Power law tails in the inventive productivity: their implications for growth∗

Mauro Rota† Francesco Schettino‡ Luca Spinesi§

March 25, 2008

Abstract

This paper inquires on the inventors’ productivity with regard to patents. The statistical methodology relies on the “Power Law” framework. We apply the Lotka’s law and the usual instruments of “Pareto Law” on NBER Patent Citations Data File, finding that patents per inventor distribution is highly uneven. In depth, we can divide the whole inventors’ population in subsets of “normal” and “important” inventors, in each period and country we considered. We also provide a theoretical interpretation of our outcomes. In particular, the results about different individual productivity in creating new ideas are shown to be tied both to updating with new ideas and to technological spillovers. Finally, both the theoretical interpretation and the empirical findings about a highly heterogeneous individual productivity could be a further candidate to explain a divergence in growth rate and income level between countries.

Keywords: Inventive Productivity, Power Law, Divergence.
JEL: O31, O33, C16.

∗With the usual disclaimers, we thank Carolina Castaldi and Fabio Clementi for helpful comments and suggestions.
†University of Rome, ‘La Sapienza’ - mauro.rota@uniroma1.it;‡Polytechnic University of Marche - f.schettino@univpm.it;§University of Macerata - luca.spinesi@unimc.it;

1
1 Introduction

Looking at chemistry publications, Lotka (1926) discovered an inverse square law of academic productivity: the number of people producing \( n \) papers is proportional to \( 1/n^2 \), i.e. for every 100 researchers with one paper, there are 100/2, with two papers, 100/3 with three, and so on. Shifting the analysis from academic to industrial research, Narin and Breitzman (1995) fitted the above law into the patents distribution of inventors. Such a parallelism is justified by the same concept of power law, i.e. a mathematical representation of a natural law, which simply states that most of the brightest ideas are the products of a few outstanding brains\(^1\). This may be due to different motives such as experience and prestige, although the main reason underpins the normal distribution of intelligence (Ernst et al., 2000). However, with the partial exception of experience, many of these explanatory factors can be hardly quantified. Moreover, while for academics the role of individual prestige can be very important, the same cannot be assumed for inventors since, to be patented, an invention must fulfill a set of objective requirements established by the patent law. This is not to say that, especially nowadays, to publish an article in a highly ranked journal is easier than to have a patent granted by the USPTO or the EPO.

Thus, the idea pushed forward by Ernst et al. (2000), according to which the distribution of patenting output should be even more concentrated than scientific publications one is hardly acceptable. Narin and Breitzman (ibid.) examined how the patents held by four large companies active in semiconductor technology (namely, AT&T, Matsushita, Fuji and Xerox) were distributed among the inventors occupied by them. Looking at the inventors productivity, they found that, in each company, the top 1% of them was from 5 to 10 times more productive than the average.

For each company, Narin and Breitzman’s findings were consistent with the above prediction, suggesting that the overall patenting performances of large firms are strongly dependent on the ingenuity and creativity of a few key inventors. Accordingly, companies should make every effort to retain and nurture them as a sort of golden eggs gooses.

The crucial role played by a narrow set of very productive inventors has\(^1\)

---

\(^1\)With a view to find an (albeit partial) explanation, de Solla Price (1976) developed a probabilistic model of academic publications the so-called square root of elitism in which, for a researcher, the probability to publish an article increases with the number of papers she/he has already published (a sort of learning by doing or Matthews effect).
been restated by Ernst et al. (2000) who, by estimating the law, for the patents of 43 German companies, find that the coefficients are very close Lokta’s predictions. However, they also find that patenting activity (measured by the raw number of patent applications) is more highly concentrated than patenting performance (measured by adjusting the number of applications for their quality). Accordingly, these authors contend that patenting quality tends to increase when the total number of patent applications decreases so that a higher patent quality can compensate for an inventor’s low patenting activity (ibid., p. 190). Since not all the patented inventions have the same value, this casts some doubts on the effectiveness of the Lotka’s law in capturing an important dimension of the inventive process.

Except the two mentioned contribution, few studies (often based on single companies) have examined the distribution of patents among individual inventors. Moreover, the functioning of the Lotkas distribution has been almost exclusively tested at company level. However, because such a distribution should capture a natural law, it should works also in other contexts. For instance, some authors have considered the scientists of different countries and found that Lotka’s law is suitable for representing the distribution of their per-capita publications. The estimation methodology of Lotka’s law could be adequate to depicts the uneven feature of the inventors productivity distribution in terms of patenting activity.

Anyway this is not the only Zipf’s law used to evaluate patenting activity. Indeed, since the pioneeristic work of Vilfredo Pareto (1897), a significant number of researchers have tried to confirm the validity of its (power) law in many fields. These applications have been particularly used in income distribution topic, deriving that, generally, the high-income range follows a Power law distribution, and the rest a lognormal (see in particular Clementi and Gallegati [7],[8]), confirming the basic assumption that, in the capitalist economy “the rich get richer and the poor get poorer”. Yet, some authors inquiring on international differences in productivity and innovative performance, tried to apply this statistical evidence in particular to patent “quality” in order to sketch out a method overcoming the patent homogeneity that comes out from a normal analysis of patent count data. Inquiring on “forward citations”, as the proxy for patent quality (see Jaffe and Trajtenberg [23]), Silveberg and Verspagen [35] divide the total set of patents (provided by NBER Patent-Citations data File) in two groups: the “normal” and the “important” ones, patents whose forward citations frequency follows a lognormal distribution belong to the former, while patents whose forward
citations frequency follows a power-law distribution belong to the latter.

All the following applications and improvements of this statistical procedure (see Castaldi and Los [5],[6]) inquire on patent quality. In this paper we want to principally inquire on the quantity feature of patenting activity. Indeed, we employ this methodology to break up the set of inventors in “normal” and “important” ones, in order to give an empirical evaluation of the patent production dynamics.

Furthermore we try to give a theoretical interpretation of these empirical findings. So, we develop a model of technology production in which the focus is on inventors that produce patents. The flow of new ideas (patents) an individual is able to introduce depends on her own labor effort and on spillovers she gains from the existing stock of knowledge, that are described with a very general law. Yet, the labor effort can be spent both in updating with other researchers’ new ideas and in creating a new idea. The distinction between the time spent either in creating or in updating with new knowledge is crucial to explain the empirical evidence bout individual productivity.

The updating function sketched in our model is similar to the idea of absorptive capacity proposed by Cohen and Levinthal (1989). In their original idea, the knowledge’s absorptive capacity by a R&D firm positively depends on its own R&D effort (expenditure) and negatively depends on the external knowledge complexity. Our idea of an individual’s updating flow cost is similar, yet we suppose that the higher the human capital an individual has the lower the cost of remaining updated, and the greater her own natural talent the smaller the updating cost. Natural talent is supposed to be normally distributed into the population, while accumulated human capital can be thought both as education (higher, graduated and post-graduated) and as on the job learning process. One more step is made allowing for a weight that completes the flow cost function. We first weight for the overall stock of patents granted, and then for individual inventor’s stock of patents.

It is shown that, as long as the overall stock of patents grows at a positive rate, the individual patenting productivity can either grow or remain constant over time. The different dynamics in individual’s productivity depends on the updating flow cost and on the degree of knowledge spillovers. In particular, by deflating the updating flow cost for each inventors’ patent stock, we guess that the updating function tend to zero at a slower pace than when it is deflated for the overall stock of knowledge. In either formulation we find the existence of top researchers whose productivity is higher and higher over time, and a larger mass of researchers whose number of new ideas per unit
of time remains constant. These results determine quite different results once aggregate productivity is analyzed. Indeed, by aggregating the flow of new patents for the mass of inventors provide us with different growth perspective for an economy. Once knowledge spillovers are large and diffuse in the economy, even if they are costly, the economy can grow at a higher pace than with localized knowledge spillovers.

The paper is organized as follows: in Section 2 we describe data and methodology we apply. Then, in Subsection 2.1 we present methodology and principal outcomes of fitting NBER Patent Data (1975-1999) by Lotka’s law, while in next Subsection, going deeper, we confirm the existence of two subsets of inventors, providing and interpreting the results. Thus, in Section 3 we give a theoretical interpretation of empirical results and, then, we conclude.

## 2 The Inventive Productivity

2 Let us define as homogenous such a market where each firm, institution or personal inventor could have the same opportunity and convenience of fill own patent application/s. The typical inventive product, patents, are linked to a special type of labour involved in patenting, inventors. The methods proposed in this section is essentially based on the the Zipf’s law principally used in scientometric literature, the Lotka law (as in Lotka 1926 and Narin and Breizman 1995). In the subsection 2.2 we offer another method relied on the power law, or Pareto law, distribution as in the recent literature (Clementi and Gallegati, 2005, Castaldi and Los 2007, Silverberg and Verspagen 2007). Both methodologies display that population of inventors distribution is uneven. The first one produces in time the same number of patents; the second one produces patents at a positive rate in time. The inventor with more skills, ability endowment, an higher stock of patents has more chances to patenting in future. They grant an higher fraction of overall patent. On contrast a large number of inventors producing the same number of patent are granted for a smaller fraction over time. Let us to go in depth showing results from data.

---

2In this, and in the following Sections and Subsections, we work on data provided by NBER Patent Citation Data File (1963-1999) as in Hall et alt. [19]
Table 1: USPTO Granted Patents by Applicant’s Country - 1975-1999

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>10,260</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Australia</td>
<td>11,386</td>
<td>1.03</td>
<td>1.96</td>
</tr>
<tr>
<td>Belgium</td>
<td>10,972</td>
<td>0.99</td>
<td>2.95</td>
</tr>
<tr>
<td>Canada</td>
<td>53,872</td>
<td>4.87</td>
<td>7.82</td>
</tr>
<tr>
<td>Switzerland</td>
<td>43,313</td>
<td>3.92</td>
<td>11.74</td>
</tr>
<tr>
<td>Germany</td>
<td>221,085</td>
<td>19.99</td>
<td>31.73</td>
</tr>
<tr>
<td>Denmark</td>
<td>6,479</td>
<td>0.59</td>
<td>32.32</td>
</tr>
<tr>
<td>Finland</td>
<td>6,984</td>
<td>0.63</td>
<td>32.95</td>
</tr>
<tr>
<td>France</td>
<td>85,398</td>
<td>7.72</td>
<td>40.67</td>
</tr>
<tr>
<td>Great Britain</td>
<td>98,012</td>
<td>8.86</td>
<td>49.54</td>
</tr>
<tr>
<td>Israel</td>
<td>7,378</td>
<td>0.67</td>
<td>50.2</td>
</tr>
<tr>
<td>Italy</td>
<td>32,433</td>
<td>2.93</td>
<td>53.14</td>
</tr>
<tr>
<td>Japan</td>
<td>421,441</td>
<td>38.11</td>
<td>91.25</td>
</tr>
<tr>
<td>South Korea</td>
<td>14,855</td>
<td>1.34</td>
<td>92.59</td>
</tr>
<tr>
<td>Netherlands</td>
<td>26,687</td>
<td>2.41</td>
<td>95.95</td>
</tr>
<tr>
<td>Sweden</td>
<td>28,286</td>
<td>2.56</td>
<td>97.56</td>
</tr>
<tr>
<td>Russia</td>
<td>6,992</td>
<td>0.63</td>
<td>98.19</td>
</tr>
<tr>
<td>Taiwan</td>
<td>19,979</td>
<td>1.81</td>
<td>100</td>
</tr>
</tbody>
</table>

Total 1,105,822 100

2.1 Evaluating Lotka’s Law

To proceed in our investigation, we used NBER patent database (1975-1999) to inquire on productivity of inventors employed in firms or institutions resident in OECDs countries, excluding those from US. This, because our intention is to compare the inventive productivity into a common and homogeneous market, the American one, in which each firm, institution or personal inventor could have the same opportunity and convenience of fill its/their patent application/s. As a consequence, since the USPTO is the ‘national’ office for American applicants, they have certainly different opportunities and scopes in filling their applications in that patent office. Following this selection criteria, we select the country whose applicants have more than 0.5% of the whole granted patents, in order to homogenize our sample. Thus, we develop our analysis on 1,105,822 USPTO granted patents, and 2,084,813 corresponding inventors (see table 2.1).

Before we proceed it is worthy to recall the generic Lotka’s law as:

\[ Y = \frac{C}{X^\alpha} \]  

where \( X \) is the number of patents, \( Y \) denotes the number of inventors with \( X \) patents, \( C \) is a constant and \( \alpha \) is the coefficient that, according to Lotka’s law, should be equal to 2.
Thus, once we have selected our granted patents sample, before we test the fit’s goodness of Lotka’s law to our data, next to the “Whole patent counts” we needed to construct “Fractional patent counts”. In fact, in the first case, for each inventor, whose name appears on a patent application, is given credit for the whole invention. But, that could generate the problem of over-counting the inventive activity, because an inventor who is single patent authorship is counted as another that produced an idea (co-invented) with other n researchers. Thus, we proceed to make our estimates also on fractional data, giving to each co-inventor an equal part of the patent. As a consequence, if 5 inventors co-developed a patented idea, they bring 1/5 credit each one, instead of 1 credit each one. As [28] notes, this was not a problem in 1926, when Lotka generates its law, because single authorship was the norm. On the contrary, since in last years multi-authorship in patents has grown rapidly, we first fractioned the whole patent counts and, after we rounded that data\textsuperscript{3}, we proceed on estimations on not over-counted data.

Figure 1 depicts the regression for the whole dataset (1975-1999) with the exception of inventors from US. From there we can observe how the difference between ‘theoretical’ Lotka’s law and our estimation is almost insignificant. At this point of our investigations, we go further by estimating the parameters of equation 1 by country (see table 2.1).

As table 2.1 shows, \(\alpha\) parameters are significant at 0.01 level for each regression, and its level is always about 2 as in Lotka’s law, and the \(R^2\) is always higher than 80, denoting the goodness of fit for each single country. Moreover, the robustness of the regression is confirmed by the outcomes of Germany, Great Britain, Japan and France that are the top USPTO applicants. In fact, while \(\alpha\) score is significantly not dissimilar than 2\textsuperscript{4}, the adjusted \(R^2\) is higher than 93%, confirming the goodness level of the Lotka’s model estimations.

2.2 Power Law Tails

The estimation methodology of Lotka’s law, showed in section 2.1, could be adequate to depicts the uneven feature of the inventors productivity distribution in terms of patenting activity. Anyway, since we have patent data by single inventors (2,084,813), we can go deeper in the investigation on

\textsuperscript{3}The fractional assignation of credits made no-sense to apply the Lotka’s law, as an inverse square law, so we need to round the data, see [28].

\textsuperscript{4}As we confirm applying Wald test on these parameters.
Table 2: Regression output (OLS) Fractional Data Counts

<table>
<thead>
<tr>
<th>Country</th>
<th>No. Patents</th>
<th>Constant</th>
<th>Obs.</th>
<th>Adjusted R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>-1.625***</td>
<td>6.241***</td>
<td>31</td>
<td>0.83</td>
</tr>
<tr>
<td>Australia</td>
<td>-2.443***</td>
<td>7.454***</td>
<td>19</td>
<td>0.90</td>
</tr>
<tr>
<td>Belgium</td>
<td>-2.449***</td>
<td>7.722***</td>
<td>22</td>
<td>0.94</td>
</tr>
<tr>
<td>Canada</td>
<td>-2.341***</td>
<td>8.693***</td>
<td>37</td>
<td>0.89</td>
</tr>
<tr>
<td>Switzerland</td>
<td>-2.047***</td>
<td>8.347***</td>
<td>53</td>
<td>0.88</td>
</tr>
<tr>
<td>Germany</td>
<td>-2.379***</td>
<td>10.671***</td>
<td>75</td>
<td>0.93</td>
</tr>
<tr>
<td>Denmark</td>
<td>-2.491***</td>
<td>7.315***</td>
<td>15</td>
<td>0.95</td>
</tr>
<tr>
<td>Finland</td>
<td>-2.308***</td>
<td>7.214***</td>
<td>20</td>
<td>0.92</td>
</tr>
<tr>
<td>France</td>
<td>-2.528***</td>
<td>9.915***</td>
<td>46</td>
<td>0.95</td>
</tr>
<tr>
<td>Great Britain</td>
<td>-2.702***</td>
<td>10.104***</td>
<td>42</td>
<td>0.97</td>
</tr>
<tr>
<td>Israel</td>
<td>-2.011***</td>
<td>6.708***</td>
<td>24</td>
<td>0.88</td>
</tr>
<tr>
<td>Italy</td>
<td>-2.249***</td>
<td>8.583***</td>
<td>42</td>
<td>0.92</td>
</tr>
<tr>
<td>Japan</td>
<td>-2.565***</td>
<td>12.159***</td>
<td>105</td>
<td>0.93</td>
</tr>
<tr>
<td>South Korea</td>
<td>-2.611***</td>
<td>8.267***</td>
<td>24</td>
<td>0.94</td>
</tr>
<tr>
<td>Netherlands</td>
<td>-2.300***</td>
<td>8.042***</td>
<td>30</td>
<td>0.87</td>
</tr>
<tr>
<td>Sweden</td>
<td>-2.452***</td>
<td>8.422***</td>
<td>25</td>
<td>0.89</td>
</tr>
<tr>
<td>Russia</td>
<td>-3.007***</td>
<td>6.366***</td>
<td>8</td>
<td>0.98</td>
</tr>
<tr>
<td>Taiwan</td>
<td>-1.825***</td>
<td>7.070***</td>
<td>38</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Significance: ***0.01,**0.05, *0.1
Figure 2: Quantile to Quantile plot (normal and lognormal distributions) - Patent per Inventor (Fractioned) 1975-1999

Our calculation on USPTO-NBER database
patents distribution form, trying to find its best approximation. In figure 2 we compare, by a quantile to quantile plot, our empirical data with theoretical normal (and lognormal) distribution.

Once we exclude that the number of patents per inventor is well described by a normal distribution (i.e. the inventing probability is not randomly distributed between inventors) we follow Clementi and Gallegati [7], transforming the original distribution, as we take the horizontal axis as the number of (fractioned and rounded) patents per inventor, and the vertical as the logarithm of the cumulate probability (figure 3), that is the probability of finding an inventor able to discover a number of new patentable idea greater or equal to \( x \) (see equation 2).

\[
P(X \geq x) = \int_x^\infty p(t)dt
\]  

(2)

According to the literature, while the first part of the distribution (the one that include the “normal” inventors) follows a two-parameters lognormal distribution (eq. 3), the second part follows a typical Power law distribution (see red line in figure 3 and upper tail 2.2).

\[
p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \left[ \frac{1}{2} \left( \frac{\log(x) - \mu}{\sigma} \right)^2 \right]
\]  

(3)

To select a suitable threshold or cutoff \( X_0 \) value separating the lognormal part from the Pareto power law tail of the empirical patents distribution per inventor - i.e. normal to “superior” inventors -, we use visually oriented statistical techniques such as the quantile-quantile plot. Since a log-transformed Pareto random variable is exponentially distributed, we conduct experimental analysis on the log-transformed data in order to obtain a fit closer to the straight line (see figure 5 for our sample).

Once we have determined the cut-off point, we go further. First, applying MLE methodology, we estimate parameters of lognormal observing and analyzing its dynamic (figure 2.2). Synthesizing the information coming from lognormal distribution, we observe Gibrat index calculated as: \( \beta = 1/ (\sigma\sqrt{2}) \). The lower is \( \beta \) the more uneven is the whole distribution, since it represents a large variance of the sample. For what concerns power law distribution indicators, we run an OLS of the logarithm of the cumulative probability on a constant and the logarithm of patents per inventor, obtaining \( \alpha = 2.752 \).
Figure 3: Cumulative Probability Distribution - Patent per Inventor (Fractioned) 1975-1999

Our calculation on USPTO-NBER database

Figure 4: Upper tail - Patent per Inventor (Fractioned) 1975-1999

Our calculation on USPTO-NBER database
Figure 5: Cut off estimation by Q-Q plot analysis 1975-1999

![Probability plot for Exponential distribution](image)

Our calculation on USPTO-NBER database

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>0.323</td>
<td>0.494</td>
<td>0.476</td>
<td>0.465</td>
<td>0.505</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.689</td>
<td>0.796</td>
<td>0.804</td>
<td>0.805</td>
<td>0.783</td>
<td>0.776</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>2.752</td>
<td>2.713</td>
<td>2.881</td>
<td>2.613</td>
<td>3.025</td>
<td>2.843</td>
</tr>
<tr>
<td>( X_0 )</td>
<td>47</td>
<td>17</td>
<td>18</td>
<td>30</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.993</td>
<td>0.987</td>
<td>0.989</td>
<td>0.991</td>
<td>0.998</td>
<td>0.998</td>
</tr>
</tbody>
</table>
and a $X_0 \approx 47$, with a high degree of goodness ($R^2 = 0.993$); generally, the higher are the cut-off point and the Pareto slope ($X_0$ and $\alpha$), the higher is the concentration of the distribution in the upper tail.

Observing the dynamical behavior in the whole period, from table 2.2 we derive that, both the distribution (proxied respectively by Gibrat $\alpha$ and $X_0$ indexes) depicts that, from 1975 to 1999 the distribution of patents per inventor had increase its uneven feature. These results could revest a high importance in terms of innovation and growth. In fact, if the “simple” existence of a Pareto distribution implies an erratic and not deterministic process in the production of new ideas (patents), so corrupting the hypothesis of diminishing returns of technology, the dynamic findings enforce the idea providing important elements in the discussion about economic convergence.

3 The Model

Let us assume continuous time and unbounded horizon. Population growth is assumed constant and equal to $n > 0$, $L_0 = 1$ for the sake of simplicity, so population at date $t$ will be $e^{nt}$. Aggregate labor supply at date $t$ will be $L_t$, partitioned into fraction $L_{Yt}$ employed in the manufacturing, fraction $L_{At}$ employed in the vertical innovation process and the remaining fraction $L_{Nt}$ employed in the creation of completely new sectors. The total labor force at date $t$ can be expressed as $L(t) = L_{Yt} + L_{At} + L_{Nt}$.

We follow the standard approach of adopting the Dixit-Stiglitz method of aggregating multisector intermediate manufacturing into a unique final output, by means of the constant elasticity of substitution (CES) composite function:

$$C_t = \left[ \sum_{i=1}^{N_t} \frac{Y_{it}^{1+\sigma}}{d_i} \right]^{1+\sigma}$$

where $N_t$ denotes the mass of intermediate goods already introduced into the economy at date $t > 0$, $Y_{it}$ is the production of variety $i$ at date $t$, and $\sigma > 1$ is inversely related to the elasticity of substitution among intermediate products. Let each variety $Y_{it}$ be produced according to a constant returns to scale production function

$$Y_{it} = A_t L_{Yit}^{\alpha} X^{1-\alpha}$$
with \( X \) denoting the fixed factor (land) whose total supply is normalized to 1, \( L_{Yi} \) is labor engaged in the production of the variety \( i \) at date \( t \), and total factor productivity index \( A_t \) captures in a synthetic way the stock of vertical innovations accumulated up to date \( t \). Concentrating on symmetry we assume \( L_{Yi} = L_{Yi} \), \( X_i = X \), and hence \( Y_i = Y_i \), \( \forall i \). Therefore we can write the aggregate final output as

\[
C_t = Y_t = A_t L_t^\alpha X^1 - \alpha N_t^\sigma = A_t (L_t - L_{Nt} - L_{At})^\alpha X^1 - \alpha N_t^\sigma \tag{6}
\]

To simplify matters, as in Howitt (2000), the mass of intermediate products grows as a result of serendipitous imitation, not deliberate innovation. Each person has the same propensity to imitate \( \beta > 0 \), thus the aggregate flow of new products is

\[
\dot{N}_t = \beta L_{Nt} \tag{7}
\]

Since population grows at the constant rate \( n \), also the number of product lines evolves at a relative pace of \( n \).

Therefore, the per capita growth rate is equal to

\[
g_{Yt} = g_{At} + \sigma n - (1 - \alpha) n = g_{At} + (\sigma + 1 - \alpha) n \tag{8}
\]

New ideas are produced using as inputs labor and the stock of ideas – summarized by the per variety productivity level and the number of product lines – which are accumulated over time with the first being paid for and the second being common property.

In the existing literature the innovative activity benefits from the knowledge spillover so that each researcher gets immediately informed, without any effort (or “waste of time”), about the evolution of the general knowledge frontier. As remarked by Aghion and Howitt (1998) and Howitt (1999), and Segerstrom (2000) innovation in one productive line is better viewed as the successful adoption of general knowledge advancement produced as a byproduct of the whole economy’s innovative efforts. We adopt this view of the innovative process, but we postulate that R&D units need to keep updated about the ongoing recent advances by spending a proportional labor cost.\(^5\)

\(^5\)See Cozzi and Spinesi (2004) about this point and the related references. In particular, we adopt the macroeconomic framework as in Cozzi and Spinesi (2004), and we provide a microfoundation of their updating sunk cost that allows the empirical findings of our contribution to be explained.
Our model’s updating labor cost is proportional to gathered information, but is inversely proportional to the economy-wide total factor productivity level. More specifically, at the date \( t > 0 \), the dynamic evolution of the stock of knowledge - i.e. the number of patents in this framework - for a researcher endowed with an exogenous ability \( \theta_j \) and with a skill level \( h_j \) follows the law

\[
\dot{A}_{jt} = \frac{\dot{A}_t}{L_{At}} = \varphi (A_t, N_t) \text{Max} \left\{ 1 - \frac{f_j (\theta_j, h_j)}{A_t} \frac{\dot{A}_t}{L_{At}} (L_{At} - \varepsilon), 0 \right\}
\]

(9)

where the flow of new patents for a generic researcher \( j \) - i.e. \( \dot{A}_{jt} \) - can be approximated by the per researcher flow of new ideas \( \frac{\dot{A}_t}{L_{At}} \). The term \( \varphi (A_t, N_t) \) is any real and positive function that captures the productivity of the accumulated knowledge stock \( A_t \) over the mass of the existing product lines \( N_t \) at time \( t > 0 \) in the improvement of productivity, \( L_{At} \) is the fraction of total labor force \( L_t \) working in R&D. In equation (9) what allows us to discriminate between the researchers’ productivity in introducing new ideas in the long term is the function \( f_j (\theta_j, h_j) \). Indeed, at a given time \( t > 0 \), the individual productivity is assumed to be not much different between researchers, so that the individual productivity can be approximated by the per researcher productivity. Over time, the flow of new ideas a researcher will be able to introduce can follow a quite different dynamic between the researchers, with a quite different long term behavior of a personal number of new patents granted to each researcher, as the data show to be in practice. The function \( f_j (\theta_j, h_j) \) represents the individual’s updating flow cost, which is tied to both her personal ability endowment \( \theta_j \) and her skill ability \( h_j \). To fix ideas, \( \frac{f_j (\theta_j, h_j)}{A_t} > 0 \) labor units have to be sunk in order to try to adopt each general technological improvement to the product/industry line in which the researcher operates.\(^6\) Every time new technological improvements occur at the economy-wide level, Howitt’s (1999) and Segerström’s (2000) sectoral adoption problem becomes a new one, which means that a second

\(^6\)As Garicano (2000, p. 878) maintains: “Workers can learn the solutions to the problems they confront at a cost. I assume that the cost of learning an interval \( A \) of problems is proportional to the size of this interval, \( \mu(A) \) (its Lebesgue measure), and call the constant period unit learning cost \( c \). For example the cost of learning all problems in the interval \( [0, Z] \) is \( cZ \).”
vertical R&D labor sunk cost has to be incurred. The sequence of successive sunk R&D labor costs per unit time is equal to the number of such sunk labor costs times the number of general technological improvements being aimed at per unit time, i.e. \( \frac{A_t}{L_{At}} (L_{At} - \varepsilon) \). This becomes a quasi-fixed flow cost for each researcher’s vertical R&D. Indeed, in order to introduce new patentable ideas, an individual should be updated of at least a fraction of other researchers’ flow of new ideas, which is captured by the term \( \frac{A_t}{L_{At}} (L_{At} - \varepsilon) \) in the equation (9).

The term \( \frac{A_t}{L_{At}} \) indicates the per researcher flow of new patents, and \( (L_{At} - \varepsilon) \) indicates the number/mass of researchers but her own research effort time, which is assumed to be an infinitesimal amount \( \varepsilon \) of the whole research effort. It is important to remark that we are not even assuming that R&D workers need to know all relevant innovations in their field in order to have a positive probability of innovating. The function \( f_j(\theta_j, h_j) \) can capture any fraction of the flow of innovations that it is necessary to know. It may well be that such a fraction is as small as one thousandth or less.

It is worthwhile noting that equation (9) captures the notion that the faster the technological frontier expands, the more flow labor effort is necessary to assimilate it, but at the same time the higher the already acquired general knowledge the easier the acquisition of new information. In fact, given the already reached level of the transmission technology, monitoring a double information flow imposes a higher monitoring time, but advances in information and communication technologies make it easier to transmit and to scan a larger mass of information per-unit time. Hence it seems quite natural to assume that quasi-fixed R&D input requirements in the information transmission process decrease in the same proportion as do the manufacturing input requirements.

The function \( f_j(\cdot) \) is assumed to be bounded below for any researcher, with \( f_j(\cdot) > 0 \) being the greatest lower bound of researcher’s \( j \) updating flow cost. Moreover, let us define the parameter \( c > 0 \) as the minimum greatest lower bound for all researchers, i.e. \( c = \min \{ f_j(\cdot) \} \). The ability endowment \( \theta_j \) can be interpreted as the individual’s \( j \) talent in finding new ideas, and in updating with the other researchers’ ideas. The skill ability \( h_j \) can be interpreted as the both the quantity and quality of human capital accumulated through schooling by the researcher \( j \) and as on the job learning process, and it represents her personal skill ability when she undertakes the research activity. The individual updating function is assumed to be a decreasing function in both ability endowment and skill, i.e. \( \frac{\partial f_j(\cdot)}{\partial \theta_j} \equiv f_j(\theta_j, h_j) < 0 \) and
The “purely innovative activity” in developing countries often – though not always - consists in the learning and adaptation to local conditions of technologies invented elsewhere. In such a case it is very natural to assume that the intertemporal learning spillover might be different from that of the leading-edge countries, because the researchers are doing qualitatively different kinds of activities. Therefore we believe that a positive feature of our formulation is the high degree of generality of function \( \varphi(A_t, N_t) \). The inspiration that local researchers get from foreign knowledge stocks is potentially complicated and heterogeneous from culture to culture, whereas the mechanical information transmission of new knowledge or new adoption flows respond in the same way to general industrial productivity nearly everywhere. Hence we believe that our formulation of R&D seems to fit also a development context in a flexible enough way as to include several different applications.

By solving equation (9) for \( \frac{\dot{A}_t}{L_{At}} \), the following dynamic law is obtained:

\[
\frac{\dot{A}_t}{L_{At}} = \frac{1}{\varphi(A_t, N_t)} + \frac{f_j(\theta_j, h_j)}{A_t} (L_{At} - \varepsilon)
\]

The long term behavior of equation (10) can be quite different depending on the behavior of the second term in the denominator of the right hand side of the same equation (10). Indeed, the first term in the denominator tends to zero in the long run. Instead the second term can have a different dynamic depending on the relative strength of its components. Both the stock of knowledge \( A_t \) and the population size tend to infinity in the long run, as long as the growth rate of \( A_t \) is positive and population grows over time. Therefore, in determining the long run behavior of equation (10), a crucial role is assumed by both the evolution and weight of the function \( f_j(\theta_j, h_j) \). As long as the updating function allowed the growth rate of the numerator in the term \( f_j(\theta_j, h_j) \left( L_{At} - \varepsilon \right) \) to be low enough relatively to the stock of knowledge growth rate, the flow of new patents would be higher and higher over time, i.e. \( \frac{\dot{A}_t}{L_{At}} \to \infty \) as \( t \to \infty \). On the contrary, as long as the updating function did not allow the growth rate of the numerator in the term \( f_j(\theta_j, h_j) \left( L_{At} - \varepsilon \right) \) to be low enough relatively to the stock of knowledge growth rate, the flow of new patents would tend to zero, and therefore a researcher would be able to introduce a constant number of patents over time, i.e. \( \frac{\dot{A}_t}{L_{At}} \to 0 \) as \( t \to \infty \). This last result is more plausible whenever a
researcher is endowed with a low talent $\theta_j$, and whenever she has accumulated a lower quantity and quality of skills in schooling, i.e. for a low $h_j$. Moreover, whenever $h_j$ represents also some learning process - such as on the job training - the evolution of such skill ability allows the updating flow cost to be reduced over time, so that an individual productivity in introducing new valuable ideas becomes higher and higher over time.

By aggregating equation (9) over the number/mass of researchers at time $t$, the following is obtained

$$\dot{A}_t = \varphi (A_t, N_t) \max \left\{ \frac{L_{At}}{A_{At}} (L_{At} - \varepsilon) \frac{\tilde{F} (\cdot) L_{At}}{A_t}, 0 \right\}$$

(11)

where $\frac{\tilde{F} (\cdot) L_{At}}{A_t} = \int_{L_{At}} \frac{f_j (\theta_j, h_j)}{A_t} dj$, with $\tilde{F} (\cdot)$ being the per researcher cumulative updating function, which can be written as proportional to the research labor size $L_{At}$ at time $t$. Because the weight of a researcher over the number/mass of researchers is assumed to be infinitesimal - being $L_{At}$ the world research labor size - the other researcher’s size $(L_{At} - \varepsilon)$ at time $t$ is well approximated by $L_{At}$, i.e. $(L_{At} - \varepsilon) \simeq L_{At}$. Therefore, equation (11) reduces to

$$\dot{A}_t = \varphi (A_t, N_t) L_{At} \max \left\{ 1 - \frac{\tilde{F} (\cdot) \dot{A}_t}{A_t}, 0 \right\}$$

(12)

where, because $c = \min \left\{ \tilde{f}_j (\cdot) \right\}$, we have that $\left[ 1 - \frac{\tilde{F} (\cdot) \dot{A}_t}{A_t} \right] < \left[ 1 - \frac{c}{A_t} \dot{A}_t \right]$. Therefore, by solving equation (12) for the growth rate of the stock of knowledge, the following inequalities are obtained

$$\frac{\dot{A}_t}{A_t} = g_{At} = \frac{1}{\varphi (A_t, N_t) L_{At}} + \tilde{F} (\cdot) < \frac{1}{\tilde{F} (\cdot)} < \frac{1}{c}$$

(13)

This holds regardless of the degrees of intertemporal spillover and regardless of the economy’s being at a steady state or out of it. Moreover, since the economy’s growth rate is always lower than $\frac{1}{\tilde{F} (\cdot)}$ in principle our result allows both for endogenous and for exogenous growth depending on the microeconomic foundations adopted in a more complete model.

So far we have assumed that researchers’ updating flow cost benefits of spillovers from all the world-wide stock of knowledge $A_t$, or however from the stock of knowledge for which this economy is specified for.
However, some empirical evidence shows the existence of localized knowledge spillovers. Studies from Jaffe (1993), Coe and Helpman (1995, 1997), Keller (2002) show the importance of the spatial proximity for knowledge diffusion and utilization. As Baldwin and Martin (2004) maintain: “The diffusion of knowledge across regions and countries does exist but diminishes strongly with physical distance which confirms the role that social interactions between individuals, dependent on spatial proximity, have in such diffusion.” A similar argument, and some empirical evidence, can be found in cities. For example Black and Henderson (1999) maintain: “An urbanized economy has different type of cities specialized in different traded goods, with city size and educational attainments varying by city type.” Moreover, specialization in some product lines can be due to the exogenous first natural characteristics of each country.

Once these empirical facts are inserted in the above modeling specification, a different result about the growth performance of the economy arises. Indeed, let us suppose that in equation (9) a researcher updating cost benefit of only very limited knowledge spillovers:

\[
\dot{A}_{jt} = \frac{\dot{A}_t}{L_{At}} = \varphi(\alpha A_t, N_t) \max \left\{ 1 - \frac{f_j(\theta_j, h_j)}{A_{jt}} \frac{\dot{A}_t}{L_{At}} (L_{At} - \varepsilon), 0 \right\} 
\]

(14)

where the differences with equation (9) concerns the spillovers term. In function \( \varphi(\alpha A_t, N_t) \), a researcher benefits of only a fraction \( \alpha \in (0, 1) \) of the stock of the existing knowledge spillovers, and the updating flow cost is reduced according to her own knowledge accumulated over time\(^7\). By following the same steps as before, the following dynamic law for an average researcher flow of new ideas is obtained

\[
\dot{A}_{jt} = \frac{\dot{A}_t}{L_{At}} = \frac{1}{\varphi(\alpha A_t, N_t)} + \frac{f_j(\theta_j, h_j)}{A_{jt}} (L_{At} - \varepsilon) 
\]

(15)

In such a case the dynamic evolution of a researcher’s flow of new ideas strongly depends on her personal ability and productivity. Indeed, as above the first term in the denominator tends to zero in the long run. Instead, the

\(^7\)This is a drastic specification for the flow updating cost adopted to render the new result more evident. Yet, also assuming that the updating cost benefits of a limited knowledge spillovers does not alter the qualitative results
second term can have a different dynamic depending on the relative strength of its components. Both the stock of knowledge $A_{jt}$ and the population size tend to infinity in the long run, as long as the growth rate of $A_{jt}$ is positive. Therefore, in determining the long run behavior of equation (15) a crucial role is assumed by both the evolution and weight of the function $f_j(\theta_j, h_j)$. As long as the updating function allowed the growth rate of the numerator in the term $\frac{f_j(\theta_j, h_j)}{A_{jt}} (L_{At} - \varepsilon)$ to be low enough relatively to her own stock of knowledge growth rate, the flow of new patents would be higher and higher over time, i.e. $\frac{A_{jt}}{L_{At}} \to \infty$ as $t \to \infty$. This is the case whenever her own stock of knowledge is higher over time. On the contrary, as long as the updating function did not allow the growth rate of the numerator in the term $\frac{f_j(\theta_j, h_j)}{A_{jt}} (L_{At} - \varepsilon)$ to be low enough relatively to own stock of knowledge growth rate, the flow of new patents would tend to zero, and therefore a researcher would be able to introduce a constant number of patents over time, i.e. $\frac{A_{jt}}{L_{At}} \to 0$ as $t \to \infty$.

However, the growth performance of an economy with localized spillovers can be quite different in the very long run than an economy with economy-wide knowledge spillovers. Indeed, by aggregating equation (14) over the number/mass of researchers at time $t$, the following is obtained

$$\dot{A}_t = \varphi (\alpha A_t, N_t) \max \left\{ L_{At} - \frac{\dot{A}_t}{L_{At}} (L_{At} - \varepsilon) \frac{cL_{At}}{A_{jt}}, 0 \right\}$$  \hspace{1cm} (16)$$

where we assume that $f_j(\theta_j, h_j) = c \forall j$, and $\bar{A}_{jt} = \max \{A_{jt}\}_{j=0}^{L_{At}}$, so that we pose the economy in the condition to have the highest growth performance, because the updating cost has a value as low as possible. In such a case $\int_{L_{At}} f_{j}(\theta_j, h_j) \, dj$ reduces to $\frac{cL_{At}}{\bar{A}_{jt}}$. Because the weight of a researcher over the number/mass of researchers is assumed to be infinitesimal - being $L_{At}$ the world research labor size - the other researcher’s size $(L_{At} - \varepsilon)$ at time $t$ is well approximated by $L_{At}$, i.e. $(L_{At} - \varepsilon) \simeq L_{At}$. Therefore, equation (16) reduces to

$$\dot{A}_t = \varphi (\alpha A_t, N_t) L_{At} \max \left\{ 1 - \frac{c}{A_{jt}} \dot{A}_t, 0 \right\}$$  \hspace{1cm} (17)$$

The stock of knowledge growth rate can be written as
**Concluding Remarks**

This paper inquire on the inventive productivity distributions of inventors employed in firms or institutions resident in OECDs countries, excluding those from US using NBER patent citations database (1975-1999). Considering the number of patents granted to individual inventor on the USPTO, the researchers’ population is divided into two well defined groups. The first one involves inventors with very low productivity, at least constant; while the second one includes researchers with increasing productivity over time. More formal we have found that the inventive productivity is distributed in the population as a log normal probability function with highly skewed tail. The last is governed by a power law or Pareto law. The skewed tail of distribution indicates that a sample of inventors in the population accounts for a very large amount of patents. In this way knowledge is concentrated in few researcher, while the largest group of inventors grants a constant number of patents over time. These results is verified at a country level over 1975-1999.

The theoretical model describes the flow of new ideas (patents) an individual is able to introduce as depending on labor effort and on spillovers one gains from the existing stock of knowledge. In articular, the labor effort can be spent both in updating with other researchers’ new ideas and in creating a own new idea.

It is shown that, as long as the overall stock of patents grows at a positive rate, the individual patenting productivity can either grow or remain constant over time. The different dynamics in individual’s productivity depends on the updating flow cost and on the degree of knowledge spillovers. In particular, by deflating the updating flow cost for each inventors’ patent stock, the updating function tend to zero at a slower pace than when it is
deflated for the overall stock of knowledge. In either formulation we find the existence of top researchers whose productivity is higher and higher over time, and a larger mass of researchers whose number of new ideas per unit of time remains constant.

These results determine quite different results once aggregate productivity is analyzed. Indeed, by aggregating the flow of new patents for the mass of inventors provide us with different growth perspectives for an economy. Once knowledge spillovers are large and diffuse in the economy, even if they are costly, the economy can grow at a higher pace in respect of localized knowledge spillovers. These results could be a further candidate to explain both the per capita output growth rate and income differences between countries. In particular, a central question looked to be that several nations were not able to “rapidly assimilate industrial technology” (DeLong 1988, p. 1148). Furthermore, the seminal papers by Lucas (1988) and Romer (1986, 1990) concern a widening in the relative income distances between rich and poor due to the nontrivial questions that technology is the key element for growth and its transfer is not inevitable and not costless as well. Moreover, Easterly and Levine (2001) have shown that technology and its spillovers matter for convergence of real per capita incomes. If spillovers are local, diffusion of technology is limited in the neighbors of innovator firm/country and its effects do not expand over the world economy. That leads to economic clusters with persistent income differences. If spillovers are large and diffuse, technology is better disposable for a larger number of agents and differences tend to diminish over time. Keller (2002) has recently estimated that technology spillovers from the most innovative countries are local, but the localization degree has fallen on 1982-1998, when within industries spillovers are considered. The theoretical results of this paper about individual productivity in creating new ideas tied to technology spillovers could be a further candidate to explain these empirical findings about divergences in growth rate and income level between countries.
References


European Economic Review 39, 859-887.


R&D, The Economic Journal, 99 (397), pp 569-596

[12] Cozzi G and Spinesi L, (2004), Information Transmission and 
the Bounds to Growth, Topics in Economic Analysis & Policy, 

Spillovers, European Economic Review, Elsevier 39 (5), pp. 859-
887.

1138-1154.


[16] Easterly W. and R. Levine (2001), It’s Not Factor Accumu-
lation: Stylized Facts and Growth Model, World Banking Eco-
nomic Review, 15, pp 177-229.

[17] Garicano L. (2000), Hierarchies and the Organization of Knowl-
874-904.


