ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 627 March 2007

Random Walk to Innovation : Why Productivity Follows a Power Law

Christian Ghiglino

Random walk to innovation: why productivity follows a power law [1]

Christian Ghiglino

Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ UK [2]

March 2007

Abstract: *In this paper we propose a mechanism generating innovations with productivity distributed according to a power law. We assume that knowledge creation occurs as new ideas are produced from combinations of existing ideas. The productivity of an innovation is determined by an unobservable intrinsic component as well as by the productivity of the parent ideas and their parents, thus generating a network of spillovers. The second important feature is that the innovator has no global information on the network of parenthood links across ideas but has access to local knowledge, as for example the list of cited references in a patent. The optimal behavior of the innovator is to "walk randomly" through the network of "citations" as this algorithm leads to selecting highly connected parent nodes. We show that the distribution of productivity resulting from this optimal behaviour follows a power law. The intuition behind the result is that the innovator focuses his efforts on strengthening local spilovers because he has no command on the other sources of productivity. When this process of innovation is embedded in a model a la Kortum (1997) balanced growth of output is generated.*

Keywords: *Economic growth, Technological progress, Innovations, Random growing networks, Ideas, Scale-free distributions.*

Journal of Economic Literature Classification Numbers: O31, O41.

---

[1] A network model of growth circulated with the title "Balanced growth with a network of ideas". However, the present model has few common points with that earlier paper besides the fact that ideas are assumed to be obtained as combinations of existing ideas. We thank A. d'Autumne, R. Becker, M. Boldrin, B. Decreuse, S. Goyal, T. Hens, C. Jones, C. Le Van, T. Mitra, P. Peretto, K. Shell, J. Stachurski and J. Zweimuller for useful comments and suggestions.

[2] E-mail: cghig@essex.ac.uk.

1 Introduction

The literature offers several models rationalizing the stylized facts of growth and innovation. However, these models of technological progress are in fact models of research effort but not of innovation. Indeed, they assume that the innovator only faces the decision of whether to innovate or to work in production but has no command on the quality of these innovations. For example, in Kortum (1997) ideas arrive at the desk of an innovator as a Poisson process and have a productivity that follow an exogenously given distribution. Kortum finds that in order to replicate the observed behavior of investment in research and productivity growth, the productivity of ideas should follow a Pareto distribution. There is some empirical evidence that the productivity of innovations follow a power law (see for example the discussion in Jones (2005)). However, we are not aware of any model generating this feature.

The goal of the present paper is then to provide a model of innovation in which the productivity of the recipes is endogenous. In fact the model follows very closely Kortum (1997). Formally, the difference is that the behavior of the innovator is further modelled and Assumption 3.2 in Kortum (1997), i.e. Pareto distributed productivity, is the result of the optimizing behavior of the innovator instead of being an assumption. The main two ingredients driving the results of the present paper is the combination of imperfect information on the productivity of ideas together with the existence of local spillover across related ideas.

In the model we assume that innovations are obtained through combinations of existing ideas. As in the case of patents and scientific papers we assume that the innovator has access to the cited references, i.e. the parents. The productivity of an innovation has three elements. Beside an unobservable intrinsic productivity random component, an idea benefits from spilovers from other ideas. As in Kortum (1997) we allow for economy-wide spillover upon which the innovator has no command. However, there is a third element upon which the innovator has some command because it is related to how the new idea is obtained. Indeed, we assume the existence of local spillovers from the parent ideas. The rationale is that "knowledge" accumulated by the parent ideas contribute to the productivity of the offspring. Furthermore, the contribution of each parent to the offspring's productivity depends on the number of projects or recipes in which the parent idea is itself involved. Clearly, in most cases the weight of each parent in the combination is not known ex-ante. It is modeled as a random variable.

As innovators are profit maximizers the objective of the innovator is to innovate in areas in which these spilovers are the greatest. However, we also assume that the innovator has no information on the global topology of the network of parenthood links (as citations). Then, due to the lack of global information, it is rational for the innovator to "walk on the network" and explore areas of greater spilovers only by "local" decisions (see Section 2.5 to get a better idea of these notions). The present paper shows that this mechanism leads the distribution of productivity to be of the Pareto type, at least asymptotically. Note that the outlined mechanism implies that there is an endogenous tendency to generate "patent classes". Loosely speaking, this is in line with the empirical evidence provided by Adams and Jaffe (1996). However, our notion of a "patent class" is fairly different.

The model also provides a way to analyze the effects of a change in the available information, due for example to a change in the policy regarding patents. One of the surprising consequences is that allowing for better global information on the links between ideas induces the innovators to excessively reward the local spillover component, an attitude leading to the collapse of the network growth. On the other hand, making the content of innovations more transparent and providing ways to reveal their productivity are factors that improve growth.

The paper has the following structure. In Section 2 the model is described while in Section 3 the property of the distribution of productivity of innovations is obtained. In Section 4 some policy implications are considered. In Section 5 the links with the literature are explained while Section 6 concludes. Finally, an Appendix contains the most important proofs.

2 The model

The general structure of the model follows very closely Kortum (1997). However, the behavior of the innovator is modelled and Assumption 3.2 in Kortum (1997) is the result of an optimization instead of being an assumption.

At date t there is a continuum of individuals, $h \in [0, L(t)]$ with $L(t)$ a non-decreasing function with $\int_{-\infty}^{t} L(t) < \infty$ for $t < \infty$. There is a continuum of consumption goods, $j \in [0,1]$. The utility of the individual from consumption is of the Dixit-Stiglitz type. Agents can chose to be an innovator or a worker in a firm in which case they receive a wage $W(t)$. The effort produced by an innovator is inelastic, fixed and constant. The total research effort in t is the labor force engaged in research $R(t)$, i.e. at each date there is a continuum of innovators h with $h \in [0, R(t)]$. The stock of past research effort is noted $K(t) = \int_{-\infty}^{t} R(s)ds$. In the model $R(t)$ is endogenous and depends on the expected profits from innovation. As in Kortum (1997) we assume that innovators come up with techniques pertaining to the production of a good of variety j, with labor as the sole input, and that innovators are not able to chose the variety j. Since all varieties are similar, we can focus on an arbitrary variety j.

At equilibrium, the return of being an innovator and that of a worker are the same. In other words the expected returns to innovation is equal to the wage rate $W(t)$. In order to be able to decide being an innovator or a worker, the agent needs to know the expected returns to innovation. This depends on the productivity of the innovation, on the probability that it will be patentable (i.e. it will be the most productive recipe) and on the time it will stay undominated. When a better technique is discovered the patent expires. Once the research effort is endogenously determined, and the distribution of productivity specified then the equilibrium growth rates can be determined. The main contribution of the paper is to endogenously determine the distribution of productivity of innovations.

2.1 The quality of ideas

Knowledge is characterized by the stock of ideas as well as their quality. The number of ideas produced depends on the amount of resources devoted to innovation. This issue is treated in Kortum (1997). Here we focus on the quality of ideas.

An idea designates a broad class of objects. It can be a sequence of operations used to produce a physical good. It may also be a sequence of assumptions, lemmas and techniques as in a scientific paper. We assume that in all cases an idea can be considered as a combination of existing ideas. Let i_t be an idea produced at date t. We note $i_t(i_t^1, i_t^2, ..., i_t^m)$ where i_t^k is an idea existing at time t used to produce idea i_t. We assume that the combinations include m parent ideas.

In order to gain intuition on the determinants of productivity it is useful to consider two examples. First, consider a recipe to produce chocolate cake. We may assume that the sequence of operations is: 1) grow chocolate plants, 2) mash the seeds with stones and hands, 3) feed the cows, 4) milk the cows, 5) put the ingredients in a casserole, 6) mix, and 7) cook with a wood fire. The present paper is an illustration of the second kind of idea. Indeed, it may be seen as a combination of: 1) the model of Weitzman on recombinant growth, 2) the Kortum's model, 3) the Pareto distribution of degrees in random growing networks, 4) the analysis of Adam and Jaffe (1996) on "patent classes", 5) the attachment algorithm in Vasquez (2003), and 6) the notion of stable distributions.

Within the present framework, the two main issues are: 1) what is the productivity of an idea? and 2) how the existing ideas are selected? The first issue is analyzed in this section while the second needs a model of the innovator. We will consider this in Section 2.3 and 2.4.

To evaluate the productivity of a recipe it is useful to consider how the productivity of a given recipe could be improved. An obvious way is to replace one operation by a more effective one. For example, to perform "the mashing the chocolate seeds" with an electric device. In our framework, this substitution is associated to the production of a new idea. Another way the productivity may increase is when one of the steps is improved without replacing any of the parent ideas. This mainly happens when that particular operation is also used in many other recipes. For example, when "mashing with stones" is used in many other situations then this operation is likely to become more effective. This type of spilovers across ideas could be called "learning-by-adoption".

We formalize the economic situation described above as follows. First, we assume that the productivity of an idea i, noted a_i, may be written as

$$a_i = \widehat{H_i}(A_i, s_{i1}, ..., s_{iN(t)})$$

where A_i is the intrinsic productivity of idea i, s_{ij} represents the contribution of idea j and $\widehat{H_i}$ is a continuous function. In this formulation all $N(t)$ available ideas at time t contribute to the productivity. Clearly the role of the existing ideas on the productivity of idea i depends on how related these are to idea i, i.e. their "distance" from idea i. At the time an idea is produced it is only linked to its parents. Consequently, we focus on the role of the immediate parents but we include the possibility of a diffuse beneficial effect of the stock of ideas. Let $M(i)$ be the set of the m parents to idea i. The contribution of the parents is assumed to be a weighted sum of their productivity, i.e. $s_{ij} = \theta_{ij} a_j$ where θ_{ij} represents the importance of parent $j \in M(i)$ in i's productivity. The $\theta_{i1}, ..., \theta_{iM(i)}$ are assumed to be non-negative, possibly random, weights adding to one.

The strength of the effect of the productivity of the parents on the productivity of idea i may also depend on the remaining existing $N(t) - m$ ideas included in H_i. We identify this effect to the diffuse economy-wide spillover. The spillover arising from all $N(t)$ ideas is noted $S(N(t))$. Then we obtain

$$a_i = S(N(t_i)) H_i(A_i, s_{i1}, ..., s_{iM(i)})$$

where H_i is a continuous function and t_i is the date of creation of idea i. For the remaining of this section we will ignore the economy-wide spilovers but they will be reintroduced in Section 3.

We further simplify the model by assuming that H_i admits an anonymous, affine and stationary representation. The overall productivity of an idea i is assumed to be

$$a_i = A_i + \sum_{j \in M(i)} s_{ij} = A_i + b \sum_{j \in M(i)} \theta_{ij} \ a_j$$

where b is a constant positive factor. The value of the a_j could be obtained with a similar expression. Note that as the parents to idea i are older then idea i these parents benefit from their own offsprings, i.e. the more recent innovations in which they are involved. Let the neighbors of node j be denoted $k \in M(j)$. To simplify, we consider uniform weights, and take expectations of A_j and a_k, denoted $\overline{A}$ and $\overline{a}$. Then

$$a_j = \overline{A} + b \sum_{k \in M(j)} \overline{a} = \overline{A} + b\overline{a}(m + k_j^{in})$$

We finally obtain

$$a_i = A_i + B_1 + B_2 \sum_{j \in M(i)} \theta_{ij} k_j^{in}.$$

where B_1 and B_2 are constants independent of i. In the present framework, the productivity of an idea depends on the number of offsprings to the parents of i. Consequently, the "potential" productivity of an idea increases with time (as the in-degree increases). However, we assume that the idea is implemented at the time it is discovered. Finally, we assume that when an idea is produced and implemented its total productivity is revealed. This information is not updated.

2.2 Knowledge structure and the costs of innovation

We assume that the innovator cannot keep track of the global evolution of the entire network and in particular that he has no knowledge of the current value of the connectivity of the existing ideas. Indeed, assume that at time t an idea $i(i_t^1, i_t^2, ..., i_t^{m-1}, i_t^m)$ is produced. Clearly, its composition and productivity are revealed and become public. However, as the network evolves the nature and number of links k_i associated to idea i change. So, the initial information is not anymore accurate at any later date. However, assuming that the innovator has no global knowledge of the network does not rule out some sort of local knowledge. We assume that once the innovator understands an idea he is able to randomly select one of the edges linking this idea to its parents. For example in the case of patents, this means that the innovator has access to the list of cited references. In this way the innovator can "walk on the network" and explore the network with only local information. We will describe this algorithm in Section 2.4.

Assuming that the composition and productivity of ideas is public does not imply that understanding and using ideas is costless. The use of ideas requires some sort of previous knowledge because it is necessary to understand the composing ideas. Therefore, when an idea is considered for a combination some resources are needed. The cost associated to this exercise is precisely the time needed to acquire the necessary specific knowledge.

The previous remarks imply that the cost of producing an idea depends on how "innovative" it is, in particular how many new ideas the innovator includes in the combination. The innovator can undertake several actions, each characterized by their own cost. We consider two polar sequences of actions.

In Scenario 1 we assume that the innovation consists of replacing idea i^m in a known existing combination produced at time $s < t$, denoted idea i_s. Formally, the innovation takes the form $i_t(i_t^1, i_t^2, ..., i_t^{m-1}, i_t^m) = i_t(i_s^1, i_s^2, ..., i_s^{m-1}, i_t^m)$. The expected costs can be decomposed into: 1) the time needed to understand the initial idea i_s in view to use it as a seed to the innovation (cost c_I), 2) the time needed to understand the new idea i_t^m (cost c_N) and, 3) the cost of changing one idea (or operation) among the m ideas (cost c_R). Finally, we assume the existence of a production cost $\bar{c}$. The total expected cost is then $\bar{c} + c_I + c_R + c_N$. In the chocolate cake example, when the hands are replaced by an electric device the cost of the new recipe includes all four costs c_I, c_R, c_N and $\bar{c}$.

In Scenario 2 the innovator selects m new ideas. This possibility is more innovative. Which of the two scenari is cheaper depends on whether the cost of understanding the "seed" plus the cost of a replacement is smaller or larger than the cost to understand $m-1$ of its m parent ideas, i.e. $c_I < (m-1)c_N - c_R$ or $c_I \geq (m-1)c_N - c_R$. We will consider both cases.

We also assume that many of the combinations are not feasible (fertile) so that the actual cost is larger than $\bar{c} + c_I + c_R + c_N$. We assume that $c_N, c_R << 1$ and that the probability of a feasible combination p is small, $0 < p << 1$. Furthermore, we assume that only once the combination is feasible the idea is produced at the cost $\bar{c}$.

.

Remark: Any existing idea $i_s(i_s^1, i_s^2, ..., i_s^{m-1}, i_s^m)$ used to produce an innovation via the substitution of its m^{th} idea, is in fact an innovation produced in some previous period s. As this can be far apart from period t the knowledge used to understand idea i_s may suffer some sort of depreciation. However, as a first approximation we neglect this type of aging of knowledge. Note that this type of time evolution is different in nature than the increase in productivity of an elementary step due to learning-by-adoption.

2.3 Innovators

In period t there is a continuum of innovators indexed by $h \in [0, R(t)]$, where $R(t)$ is endogenously determined. Note that as the stock of discovered ideas is known to everybody it is irrelevant to keep track of who discovered what so that the age of the innovator is irrelevant. We assume that the innovator has access to the credit market, so he is only concerned by the expected profit of innovation. We also assume that the innovator does not choose the sector j.

As the innovator maximizes the expected profits of his activity we need to find the expected duration of the monopoly and the probability $\pi(t)$ that the idea is patentable. This can be computed as in Kortum (1997). Then the expected return of an innovation is $\pi(t)V(t)$ where $V(t)$ is the the expected value of a patent discovered at date t. The behavior of the innovator is

$$Max E_t[\pi(t)V(t) - c(t)]$$

where $c(t)$ is the cost.

The cost $c(t)$ depends on the way the idea is obtained. As seen in the previous section, we mainly focus on two polar cases. When $c_I < (m-1)c_N - c_R$, Scenario 1 minimizes the costs. In this case the innovator start by selecting at random one of the existing ideas to produce good j. This idea is used a "seed" and it is "muted" by selecting one of the m parent ideas and replacing it with some other existing idea. More precisely, the $m-1$ elements of an existing idea $i(i^1, i^2, ..., i^{m-1}, i^m)$ are combined together with an existing idea i_t^m so that the new idea is now $i_t(i_t^1, i_t^2, ..., i_t^{m-1}, i_t^m)$ with $i_t^1 = i^1, i_t^2 = i^2, ..., i_t^{m-1} = i^{m-1}$. Note that the indices $1, ..., m$ are randomly numbered so that choosing the idea m is without loss of generality. Note also that to select a popular idea as the initial "seed" to the innovation would not increase the expected productivity of the final innovation. Finally, if $c_I < (m-1)c_N - c_R$ then the costs are minimized with Scenario 2. In this case, all m parents are chosen independently. We will see in the next section how these ideas are chosen.

Because not all combinations are fertile the cost of innovation is likely to be larger than the one associated to the two Scenaria described above. Indeed, when a combinations is unfeasible or unproductive the innovator is forced to proceed to further replacements. Independently of the chosen Scenario these further successive replacements affect the same parent. Indeed, provided, all the costs $\bar{c}, c_I, c_R$ and c_N are non zero, replacing one idea at a time is a way to come up with an innovation at the minimal cost. Furthermore, there is no benefit of making simultaneous substitutions as the quality of the innovation is ex-ante independent of how many changes are made to the original innovation. The single replacement appears to be the optimal strategy. However, all the results would hold if the innovator would randomly select which of the parent to substitute at each step (provided $pm << 1$). Finally, note that because replacing ideas is not costless a single sequential replacement at a time maximizes the number of feasible ideas for a given cost.

What is the average cost of a feasible invention? We may normalize this average cost of creation to unity. As the probability of a combination to be feasible is small, $p << 1$, a large number of unfeasible combinations is tried before a feasible one is obtained. Consequently, the process can be approximated by a Poisson process of parameter λ = average number of feasible combinations per unit interval when walking on the network = 1.

Consequently, even though the model is in continuous time, on average one node per unit time interval is added to the network of ideas for producing good of variety j. In other words, ideas arrive to each individual researcher as a Poisson process normalized to unity. Finally, this implies that the number of techniques for producing good j discovered between t and s is a Poisson distribution with parameter $K(s) - K(t)$ where $K(t)$ is the stock of past research evaluated at time t.

2.4 The innovator's optimal behavior

There is still the question of how to come up with the best possible innovation. Assume that $c_I < (m-1)c_N - c_R$ (the other case can be treated similarly). In this case, starting from a known existing idea and proceeding to one substitution at a time, minimizes the costs. However, the quality of the innovation depends on how the m^{th} idea is chosen. Given the ignorance on the present productivity of ideas, the best action is the one that leads to the node (idea) with the highest in-degree. However, because of the lack of global visibility on the network links it is not possible to pick this node directly. The innovator needs to proceed in an indirect way.

A possibility is to select a node at random. However, the agent can do better. Indeed, we assume that once the innovator understands a node he has the possibility to follow one of the incoming edges connecting the node with the rest of the network. We assume that this occurs with a probability, or efficiency, $0 \leq q_c \leq 1$. For example, in the case of patents the innovator can follow the links as indicated by the list of cited references. Given the available information, the optimal choice is to randomly select one of these edges, at least when possible. Finally, as not all combinations are feasible, the innovator continues his walk on the network until he reaches a compatible combination. Intuitively, this simple algorithm favours nodes with higher in-degree because there are more paths leading to them.

The walking algorithm: The rule to find a m^{th} idea is as follows:

- Step 1: The innovator picks a node by a random draw.
- Step 2: With probability q_c , $0 \leq q_c \leq 1$, the innovator follows one of the edges incoming into the node under consideration. With probability $1 - q_c$ the innovator jumps (for example by mistake) to a random node and the search stops.

- Step 3: With probability p, $0 < p < 1$, the node attached at the other end of the edge is selected and the search stops. With probability $1-p$ the innovator go to Step 2.

Similar rules have been considered by several authors (e.g. Vazquez (2003), Saramaki and Kaski (2004), Evans and Saramaki (2005) and Jackson and Rogers (2006)). When there are no mistakes, i.e. $q_c = 1$, it has been shown that in the limit case where the mean-field approach is valid the local algorithm implements the linear preferential attachment rule: the i_t^m idea is chosen as if a node j of degree k_j would be selected according to a probability Π_j of the form

$$\Pi_j = \frac{k_j}{\sum_{j'} k_{j'}}$$

where $j' = 1, ..., N(t)$ spans over the entire network. An important result of random growth literature is that linear preferential attachment leads, at least asymptotically, to a scale free distribution in the nodes' degrees. Note that the distinction between in-degree and total degree is irrelevant at this stage.

In the present framework the algorithm also induces a cumulative distribution for the incoming degrees which is asymptotically scale free.

Lemma 1 *Given that the nodes' degrees are not correlated, the individually optimal process of innovation is such that*

a) ***Choice of the "seed" idea.*** *If $c_I < (m-1)c_N - c_R$ then the innovator selects one existing idea at random and keeps $m-1$ of its parents, each selection occurring with a probability q_c. When a parent is not selected, an existing random idea is selected with probability $1-q_c$. If $c_I \geq (m-1)c_N - c_R$ then the initial "seed" idea is obtained choosing $m-1$ ideas according to Step 1 and Step 2 of the walking algorithm modified such that $p = 1$.*

b) ***Choice of the m^{th} idea.*** *The m^{th} idea is chosen according to the Steps 1, 2 and 3 of the walking algorithm described above.*

c) ***Degree distribution.*** *Assume that the number of innovators grow at a constant rate g. Assume also that the innovator behaves optimally as in a) and b). Then, within the validity of the mean-field approximation,* *the resulting network of ideas is such that the in-degrees of the nodes follow a translated power law*

$$P[k_i^{in} \leq k] = 1 - (\frac{c_1}{k + c_1})^{\xi}$$

where c_1 *a strictly positive parameter. If* $c_I < (m-1)c_N - c_R$ *then* $\xi = \log(1+g)[g[(q_c\frac{m-1}{m} + pf(q_c)]]^{-1}$ *while otherwise* $\xi = \log(1+g)[g[(q_c + pf(q_c)]]^{-1}$.

Proof: See Appendix.

We would like to make some remarks.

i) In both Scenario 1 and Scenario 2 the distribution of degrees is a displaced power law. Many other scenarios would also generate power laws as can be seen from the proof of Lemma 1. The robust appears quite robust to changes in the cost structure and in the search algorithm.

ii) The exponent of the Pareto distribution obtained in the Lemma depends the probability of mistake (q_c), the fertility (p) and the growth rate of the research effort (g). However, even in a deterministic case the power law of productivity would appear.

iii) The applicability of Lemma 1 is not restricted by the assumption of absence of correlation between the degrees. Strictly speaking correlations do exist even in the standard Barabasi and Albert model for finite size but disappears as the network grows. A similar pattern is observed by Vazquez (2003) and Evans and Saramaki (2005) for networks generated by "walks on the net" with and without errors. The absence of correlation between the nodes' degrees is then a good approximation for the type of networks considered here.

iv) In the "walking algorithm" as soon as the walk on the network is interrupted and the innovator jumps to a random node the search stops. We could instead assume that the search continues after the jump. The network would evolve as in the case $q_c = 1$.

As a general remark, note that the search algorithm is the optimal rule given the cost structure and rewards. However, we could also take linear preferential attachment as a primitive, i.e. a property of the innovator's behavior. Indeed, linear preferential attachment can be given a direct intuition not based on local knowledge: in the presence of global but noisy knowledge one may assume that ideas that have been often used in the combinations are likely to be more productive, everything else being equal. Consequently, popularity or visibility, is a key property the uniformed innovator should look for.

3 The law for productivity

The productivity of an idea is a function of the intrinsic productivity of the idea and of the weighted sum of the in-degrees of the neighbors, all random variables. As only the asymptotic behavior of the a_i will matter in the analysis (see below), either the in-degree, k_j^{in}, or the total degree, $k_j = m + k_j^{in}$, may be considered. Without lack of generality, we adopt the in-degree as it characterizes the number of new recipes using or referring to idea j. Ignoring the diffuse spillover, productivity is formally defined as

$$a_i = A_i + B_1 + B_2 \sum_{j \in M(i)} \theta_{ij} k_j^{in}$$

Note that the idea is implemented at the time it is produced, so any idea i has exactly m neighbors all of them being "parents". The results in Lemma 1 indicates that the random network grows in a way such that the cumulative distribution of degrees is a translated power function with a shift c_1 and an exponent ξ. Furthermore, the way the parent ideas $j \in M(i)$ are chosen implies the absence of correlations across them. Consequently, the non-intrinsic part of the a_i is a sum of independent random variables distributed according to power laws with the same exponent. Indeed, the translation parameter c_1 may simply be integrated in the B_1 constant. It is also important to note that the sum involves a finite but possibly large number of random variables.

The analysis in Kortum (1997) shows that to obtain results on the long run behavior of the economy it is not necessary to know the exact distribution of the a_i's. Indeed, the observed trend in productivity and in research are obtained in Kortum's model provided the search distribution $F(a) = P[a_i \leq a]$ is in the basin of attraction of the Frechet extreme value distribution. This is the case if and only if the distribution has unbounded upper support and is asymptotically a power law. The next Lemma is therefore relevant.

Lemma 2 *Assume that the conditions ensuring the validity of Lemma 1 hold and that* $\xi \in (0, 2)$. *Let* m *be the number of parental ideas used in innovation. In the case with deterministic weights* θ, *the productivity of ideas follows a distribution* $F^m(a) = P[a_i \leq a]$ *with unbounded upper support and such that for all* c

$$\lim_{m\to\infty} \lim_{a\to\infty} \frac{P[a_i > a]}{P[a_i > ca]} = \lim_{m\to\infty} \lim_{a\to\infty} \frac{1 - F^m(a)}{1 - F^m(ca)} = c^{\xi}$$

When the weights are random the same result holds provided there exists $\delta > 0$ *such that* $E\theta_{ik}^{\xi+\delta} < \infty$ *for each* $1 \leq k \leq m$.

Proof: See the Appendix. Q.E.D.

The limit on m is due to the fact that the productivity of an innovation is related to the productivity of its parents and these follow translated power laws and that for large m the sum of power laws can be approximated by a power law. From Lemma 2 we get the following Theorem.

Theorem 1 *Subject to the validity of the conditions stated in Lemma 1, the productivity of innovations follows asymptotically in* a *and* m *a Pareto distribution*

$$P[a_i > a] \simeq a^{-\xi}$$

where $\xi = \log(1 + g)[g[(q_c + pf(q_c)]]^{-1}$.

The result in Theorem 1 is very robust as it is driven by the fact that the innovator has only local knowledge and that the expected productivity increases with the in-degree of the parents. Indeed, this provides the innovator with incentives to replicate linear preferential attachment, which is known to be responsible for the power law for the degrees. It should be noted that the fact that the tail is a power law does not depend on the exact form of the beliefs of the innovator, as long as he believes in the existence of positive externalities.

As in Kortum (1997), an economy-wide externality associated with the total stock of ideas, can be reintroduced. This produces a second channel through which the productivity can grow. His model correctly predicts the data provided the externality has the form $S(N) = N^{\gamma}$ where N is the stock of ideas and $0 \leq \gamma < \infty$. Note that the existence of this externality which increases the probability that a given level of efficiency is obtained is not required for growth but improves the fit of the model with the observed trends in the data. In the present model the global externality affects the productivity of an idea i in the following way

$$a_i = N^{\gamma\xi}[A_i + B_1 + B_2 \sum_{j \in M(i)} \theta_{ij} k_j^{in}]$$

where γ is a parameter describing the economy-wide spillover. The results in Lemma 2 are easily adapted to include this term. We then obtain the following result.

Corollary 1 *Subject to the validity of the condition in Theorem 1, Proposition 3.3 and Proposition 3.4 in Kortum (1997) hold without the need of Assumption 3.2. In particular, the long run productivity growth rate takes the form* $(1+\gamma)n/\xi$ *where* n *is the population growth rate and* γ *characterizes the global spillover associated to the stock of research effort. The value of* ξ *linearly depends on* $\log(1+g)/g$ *where* g *is the equilibrium growth in research effort.*

The Corollary shows that long run productivity growth is $(1+\gamma)n/\xi$. For large m, in the present model $\xi = \log(1+g)[g[(q_c+pf(q_c)]]^{-1}$. Consequently, the value of the growth rate of output depends on the growth of research effort which is endogenously generated (this can easily be computed using Proposition 3.4 in Kortum (1997)). The probabilities p and q_c affect the value of the parameter ξ so that, according to the expression given above, can then take any value between 0 and ∞. As in Kortum (1997) the arbitrage between ξ and γ allows to approach the observed level of research intensity and the renewal data compatible with observed aggregate hazard rate.

Finally, the additive specification for the productivity of ideas introduced in Section 2 may seem inappropriate because in the limit the role of the intrinsic productivity of the idea vanishes. We Here we consider instead the case in which the productivity has the form

$$a_i = N(t)^{\gamma\xi} H_i(A_i, a_{i1}, ..., a_{iN_i}) = N(t)^{\gamma\xi}[\delta_i + \overline{\delta}][B + b \sum_{j \in M(i)} \theta_j k_j^{in}]$$

The nice feature of this formulation is that the intrinsic productivity of the idea remains relevant even in the limit of large degrees. The result of Theorem 1 and Corollary 1 remain true.

Corollary 2 *Let* $a_i = N(t)^{\gamma\xi}[\delta_i + \overline{\delta}][B + b\sum_{j\in m(i)} \theta_j k_j^{in}]$. *Subject to the validity of the condition in Theorem 1, Proposition 3.3 and Proposition 3.4 in Kortum (1997) hold without the need of Assumption 3.2. In particular, the long run productivity growth rate takes the form* $(1+\gamma)n/\xi$ *where* n *is the population growth rate and* γ *characterizes the global spillover associated to the stock of research effort. The value of* ξ *linearly depends on* $\log(1+g)/g$ *where* g *is the equilibrium growth in research effort.*

Proof: See Appendix. Q.E.D.

4 Policy implications

The model can be used to analyze the effects of possible government policies designed to increase growth. In the next few lines we focus on three options.

A first possible policy would be to improve the public visibility of the network of innovations. For example, ranking of innovations reflecting their popularity and the relevance of the sources citing them could be produced, as is the case for scientific papers. Clearly, whenever the innovator has some, possibly noisy, global knowledge about the degree of the nodes, this information would used by the innovator to favor high degree nodes. The induced choice rule rewards more than linearly the degree. As a first approximation we may consider that this non-linear attachment rule produces a pattern of growth similar to the one associated to the "superlinear preferential attachment" rule $\Pi_i \sim k_i^\eta$, with $\eta > 1$. It is known that the behavior of the network depends on how far η is from unity. For large values, a winner-takes-all pattern arise in which almost all nodes have m links and a gel node has all the remaining links, i.e. the network evolves toward a gelation pattern. An example of such an idea could be the "wheel". With superlinear preferential attachment, the idea of the wheel would be used again and again in almost all combinations. Consequently, its degree would increase rapidly while the remaining ideas would have a small degree. The productivity of new ideas will suffer from the fact that the $m-1$ other ideas have to be chosen from the pool of ideas with low degree. It is likely that the economy would eventually stop to grow. Clearly, this version of the model should be considered with "granum salis". In particular, it is likely that before reaching the extreme gelation pattern the marginal positive effect of local spilovers would vanish.

In the model, knowledge is public. This hypothesis, which implies an extreme intertemporal spillover, is standard in the literature, e.g. Aghion and Howitt (1992). However, it can be argued that each innovator knows better the ideas he discovered himself. In this situation, the cost of replacing an idea in i_t to produce i_{t+1}would be lower when the innovator discovered himself idea i_t. This would induce the innovator to keep trying innovating in areas close to his own prior ideas. As for each given innovator this set is small, eventually the innovators will run out of new combinations. As new combinations become more expensive growth would slow down. Consequently, policies aiming at making public the information concerning the content of patents are expected to be useful for long run growth.

A policy which is expected to be beneficial in all situations is one intended to facilitate the testing of ideas, i.e. reveal their productivity. Implementing and testing ideas typically requires large quantities of both physical and human capital. This fact has been interpreted by the literature as an indication that resources should be devoted in priority to implementation rather than to pure research. However, growth could be indirectly stimulated by investments in testing techniques. In this perspective, one of the benefits of the "IT revolution" would be to ease the testing of ideas through "cheap" simulation. In other words, growth would benefits from any policy sustaining investments in this type of "IT".

5 Related literature

A few papers try to propose a microeconomic model of technological progress based on combinations of ideas. Similarly to the present paper, Weitzman (1998) assumes that technological progress occurs through (re)combinations of existing ideas. However, in his model progress is only associated to the number of ideas, i.e. productivity depends on the stock of innovations. Note that this assumption is very common in the "endogenous growth literature" and is similar to our way to model economy-wide spilovers but not the local spilover effect. Another feature of the model proposed by Weitzman is that ideas combine automatically, without any selection. In his model, what shapes the dynamics of output is the resource requirements associated to the production of viable innovations and not the distribution of their productivity.

In our framework, the selection and quality of existing ideas play a major role. In this respect the present paper is in the spirit of models including a technology space. Several papers model progress as a selective process via the introduction of a distance between firms or recipes. In particular, Auerswald, Kauffman, Lobo and Shell (2000) assume that there is an existing distribution of uncovered recipes and that trials reveals its topology. The distance between two recipes matters in the sense that the closest ones have more chances of being uncovered. Similarly to the present paper, in Auerswald et al. externalities arise between close recipes. The model describes well learning curves but as the technology landscape is exogenously given and finite the model is not helpfull to explain sustained long-run growth.

In Olsson (2000) and Olsson (2005) recipes are points of the positive orthant of a Euclidean space. The available knowledge is then a subset of the positive orthant. The boundary of this set defines a technological frontier. Incremental progress, or normal science, consists in convexifying the set of available knowledge. The idea is appealing but the nature of the technological space is ad-hoc. On the other hand, Peretto and Smulders (2002) assume that the size of the market increases the distances between firms because these become more specialized. As in our model, spilovers exist but these are reduced when the general increase in the distances occur. The model is able to eliminate the disturbing scale effect but not to fully endogeneise the distribution of productivity.

Finally, related to our paper is also the standard quality ladder literature. Indeed, at least since Schumpeter, the literature has considered models in which progress is associated to sequential innovations in a partially ordered network of commodities. Kelly (2001) proposes such a model of linkage formation. As in the other models described above, there is no natural notion of distance across innovations. For example, Kelly (2001) simply assumes that innovation in Sector j is affected by research in Sector j-1 and j+1. The ladder is exogenous.

6 Conclusion

The paper proposes a model in which technological progress is endogenous and the outcome is compatible with the available data on output growth, research effort and research productivity. The basic ingredients are that: 1) new ideas are combinations of existing ideas, 2) there exists positive spillovers across related ideas, and 3) the innovator has no global knowledge of the network of ideas but is able to follow, at some cost, the parenthood links once he understands an idea. The obtained productivity pattern is driven by the fact that the innovator focuses his efforts on strenghtening local spilovers through appropriate combinations because he has no command on the other factors affecting the productivity of the ideas, and therefore their expected returns.

The model can be used to analyze the effects of possible government policies designed to increase growth. The model surprisingly predicts that a policy that improves the public visibility of the network of innovations, as ranking of innovations, would not be beneficial. On the other hand, policies that reduce the cost of testing new ideas, as stimulation of the IT sector devoted to simulation, are expected to be very beneficial for growth. Finally, the assumption of public information is crucial. Any policy reliving imperfect information on the content of innovations would be beneficial.

7 References

1. Adams, J., and A. Jaffe, Bounding the Effects of R & D: An investigation using matched establishement-firm data. RAND Journal of Economics (1996) 27(4), 700-721,

2. Aghion, P., and P. Howitt, A model of growth through creative destruction, Econometrica (1992) 60, 323-351.

3. Auerswald, P., S. Kauffman, J. Lobo and K. Shell, The production recipe approach: An application to learning by doing, Journal of Dynamics and Control, (2000), 24(3), 389-450.

4. Breinman, L., On some limit theorems similar to the arc-sin law. Theory of Probability and its Applications (1965), 10(2), 323-331.

5. Davis, R. A. and Resnick, S.I. Limit theory for bilinear processes with heavy-tailed noise. The Annals of Applied Probability (1996), 6(4), 1191-1210.

6. Evans, T. S. and Saramaki, J.P., Scale Free Networks from Self-Organisation, Phys. Rev. E 72 (2005) 26138, http://arxiv.org/abs/cond-mat/0411390.

7. Goovaerts, M., R. Kass, R. Laeven, Q. Tang, and R. Verrnic, The tail probability of discounted sums of Pareto-like losses in insurance, Scandinavian Actuarial Journal, 6 (2005), 446-461.

8. Jackson, M., and B. Rogers, Meeting strangers and friends of friends: how random are social networks, mimeo, 2006.

9. Jones, C., The Shape of Production Functions and the Direction of Technical Change, Quarterly Journal of Economics, May 2005, Vol. 120 (2), pp. 517-549.

10. Kelly, M., Linkages, Threshold, and Development, Journal of Economic Growth, 6 (2001) 39-53.

11. Kortum, S., Econometrica, 65 6 (1997) 1389-1419.

12. Olsson, O., Knowledge as a Set in Idea Space: An Epistemological View on Growth, Journal of Economic Growth 5: 253-275 (2000).

13. Olsson, O., Technological Opportunity and Growth, forthcoming in Journal of Economic Growth (2005).

14. Peretto, P. and S. Smulders, Technological Distance, Growth and Scale Effects, The Economic Journal, Vol. 112, 603-624, (2002).

15. Samorodnitsky, G. and Taqqu, M.S., Stable non-gaussian random processes: Stochastic models with infinite variance (New York, Chapman and Hall 1994).

16. Saramaki, J., and K., Kaski, Scale-free networks generated by random walkers, Phys. A 341 80 (2004)

17. Vazquez, A., Growing networks with local rules: Preferential attachment., clustering hierarchy, and degree distributions, Phys. Rev. E, 67, 056104 (2003)

18. Weitzman, M., Recombinant Growth, Quarterly Journal of Economics, (1998) 113, 2, 331-360.

19. Zaliapin, I. V., Y. Y. Kagan, and F. P. Schoenberg, Approximating the distribution of Pareto sums, Pure and Applied Geophysics, 162 (2005) 1187-1228.

8 Appendix

8.1 Proof of Lemma 1

Concerning part *a)* of the Lemma, the innovator selects at random the "seed" because selecting a popular idea instead would not increase the expected productivity of the final innovation. Concerning part *b)*, the optimality of the search algorithm is a consequence of the lack of global information, the structure of costs and the existence of local spilovers affecting the productivity of the parents. Concerning part *c)* of the Lemma, we first assume that the mean-field approximation is valid (as in Theorem 1 in Jackson and Rogers (2006)). The proof depends on whether Scenario 1 or Scenario 2 minimize the costs.

8.1.1 Case 1: $c_I < (m-1)c_N - c_R$.

The innovator first selects at random an existing idea i and aims at keeping $m-1$ of its m parents. However, each of these parents will effectively be reached and selected with a probability q_c while with probability $1-q_c$ the innovator selects another idea at random. Then the probability that a node j gets a new attachment in this way is $\frac{m-1}{m}\frac{q_c}{N(t)}k_j^{in}$ where $N(t)$ is the stock of nodes (ideas) at time t. Indeed, j has k_j^{in} incoming edges connected to neighboring ideas that have j as a parent and the probability to hit a given node with a random draw is $\frac{1}{N(t)}$. Finally, $\frac{m-1}{m}$ corresponds to the fact that the parent is disregarded with a probability $\frac{1}{m}$. Second, node j may be selected by a random draw when the innovator ignores one of the $m-1$ parents as described above. Furthermore, j may also be selected when the the innovator stops (by mistake) his walk on the network to find the m^{th} node. The probability that this occurs is $(1-q_c)((m-1)+1)\frac{1}{N(t)}$. Thirdly, a node j may also be selected when the walk on the network passes through node j and ends exactly on node j (with probability p). This is possible only if the walk passes through one of its k_j^{in} parental neighbors, say node jj. Now, node jj may be reached in several ways. First, it may be selected directly by a random draw, with probability $\frac{1}{N(t)}$. The probability that node j gets a new link through this channel is then $p\frac{1}{N(t)}k_j^{in}$. Node jj may also be selected as a result of a walk along one of its incoming edges. Assuming that a parental neighbor of jj was directly hit by the random draw, the probability would be $\frac{1}{N(t)}k_{jj}^{in}\, q_c p k_j^{in}$. Taking expected values (as allowed by the mean field approximation) we get $\frac{1}{N(t)}q_c p m k_j^{in}$. Indeed, the expected value for k_{jj} is given by $2mN(t)/N(t)$ because m edges are created when each of the $N(t)$ nodes are added. Finally, as $k_{jj} = m + k_{jj}^{in}$ we get that $E[k_{jj}^{in}] = m$. In fact, the same logic applies to the parents of the parents of the parents and so on. Finally, the exact value of the overall probability that node j is selected is given by the sum over the entire network of all walks ending on node j. In other words, the expression is a series in p_c.We denote this series by $f(p_c)$, the first two terms being those calculated above.

Then in the mean-field approximation the expression for the rate of change of the inward degree $k_i^{in}(t)$ of node i at time t is

$$\frac{dk_i^{in}(t)}{dt} = \overset{\bullet}{N}(t)\left[\frac{m-1}{m}q_c\frac{k_i^{in}(t)}{N(t)} + m(1-q_c)\frac{1}{N(t)} + p\frac{k_i^{in}(t)}{N(t)}\left[f(q_c)\right]\right]$$

where $\overset{\bullet}{N}(t)$ is the rate at which new nodes are added to the network.

There is a mass $R(t)$ of innovators at time t. There is also a continuum of goods indexed by $j, j \in [0,1]$. As innovators are assigned randomly to each good, there are $R(t)$ innovators trying to come up with an innovation pertaining to the production of good j. We assume that number of innovators grow at a constant rate g, i.e. $\frac{\overset{\bullet}{R}(t)}{R(t)} = g$ or $R(t) = R(0)e^{gt}$. As each innovator produces one innovation per unit interval in the continuum formalization they innovate at a unit rate. The stock of innovations is $N(t) = \int_0^t R(s)ds = R(0)\frac{1}{g}[e^{gt} - 1] \simeq R(0)\frac{1}{g}e^{gt}$ and $\overset{\bullet}{N}(t) = R(0)e^{gt}$. So, $\overset{\bullet}{N}(t)/N(t) \approx g$. Consequently, in the mean-field approximation the expression for the rate of change of the in-degree $k_i^{in}(t)$ of node i at time t is

$$\frac{dk_i^{in}(t)}{dt} = gq_c\frac{m-1}{m}k_i^{in}(t) + gm(1-q_c)\frac{1}{N(t)} + gp\frac{k_i^{in}(t)}{N(t)}\left[f(q_c\right]$$

Collecting the terms we get

$$\frac{dk_i^{in}(t)}{dt} = gm(1-q_c) + g[q_c\frac{m-1}{m} + pf(q_c)]k_i^{in}(t) \equiv c_1k_i^{in}(t) + c_2$$

By construction a node i born in period t_i has a nill in-degree in period t_i, i.e. $k_i^{in}(t_i) = 0$. The solution to the differential equation is of the form

$$k_i^{in}(t) = \left(\frac{c_2}{c_1}\right)e^{c_1(t-t_i)} - \frac{c_2}{c_1}$$

where c_1 and c_2 are the time independent constants defined above. Then it can be seen (see e.g. Lemma 1 in Jackson and Rogers (2006)) that the associated cumulative distribution for the in-degree is

$$P[k_i^{in} \leq k] = 1 - \left(\frac{\frac{c_2}{c_1}}{k + \frac{c_2}{c_1}} \right)^{\xi}$$

with $\xi = [\log(1+g)]/c_1 = \frac{\log(1+g)}{g[(q_c \frac{m-1}{m} + pf(q_c)]}$.

8.1.2 Case 2: $c_I \geq (m-1)c_N - c_R$.

The innovator creates a new idea by selecting m existing idea, each with a probability q_c. In this case the first term in the differential equation becomes simply because $\frac{mq_c k_i^{in}(t)}{mN(t)} = \frac{q_c k_i^{in}(t)}{N(t)}$.

$$\frac{dk_i^{in}(t)}{dt} = gm(1-q_c) + g[(q_c + pf(q_c)]k_i^{in}(t) \equiv c_1' k_i^{in}(t) + c_2'$$

The type analysis of Case 1 gives the same cumulative function $P[k_i^{in} \leq k]$ as above but with c_1' and c_2' as constants. Q.E.D.

8.2 Proof of Lemma 2

Consider first the case in which the weights θ are deterministic. Without lack of generality assume that all weights are identical. Then the m random variables follow independent and identical translated power laws. The translation parameter c_1 may be included in the constant term B_1. Note that c_1 does not depend on m for large values of m. The productivity of ideas then follows a sum of standard power laws. For sufficiently large m, the sum of the Pareto distribution may be approximated by a stable distribution. Indeed, both Condition i) and ii) in Theorem 1.8.1 in Samorodnitsky & Taqqu (1994)) hold. In particular, $a^{\xi}[1 - F_K(a) + F_K(-a)]$, where F_K is the cumulative function of each power law of parameter ξ, is slowly varying at infinity. On the other hand, Theorem 1.1.2 in Samorodnitsky and Taqqu (1994) shows that stable random variables have an asymptotic behavior equivalent to a power law. For $0 < \xi < 2, \xi \neq 1$, the asymptotic behavior of the tail probability of an $\xi-$stable distribution is given by

$$F_{\xi}^{(0)}(a) = 1 - [a^{\xi}\Gamma(1-\xi)\cos(\pi\xi/2)]^{-1}$$

as can be seen from Samorodnitsky & Taqqu (1994), Property 1.2.15 (see also Zaliapin et al. (2005)). Therefore

$$\lim_{a\to\infty}\frac{1-F(a)}{1-F(ca)}=\frac{[a^{\xi}\Gamma(1-\xi)\cos(\pi\xi/2)]^{-1}}{[(ca)^{\xi}\Gamma(1-\xi)\cos(\pi\xi/2)]^{-1}}=c^{\xi}$$

When the weights are random the result is obtained using Breinman (1965) and Lemma 2.1 in Davis and Resnick (1996). Q.E.D.

Remark: The distribution of a sum of m Pareto distributed random variables can also be approximated, for sufficiently large m, by noticing that the largest of the realizations of the individual variables has the same magnitude as the sum (see Zaliapin et al. (2005), Section 3).

8.3 Proof of Corollary 2

The density of a product of two random variables $Z = XY$, with $f(x)$ and $h(y)$ being their densities, is given by

$$g(z)=\int_0^{\infty}\frac{1}{x}f(x)h(\frac{z}{x})dx$$

Let $f(x)$ be the density of the intrinsic productivity. First, assume that the distribution of the spilovers is exactly a power law so that $h(y)=(\frac{y}{\gamma_k})^{-\xi-1}$. In this case we get

$$\begin{aligned} g(z) &= \int_0^{\infty}\frac{1}{x}f(x)h(\frac{z}{x})dx=\int_0^{\infty}\frac{1}{x}f(x)(\frac{z}{\gamma_k x})^{\xi-1}dx= \\ &= z^{\xi-1}\int_0^{\infty}\frac{1}{x}f(x)(\frac{1}{\gamma_k x})^{\xi-1}dx\sim z^{\xi-1}\end{aligned}$$

The argument needs to be modified as the distribution of spilovers is not an exact power law. However, we have seen that for large values the dominant term of the cumulative function is

$$F_{\xi}^{(0)}(a)=1-[a^{\xi}\Gamma(1-\xi)\cos(\pi\xi/2)]^{-1}$$

Therefore, for large values the density function is also a power law of degree $a^{-\xi-1}$. The previous result then applies asymptotically. QED