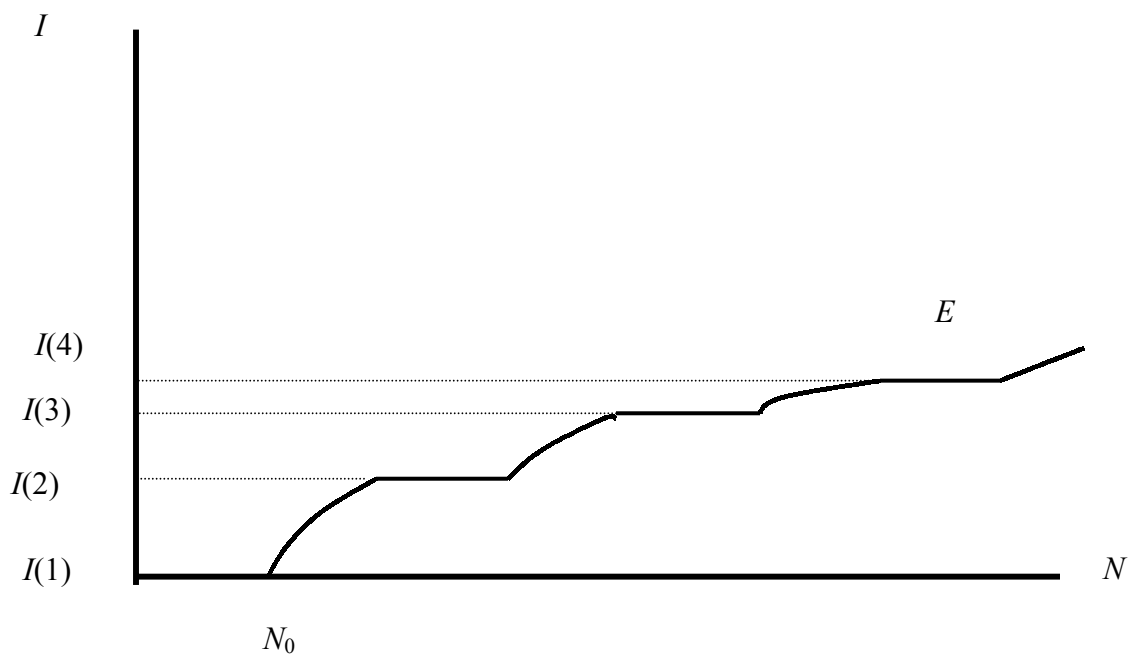


Figure 2

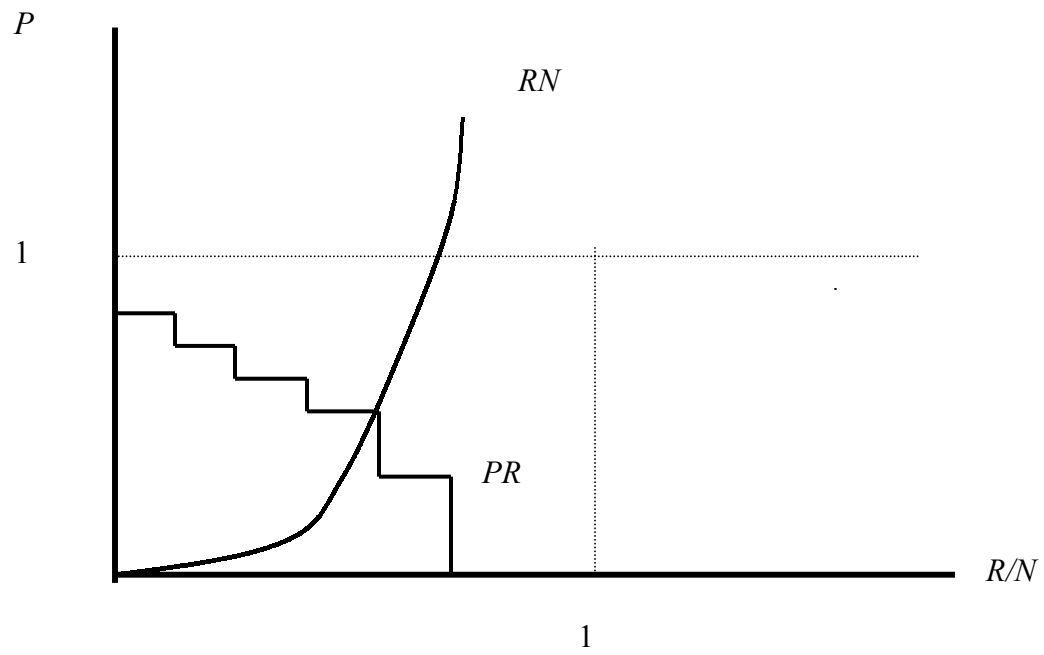


Figure 3