Patent Citations, the Value of Innovations and Technological Trajectories

*Bulat Sanditov**
MERIT, P.O. Box 616, 6200MD Maastricht.

14th August 2005

Abstract

Patent citations as a proxy for the value of innovations has received considerable attention in the recent years. This paper examines how the framework of technological trajectories can be applied to explain the distribution of patent values. A simple model based on generalized Polya urn processes is proposed, and it is shown that the model fits empirical distribution of patent citations (USPTO and EPO data) surprisingly well.

1 Introduction

It is well recognized these days that only efficient production, accumulation, and utilization of technological knowledge can ensure long term economic growth. Planning and implementing R&D programmes have become a routine task for many governments and companies around the world. Therefore the knowledge about the distribution of returns from R&D is of great practical importance.

The main problem hindering research in this direction has been scarcity of data on R&D. However with the arrival of new data, particularly patent data, and advances in methodology the field

*Preliminary draft. Please, do not cite without permission.

is rapidly expanding. The evidence accumulated in recent years confirm earlier findings and are univocal on the overall features of the distribution of the innovation values: it is highly skewed with most of the innovations having value close to zero, and few innovations scoring very high, a fact that has direct implications for planning and evaluation of innovation policies and firm strategies (Scherer and Harhoff 2000).

Although the extreme skewness of the distribution is now a well established fact, the precise form of the distribution of the innovation values is still under debate. In particular, there is a controversy about the right tail of the distribution. Based on the results of a survey of holders of German patents Harhoff, Scherer, and Vopel (1997) report that the the best fit for the tail (defined as innovations with values over DM 23,000) is obtained with lognormal distribution (vs. Pareto and Singh-Maddala distributions). On the contrary, applying techniques of extreme-value theory to the set of different data on the innovation values, Silverberg and Verspagen (2004) demonstrate that if the lower bound of the tail is set correctly, the tail is fitted better with Pareto distribution (rather than with lognormal distribution).

So far, research in this direction has been focussed on the properties of the distribution. In this paper I propose to approach the problem from the other end: instead of questioning what is the exact form of the distribution of innovation values I inquire about the *process* that generates the distribution. I will argue that the evolutionary theory of technical change is helpful in understanding the dynamics of innovation values.

My argument proceeds along the following lines. According to evolutionary theory the process of technical change is incremental and follows "technological trajectories". The latter implies path-dependency in development of technology: success of the innovation, representing a particular approach in solving certain type of engineering problems, might lead to lock-in, i.e. for a period of time this type of problems will be approached in the similar way. Moreover, success of the innovation, related to resolution of an important design problem, plays a role of "focusing device" (Rosenberg 1969): it directs innovative search to the areas of "technology space" where the given innovation may lead, and as a result, stimulates the flow of the inventions based on the technology represented by the innovation. Assuming that the value of the innovation depends on the scope of the problems it can be applied to, we can expect that the more valuable the innovation is, the more likely it is to be employed in consequent innovations, and, as a result, the more valuable it will become.

To formalize this intuition I propose a simple model based on generalized Polya processes that takes into account the path-dependent nature of technical change, and show that the model fits the distribution of patent citations (a measure of the value of patented inventions) very well.

The paper is structured as follows. In the next section I review the literature on how the value of innovations is related to the characteristics of the process of technical change, and the measures of the value of innovations with particular attention to patent citations that I use later to test my model. Then I formulate the model and describe the data I use to test the model. The results of fitting the distribution of patent citations are presented in Section 5. Section 6 discusses several aspects of the model. The last section concludes the paper.

2 Literature

Value of Innovations and Direction of Technical Change

It is almost obvious and self-evident that the value of an innovation depends on the characteristics of the technical change, and therefore the value cannot be defined out of the context of the history of the technology. This fact is not very important for a retrospective judgement (because the history has been already realized), however if we aim at a dynamic view of technical change, then before questioning what is the distribution of the value of innovation, we must try to answer what is the *process* that makes some innovations more important than others.

The answer to this question depends on what kind of picture of the technical change one has in mind. For a simple linear model of technological progress which seems to dominate modern growth literature, the answer is straightforward: technological progress is nothing but expansion of the production set, therefore the innovation offering the highest reduction of production costs (for a process innovation), or/and higher quality of the product (for a product innovation) will have the highest value which is steadily decreasing as new, better methods of production keep coming (e.g. Aghion and Howitt 1992). Assuming that the gain in the productivity is distributed according to some probability law, the uncertainty surrounding the value of the innovation is contained in the demand and the rate at which innovations arrive. Moreover most models assume that demand does not change over time (embedded in the utility function of a representative consumer). Therefore *ex-post* distribution of the values can be inferred from the distribution of the productivity gains and the rate of the technical change. There is no place for the direction of technical progress, because in the linear model there is only one (toward increasing productivity).

On the other hand, according to the evolutionary tradition in the economics of technology, the process of technical change follows path-dependent "technological trajectories" punctuated by discontinuities of "natural trajectories"/ "technological guideposts"/"technological paradigms" (Nelson and Winter 1977, Sahal 1981, Dosi 1982). Most of the time we expect to observe relatively stable clusters of (interlinked) technologies, with more valuable core technologies in centre of each cluster.

The clustering of the patented inventions in technological space has been analyzed with the use of patent documents through IPC classification, patent citations (in bibliometric style), and textual analysis of the patent documents. Pier, Rost, Teichert, and von Wartburg (2003) use (EPO) patent citation data to decompose the "technological blob" of mobile telecommunication. Huang, Chen, Yip, Ng, Guo, Chen, and Roco (2003) use longitudinal patent data for nanoscale science and engineering to make country, institution and technology field comparisons. They employ both content map analysis and patent citations. On time-series content maps they observe several dominant topics occupying different periods of time. Graff (2003) surveys the use of patent data for identification of micropatterns in innovations in agricultural technology.

There are (at least) two factors behind path-dependency in technical change which tend to 'bunch' technologies together: (a) complementarities between contemporary technologies, and (b) localization of the search in the technological space, due to the bounded rationality of agents.[1]

In the study of interdependencies between technologies in the American economy Nathan Rosenberg notes

> Inventions hardly ever function in isolation. Time and time again in the history of American technology, it has happened that the productivity of a given invention has turned on the question of the availability of complementary technologies. Often technologies did not initially exist, so that the benefits of potentially flowing from invention A had to await the achievements of inventions B, C, and D. These relationships of complementarity therefore make it exceedingly difficult to predict the flow of benefits from any single invention and commonly lead to postponement in the flow of such expected benefits. (Rosenberg 1982, p.56)

[1] For different technologies relative importance of systemic and cognitive factors mentioned here may differ.

Rosenberg supports this thesis with a number of examples from the history of transport sector, agriculture, electricity, machine tools, metallurgy *etc.*

Silverberg and Verspagen (2005) point out that even such seemingly simple invention as a bicycle is, in fact, a collection of many related inventions including "pneumatic tyres, ball bearing (and thus precision machining, the precision grinding machine ...) [...] without which bicycle boom of the 1890s would have been unthinkable". They formulate and examine a model in which a new technology becomes feasible only if it has links with the technologies already in use. In this model the importance (value) of the technology depends not only on the increase in the productivity this technology offers, but also on whether this technology can make other technologies available, i.e. on the direction of technical change.

From the very beginning, modern evolutionary economics has recognized that economic agents are characterized by bounded rationality (Nelson and Winter 1982), i.e. their behaviour is governed not by full optimization over the complete set of control variables, but by the process of trial-and-error, some search heuristics (embedded in the routines of an organization), or optimization over a subset of control variables in a limited domain. A range of different models of the search process can be found in the current evolutionary economics literature (Frenken 2004, for a survey): simple trial-and-error similar to the learning in an evolutionary game, genetic algorithms when strategies are coded in binary strings, and a new string arises through recombination of parent strings (Birchenhall 1995, Dawid 1999), NK-models of search on technological landscapes, an application of the cutting-edge theories from the evolutionary biology (Kauffman 1993, Frenken and Nuvolari 2004), and simulated annealing, a method of combinatorial optimization originated from modelling thermodynamic systems in physics (Cooper 2000).

It is worth emphasising that given our research question we shall view the search process not at the level of individual agents performing their search on their own, but as a process that involves the whole technological community; this community includes inventors, firms, government labs, academicians and the like. There is an obvious parallel with the sociology of science, in particular, with Thomas Kuhn's "scientific paradigms". During the stable phase of the development of a technology researchers and engineers have a number of standard approaches to solve standard problems shared by the community. To solve a particular engineering problem means to find an appropriate standard solution (design) and to adjust it to the problem (Cooper 2000).

Furthermore, the *research agenda* (i.e. what needs to be improved, what can be achieved with available techniques *etc.*) is also shared at the community level. As a result, the direction of innovative search is framed by the current state of the technology and hence depends on the previous success (in terms of both technological achievements and commercial benefits). Such a picture of innovative search goes along with the views of Rosenberg (1969, 1974) who sees inventive activities as focused on a set of related engineering problems ("focusing devices/technological imperatives") which result in "compulsive sequences" of innovations over time.

The evolutionary view of technical change does not contradict the intuition one can get from patent citation literature. Trajtenberg (1990) explains that "if citations keep coming, it must be that the innovation originating in the cited patent had indeed proven to be valuable". Somewhat similarly Harhoff, Narin, Scherer, and Vopel (1999) word it as " it is reasonable to suppose that the prior inventions cited in new patents tend to be the relatively important precursors that best define the state of the art. The broader the shoulders, the more likely they are to be cited".

Summarizing at this point, we expect to see clusters of related technologies with the set of "standard solutions" in the centre of each cluster. The value of an innovation depends on the direction of technical change: *a successful innovation is likely to be replicated time and time again*, either because it is a key to link new technologies with existing ones as in the model of Silverberg and Verspagen (2005), or because boundedly rational agents use it as a starting point in the process of search in technology space. We can hypothesize that the patent citations capture this relationship between innovations.[2]

Moreover, success in solving an important design problem attracts more innovative search in the related area of the technology space. Search aims both to improve the solution and to explore the area of the technology space opened by the innovation. Increasing intensity of the search, in turn, leads to an increasing flow of innovations based on the given innovation, and expands the scope of the problems to which the technological knowledge underlying the innovation can be applied. Therefore, other things equal, the path-dependent nature of technical change implies a path-dependent dynamics of innovation values.

Now, let us turn to the issue of how we can trace the path-dependent nature of technical change in the patent data.

Patent citations and the value of innovations

There are several ways to assess the values of patented inventions. Pakes and Schankerman (1984), Pakes (1986), Schankerman and Pakes (1986) have employed data on patent renewal to estimate the characteristics of the values of the patent rights. Lanjouw, Pakes, and Putnam (1998) extended this framework in order to utilize data on the applications for a patent (related to the same invention) in different countries ("family size"). Another approach to assessing the value of patents is to use the stock market

[2]Citations can be made by inventors themselves or added by their attorneys or patent examiners.

valuation of a company to which the patents have been granted (Griliches 1981, Pakes 1985, Hall, Trajtenberg, and Jaffe 2001, among others). Yet another stream of research that has proved to be very productive is to utilize information contained in the patent documents themselves (number of citations, number of claims, number of IPC classes). In particular, citations-based indices have been very successful (Trajtenberg 1990, Trajtenberg, Henderson, and Jaffe 1997, Harhoff, Narin, Scherer, and Vopel 1999, Jaffe and Trajtenberg 2002). Finally, in the recent paper Harhoff, Scherer, and Vopel (2003) found that outcomes of opposition against patent grants proved to be highly informative for predicting the value of the patent rights (taken from a survey of holders of German patents).

Results of most studies indicate that the measures of the patent values mentioned above are mutually coherent, and more importantly, most of the measures correlate well with the value of the patents inferred from the direct surveys of the patent holders. Harhoff, Narin, Scherer, and Vopel (1999), Harhoff, Scherer, and Vopel (2003) tested a set of different measures of patent quality as predictors of the patent value obtained from the survey of holders of German patents and found that forward citations, family size, outcomes of opposition proceedings, and whether patents were renewed to a full-term correlate well with the patent values.

Among other measures of the patent quality measures based on citations (in particular, forward citations) are appealing for a number of reasons. First, as has already been mentioned, most studies suggest that the number of citations a patent has received (*forward citations*) is a good proxy for social and private returns to the innovations. Second, all information needed for the construction of appropriate measures is contained within the patent document. Third, modern software and publicly available computer-readable data make it easy to construct these measures tailored to different patent classes, institutions, countries *etc.*

Furthermore, one might interpret patent citations to prior art as "paper trails" of knowledge spillovers (Jaffe and Trajtenberg 2002). Such interpretation of patent citations led to a prolific research avenue in different areas of innovation studies ranging from spatial economics (Jaffe, Trajtenberg, and Henderson 1993), to university-industry links (Henderson, Jaffe, and Trajtenberg 1998), and social network analysis (Balconi, Breschi, and Lissoni 2004).

Nevertheless, patent citation data should be used with some caution. First, there is a problem with the "benchmarking" of citation data (Hall, Jaffe, and Trajtenberg 2001). For example, if we are to compare two patents taken in different years, and suppose that the older patent has received more citations than the other one, then it is not clear if it is because the old patent is more valuable, or simply because, it had more chances to be cited (truncation problem). It is also important to keep in mind that the stock of patents is rapidly growing; hence, other things equal, the earlier patent have higher chances to be cited, then the patents which were taken later. Furthermore, changes in practices in Patent Offices and in patenting strategies of firms may lead to additional complications for an intertemporal comparison.

Second, results of surveys of innovators have cast some doubts that patent citations represent "paper trails" of *direct spillovers* i.e. the fact that the owner of a citing patent learned about the innovation contained in the cited patent from the cited patent itself or from the holder of this patent prior to the invention (Jaffe, Trajtenberg, and Fogarty 2000). It is not rare that innovators have learned about the predecessors of their patents only at the stage of patenting. Many citations have been added by patent examiners, or innovators' attorneys, and hence cannot be regarded as an evidence of direct spillovers.

The problem with benchmarking can be resolved if we limit comparison of patents to one cohort, i.e. to patented inventions made in more or less the same time, provided that by the time of observation the patents have accumulated enough citations. This, in turn, raises a question about dating the patents. A patent document published by a Patent Office of interest, in our case - the United States Patent and Trademark Office (USPTO) for the NBER dataset, and the European Patent Office (EPO) for the CESPRI dataset, contains several dates: *priority date* - the date when the patent was applied to the Patent Office in any jurisdiction; *application date* - the date when the inventor filed the documents for a patent to the Patent Office of interest; *grant date* - the date when the Patent Office issued the patent to the inventor. We are interested in the date closest to the time of invention, which is the priority date for the EPO patent data, and application date for the USPTO data[3]. Therefore to avoid the problem with benchmarking we shall select patents with the priority/application date within a small period of time (one year seems to be appropriate time span).

The concern about whether a patent citation represents a direct spillover, or it is evidence of an indirect spillover coming through the "word-of-mouth" via the social network of inventors (Breschi and Lissoni 2004), is not essential for our purposes, as far as the citation correctly traces the lineage of the technologies, i.e. it links related technologies and establishes the precedence of the inventions.

3 The Model

The model is based on the evolutionary view of technical change and patent citation literature outlined in the previous section. According to the evolutionary theory the value of an innovation depends on how well the innovation is embedded in the current

[3]NBER dataset provides no priority dates, however if we consider a cohort of patents issued to the US inventors the difference in the dates is likely to be small.

"technological paradigm", i.e. on the frequency with which the technological knowledge underlying the innovation is utilized in the consequent development of the technology. We also assume that a citation received by a patent documents an instance when the piece of knowledge represented by the patent has been used. Thus, in accordance with the patent citation literature we can state

Assumption 1 *The value of a patented invention is reflected by the number of citations received by the patent: the higher is the number of citations, the more valuable the invention is.*

Furthermore, a successful innovation might work as a "focusing device" for the consequent innovative search. It is reasonable to assume that the impact of innovation on the direction of development of technology depends on the current value of the innovation. The more valuable the innovation is, the more likely the particular piece of technological knowledge represented by the patent will be utilized in consequent innovations, and, as a result, the more valuable it will become. According to Assumption 1 the growing importance of the innovation will be reflected in the frequency of citations the patent will receive. Therefore,

Assumption 2 *The higher is the value of a patented invention, the more likely it is to be used by consequent innovations, the more valuable it will become, and the more citations it will receive.*

The model can be formalized as follows. Consider N patents at time $t = 0$ indexed by i, $i \in \{1, \ldots, N\}$. At time $t = 0$ the patent i has value $v_{i,0}$, reflected by the number of citations it has received, $c_{i,0}$. Each moment in time one citation is made.[4] The probability that the patent i is cited is proportional to the value of the technology the patent i represents, $v_{i,t}$

$$p_{i,t} = \frac{v_{i,t}}{\sum_{j=1}^{N} v_{j,t}} \tag{1}$$

[4]Time in this model is measured in citations. It is not the same as the calendar time.

A citation received by patent i implies that technology i has been used, and reflects the increase in the value of the patent, i.e. $v_{i,t+1} > v_{i,t}$ if $c_{i,t+1} = c_{i,t} + 1$. We consider two "value functions" mapping values into citations: a linear function

$$v(c_{i,t}) = v_0 + c_{i,t}, \tag{2}$$

and non-linear function in the form

$$v(c_{i,t}) = v_0 + {c_{i,t}}^{\alpha}. \tag{3}$$

Inserting (2) into (1) we can rewrite it as

$$\Pr(c_{i,t+1} = c_{i,t} + 1 | c_{1,t}, \ldots, c_{N,t}) = \frac{v_0 + c_{i,t}}{V_0 + t}, \tag{4}$$

where $V_0 = \sum_{j=1}^{N}(v_0 + c_{j,0})$, i.e. the sum of the patent values at time $t = 0$.

For the non-linear value function (3) we have

$$\Pr(c_{i,t+1} = c_{i,t} + 1 | c_{1,t}, \ldots, c_{N,t}) = \frac{v_0 + {c_{i,t}}^{\alpha}}{v_0 N + \sum_{j=1}^{N} {c_{j,t}}^{\alpha}}. \tag{5}$$

Formulas (4) and (5) define stochastic processes that belong to the class of generalized Polya processes (finite case). Early applications of Polya processes in economics go back to the works of an IIASA group in the 1980s (Arthur, Ermoliev, and Kaniovski 1983). The recent revival of the interest in the Polya processes was induced by the rapidly growing literature on the evolution of networks originating from the studies of WWW, but spread into a number of disciplines (physics, ecology, molecular biology, sociology *etc.*).

The generalized Polya process (Chung, Handjani, and Jungreis 2003) can be defined as follows

Definition 1 *For fixed parameters, $\alpha \in R$, $0 \leq p < 1$ and a positive integer $N > 1$, begin with N bins, each containing one* *ball and then introduce balls one at a time. For each new ball, with probability p, create a new bin and place the ball in that bin; with probability* $1 - p$*, place the ball in an existing bin, such that the probability that the ball is placed in a bin is proportional to* k^{α}*, where k is the number of balls in that bin.*

For a *finite Polya process* $p = 0$, i.e. no new bins are created. If $p > 0$ we have an *infinite Polya process*. Parameter α describes the type of feedback: it is said that there is *positive feedback*, if $\alpha > 1$, *negative feedback* if $\alpha < 1$, and *linear feedback* if if $\alpha = 1$. The case of $\alpha = 1$ and $p = 1/2$ is often referred in the literature as the *preferential attachment scheme* (Albert and Barabasi 2002, Barabasi 2002).

The infinite process with different types of the feedback function has been studied extensively in the context of network growth (mostly to explain the distribution of nodal degrees). In particular, the preferential attachment scheme has received a lot of attention. The nodal degrees of the resulting graph, so called scale-free network, are distributed according to a power (Pareto) law, which is often seen as an indication of self-organization and can be observed in nature in a variety of situations (Barabasi 2002). However for our purposes we shall limit our attention to the finite case.

For the finite process with linear feedback ($p = 0$, $\alpha = 1$) such as one defined by (4) it is possible to show that as time (the number of balls) goes to infinity, the proportions of the balls in the bins (*a.s.*, almost surely) approach their limits X_i, $i \in \{1, ..., N\}$, which are distributed uniformly on the simplex $\{(X_1, ..., X_N) : X_i > 0, X_1 + ... + X_N = 1\}$ (Chung, Handjani, and Jungreis 2003, Theorem 2.1).[5] It follows that, the distribution of the proportions has an exponential tail in drastic contrast with the infinite case

[5]The processes defined by equations (4) and (5) are not exactly the same as the process in Definition 1, because, in general, v_0 is not equal to 1. However it does not affect the results for the limit distributions mentioned here.

mentioned above.

The limit distribution of proportions is different for the other types of the feedback. For negative feedback, $\alpha < 1$, balls are distributed equally among bins, i.e. $X_i = 1/N$ for any $i \in \{1, ..., N\}$. If positive feedback is the case as in the process defined by (5), then $X_i = 1$ for one bin and $X_i = 0$ for the other balls, i.e. a "winner takes all" situation (Chung, Handjani, and Jungreis 2003, Theorem 2.2). The latter case is interesting, we may expect to see long and probably fat tails at any finite time.

The results for the *limit* distribution ($t \to \infty$) mentioned above are indicative for what we can expect for the *asymptotic* distribution, $(x_1^t, ..., x_N^t)$ for $t >> N$: in the case of linear feedback the distribution the tail of the distribution is decreasing exponentially, in case of the positive feedback we may expect to see heavier tails. However, a distribution arising at finite time (the number of citations in our case) which we are interested in can be quite different from the limit distribution.

Most studies of Polya processes focus either on the asymptotic distribution ($t >> 1$, when the initial conditions are not important) for the infinite case (Albert and Barabasi 2002, Krapivsky, Redner, and Leyvraz 2000), or on the limit distribution for the finite case and the rate of convergence toward the limit distribution (Bassanini and Dosi 1999). To the best of my knowledge there are no general results concerning the distribution at any given period of time. In the Appendix to this paper using the rate equation describing the evolution of the distribution of the number of the balls in a bin I derive the recursive formula for the distribution at any given t. In case of the linear feedback the solution can be found in a closed form (formula 9). For the case of the non-linear feedback there is no solution in the closed form and therefore we have to rely on the results of simulations.

4 Data

The NBER dataset created by Hall, Jaffe, and Trajtenberg (2001) contains data for all utility patents granted by the USPTO from 1963 to 1999 (about 3 million patents) and all citations made by patents granted from 1975 to 1999 (about 16 million citations).

For my study I have chosen patents applied in 1989, similar to the cohort used in (Silverberg and Verspagen 2004). Since the NBER dataset contains no priority dates, there might be a problem with dating the patented inventions related to the patents applied earlier in other (then the USA) countries, because for these inventions the date of invention (which we are interested in) is likely to distant from the date of application to the USPTO. Moreover, patent citations might have a 'home country bias' i.e. other things equal there may be a bias towards citing the patents granted to US inventors. These problems can be reduced if we restrict our focus to the patents issued to the US inventors (first inventor), assuming that before applying to other Patent Offices US inventors are more likely to apply for a patent at USPTO. In addition, it also helps to avoid a potential complication due to home country bias. This leaves 50,263 patents in the 1989 cohort and 341,365 citations received by these patents from 1990 to (including)1999. The distribution of the citations from 1989-1999 is shown at the left diagram of Figure 1 (blue triangles). The distribution is highly skewed with the large share of patents having received near zero citations. It also has a long and heavy tail. The most cited patent has received 245 citations.

To perform the simulations and fitting we need the initial distribution of patent citations. Left diagram of Figure 1 also shows the distribution of the patent citations in 1989, i.e. citations within the cohort (green circles). In total there are 3,434 citations unequally distributed among the patents. Most of the patents 47,472 (94.4%) have no citation, the maximum number of citations received by a patent is 7.

I also used the data on patents granted has been collected at CESPRI. Similarly, I limit the scope of my study to the patents with the priority year 1989. The EPO cohort of 1989 contains 61,799 patents. Due to differences in the citations practices adopted by EPO and USPTO the average number of citations per patent for the EPO patents is lower than for the USPTO patents (e.g. Breschi and Lissoni 2004), therefore for the EPO cohort of 1989 the total number of citations received is much lower than for the USPTO patents, by the end of 1999 the patents have received 99,684 citations. The most cited patent receiving 82 citations. The number of citations internal to the cohort is 1,591, with the most cited patent having received 6 citations in 1989. Both distributions of patent citations in 1989 and 1999 are shown at the right diagram of Figure 1.

5 Results

Linear feedback For a linear feedback (4) the dynamics of the number of the patents with zero citations, n_0, is (formula (10) in the Appendix)

$$n_{0,t} = N\left(1 + \frac{t}{v_0 N + t_0}\right)^{-v_0},$$

where $t_0 = \sum_{i=1}^{N} c_{i,0}$, i.e. the total number of citations at $t = 0$ (citations within the cohort). Therefore fitting v_0 (the only parameter in the linear model (4)) can be done using only values for $n_{0,t}$ (instead of fitting whole distribution). Fitting the cohort of the USPTO patents gives $v_0 = 1.1$. The results of fitting are shown at Figure 2.

Now, with the initial distribution of patent citations and the estimated value of v_0, using equation 11 we can predict the distribution of the frequency of the number of citations by the end of 1999 ($t = 337,931$). The resulting distribution is shown at Figure 3.[6]

First, note that for the patents with small number of citations, the fit is good (especially if consider that we have only one parameter in the model). It indicates that the function of preferential attachment is close to linear in the region of small number of citations, c, where most of the distribution resides (98% of patents have not more than 30 citations).

Second, the tail of the actual distribution is obviously heavier than the linear model predicts (Figure 3). Indeed, a linear value function generates distributions with exponential tails, while the actual distribution has a Pareto-type shape for large values of c. Thus we might expect that the function of preferential attachment underlying the actual distribution of patent citations is superlinear (convex). The nonlinearity leads to (a) effective "freezing" of the low end of the distribution at large t, because the probability that a patent with small number of citations receives additional citations is falling rapidly (faster than t^{-1}); and (b) depletion of the middle of the distribution, and as a result "fatter" tails.

Non-linear feedback The results for fitting the distribution of USPTO patent citations with simulated distribution, in case of non-linear feedback in the form (5) are shown at the top diagrams of Figure 4. The values of parameters providing the best fit are $v_0 = 2.0$ and $\alpha = 1.26$. As one can see the simulated distribution fits observed distribution very well for most values of c (there is an overshooting at $c = 1$).

Fitting the EPO data using the same procedure gives $v_0 = 1.1$ and $\alpha = 1.3$. A fit (in linear and double logarithmic scale) is shown at the two bottom diagrams of Figure 4. The lower "propensity to cite" of EPO patents mentioned above reveal itself

[6]Fitting the distribution of the EPO citations produces similar results and is not reported here.

in the lower value of parameter v_0. However, and more important, the value of the parameter α describing non-linearity and controlling the shape of the middle range and the tail of the distribution is not that different from the value of α providing the best fit for the USPTO cohort.[7]

Figure 5 shows quantile-quantile plots (QQ-plots) for the simulated distributions vs. observed distributions. If the data falls on 45^0 line of a QQ plot, it means that the distributions underlying the samples of observed and simulated are identical. As one can see from the Figure 5 the quantiles of the simulated distribution for the EPO cohort are lying on the 45^0 line until approximately 40 citations, which is 99.99-percentile of the observed sample (30 citations is the 99.96 percentile). For the USPTO cohort reasonable fit is achieved from 0 to about 150 citations which includes 99.98% of patents (100 citations correspond to 99.92 percentile).

The simulated distributions have fat tails. The tail index of the distribution (the exponent in the Pareto distribution describing the tail) can be estimated using the Hill estimator (Hill 1975)

$$\gamma_{N,k} = \frac{1}{k}\sum_{i=1}^{k}(\ln c_{(i)} - \ln c_{(k+1)}),$$

where $c_{(1)} \geq c_{(2)} \geq \cdots \geq c_{(N)}$ denote order statistics. A Hill plot, the diagram of the inverse of the Hill estimator, $1/\gamma_{N,k}$, vs. the rank of the observation, k, can be used to learn about the tail index and the cut-off value of the tail: the value of $1/\gamma_{N,k}$ at which the plot stabilizes provides an estimate for a tail index, and the value of the corresponding order statistic gives the cut-off value for the tail. The Hill plots for observed and simulated data for the USPTO and the EPO cohorts are shown at Figure 6

[7]Experiments with fitting the whole 1989 cohort of the USPTO patents without selection on the country of the first inventor (96,077 patents), and the cohort of the USPTO patents applied for in 1975 (with citations received from 1975 to 1999) gives slightly different values of v_0, but rather robust on the value of parameter $\alpha \approx 1.2$–1.3.

(left: USPTO data $v_0 = 2.0$, $\alpha = 1.26$, right: EPO data $v_0 = 1.1$, $\alpha = 1.3$). The plot stabilizes at value of α somewhere between 3.0 and 4.0 for both cohorts, i.e. the exponent in the Pareto distribution exceeds 2.0 therefore the distribution has finite mean and variance.

It is interesting to compare our results with the results of Trajtenberg (1990) who used value functions similar to (2) and (3) (with fixed $v_0 = 1$) in the study of patents in CT scanner technology for construction of weighted patent counts (WPC). He found that WPC have significant (cross-time) correlation with the social value of innovations estimated via demand for the new models of scanners. For the non-linear WPC the best results were obtained with $\alpha = 1.3$ and $\alpha = 1.5$. Although our data and approach are rather different, the value of the parameter α providing the best fit falls in the same range.[8]

It is also worth mentioning that Hall, Trajtenberg, and Jaffe (2001) in their study of the impact of company's stock of patents on the market valuation of the company found that the relationship between the market valuation and the number of citations received by the patents owned by the company is non-linear - while the impact of citations is not significant for patents with low number of citations, the magnitude of the effect becomes significant as the number of citations grows. A comparison between the results of the simulations with the linear (4) and non-linear (4) models lead to the conclusion that the value function $v(c)$ is non-linear.

6 Discussion

Let us turn to limitations and possible extensions of the model. First, I would like to elaborate on the problem of "intrinsic val-

[8] According to (Trajtenberg 1990) the difference in correlation between $\alpha = 1.3$ and $\alpha = 1.2$ are only several percentage points.

ues" of inventions and their relationship with the productivity gain. Then I will make a brief remark on the omission of the variation mechanism. At the end of the section I will discuss the use of technological fields conveyed by patent classification to describe the path-dependent process of technical change.

Many models concerning technical change emphasise that the main characteristic of an innovation is the increase of the productivity which this given innovation offer once it is adopted. From this perspective the "intrinsic value" of an innovation is already predetermined and mostly (if not solely) depends on the productivity gain which is assumed to be distributed according to some probability law. Therefore, there is no question about the process that govern the dynamics of the values, but an inquiry about the distribution of patent values can be safely reduced to the question about the exact form of the distribution of the productivity gains. This view is in sharp contrast with the model proposed in this paper. Indeed, the model assumes that all innovations are "born" equal, and it is selection which following the evolutionary theory of technical change that generates the differences in the values.

Surely, the value of an innovation depends on many factors besides the direction of technical change, including productivity gain, but also demand for the new product, advances in science and so on. Acknowledging the importance of factors other than the productivity gain, I shall remark on the latter, primarily because as has already been mentioned, most models take it as a premise.

It is certainly true, that if we consider a range of alternative technologies which were developed some time in the past to address a certain design problem, then a technology dominating the market at present is more likely to be more efficient. However, to conclude that at the time of invention it had higher "intrinsic value" related to its efficiency in comparison with other alternatives might be an unjustified stretch.

From the history of technology we know many examples when with respect to productivity a newly born technology had been inferior to the existing one and only incremental improvements over a long period of time let these technologies prevail.[9] The reason why innovators spent their time on working with seemingly inferior technologies is that these technologies, while being less productive, offered a basis for a technological breakthrough, and reasons why these technologies have surpassed the alternative designs are rooted in the complementarities between technologies (Rosenberg 1982).

Let me illustrate this point with the results from the percolation model of Silverberg and Verspagen (2005) mentioned in Section 2. Consider the technology space in a form of two-dimensional lattice, with the vertical dimension representing productivity (with more productive technologies at the top and less productive ones in the lower part of the lattice). A technology becomes available only when at least one adjacent technology is already in use. Initially agents know only technologies at the bottom. Growth in such a model is the process of percolation from the bottom to the top of the lattice. If all technologies had the same probability to be discovered, growth would occur along a line(s) connecting bottom and top of the lattice. However, linear growth is prevented by a random "landscape": each point of the lattice representing a certain technology has different probability to be discovered. Consider the extreme case when the probability of discovering technology which is the next on the "linear expansion path" is zero. If the search were constrained to the area of the technology space just above the most efficient current technology, then the technological progress would cease forever. Nevertheless, it proceeds due to the agents who keep searching in areas of less productive technologies which at the

[9] We can only guess how many potentially valuable technologies never made it through.

end results in finding a "side-path". What is important in our context is that most productive technologies does not have to be the most promising, and once the growth is stuck it is less productive technology that can make a difference, if it can lead out of the "deadlock".

Therefore judgments about "intrinsic value" *ex-post*, conditional on the success or failure of technologies might oversimplify the complex picture of technical change. The *ex-post* value as measured by patent citations is the result of a path-dependent process and reflects different factors such as productivity gains, demand conditions, complementarity with other innovations, and some (mis)fortune. It hardly can be reduced to the productivity gain alone.

Having stated that, I nevertheless shall note that the approach presented in this paper can be (and should be) improved. Emphasising importance of path-dependency in the evolution of innovation values, I have omitted the fact that the innovations in consideration (patent from the 1989 cohorts in our case) were not born in vacuum, but also were a consequent development of some earlier technologies. Once we assume that the current value of an innovation depends on how well the innovation is embedded in the current "technological paradigm" reflected by the number of "forward citations", we can make one step further and assume that the initial value of an innovation i, $v_{i,0}$, depends on how well the innovation was embedded in the paradigm at the time of the invention, and hypothesize that "backward citations" i.e. citations made by the patent convey some information about it.[10]

Another limitation of the study presented here is that focusing on the patents from one cohort I have restricted the scope of the analysis to the *selection mechanism*, omitting the other main component of the evolutionary process - the *variation mechanism*,

[10]Deng et al. (1999) report that the number of backward citations is indicative for the value of innovations.

the mechanism that generates new technologies and leads to the discontinuities in the technological trajectories. Some features of the variation can be traced in the patent data. For example, "aging" of patents, i.e. the decline in the rate of receiving citations with time may be a reflection of the shifts in technological trajectories in different subfields of the technology. Moreover, the process of formation of new technologies might be reflected in the patent classification, the point to which we will come late in this section. However, to conduct the study of variation mechanism based on patent citations, one has to find a solution for the problem of "benchmarking" mentioned in the Section 2, because such a study cannot be done without inter-cohort comparisons.

There is also a problem related to the fact that patents which we consider do not belong to the same technological field. It raises two issues. First, it is well known that different industries have different "propensity to patent". Therefore, it may be that the actual distribution of patent citations is sheer reflection of this fact rather than a result of the path-dependent process similar to one proposed in this paper. Underlying this question is a suspicion that if we restrict our analysis to one technology then the shape of the distribution of patent citations may be quite different from the shape of the overall distribution. On the other hand, if the model is correct, then at the level of a patent class (or related patent classes) we expect to see the distribution of patent citations similar to one on the level of the whole cohort.

First, note that there are, indeed, differences in the average number of citation in different classes that can be attributed to differences in the "propensity to patent" across industries: distributions of patent citations from different patent classes occupies different range of distribution. For example, the USPTO patents related to data processing (USPTO patent classes 700-714) are on average more heavily cited through 1989-1999. However, inspection of the distribution of the patent citations within the same patent class (or related patent classes) reveals the picture similar to one we have seen at the level of the whole cohort.

Figure 8 shows the distribution of patent citations for USPTO patents related to data processing (classes 700-714) applied in 1989. As one can see it is strikingly resembles distribution of patent citations for the whole cohort of 1989 (Figure 1): it is highly skewed and has a long tail. The best fit is obtained with parameters $v_0 = 4.0$ and $\alpha = 1.26$. The value $v_0 = 4.0$ is twice as high as the fitting value for the whole cohort ($v_0 = 2.0$), which reflects the fact that through 1989-99 the patents in data processing have been cited more frequently than patents from other classes. However the parameter $\alpha = 1.26$ controlling the shape of the centre and top end of the distribution is the same as for the whole population, which implies that the functional form of the value function (except the shift of intercept) is the same as for the whole cohort.

The second issue concerning the technological field is related to the boundaries within which a company can reallocate its R&D activities responding to shifts in technological trajectories. The model assumes that exploring opportunities opened by previous inventions a company chooses to search in the area of the technology space that is "popular". It does not contradict economic intuition, when the reallocation is to take place within the same technological field, however if it is to be done across industries, then, at least, one need an explanation: afterall, why a company producing, say, domestic appliance should be investing in nanotechnology?

There are two reasons which could partially justify assumptions of the model. First, most patents is assigned to large diversified companies (such as IBM), or industrial conglomerates involved in innovative activities in many R&D intensive sectors, and reallocation of innovative activities by such company corresponds to the reallocation of R&D budgets within a company or a conglomerate. Second, highly cited patents are likely to be related to General Purpose Technologies (GPTs) (Hall and Trajtenberg 2004), the technologies which penetrate most sectors of the economy. Therefore companies in different sectors may be involved in adapting a GPT to their needs, and it is reflected in the observed pattern of patent citations.

However reasonable it seems, the model needs to be modified to take into account the fact that many economic entities performing R&D are specialized in certain sectors. For example, we may change the model in such a way that the selection of the technology from which to start the R&D search, and as a result which patent will be cited is done in two steps. At the first step, the sector in which a new patent is to be taken will be selected, and then a particular technology (represented by the corresponding patent), which is to be used as a starting point will be chosen on the basis of the values of technologies in this field.

This modification of the model, in turn, opens a question of how to choose the technological field for a new patent. For that we can use information about the patent classes (as a representation of separate fields). The problems concerning the use of patent classifications are discussed below. If the selection of the technological field is done in the way similar to the one which we use in our model of patent citations, i.e. the probability of a patent to appear in a certain sector is a function of the number of patents in this field in comparison to the whole stock of patents in all patent classes, then we would have some kind of a "nested" Polya process.[11]

Our model states that the more R&D have been done in a certain field (resulting in more patented inventions), the more R&D effort will be directed to this field in close future. Translating the

[11] If both value functions at both stages of selection are linear then the two-stage process is observationally equivalent (i.e. distribution of citations is the same) to one in simple one-stage citations model (4).

assumptions of the model into the context of patent classes we would expect that the larger is the share of a patent class in the stock of patents, the higher is the probability that the next patent will appear in this patent class. [12] To check if this intuition is correct, in Figure 7 (left diagram) I plot the share of a USPTO patent class (417 patent classes) in the stock of the patents applied in each of the years 1989-1999 ($n_{i,t}/n_t$) against the share of the patent class in the stock of the patents from 1963 to the respective year ($N_{i,t}/N_t$). Take for example, 932 patents applied in 1989 in the patent class 29 "Metal working"($n_{29,1989} = 932$). In 1989 the number of patent applications to all classes, n_{1989}, was 96,077, it gives us the share of the class 29 in the stock of all patents applied in 1989, $n_{29,1989}/n_{1989} = 932/96,077 \approx 0.0097$. Now, from 1963 to (not including) 1989 there were 20,323 patent applications in the class 29 ($N_{29,1989} = 20,323$). The total stock of all patents from 1963 to 1989 is $N_{1989} = 1,878,708$, therefore the share of the class 29in the total stock of patent applications in 1989-1999 is $N_{29,1989}/N_{1989} = 20,323/1,878,708 \approx 0.0108$. As one can see from Figure 7 the observations reside close to 45^0 line.[13]

Figure 7 (diagram on the right) shows the evolution of several patent classes (circles mark points in 1989). Generally we can divide all patent classes into three broad categories according to their growth patterns: mature technologies with stable shares, old technologies with shrinking shares, and new technologies with growing shares. As one can see from Figure 8 patent class 29 "Metal Working" containing 29,858 patents, or 1.02% of all patents from 1963-1999, has a stable share in the total patent stock, and the rate of arrival of new patented inventions in this class is proportional to its share. Classes 435 ("Chemistry:

[12] This maps exactly into the model but at a higher level of aggregation.

[13] The slope in double logarithmic scale slightly exceeds 1.0 indicating a superlinear relationship between the variables, akin to the non-linear model discussed earlier.

Molecular Biology and Microbiology", 30,257 patents or 1.03%), 436 ("Chemistry: Analytical and Immunological Testing", 6,998 patents or 0.24), and 514 ("Drug, Bio-Affecting and Body Treating Compositions", 58,062 patents or 1.99%) are rapidly expanding through 1963-1999. At the same time the shares of class 12 ("Boot and Shoe Making", 1,251 patents or 0.04%) and class 66 ("Textiles: Knitting", 3,846 or 0.13%) are going down. These developments in patenting activities are not necessarily related to the current shares of the corresponding industries in the total output, but we might expect that they reflect long-term trends in the economy.

The distribution of the USPTO patents in the patent classes and the results of fitting with lognormal and gamma distributions is shown at Figure 9. The results of fitting, goodness-of-fit statistics, and comparison of observed and estimated quantiles are reported in Table 1–3. Although goodness-of-fit for both lognormal and gamma distributions are reasonable ($p < 0.01$), the Gamma distribution is marginally better. Notice, that formula (9) derived for the finite Polya process with linear feedback predicts the distribution close to gamma distribution.[14].

Coming back to our research question, this information could be used for building a two-stage model as described above. However, there are also some difficulties here related to patent classification. First, there is inherent ambiguity to which industry (and related patent class) an invention should be assigned. An invention may be assigned to a class on the basis of the industry from which it originated, the industry that will produce the new product, or the industry which will use it (Griliches 1990). As a result, developments of the same technology may be divided among different patent classes. Another problem, also related to the interconnections between patent classes, is that a patent class hardly can represent a whole industry or a sector. There-

[14] to see this one can use the representation of Beta function through Gamma functions

fore to proceed with a two-stage model one has to decide how to aggregate classes into industries.[15]

Notice also, that the patent classification is evolving with the technology: new patent (sub)classes are being added, reclassified etc. Relying on the current classification one necessarily has some kind of bias when making judgement about past inventions. For example, if one is to use current classification on some fine level, say, 6-digit subclasses, then one might be surprised by discovering that a number of subclasses were unpopulated back in the 70s. These subclasses have been added as the corresponding technology came into being. In terms of the model the situation with addition of patent (sub)classes should be modelled with an infinite Polya process, where new bins are constantly being added. It is also possible to use more general class of the processes, Yule processes (Yule 1925), earlier applied in evolutionary biology. An advantage of the Yule process is that it not only accounts for the addition of the new classes, but also traces the lineage of the evolutionary tree.[16]

7 Conclusions

Taking the prospective of the evolutionary theory of technical change I argued that the value of innovations depends on the direction of technical change, and patent citations reflect the path-dependencies in the development of technology. Innovations well embedded in the current "technological paradigm" have higher value and play a role of "focusing devices" shaping direction of innovation search. As a result the higher is the value of an innovation the more likely it is to be used as a starting point for consequent innovations and the more valuable it will become.

[15] Several ways to do it are outlined in (Hall and Trajtenberg 2004).

[16] For a description of the Yule processes and the properties of the distributions generated in these processes see (Newman 2005).

To formalize this argument I proposed a simple model based on generalized Polya processes with linear and non-linear (positive) feedback.

Using NBER and CESPRI data on patents granted by the USPTO and EPO I have shown that the model does produce distribution of patent citations close to the observed one. The model with a linear feedback predicts correctly the distribution of patent citation for patents with relatively small number of citations (about 95% of the distribution). The model with non-linear feedback predicts the distribution of patent citations correctly for the whole range of citations. Simulated distributions do have fat tails (as the observed distributions). Interestingly, the exponent in the feedback function providing the best fit is in the same range as the estimate of Trajtenberg (1990) obtained in a different context.

Aknowledgements

I am grateful to Stefano Breschi, Robin Cowan, Clemens Kool, Francesco Lissoni, Franco Malerba, Pierre Mohnen, Gerald Silverberg, Bart Verspagen, Lorenzo Zirulia, and participants of MERIT/UNU-INTECH seminar for their helpful comments. Part of the research has been carried out while visiting CESPRI (Università Bocconi).

References

Aghion, P., and P. Howitt (1992): “A Model of Growth through Creative Destruction,” *Econometrica*, 60(2), 323–51.

Albert, R., and A. Barabasi (2002): “Statistical mechanics of complex networks,” *Reviews of Modern Physics*, 74, 47–97.

Arthur, W. B., Y. M. Ermoliev, and Y. M. Kaniovski (1983): “A Generalized Urn Problem and its Applications,” *Kibernetika*, 19, 49–56.

Balconi, M., S. Breschi, and F. Lissoni (2004): “Networks of Inventors and the Role of Academia: An Exploration of Italian Patent Data,” *Research Policy*, 33(1), 127–45.

Barabasi, A. (2002): *Linked. The New Science of Networks.* Perseus, New York.

Bassanini, A. P., and G. Dosi (1999): "Competing Technologies, Technological Monopolies, and the Rate of Convergence to a Stable Market Structures," *mimeo.*

Birchenhall, C. (1995): "Modular Technical Change and Genetic Algorithms," *Computational Economics*, 8(3), 233–53.

Breschi, S., and F. Lissoni (2004): "Knowledge networks from patent data: Methodological issues and research targets," *CESPRI Working Papers*, 150.

Chung, F., S. Handjani, and D. Jungreis (2003): "Generalizations of Polyas urn problem," *Annals of Combinatorics*, 7, 141–154.

Cooper, B. (2000): "Modelling Research and Development: How Do Firms Solve Design Problems?," *Journal of Evolutionary Economics*, 10(4), 395–413.

Dawid, H. (1999): *Adaptive learning by genetic algorithms: Analytical results and applications to economic models.* Springer, Heidelberg and New York, second edition edn.

Dosi, G. (1982): "Technical paradigms and technological trajectories: A suggested interpretation of the determinants of technical change," *Research Policy*, II(3), 147162.

Frenken, K. (2004): "History, state and prospects of evolutionary models of technical change: a review with special emphasis on complexity theory," in *Complexity Focus*, ed. by J. Casti. Exystence, Turin.

Frenken, K., and A. Nuvolari (2004): "The early development of the steam engine: An evolutionary interpretation using complexity theory," *Industrial and Corporate Change*, 13(2), 419–450.

Graff, G. D. (2003): "Observing Technological Trajectories in Patent Data: Empirical Methods to Study the Emergence and Growth of New Technologies," *American Journal of Agricultural Economics*, 85(5), 1266–74.

Griliches, Z. (1981): "Market value, R&D, and patents," *Economics Letters*, 7(2), 183–187.

——— (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28(4), 1661–1707.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001): "Market Value and Patent Citations: A First Look," *NBER Working Papers*, 8498.

Hall, B. H., and M. Trajtenberg (2004): "Uncovering GPTs with Patent Data," *NBER Working Papers*, 10901.

Hall, B. H., M. Trajtenberg, and A. B. Jaffe (2001): “The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools,” *NBER Working Papers*, 8498.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999): “Citation Frequency and the Value of Patented Inventions,” *Review of Economics and Statistics*, 81(3), 511–15.

Harhoff, D., F. M. Scherer, and K. Vopel (1997): “Exploring the tail of patent value distributions,” *Center for European Economic Research Discussion Paper*, 97-30.

——— (2003): “Citations, Family Size, Opposition and the Value of Patent Rights,” *Research Policy*, 32(8), 1343–63.

Henderson, R., A. B. Jaffe, and M. Trajtenberg (1998): “Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988,” *Review of Economics and Statistics*, 80(1), 119–27.

Hill, B. (1975): “A simple general approach to inference about the tail of a distribution,” *Annals of Statistics*, 3(5), 1163–1173.

Huang, Z., H. Chen, A. Yip, G. Ng, F. Guo, Z.-K. Chen, and M. C. Roco (2003): “Longitudinal Patent Analysis for Nanoscale Science and Engineering: Country, Institution and Technology Field,” *Journal of Nanoparticle Research*, 5, 333–363.

Jaffe, A. B., and M. Trajtenberg (2002): *Patents, citations, and innovations: A window on the knowledge economy.* MIT Press, Cambridge and London.

Jaffe, A. B., M. Trajtenberg, and M. S. Fogarty (2000): “Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors,” *American Economic Review*, 90(2), 215–218.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993): “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *Quarterly Journal of Economics*, 108(3), 577–98.

Kauffman, S. A. (1993): *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, Oxford and New York.

Krapivsky, P. L., S. Redner, and F. Leyvraz (2000): “Connectivity of Growing Random Networks,” *Phys. Rev. Lett.*, 85(21), 4629–4632.

Lanjouw, J. O., A. Pakes, and J. Putnam (1998): “How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data,” *Journal of Industrial Economics*, 46(4), 405–32.

Nelson, R., and S. Winter (1977): "In Search of a Useful Theory of Innovation," *Research Policy*, 6, 36–76.

Nelson, R. R., and S. Winter (1982): *An Evolutionary Theory of Economic Change*. The Belknap Press of Harvard University, London.

Newman, M. (2005): "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, (forthcoming).

Pakes, A. (1985): "On Patents, R&D, and the Stock Market Rate of Return," *Journal of Political Economy*, 93(2), 390–409.

Pakes, A., and M. Schankerman (1984): "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources," in *R&D, Patents, and Productivity*, ed. by Z. Griliches, NBER Conference Report, pp. 73–88. University of Chicago Press, Chicago and London.

Pakes, A. S. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54(4), 755–84.

Pier, H., K. Rost, T. Teichert, and I. von Wartburg (2003): "H04Q-007 - Deconstructing the Technological Blob of Mobile Telecommunication. An Application of Social Network Analysis for Categorizing Technological Fields," in *ZEW Workshop "Empirical Innovation and Patenting"*, Mannheim, Germany. March 14-15.

Rosenberg, N. (1969): "The Direction of Technological Change: Inducement Mechanisms and Focusing Devices," *Economic Development and Cultural Change*, 18(1), 1–24.

——— (1974): "Science, Invention and Economic Growth," *Economic Journal*, 84(333), 90–108.

——— (1982): *Exploring the black box: Technology, economics, and history.* Cambridge University Press, Cambridge; New York.

Sahal, D. (1981): *Patterns of Technological Innovation.* Addison-Wesley, New York.

Schankerman, M., and A. Pakes (1986): "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period," *Economic Journal*, 96(384), 1052–76.

Scherer, F. M., and D. Harhoff (2000): "Technology Policy for a World of Skew-Distributed Outcomes," *Research Policy*, 29(4-5), 559–66.

Silverberg, G., and B. Verspagen (2004): "The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance," *MERIT/Infonomics research memoranda*, 2004-021, forthcoming in the Journal of Econometrics.

——— (2005): "A percolation model of innovation in complex technology spaces," *Journal of Economic Dynamics and Control*, 29(1-2), 225–244.

Trajtenberg, M. (1990): "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, 21(1), 172–87.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997): "University versus Corporate Patents: A Window on the Basicness of Invention," *Economics of Innovation and New Technology*, 5(1), 19–50.

Yule, G. (1925): "A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis.," *Philos. Trans. R. Soc. London B*, 213, 2187.

Appendix

Consider N patents indexed by $i \in \{1, \ldots, N\}$. Let $c_{i,t}$ denote the number of citations the patent i received at time t. The distribution of the patents at time t is $n(t) = (n_{0,t}, n_{1,t}, \ldots)$, where $n_{c,t}$ is the number of patents cited c times. Further assume that the probability of a patent i to be cited is proportional to the current value of the patented invention which is related to the number of citations the patent received, c_i, as $v_{i,t} = v(c_{i,t})$. Then the *rate equation*[17] describing the dynamics of the system is

$$n_{c,t+1} - n_{c,t} = \frac{v_{c-1}}{\sum_{j=0}^{c_{max}(t)} v_j n_{j,t}} n_{c-1,t} - \frac{v_c}{\sum_{j=0}^{c_{max}(t)} v_j n_{j,t}} n_{c,t}, \tag{6}$$

where $v_c = v(c)$, and $c_{max}(t)$ is the maximum number of citations a patent may have received by time t ($c_{max}(t) = c_{max}(0) + t$). The first term in RHS (6) describes an increase in the number of the patents with c citations due to citing of a patent with $(c-1)$ citations. The second term in (6) is a loss term due to citing of a patent with c citations. Since there is no arrival of new patents (the number of patents in the cohort is fixed), for patents with no citations ($c = 0$) the first term is equal to zero, i.e.

$$n_{0,t+1} - n_{0,t} = -\frac{v_0}{\sum_{j=0}^{c_{max}(t)} v_j n_{j,t}} n_{0,t}. \tag{7}$$

At time $t = 0$ the distribution of patents is

$$n(0) = (n_0^0, \ldots, n_{c_{max}(0)}^0, 0, \ldots).$$

Equations (6), (7), and the initial condition define the evolution of the system.

[17] deterministic equation that describes evolution of expected values

Linear feedback In case of the linear value (preferential attachment) function (2) we can find a closed-form solution. The sum in the denominator of the equation (6) is

$$\sum_{j=0}^{c_{max}(t)} (v_0 + j) n_{j,t} = V_0 + t, \tag{8}$$

where V_0 is the sum of the values of the patented inventions at $t = 0$, i.e. $V_0 = v_{1,0} + \dots v_{N,0}$. Note that

$$V_0 = \sum_{j=0}^{c_{max}(0)} (v_0 + j) = v_0 N + t_0,$$

where t_0 is the total number of citations at $t = 0$, i.e. citations within the cohort.

For the sake of simplicity assume that at $t = 0$ patents have no citations ($n_0^0 = N$, and $n_c^0 = 0\ \forall c \neq 0$, i.e. $t_0 = 0$). The evolution of the system can be analyzed drawing a binomial tree: each node ot the tree represents $n_{c,t}$, the probability of transition from node (c, t) to the node $(c + 1, t + 1)$ is equal to the probability of citing a patent with c citations at time t. Given (6) it is clear that the probability to arrive to node (c, t) from the origin $(0, 0)$ does not depend on the path chosen.[18] Taking into account this fact the further derivation of the distribution $n(t)$ becomes trivial, for $c \geq 1$ the result is

$$n_{c,t} = N \binom{t}{c} \frac{\prod_{i=0}^{c-1}(v_0 + i) \prod_{i=0}^{t-c-1}(V_0 - v_0 + i)}{\prod_{i=0}^{t-1}(V_0 + i)} = \\ = N \binom{t}{c} \frac{B(V_0 + t, v_0 + c)}{B(V_0, v_0)}, \tag{9}$$

where $B(x, y)$ is a Legendre beta-function. For patents with no citations $(c = 0)$

$$n_{0,t} = N \prod_{i=0}^{t-1} \left(1 - \frac{v_0}{V_0 + i}\right).$$

Provided that $V_0 = v_0 N >> v_0$ we can approximate it as

$$\ln n_{0,t} = \ln N + \sum_{i=0}^{t-1} \ln \left(1 - \frac{v_0}{V_0 + i}\right) \approx \ln N - \sum_{i=0}^{t-1} \ln \frac{v_0}{V_0 + i} \approx \\ \approx \ln N - v_0 \ln \left(1 + \frac{t}{V_0}\right),$$

[18] no such luck for a non-linear feedback...

therefore

$$n_{0,t} = N\left(1 + \frac{t}{V_0}\right)^{-v_0}. \qquad (10)$$

Since the linear model has only one parameter (v_0), fitting of this parameter can be done with the equation (10) using only data on $n_{0,t}$.

The equation (9) can be rewritten in a more convenient recursive form as following ($c \geq 1$)

$$\ln n_{c+1,t} = \ln n_{c,t} + \ln \frac{(t-c)(v_0+c)}{(c+1)(t-c+V_0-v_0-1)}. \qquad (11)$$

The system of (10) and (11) provides us with the distribution of patent citations at any t.

The linearity of the process allows us to extend the solution to the general case of initial conditions. The resulting distribution is simply a superposition of the distributions, which one can derive from the analysis of binomial trees with origin in $(c, 0)$ where $0 \leq c \leq c_{max}(0)$.

Figures and Tables

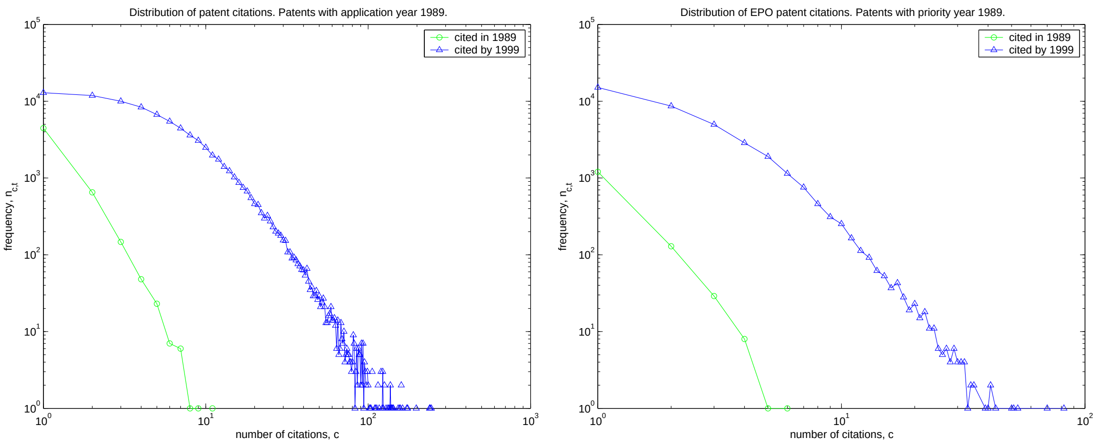


Figure 1: The distributions of patent citations. **Left:**USPTO cohort, application year 1989. **Right:** EPO cohort priority year 1989.

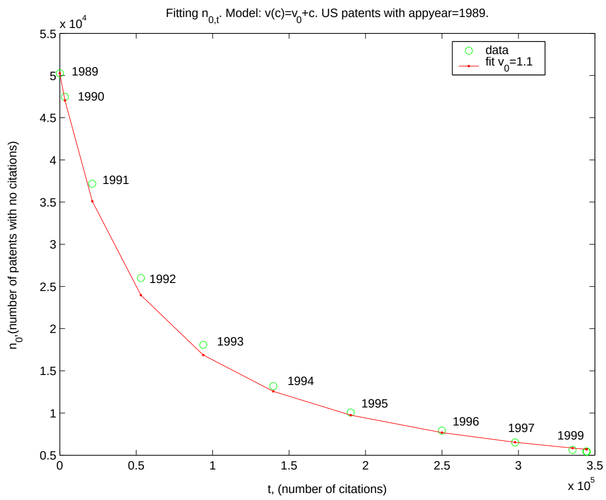


Figure 2: Fit $n_{0,t}$ for the USPTO cohort with linear model: $v_0 = 1.1$.

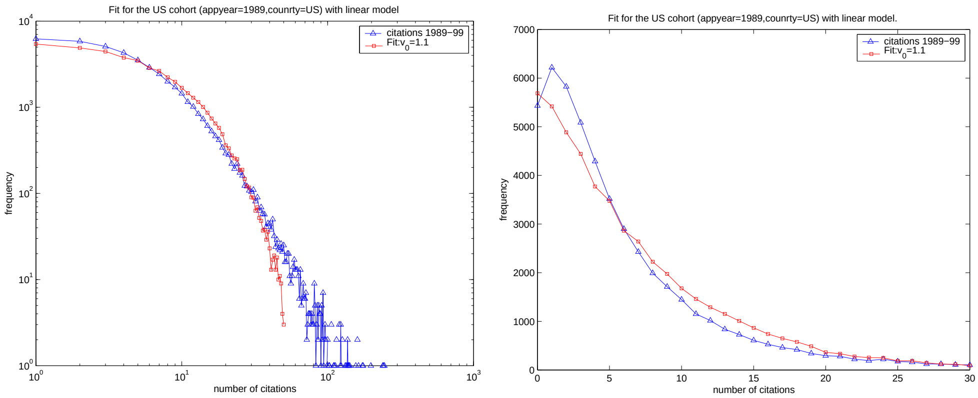


Figure 3: Fit of the distribution of patent citations for the USPTO cohort with linear model: $v_0 = 1.1$.

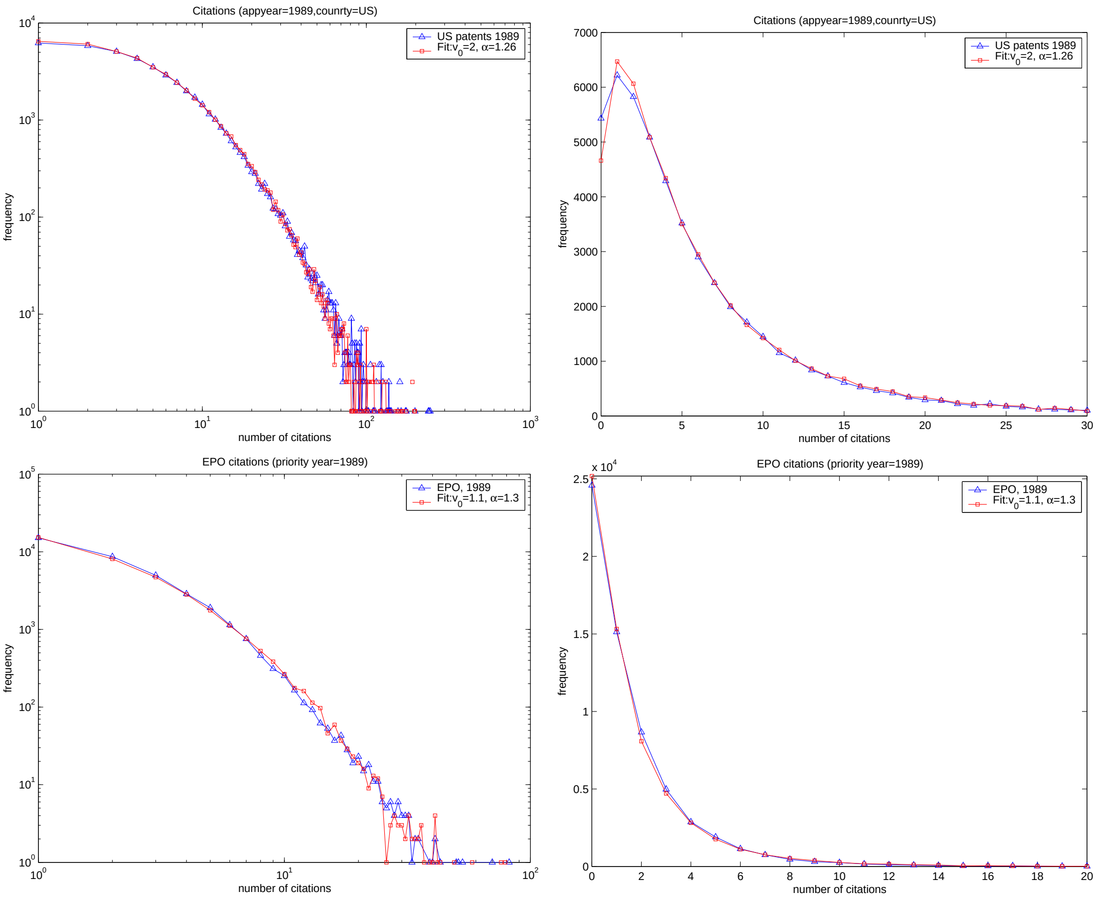
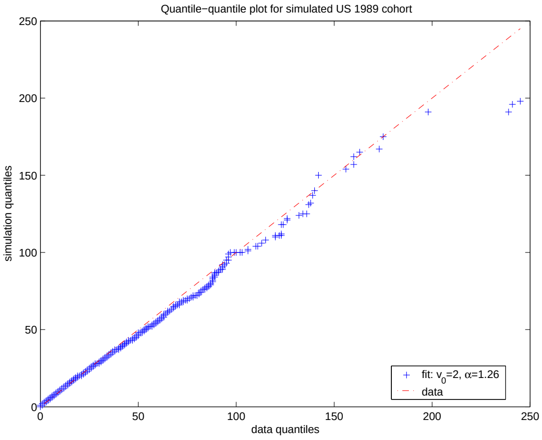


Figure 4: Fit of the distribution with non-linear model. **Top:** USPTO cohort $v_0 = 2.0$, $\alpha = 1.26$. **Bottom:** EPO cohort $v_0 = 1.1$, $\alpha = 1.3$.



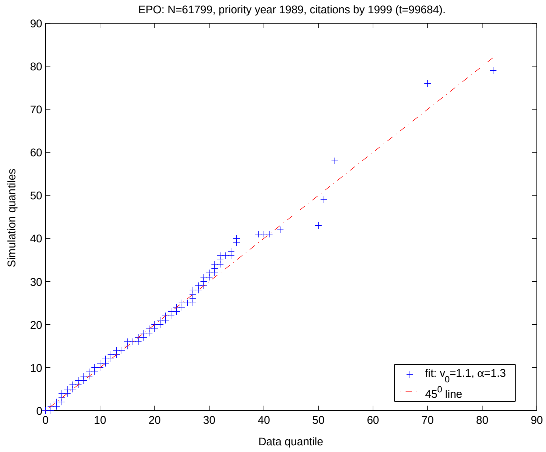
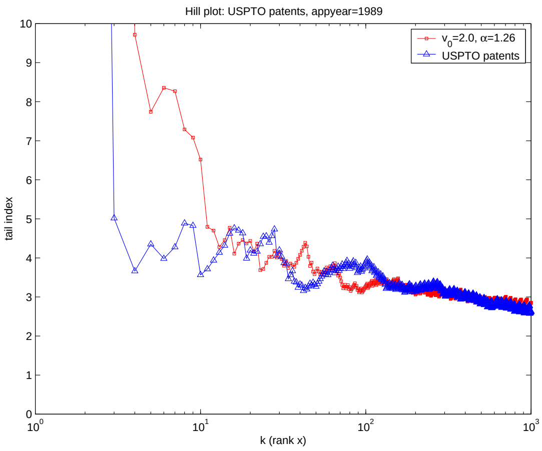


Figure 5: QQ-plots for the simulated distributions, non-linear model. **Left:**USPTO cohort $v_0 = 2.0$, $\alpha = 1.26$. **Right:** EPO cohort $v_0 = 1.1$, $\alpha = 1.3$.



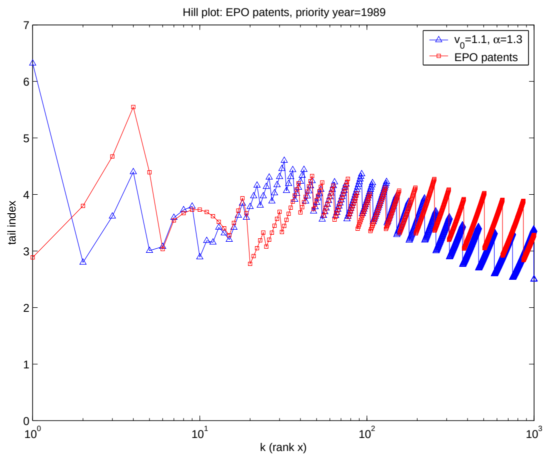


Figure 6: Hill plot for the observed and the simulated distributions, non-linear model. **Left:**USPTO cohort $v_0 = 2.0$, $\alpha = 1.26$. **Right:** EPO cohort $v_0 = 1.1$, $\alpha = 1.3$.

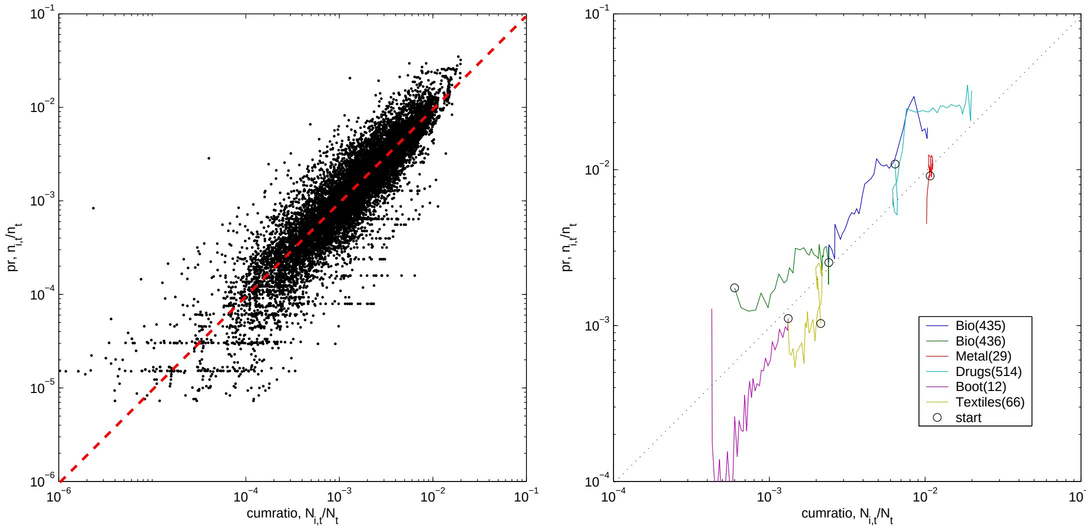


Figure 7: **Left:** Share of a USPTO patent class in the stock of patents applied in year t (from 1990 to 1999), $n_{i,t}/n_t$ vs. the share of a patent class in the whole stock of patents applied since 1963,$N_{i,t}/N_t$. **Right:** Diagram $n_{i,t}/n_t$ vs. $N_{i,t}/N_t$ for several USPTO patent classes. Circles mark positions in 1989. Classes: 29-Metal working, 435-Chemistry: Molecular Biology and Microbiology, 436-Chemistry: Analytical and Immunological Testing, 514-Drug, Bio-Affecting and Body Treating Compositions, 12-Boot and Shoe Making, 66-Textiles: Knitting.

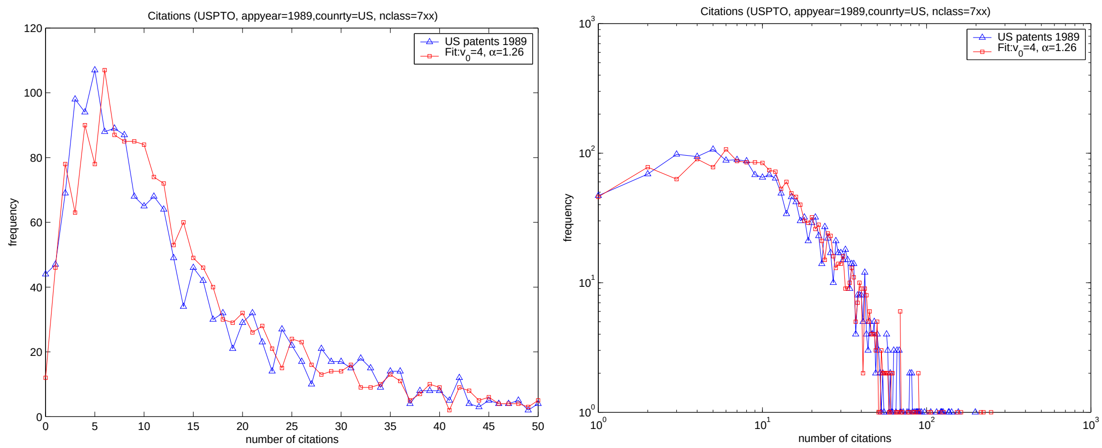


Figure 8: The distribution of patent citations for patent classes 700-714 (Data processing) with application year 1989, and the USA as a country of the first inventor.

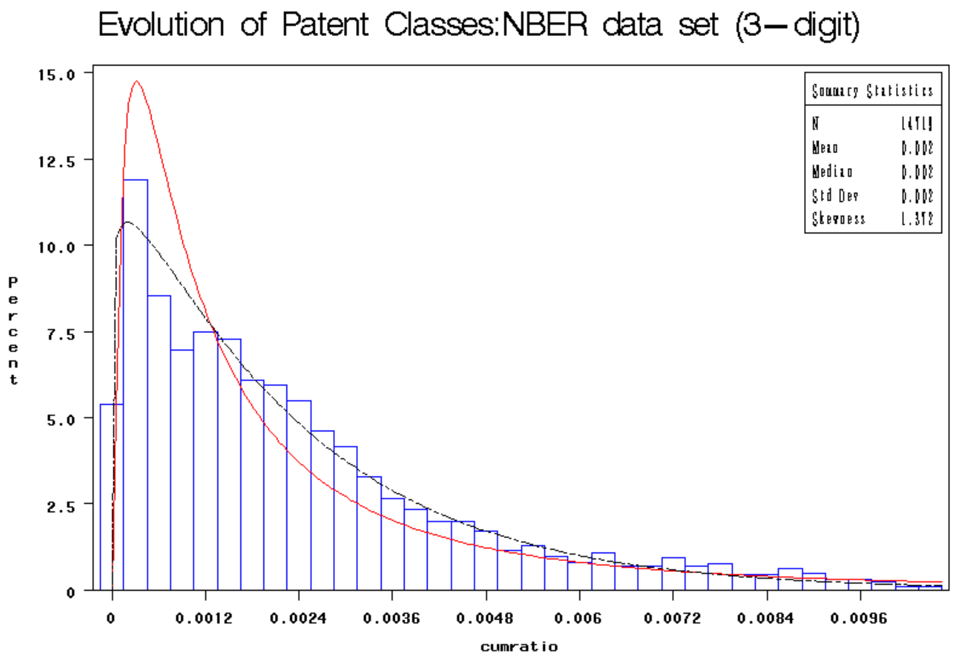


Figure 9: The distribution of USPTO patent class sizes. Patents granted from 1963-1999. Fit: red solid line - LogNormal, black dash-dot line - Gamma.

Table 1: Fitting for the Distribution of Patent Class Sizes (NBER dataset, 3-digit) with Lognormal and Gamma Distributions.

Parameters	Lognormal		Gamma	
	Symbol	Estimate	Symbol	Estimate
Threshold	θ	0	θ	0
Scale	ζ	-6.57174	σ	0.002175
Shape	σ	1.221316	α	1.088043
Mean	0.00295			
Std Dev	0.005475			

Table 2: Goodness-of-Fit Tests for the Distribution of Patent Class Sizes.

Test (Statistic)		Lognormal		Gamma	
		Statistic	p Value	Statistic	p Value
Kolmogorov-Smirnov	(D)	0.091486	< 0.010	0.0234421	< 0.001
Cramer-von Mises	(W-Sq)	35.711508	< 0.005	1.6382987	< 0.001
Anderson-Darling	(A-Sq)	204.837144	< 0.005	10.8396215	< 0.001

Table 3: Quantiles for for the Observed and Estimated Distributions of the Patent Class Sizes.

Percent	Observed	Lognormal	Gamma
1.0	0.00003	0.00008	0.00003
5.0	0.00014	0.00019	0.00015
10.0	0.00026	0.00029	0.00029
25.0	0.00073	0.00061	0.00074
50.0	0.00178	0.00140	0.00169
75.0	0.00326	0.00319	0.00328
90.0	0.00548	0.00669	0.00534
95.0	0.00717	0.01043	0.00688
99.0	0.00919	0.02398	0.01045