Pricing Patents through Citations

Fernando Leiva B.*

September 2005

Abstract

This paper provides formal treatment to the idea of patenting as a form of market stealing between R&D firms. It extends the creative destruction literature by allowing innovations to build off each other forming a network of ideas. Patent citations keep track of this network. The theory maps the distribution of productivities in the development of new ideas onto the distribution of patent values through patent citations. If productivities are drawn from a Pareto-Levy distribution then the distribution of patent values also falls within this class. The theory is then applied to data on US patent citations. Model-based valuations support the assumption of Pareto distributed productivities. As an added feature, model-based valuations outperform citation counts (the traditional measure) as a proxy for patent values.

*University of Iowa. Email: fernando-leiva@uiowa.edu. I would like to thank faculty and students at the University of Iowa and the University of Rochester and seminar participants at the University of British Columbia and the IIES at Stockholm for their helpful comments. I'd also like to thank Luis Fornero for taking time off to give me computational advice and Yigit Saglam for excellent research assistance.

1 Introduction

In conventional models of R&D, firms face the choice of time, effort, or money to be spent on enhancing technology in order to maximize profits from a given invention. In sharp contrast, true "research and development "undertaken by firms is mostly an activity where technological improvement is reduced to a minimum. Firms have routinized the process of R&D to the point where they can easily target a given level of technological improvement they want to achieve in a certain period of time. Thus, true R&D expenditures concentrate on developing the market for an invention, as much and as quickly as possible.[1] What is more, the bulk of applied inventions consists of smart recombinations of prior inventions, aided (only sometimes) by the latest basic knowledge. A clear example is the case of pharmaceuticals. New drugs are repeatedly found to be a minor change to already existing drugs. Often called "me-too " drugs, these new drugs are designed specifically to (at least partially) capture the market of the old ones.[2] Technological change, if at all, is often just a by-product and not the means to profit maximization.

Another important aspect of the innovation process is the fact that inventions build off each other forming a network-like structure of recombinations of past ideas. Patented innovations are a well documented example: Patents are the result of R&D, they also build off each other forming a network and, most importantly, this network can be traced by using patent citations.

Patent citations relate new developments to the previous works upon which these developments are built (as do, for example, academic citations). The observable result is a network of links and links of links and so on, the patent citations network. The difference with citations between academic publications is that in this case the network establishes links between marketed innovations that have been granted a monopoly in the production process of the good(s) associated to them. Thus, patent citations provide more information than just a prior-to-new-art link. Turning back to the pharmaceuticals example, the new me-too drug would cite the older ones, acknowledging them as prior art. But because the new drug was designed to steal markets from the older ones then the citations take on another meaning: They become a crude form of book-keeping of the market stealing activity taking place.

In line with the above arguments, I model R&D as a process where firms only spend resources to increase the size of the market for a given invention. In other words, technological change is not the driving force. I model it as a zero sum game in the sense that for every market captured by some new patent, there is an old one that loses it due to its inferior technology.

In order to deliver network-like features, in the model market stealing through innovation is only partial. This means that patents do not necessarily become obsolete when new technology arrives. Then, by assuming that patent citations keep track of the market stealing, the model can price any individual patent through the tree of citations that it generates. I test this implication by applying it to actual data on US patent citations. I find that the values delivered by the model outperform citations counts (the proxy for patent values used by the literature) both in terms

[1]For a very good descriptive account see Baumol 2004.
[2]For more on the case of pharmaceuticals, see for example Angell 2004.

of the overall distribution of values and their precision.

The key result in terms of the distribution of patent values is that the distribution of values delivered by the model is strikingly close to the Pareto-Levy distribution. This result is consistent with Harhoff et al. (1997) and Sanders, Rossman and Harris (1958). Both are surveys that ask patent holders to estimate the value of their inventions. In both cases the Pareto-Levy and the Log-Normal distributions provide the the best fit to their data. On the other hand, raw citation counts produce distributions that are far from the heavy tailed Pareto-Levy distribution.

The precision of valuations is measured via renewals (a set of fees that the patent holder must pay in order to avoid early expiration). The key result here is that highly valued patents (by the model) are renewed to full term more often than highly cited patents, indicating that model-based valuations are more precise than simple counts.

There is an extensive applied literature on patent citations and their use as indicators of patent quality.[3] However, the methodology applied hasn't gone beyond counting the number of citations a patent receives to proxy for patent quality. Counting citations avoids taking a stand on questions such as how and why citations arise and what type of information they convey. Focusing on simple counts deliberately ignores any added information within the network of citations. Thus, an additional feature of this paper is that it shows how by adding structure -a theory on the role of citations- one can extract more information from the citations data.

The concept of market stealing through innovation is not new. Models of creative destruction are abundant in the literature. In them, typically, a firm does R&D, which, in turn leads to an expansion in technology. Then, through a patent, the firm appropriates the benefits of the new technology. The appearance of new technology renders previous ideas obsolete. Innovations are either good-specific (quality ladders) or they span the whole set of goods (TFP changing).[4] But these models overlook a subtle, but very important feature of the innovation process, precisely the fact that inventions build off each other forming a network of links between past and present ideas. This has created a divorce between the theory of R&D and the applied work involving the use of data on patent citations.

The theory developed here explicitly describes the R&D process in terms of a sequence of forward-looking and profit-maximizing decisions where citations arise as a natural by-product of interlinking innovations. It takes the book-keeping argument to an extreme by assuming that all citations represent some degree of market stealing. What the theory is doing is simply to map the space of ideas into the space of patent values (or profits) through citations. With only a few additional assumptions, the theory can be taken to the data, as suggested above. Thus, findings within the data like, for example, a Pareto-Levy distribution of patent values can be mapped back into the space of ideas in the form of a Pareto Levy distribution of productivities in the development of ideas.

[3]See for example Harhoff et.al (1997, 1999 & 1999a) Trajtenberg (1990) and Hall et.al (2000, 2001).

[4]They may also consist of entirely new goods, unrelated to the past (variety growth). Classic examples come from the endogenous growth literature. They include Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), to name a few.

Section 2 presents the R&D model. It is capable of creating a network of citations portraying the actual U.S. patent citations network. It also ties together the distribution of productivities in the development of ideas and the distribution of patent values under the same class (the Pareto-Levy). In section 3 the model is applied to the data on US patent citations. Results support the initial assumption on the distribution of productivities and compares model-based valuations to citation counts. Finally, section 4 presents conclusions and suggests possible extensions.

2 The Model

There is a continuum of goods in the economy. A patent allows a firm to collect monopoly profits on a set of goods with positive measure. Firms go through two stages before being granted a patent: Invention and development. Invention is the process of studying old technology and creating new one. Development is the stage where applications to the new technology are found in the form of more goods covered by the patent.

During the invention stage every patent studied gives the inventor an opportunity to create new technology. If the firm fails to create new technology then a new patent has to be studied. The process goes on until the firm eventually succeeds. Once it does, the firm enhances the technology to the point where it reaches the patentability standards of the patent office and then it goes on to the development stage. All patents studied during invention are considered prior art. They are consequently cited by the inventor when applying for the new patent. In an intermediate step before development there is a give and take process with the patent office where the firm spends the least possible resources to meet the standards of patentability. This process is stylized as reaching at least a given productivity jump with respect to the previous technology in the production of the good (as in most quality ladder models). Profit-maximizing firms always choose this lower bound as their target, with the costs of achieving it being the same for all firms. Moreover, the jump is normalized to be such that profits per unit of time equal one for any good subject to a patent monopoly. This avoids any growth accounting and focuses the attention on market stealing (the development stage) exclusively.[5]

Development is the process by which applications to the new technology are found. Strictly speaking, it is the process by which the new technology is applied successfully to enhance the productivity in the production of more and more goods. The firm hires labor to find applications for the new technology. Every time a new market is successfully captured, some old patent may be losing that market. A new patent is assumed to steal markets from those patents that were studied during invention. Since studied patents are cited, then a patent captures markets from its cited patents. This assumption gives citations the double role mentioned above: they connect prior art with new technological developments, but also constitute a paper trail of the market stealing effect of R&D. Heterogeneity is introduced by assuming that with some inventions it is easier to find applications than with others.

[5]For a theoretical approach that features both market stealing and endogenous growth see Leiva 2004

The remainder of this section briefly describes the timing of events and each of the two stages an R&D firm goes through in more detail. It then shows how the profits of a single firm evolve, once the firm has chosen its initial market share. This is used to state the firm's maximization problem. Next, it solves for the symmetric equilibrium. In a symmetric equilibrium, the distribution of patent values is stationary because all firm act and expect other firms to act in the same way. Finally, it shows the conditions under which the stationary distribution of patent values is of the Pareto-Levy class.[6]

2.1 Timing

At any given period i, there is only one firm doing R&D.[7] The firm first goes through the invention stage. The outcome is a set of cited patents $M(i)$, where $\#M(i) = m_i$. Then the firm draws its idiosyncratic productivity in market stealing B_i. Finally it decides how much labor L_D to hire for the development of the patent. The patent is then granted and the firm begins production of the goods associated with its patent (profits are normalized to one per unit of time). Then, for the next T periods the firm keeps producing until eventually it loses its entire market share. This happens because future patents improve on the goods produced by the firm. The market share of patent i is gradually reduced until it disappears either because new patents arriving wipe it out or because the patent expires (at time $T + i$).[8]

2.2 Invention and Research Stage

The invention stage is a mechanical stage. A patent is studied and then an experiment is run. If the experiment succeeds, new technology is created and the firm goes on to the next stage. If the experiment fails, a new patent is studied and a new experiment is run. This process goes on until an experiment succeeds. All patents studied are cited. The number of citations made by any patent i is denoted $m(i)$ and the set of patents cited by patent i is denoted $M(i)$. Experiments are assumed to be i.i.d.[9] The probability of success is denoted p. The probability that patent i

[6]This distribution is of particular interest because it is the one delivered by applying the model to the data.

[7]Thus, i denotes both a given time period and the patent that was created at that time period. There is no gain in assuming that many firms can produce a patent at the same time. It would complicate the algebra and impose unnecessary restrictions on citations. Instead, it is simpler to think of a time period as very short in length.

[8]When a patent with a positive market share expires, the markets for all the remaining goods covered by the patent become perfectly competitive. Thus, profits for those goods fall to zero until some new patent improves on them. This means that at any point in time there will be a set of goods protected by patent monopolies and another (its complement) subject to perfect competition. The measure of each set will fluctuate depending on the market share of the incoming patent, the market share of the expiring patent and the status of each of the patents cited by the incoming patent (whether these patents still retain a positive market share or not). How the measures of these two sets fluctuate has no influence on the expected profits of any given patent because the likelihood of being cited depends exclusively on the current market share of a patent.

[9]Altering this assumption would only modify the distribution of the number of citations made by a patent.

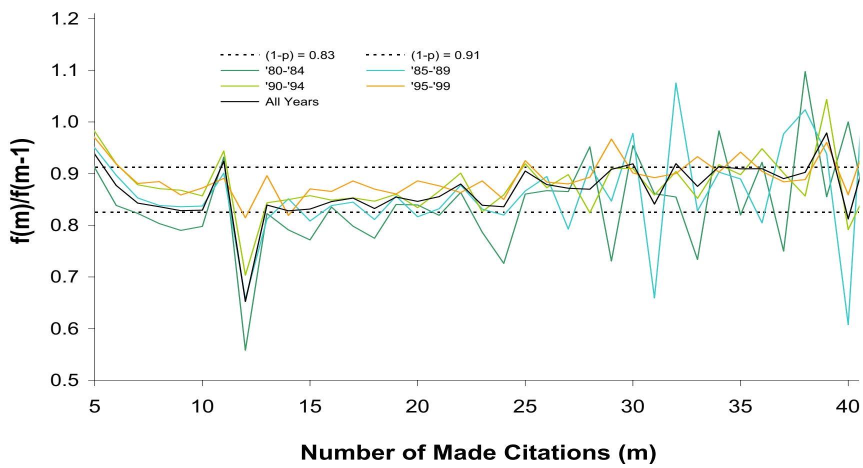

Figure 1: Ratio of made citation frequencies

cites m patents is therefore:

$$\Pr\left(m\left(i\right)=m\right)\equiv p_m=p\left(1-p\right)^m.$$

To give a sense of how realistic this assumption is notice the following:

$$\forall m, \qquad \frac{p_m}{p_{m-1}}=\left(1-p\right).$$

The plot of $\frac{f_m}{f_{m-1}}$ in Figure 2.1 shows the ratio of frequencies of the number of made citations by five cohorts of patents. The thicker line depicts the ratio for all patents pooled together.[10] The ratio of frequencies fluctuates between 0.83 and 0.91 for the range of made citations between 5 and 40. 95% of patents make a number of citations within that range. This would be consistent with assuming the probability of success p to be within the range of 0.17 and 0.09.[11]

[10]The data was taken from Hall, B., Jaffe, A., Trajtenberg, M.(2001). It includes *all* U.S. patents applied for between 1980 and 1999.

[11]For patents with less than five made citations the relationship does not hold; the ratio is higher. Interpreted in terms of the model it simply suggests that the initial experiments have a lower likelihood of success. If incorporated into the model results do not change.

Patents to be studied are assumed to be randomly selected. The probability of any patent being selected for studying is equal to the market share of that patent. If more than one patent has to be studied then subsequent independent draws determine new patents to be studied. Random selection (non-directed invention) is a convenient though unrealistic assumption. Together with independence, it helps keep the invention stage simple and tractable.[12] In equilibrium, firms are indifferent between choosing any two patents to study. Thus, random selection of studied patents results in a consistent model where the evolution of revenues and, thus, a symmetric equilibrium are well defined.[13]

2.3 Development Stage

A patent is assumed to be applicable to a set of goods within the interval $[0, 1]$. Profits for any good within the interval are normalized to one per unit of time. The market share (and size) of any patent i is equal to the measure of goods it is applied to. Since patents have different market shares at different points in time, the market share of patent i at time t is denoted $\mu_t(i)$. One patent arrives per unit of time. Thus, patent i also refers to the patent created at time i. The initial market share of a patent $\mu_i(i)$ is denoted simply $\mu(i)$.

In order to have positive profits, a new patent must capture a positive measure of markets (goods). It steals markets from the patents that were studied during the invention stage. To do this the firm invests resources (labor L_D) into finding these markets. The process of market stealing is governed by the following function:

$$\mu(i) = B_i {L_D}^{\nu}, \quad \nu \in (0, 1).$$

There is an idiosyncratic productivity shock B_i. It is revealed after the invention stage but before the firm chooses the amount of labor into development. Marginal returns to market stealing are decreasing in the only input, labor.[14]

With the choice of the amount of labor, a firm's initial market size is determined. After that, the patent is granted and the firm starts the production of the goods it captured. In subsequent periods, as long as the patent is not cited by the incoming ones, it does not lose any market share and profits remain the same. Whenever it is cited it loses part (or all) of its market share and never recovers it. When

[12] It also delivers a simple and tractable way of applying the model to the data.

[13] A more realistic assumption would be to define a sense of proximity in some interesting way and then to correlate subsequent draws to previous ones. There is a vast literature, largely unrelated to economics that studies networks through basic assumptions on link formation. It has a strong resemblance to the process of citations formation described here. A recent survey by Jackson (2005) provides a comprehensive account of this literature and its application to economic problems. Moreover, he suggests a specific example, a hybrid model of network formation (a combination of random graph theory with preferential attachment models) that, appropriately modified, could be used to enrich the invention process developed here.

[14] This guarantees that the invention and the development stage remain independent. Notice that neither the number of patents studied nor the value of the patents studied during invention determine the market stealing potential of a firm. In other words, studying an additional patent or a more profitable one during invention has no effect on the productivity in market stealing during development.

the patent expires the firm loses its monopoly in the production of whatever the remaining markets are at that point. Profits drop to zero.

Even though the choice of labor into R&D is a one-time decision, it involves an evaluation of expected future profits. The evolution of profits depends on the actions of future firms doing R&D. Two important assumptions are made on the process of market stealing, that have implications on the way profits evolve:

i *Market stealing is limited to patents studied during invention stage.*

ii. *Total market stolen is divided equally among cited patents.*

Assumption *(i)* allows for citations to become a paper trail of market stealing activity. It also allows a firm to easily compute the likelihood of being cited by future patents: the firm only needs to know the random selection process that occurs during invention to obtain the likelihood of being cited in the future.

In the event of being cited by a future patent, assumption *(ii)* allows the firm to easily compute the likelihood of losing a certain market share. All the firm needs to know is the (stationary) distribution of initial market shares and the distribution of the number of cited patents by incoming patents.[15] Assumption *(ii)* can be obtained instead as a result by adding a little more structure to the problem: A set-up that is symmetric in cited patents. In the appendix, an alternative formulation for the market stealing production function is developed. There, a firm has the option of stealing a different amount of market share from different citations but chooses not to.

2.4 Evolution of Profits

In order to state the maximization problem of the firm we need to show how profits evolve once the initial market share of a patent is chosen. Since a patent expires after T years, then:

$$\Pi\left[\mu\left(i\right)\right] = \sum_{t=i}^{T+i} \frac{\mu_t\left(i\right)}{\left(1+r\right)^{t-i}}.$$

The problem patenting firms face every period does not depend on time. It does not even depend on any state variables from past periods. The decision is intra-temporal. Still, it is not trivial to obtain the expected profits from selecting any given initial market share. Once the initial market share for a patent is selected the firm has no control over the evolution of the patent's market share. It will depend on whether the new patents cite it (and therefore reduce its market share) or not.

Any sequence of market shares between creation and termination dates is possible, as long as it is a non-increasing function of time and the initial market share is

[15] An exception to this assumption occurs when the market share a firm would like to steal from a citation is greater than the market share available to steal. In that case the firm is assumed to complete the difference with goods from expired patents.

the one selected by the firm. Formally, a sequence of market shares for, say patent s is defined as:

$$\boldsymbol{\mu}_s \equiv \{\mu_t\}_{t=s}^{T+s},$$

where $\mu_s = \mu(s)$ is the only choice of the firm.

Formally, to calculate the expected value we need to know the density function of all possible sequences $\boldsymbol{\mu}_s$. The likelihood of any given sequence $\boldsymbol{\mu}_s$ will be denoted $g(\boldsymbol{\mu}_s)$ and is given by the following expression:

$$g(\boldsymbol{\mu}_s) = \prod_{t=1+s}^{T+s}\left[\sum_{m=0}^{M} p_m \left\{g_1(m, \mu_{t-1}, \mu_t) + g_2(m, \mu_{t-1}, \mu_t) + g_3(m, \mu_{t-1}, \mu_t)\right\}\right],$$

where:

$$\begin{aligned}
g_1(m, \mu_{t-1}, \mu_t) &= [1 - I(\mu_{t-1} - \mu_t)][1 - \mu_{t-1}]^m, \\
g_2(m, \mu_{t-1}, \mu_t) &= I(\mu_t)\, I(\mu_{t-1} - \mu_t)\left[1 - (1 - \mu_{t-1})^m\right] f_E(m[\mu_{t-1} - \mu_t]), \\
g_3(m, \mu_{t-1}, \mu_t) &= [1 - I(\mu_t)]\left[1 - (1 - \mu_{t-1})^m\right][1 - F_E(m[\mu_{t-1} - \mu_t])],
\end{aligned}$$

and where

$$\begin{aligned}
I(x) &= \begin{cases} 1 \text{ if } x > 0, \\ 0 \text{ if } x = 0. \end{cases} \\
F_E(x) &\equiv \text{Expected distribution function of initial market shares of a new patent.}
\end{aligned}$$

The interpretation of $g(\boldsymbol{\mu}_s)$, going from left to right, is the following: It is the product of T independently distributed random events, each of which accounts for the appearance of a new patent that can either cite patent s or not. For any given new patent that comes in, the likelihood of citing patent s depends on how many citations the new patent makes. That is why for any new patent we have to consider the summation over all possible number of citations made m (mutually exclusive events). The likelihood of making any number of citations is captured by the probability function p_m.

Within a given sequence, say for instance that the time t patent produces m citations. Then if the sequence $\boldsymbol{\mu}_s$ is such that $\mu_{t-1} - \mu_t = 0$, we are interested in the likelihood of our patent not being cited (since it did not lose market share between $t-1$ and t). This is captured by $g_1(m, \mu_{t-1}, \mu_t)$. If, on the other hand $\mu_{t-1} - \mu_t > 0$ then our patent lost market share between $t-1$ and t. The functions $g_2(m, \mu_{t-1}, \mu_t)$ and $g_3(m, \mu_{t-1}, \mu_t)$ capture this event. If the remaining market share is still positive after the citation by patent t then the likelihood of this event is given by $g_2(m, \mu_{t-1}, \mu_t)$. It is equal to the product of two terms. The first one is the likelihood of patent s being cited by patent t, that is: $1 - (1 - \mu_{t-1})^m$. The second one is the likelihood of the initial market share of patent t being exactly such that it steals $\mu_{t-1} - \mu_t$ from patent s, that is: $f_E(m[\mu_{t-1} - \mu_t])$.[16] We multiply the terms because these are independent events. Finally, for the case when the remaining market share is zero ($\mu_t = 0$) then $g_3(m, \mu_{t-1}, \mu_t)$ captures the added fact that the market share of patent t can also be higher than what is needed to

[16] In a symmetric equilibrium, $f_E(\mu) = f(\mu)$, that is, the expected distribution of initial market shares is equivalent to the actual distribution of initial market shares of patents.

steal $\mu_{t-1} - \mu_t$ from patent s. That is why we use $1 - F_E\left(m\left[\mu_{t-1} - \mu_t\right]\right)$ instead of $f_E\left(m\left[\mu_{t-1} - \mu_t\right]\right)$.

The expected value of future profits is, thus:

$$\Pi_E\left[\mu\left(s\right)\right] \equiv E\left\{\Pi\left[\mu\left(s\right)\right]\right\} = \int_{\forall \boldsymbol{\mu}_s} g\left(\boldsymbol{\mu}_s\right)\left[\sum_{t=s}^{T+s} \frac{\mu_t}{\left(1+r\right)^{t-s}}\right] d\boldsymbol{\mu}_s.$$

2.5 Maximization Problem of the Firm

Expected net profits $\pi\left(i\right)$ for patent i are the expected present value of the flow of future profits minus the one time labor into development costs and invention costs. Invention costs are just some negligible fixed cost c per cited patent. Formally:

$$\pi\left(i\right) \equiv \max_{L_D}\left\{\Pi_E\left(B_i {L_D}^{\nu}\right) - wL_D - cm_i\right\}.$$

The FOC is:

$$\Pi_E'\left(\mu\left(i\right)\right)\nu B_i {L_D}^{\nu-1} - w = 0.$$

We want to have $\mu\left(i\right)$ as a function of B_i exclusively. Denote this function $\mu\left(B\right)$, after dropping subscript i. By replacing L_D for $\left(\mu B^{-1}\right)^{\frac{1}{\nu}}$ in the above:

$$\Pi_E'\left(\mu\right)\nu B^{\frac{1}{\nu}} \mu^{\frac{\nu-1}{\nu}} = w.$$

This implicitly defines $\mu\left(B\right)$.

Upon assuming a given distribution of productivities $P\left(B\right)$, the shape of $\mu\left(B\right)$ given by the above equation will determine the shape of the equilibrium distribution of initial market shares $F\left(\mu\right)$. Once we have this distribution we can obtain the distribution of patent prices (the expected present value of future profits), $H\left(\Pi_E\right)$, by using the function of expected profits with respect to initial market share $\Pi_E\left(\mu\right)$.

2.6 Symmetric Equilibrium

Exogenous to this sector of the economy are wages w, the interest rate r, the probability of success during invention p, the distribution of productivities during development (the c.d.f denoted $P\left(B\right)$ and the p.d.f. denoted $p\left(B\right)$), the concavity parameter of the market development production function ν, and the time until expiration T. All other functions and distributions are obtained endogenously in equilibrium. They are the probability function of made citations p_m, the optimal choice of initial market share as a function of productivity in market stealing $\mu\left(B\right)$, the distribution of expected and actual initial market shares $F_E\left(\mu\right)$ and $F\left(\mu\right)$ (density functions $f_E\left(\mu\right)$ and $f\left(\mu\right)$), the function of expected profits as a function of initial market share $\Pi_E\left(\mu\right)$, and the distribution of patent values $H\left(\Pi_E\right)$ (density function $h\left(\Pi_E\right)$). Assuming well behaved functions (namely, that $\mu\left(B\right)$ and

$\Pi_E(\mu)$ are continuous and strictly increasing and thus, invertible) then a symmetric equilibrium is defined as follows:

Given (exogenous) initial conditions $< w, r, p, \nu, P(B), T >$, a symmetric equilibrium in this economy consists of $< p_m, \mu(B), F_E(\mu), F(\mu), \Pi_E(\mu), H(\Pi_E) >$ where:

1. $p_m = (1-p)^m p.$

2. $\mu(B)$ solves $\Pi'_E(\mu)\, \nu B^{\frac{1}{\nu}} \mu^{\frac{1-\nu}{\nu}} = w.$

3. $f_E(\mu) = f(\mu) = p(B(\mu))\, B'(\mu), \quad$ where $B(\mu)$ is the inverse of $\mu(B)$.

4. $\Pi_E(\mu) = \int_{\forall \boldsymbol{\mu}} g(\boldsymbol{\mu}) \left[\sum_{t=0}^{T} \frac{\mu_t}{(1+r)^{t-i}}\right] d\boldsymbol{\mu},$

   where

$$g(\boldsymbol{\mu}) = \prod_{t=1}^{T} \left[\sum_{m=0}^{M} p_m \left\{g_1(m, \mu_{t-1}, \mu_t) + g_2(m, \mu_{t-1}, \mu_t) + g_3(m, \mu_{t-1}, \mu_t)\right\}\right],$$

   and

$$\begin{aligned} g_1(m, \mu_{t-1}, \mu_t) &= [1 - I(\mu_{t-1} - \mu_t)]\, [1 - \mu_{t-1}]^m . \\ g_2(m, \mu_{t-1}, \mu_t) &= I(\mu_t)\, I(\mu_{t-1} - \mu_t)\, [1 - (1 - \mu_{t-1})^m]\, f(m[\mu_{t-1} - \mu_t]) . \\ g_3(m, \mu_{t-1}, \mu_t) &= [1 - I(\mu_t)]\, [1 - (1 - \mu_{t-1})^m]\, [1 - F(m[\mu_{t-1} - \mu_t])] . \end{aligned}$$

5. $h(\Pi_E) = f[\mu(\Pi_E)]\, \mu'(\Pi_E), \quad$ where $\mu(\Pi_E)$ is the inverse of $\Pi_E(\mu)$.

Notice that knowing the functional form of $\Pi_E(\mu)$ would avoid having to solve for a fixed point problem. The object of interest is the distribution of patent prices $H(\Pi_E)$. Eq-5 states that to obtain $H(\Pi_E)$, the distribution of initial market shares (Eq-3) and the function relating patent prices to initial market shares $\Pi_E(\mu)$ (Eq-4) are required. To obtain the distribution of initial market shares, the distribution of productivities $P(B)$ (exogenous) and the function relating productivity and initial market share $\mu(B)$ (Eq-2) are required. Finally, to obtain the latter the function relating patent prices to initial market shares $\Pi_E(\mu)$ (Eq-4) is required. Simply put, once we find the functional form of $\Pi_E(\mu)$ we can first solve for Eq-2, then for Eq-3 and then for Eq-5.

2.7 Linear Pricing Function

This section provides strong evidence supporting the conjecture of a linear pricing function $\Pi_E(\mu)$. For restricted parameter values, this result is proven by using a limiting argument. For the general case, a numerical method for obtaining $\Pi_E(\mu)$ is developed and used. The results confirm the conjecture of a linear pricing function. Finally, all equilibrium equations are solved for, assuming the exogenous distribution of productivities is of the Pareto-Levy class. The result is that both the initial market share and patent value distributions are also of the Pareto-Levy class.

Conjecture. $\Pi_E(\mu) = A\mu$.

Under specific assumptions and through a limiting argument this conjecture can be proven to be true. The proof requires the average market share, or at least the average market stealing to be infinitesimally small. This is a natural assumption, given that the number of non-expired patents in the U.S. is greater than 2,000,000 (implying market shares in the order of a thousandth of one percent).

Proposition 1. *If patent protection lasts only two periods ($T = 1$) and the average value of a patent is infinitesimally small then* $\Pi_E(\mu) = A\mu$.

The proof is found in the appendix. The intuition is as follows:

Assume $T = 1$ and, for simplicity that the amount of market share stolen by a future patent is fixed at, say, x (in reality it is a random variable that the firm cannot influence). Then

$$\Pi(\mu) = \mu + \frac{\mu'}{1+r},$$

where next period's market share μ' is either μ when not cited or $\mu - x$ when cited. This means

$$E[\mu'] = (1-\mu)\mu + \mu(\mu - x) = \mu(1-x).$$

This is a linear function of μ, so expected profits $\Pi_E(\mu)$ are also a linear function of μ.

Even though we have proven linearity for a particular case, linearity still remains a conjecture for a more general case. We can calculate $\Pi_E(\mu)$ numerically by guessing an initial function with some intuitive appeal and then verifying if on average it is a good indicator of the present value of benefits by simulating the model repeatedly. Once we find $\Pi_E(\mu)$, obtaining $g(\boldsymbol{\mu})$ becomes irrelevant and obtaining the other equilibrium functions is trivial. The simulation algorithm proposed is the following:

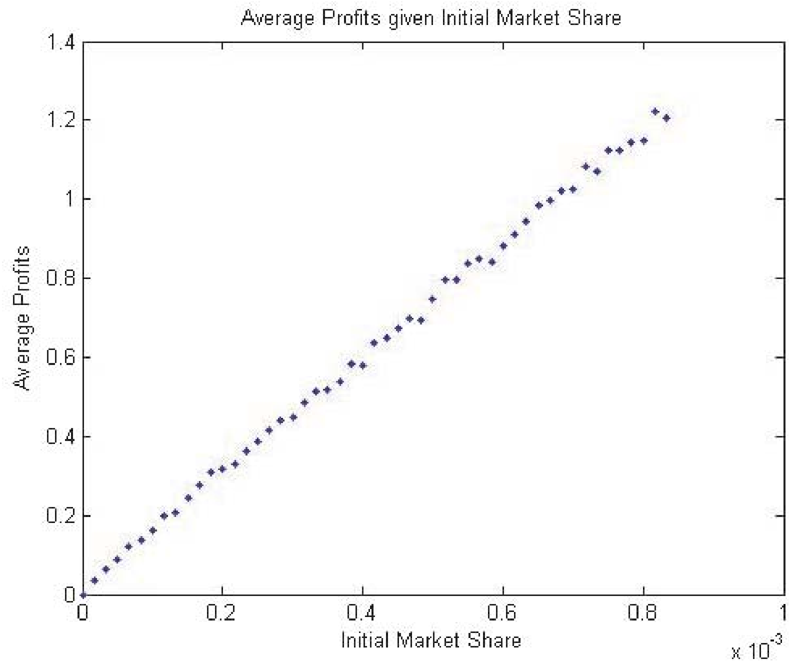

Figure 2: Average profits given initial market share

i. *Guess* $\Pi_E(\mu)$*using an nth order polynomial.*

ii. *From the R&D firms FOC obtain* $\mu(B)$.

iii. *Simulate the model for a large number* $\overline{N}$ *of patents.*

iv. *Calculate* (μ_i, Π_i), $\forall i \in \{\underline{N}, \overline{N}\}$.

v. *Regress* Π_i *on the nth order polynomial in* μ_i.

vi. *Update parameters. Continue until convergence criterion is met.*

The results from simulating the model are shown below. They are robust to varying the exogenous parameters and the exogenous distribution of productivities. Figure 2.2 has initial market share in the x axis and mean profits in the y axis. The mean is taken over a sample of 200 patents of equal initial market share, for every possible initial market share.

Graphically there is barely a difference between fitting the results to a 1st, 3rd or 5th degree polynomial. The R-Squared is 0.9994 for the linear fit and does not improve before the 4th decimal when fitting it to a 3rd or 5th degree polynomial. We can safely assume that the conjecture holds for expected initial market shares in the order of one hundredth of one percent.

With a linear pricing function the problem simplifies to the extent that we can find the exact distribution of initial market shares, as well as the distribution of patent values (present value of the flow of future profits) and productivities in development. Moreover, the following proposition links all three distributions under the same class:

Proposition 2. *If researchers draw productivities from a Pareto distribution then the equilibrium distribution of expected profits is Pareto.*

The proof is very simple; it is also found in the appendix. It relies on the choice of a power function as the market stealing production function.

If the assumption on the distribution of productivities is true, then when taken to the data the actual distribution of patent values should be Pareto. This is not true for other distributions. For example, a uniform distribution of productivities would yield a different distribution of patent values and that distribution would not be supported by the data.[17]

3 Retrieving the value of a patent

We have explored R&D and patenting from the point of view of a profit maximizing firm. The by-product of such behavior is a network of citations. This network not only informs about prior art in the development of new marketable innovations but is also a paper trail of market stealing activity. It is possible for the external observer to re-trace this activity. What follows is the derivation of an algorithm that delivers patents' initial market share by following the trail of stolen markets.[18]

3.1 Market stealing process

Recall that almost any pair of patents (i, j), where patent i cites j, are related by the following equation:

$$\mu(i, j) = \frac{\mu(i)}{m_i},$$

the exception being when patent i is the last citation received by patent j. In that case the relationship is given by:

$$\mu(i, j) = \min\left\{\frac{\mu(i)}{m_i}, \mu_{i-1}(j)\right\},$$

the reason being that the market share of patent i may be lower than the desired level of market stealing by the firm producing patent i.

[17] What should not be interpreted is that a Pareto Distribution is the only possible distribution that can be mapped back and forth from values to productivities. In principle, other distributions could have the same feature. But, as will be seen in the applied section, the data does not support any distribution. I conjecture that if at all, the data can only be consistent with other heavy tailed distributions.

[18] To retrieve the actual price of the patent at its conception, one would simply have to apply the (linear) pricing function to the obtained initial market share. Thus, there is no substantial difference between finding the initial market share and the actual price.

Take any patent j. The number of citations received by j will be denoted $n(j)$. Let $\{i_k\}_{k=1}^{n(j)}$ be the set of patents that cite j, ordered from first to last. At any point in time t before the last citation occurs we know the market share of patent j is still strictly positive (or else it would have not received that last citation further on). Moreover, the value will be given by the following equation:

$$\begin{aligned} \mu_t(j) &= \mu(j) - \sum_{i=j+1}^{t} \mu(i,j) I(i,j), \quad \text{where } I(i,j) = \begin{cases} 1 \text{ if } j \in M(i), \\ 0 \text{ otherwise.} \end{cases} \\ &= \mu(j) - \sum_{i=j+1}^{t} \frac{\mu(i)}{m_i} I(i,j). \end{aligned}$$

Right before receiving the last citation, the market share will thus be:

$$\mu_{i_k-1}(j) = \mu(j) - \sum_{k=1}^{n(j)-1} \frac{\mu(i_k)}{m_{i_k}},$$

rearranging terms we can obtain the market share of a patent at its conception as a function of the values of future citations plus the value right before receiving its last citation:

$$\mu(j) = \mu_{i_{n(j)}-1}(j) + \sum_{k=1}^{n(j)-1} \frac{\mu(i_k)}{m_{i_k}}.$$

Notice that the right hand side of the expression is just a sum of values created at a future date. If we know the values of the citing patents, then the second term of the right hand side (the summation) will be known as well. The first term, the market share right before the last citation, will be inferred using the equilibrium distribution of values for all patents and the history of citations received. A detailed explanation on how to obtain it is found in the appendix.

3.2 Taking the Model to the Data

Denote the price assigned to patent j as $\widehat{\mu}(j)$. Then we know:

$$\widehat{\mu}(j) = E\left[\mu_{i_{n(j)}-1}(j)\right] + \sum_{k=1}^{n(j)-1} \frac{\widehat{\mu}(i_k)}{m_{i_k}},$$

that is, the price assigned to j depends on the prices assigned to all its received citations. The valuation (or pricing) algorithm starts from the latest patent and runs back in time until the earliest patent is priced.[19] It uses the information recursively. The appendix shows that, to be implemented, the algorithm requires, for every patent j, the following information as well: The number of citations occurring between the k^{th} and the $k+1^{th}$ citation received by j (denoted $\widehat{m}_{k,j}$), the number of citations made by j (denoted m_j) and the shape of the distribution of patent values in order to obtain $E\left[\mu_{i_{n(j)}-1}(j)\right]$.

The valuation algorithm can be summarized as follows:

[19] Only relative prices can be assessed because there is no monetary unit being carried.

Starting from the latest on to the earliest patent:

i. *Obtain* $\{m_j\}_{\forall j}$.

ii. *Obtain* $\{\widehat{m}_{i,j}\}_{\forall j, \forall i \in N(j)}$.

iii. *Obtain* $\widehat{\mu}(j) = E\left[\mu_{i_{n(j)}-1}(j)\right] + \sum_{k=1}^{n(j)-1} \frac{\widehat{\mu}(i_k)}{m_{i_k}}$.

iv. *Add* $\frac{\widehat{\mu}(j)}{m_j}$ *to the value of all citations made by patent* j.

v. *Apply* $\Pi_E\left[\widehat{\mu}(j)\right]$ *to obtain the price of patent* j.

That is, (from latest to earliest) obtain the number of citations made by the patent. Then obtain the number of failures between every success for all citations received by the patent. Then obtain a predicted value according to the formulas of the previous section. Then pass the appropriate fraction of the predicted value on to all citations made by the patent. Finally apply the pricing function (price as a function of initial market share) to obtain the price.

The main characteristic of this algorithm is that in order to valuate a given patent, it will use the citations received by that patent, the citations of the citations, and so on. In sum, it will use the entire tree of citations that the patent generates. However, because of how the model was designed, backward citations (citations made by a patent) do not provide any information about the patent's value. We do know that the sum of the stolen values from all citations made would allow us to retrieve a patent's value. But as outside observers we do not know nor have any way of inferring exactly how much the patent steals from those patents it cites. Two crucial assumptions are responsible. The first one is the independence of the market stealing productivity random process. Recall that firms draw independently their market stealing productivity from a distribution that is invariant. Thus, researchers have no incentives to act strategically when choosing which patents to study. The second one is the assumption on the market stealing production function. Since total market share stolen depends only on total effort into market stealing, then it does not pay off to study more than the necessary patents during invention. To sum up, the only way we could obtain information from past citations is if firms acted strategically during invention by choosing the number or the value of the patents to study. Since this does not occur, the past contains no information about the value of a patent.

The above argument should not be taken as a position defended in this work. On the contrary, the next step is to evaluate to what extent this crucial result holds true. This valuation algorithm may still be derived in a setting where firms do research strategically and, therefore, past citations do provide information on patent value. This suggests that it should be possible to test the non-directed research assumption.[20]

[20]Previous studies find ambiguous results on this topic: Lanjow and Schankerman (1997) find backward citations to provide no information in a study on patent litigation. Harhoff et al. find

Another important characteristic of the valuation process is that the precision of values increases as we go further back in time. Patents closest to the present have had little time to be cited. Thus, the recursive part of the algorithm will most likely have little weight on the total value. And even when it does have a big weight, it will depend on values of patents where the recursive part of their value has little weight on their own total values. On the other hand, a patent that is, say, 30 years old is not likely to receive any more citations than the ones it already has. Since the valuation algorithm is only forward looking, then it will do a much better job on a 30 year-old patent than it will on a very recent one. This fact becomes apparent in the following section, where the model is taken to data on U.S. patent citations.

3.3 The U.S. Patent Citations Data

The patent citations data set was created by Hall, Jaffe and Trajtenberg. It comprises all U.S. patent citations made between 1975 and 1998. This allows researchers to easily track the citations received by any given patent. Otherwise, one would have to check one by one all future patents, just to know the set of citations received by a single patent.[21] The data set consists of around 3 million patents and 16 million citations. It runs reliably from 1975 to 1998 but also includes patents granted before 1975 (though not all patents belong to the data set for those years).[22]

The received citation totals distribution for each annual cohort of patents running from 1975 to 1998 are shown, for selected cohorts, in figure 3.

The distribution of received citations converges to the distributions of the 1970's patents. This suggests that the assumption in Hall, Jaffe, and Trajtenberg (2001) about the "lifetime" of a patent being no higher than 35 years was a sensible one. Moreover, the distribution of received citations barely changes after the first 12 years since the application date (any patent applied for before 1987). This suggests that even before the patent expires (17 years if renewed to full term) we should not expect a patent to receive many more citations. This is consistent with this model). Hall, et.al. find that more than 80% of all the citations received by a patent occur before the first 17 years.

The distribution of total citations received would not pass a graphical test for being of the Pareto type. If the distribution was Pareto then we should observe a downward-sloping linear relationship between the number of received citations and frequency. The concave shape we observe implies the distribution is less skewed than a Pareto.

3.4 Assessing the Private Value of Patents

Harhoff, Scherer and Vopel conducted a survey of German patent holders on patent values to asses the predictive power of numerous patent indicators. They find that

that the number of backward citations is positively related to patent value. The issue on the quality of backward citations has not been explored.

[21]Needless to say, the work here is highly indebted to the incredible job done by the authors and their associates.

[22]The data set can be downloaded from the NBER web-page at www.nber.org

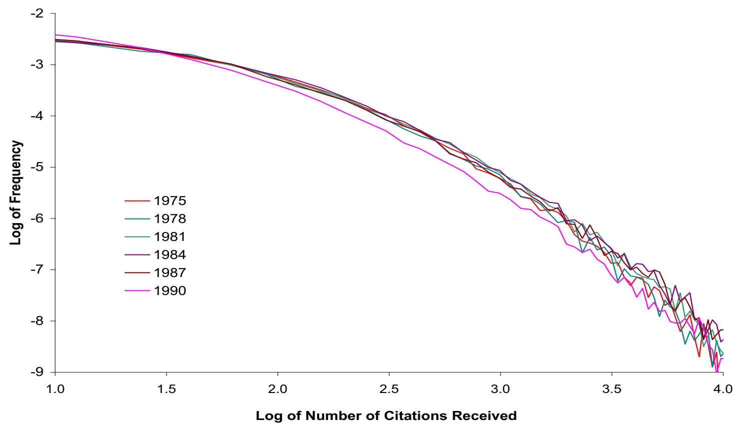

Figure 3: Distribution of received citations for selected cohorts

forward citations are positively related to a patent's value, a feature the model presents here. They also find that the distribution of values of the 1977 German patents, as reported by the patent holders, is highly skewed and found the log-normal and the Pareto to be the best fits to their survey data.

This section shows the results of applying the model to data in order to assess the private value of patents. By construction, '99 patents are given (almost) equal values. Then, all patents are recursively valuated as explained in previous sections.[23] The implication for the distribution of values of any given cohort is clear: Starting in 1999 with a degenerate distribution of patent values and as we go further back time, the distribution should converge to the underlying true distribution (assuming the model specification is correct).

The model's ability to price patents is assessed along two dimensions: First, at the distributional level, valuations by cohort do converge to a particular distribution, one with a thick tail. A frequency plot by cohort on a double logarithmic scale shows how the distribution of patent values by cohort converges to a straight line,

[23] The valuation algorithm used in the applied section is a simplified version of the one derived by the model. It preserves the recursive structure but simplifies the initial condition

$$E\left[\mu_{i_{n(j)}-1}(j)\right]$$

to a function of the application year and the number of citations received. Even though this has an effect on the valuations of the earlier cohorts, the older cohorts are robust, even to extreme changes in the initial conditions.

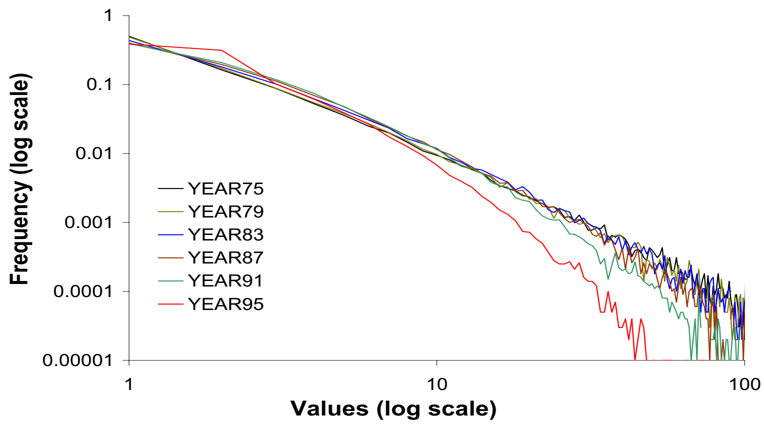

Figure 4: Distribution of patent valuations for selected cohorts

indicating that the true underlying distribution is of the Pareto-Levy type. Second, in terms of the precision of valuations, model-based valuations are more precise in predicting renewals than simple counts are.

Figure 4 plots the tail of the patent value distributions of selected cohort years. Plotted values range from 1 to 100 (35th to 99.98th percentile in 1995 and 40th to 99th percentile in 1975, for example).[24] Cohorts are equally spaced to give a sense of how rapid the convergence is to the underlying distribution. For example the distribution of values of the '95 cohort curves out more than the rest. Such a cohort had only 4 years to depart from the degenerate distribution of the '99 cohort. The cohorts of 1975, 1979 and 1983 are barely distinguishable.

Figure 4 also shows how the distributions by cohort converge to a straight line. The '75 cohort is the closest to a straight line. Assuming the true distribution of values depicted in double logarithmic scale is a straight line then it is (or at least the tail of the distribution is) of the Pareto-Levy class. This finding is consistent with assuming that researchers draw their productivity into market stealing from a Pareto distribution.

We now turn to the precision of model-based valuations as a measure of patent values. The correlation between assigned values and citation counts is around 60%, a low enough correlation to generate two very different rankings of patents, but

[24]Because we only use citations to predict patent prices, the levels are not informative. Only relative values are assessed.

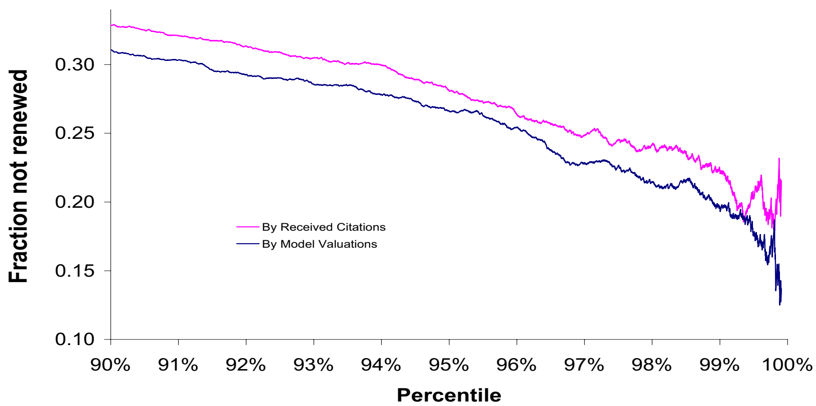

Figure 5: Non-renewal frequency of 1983 patents

high enough to suggest that both measures are reliable indicators of quality. The precision of these two indicators is assessed by using patent renewal data on all patents granted in 1983.

Renewals are a series of payments that the owner of a patent has to go through in order to prolong the life of its patent. For the cohort of 1983 these renewals took place after the third, seventh and eleventh year of the life of the patent. Failure to pay the corresponding fee at any given year resulted in an automatic termination of the patent. Thus, it is natural to presume that patents that were renewed to full term were more valuable than those that were not.

The total number of patents granted in 1983 was 54554. Half of them were not renewed to full term. When controlling for patent quality (both by means of citation counts or assigned values), non-renewals drop significantly. For example, for the 500 most valuable patents -the 99th percentile- the proportion of non-renewals drops to 22% when using citation counts as a proxy for value and to 19% when using model valuations. Figure 5 depicts the non-renewal proportions for different cut-offs starting from the 90th percentile and all the way up to the top 25 patents. Because both quality criterions yield different orderings, the curves are also different. Both curves show a decreasing trend because as the cut-off patent quality rises, the lower quality patents are being left out, that is, those less likely to have been renewed to full term. Notice too that for cut-offs above the 90th percentile, the proportion of non-renewals is higher for citation counts than for valuations.[25] It is an indication that model-based valuations outperform citation counts as a measure of patent quality.

4 Conclusions and Extensions

The model developed here serves a double purpose. On the theoretical side, it enriches the literature on creative destruction by allowing for partial creative destruction in a way that is consistent with actual R&D performed by firms. It allows a mapping between the space of ideas and the space of profits through patent citations. On the applied side, this mapping is taken to the data. The result is not only a confirmation of the initial mapping (Pareto-distributed productivities to Pareto-distributed patent values); it is also a direct contribution to the literature that uses patent data to provide measures of economic activity. Model-based patent values provide a better proxy for patent quality than simple citation counts do.

This approach is the first to use the network as a whole in order to assign individual values to patents. The recursive formulation arising from the model is simple and extremely tractable. It gets around most of the problems faced by simple counts and it does not use any information outside of that provided by the network of citations.

Far from being a complete framework for the use of patent citations, this is as a starting point. The patent citations network is full of valuable information. Models that take a stand on what citations mean (whatever it may be) have a better chance of delivering relevant testable implications. This model in particular may be used as a benchmark to create richer ones such as those where renewals, number of claims or backward citations are taken into account. Another potentially productive avenue is to try more elaborate search processes than the one used here. This may help in studies that relate the characteristics of the network of patent citations to technological spill-overs and technological change. Finally, other extensions could involve different assumptions on the R&D process such as the building of knowledge capital by firms. This would allow for the analysis of industry and firm dynamics, giving backward citations a role too.

[25] The two curves converge at the 50th percentile, where the number of received citations is around 5 for both orderings. This is not surprising since, by construction, the algorithm works better for patents that receive many citations. Of course, the bulk of the total value (more than 60%) is captured by the top 10% of depicted in the graph.

Appendix

Endogenous choice of symmetric market stealing

Here, instead of choosing the aggregate level of market stealing, a firm faces individual market stealing production functions for each of its cited patents. Thus, a firm optimally choose labor into market stealing $L_D(j)$ for any given citation j. Assume a firm decides to hire L_D units of labor in total. The optimal allocation of labor into market stealing from different citations is given by the following:

$$\mu(i) = g(L_D, B_i) = \max_{\{L_D(j)\}_{\forall j \in M(i)}} \left\{ \sum_{\forall j \in M(i)} \mu(i,j) \right\}, s.to: L_D = \sum_{\forall j \in M(i)} L_D(j),$$

where $\mu(i,j)$, defined as the market that patent i steals from patent j, is assumed to have the following functional form:

$$\mu(i,j) = B_i A(m_i) L_D(j)^{\nu}.$$

Two things should be noticed. The first one is that the symmetric set-up ensures equal market stealing from all cited patents, that is:

$$L_D(j) = \frac{L_D}{m_i}.$$

The second, that assumption *(iii)* places a restriction on $A(m_i)$. We need $A(m_i)$ to be such that total market share does not depend on the number of cited patents:

$$\begin{aligned} \mu(i) = g(L_D, B_i) &= \sum_{\forall j \in M(i)} B_i A(m_i) \left(\tfrac{L_D}{m_i}\right)^{\nu} \\ &= {m_i}^{1-\nu} B_i A(m_i) {L_D}^{\nu} \\ \Rightarrow A(m_i) &= {m_i}^{\nu-1}, \end{aligned}$$

replacing with these two conditions we obtain:

$$\mu(i) = B_i {L_D}^{\nu}.$$

Besides being able to obtain equal market stealing as a result, this set-up suggests natural ways of testing the underlying assumptions. If we want to test the validity of the assumption that market stealing depends on aggregate effort into stealing (and not on individual effort) we can generalize $A(m_i)$ to be equal to ${m_i}^{\zeta(\nu-1)}$. That way if $\zeta < 1$ we would know that more citations made by a patent imply more value for the patent. This could give firms incentives to act strategically during invention (by studying more patents than the ones they actually need). Another test can be created by perturbating the symmetry of the set-up. The simplest one would be to have the productivity in individual market stealing be idiosyncratic to the cited patent, that is having B_{ij} instead of B_i, where:

$$B_{ij} = B_i + \varepsilon_j \qquad \text{such that } E(\varepsilon_j) = 0 \text{ and } \varepsilon_j \sim i.i.d., \forall j.$$

If these perturbations are only observed by the patenting firm after the choice of Labor into development, then the symmetry in the choice of labor remains. If, on the other hand, they observe it prior to the choice of labor, symmetry no longer holds.

Proof of Proposition 1

Proposition 1. *If patent protection lasts only two periods$T = 1$)and the average value of a patent is infinitesimally small then* $\Pi_E(\mu) = A\mu$.

Proof. Assume $T = 1$, then

$$\Pi(\mu) = \mu + \frac{\mu'}{1+r},$$

where μ' is next period's market share (afterwards it expires and profits drop to zero, regardless of the market share), which means:

$$\mu' = \begin{cases} \mu & \text{if the patent is not cited} \\ \mu - \min\{x, \mu\} & \text{if the patent is cited} \end{cases}$$

here x is a random variable representing the amount that the new patent attempts to steal from the old one. The minimum operator ensures that the new patent does not steal more than what the old one has to offer. This amount x depends on two things: the distribution of citations made by the new patent and the distribution of initial market shares the new patent may have. What comes out from combining the two is some distribution of attempted market stealing that is denoted $\widetilde{f}(x)$. The only relevant feature is that the expected value of x cannot be greater than the expected initial market share of a patent (by construction, a patent's initial market share is always greater than what it steals from any given patent). we are interested in the expected value of a patent:

$$\Pi_E(\mu) = \mu + \frac{E[\mu']}{1+r},$$

where:

$$E[\mu'] = (1-\mu)\mu + \mu\left[\mu - \int \widetilde{f}(x)\min\{x, \mu\}\,dx\right].$$

The first term on the right hand side indicates that the market share in the second period will remain exactly as the first period's market share with probability $1-\mu$ (if the patent is not cited). The second term states that with probability μ the patent is cited and it loses an expected amount given the integral. Notice:

$$\mu > E[\mu'] > \mu\left[1 - \int \widetilde{f}(x)\,x\,dx\right] = \mu\left[1 - E(x)\right],$$

implying that

$$\mu\left(1 + \frac{1}{1+r}\right) > \Pi_E(\mu) > \mu\left(1 + \frac{[1-E(x)]}{1+r}\right),$$

so the expected profits function is a function sandwiched between two linear functions. If these two linear functions are sufficiently close we know that $\Pi_E(\mu)$ is, for practical purposes, linear. Formally, we have

$$\lim_{E(x)\to 0}\{\Pi_E(\mu)\} = \mu\left(1+\frac{1}{1+r}\right) = \mu A,$$

In words, if the expected attempted stolen market x is infinitesimally small then the price of a patent is linear in initial market share. We can guarantee that $E(x)$ will be very small by assuming that the expected initial market share is also infinitesimally small. ■

Proof of Proposition 2

Proposition 2. *If researchers draw productivities from a Pareto distribution then the equilibrium distribution of expected profits is Pareto.*

Proof.

1.	w	$= A\nu B {L_D}^{\nu-1}$	(FOC of firm).
2.	μ	$= B {L_D}^{\nu-1}$	(Market stealing prod. fn.).
3.	B	$\sim Pareto\,[a,b]$	(By assumption).

By 1-2 we can show that

$$B(\mu) = \left(\frac{A\nu}{w}\right)^{\nu}\mu^{1-\nu},$$

because $B(\mu)$ is strictly increasing and continuous, then the distribution of μ will be:

$$\begin{aligned} F(\mu) &= 1-b^a\left[\left(\tfrac{A\nu}{w}\right)^{\nu}\mu^{1-\nu}\right]^{-a} \\ &= 1-b^a\left(\tfrac{A\nu}{w}\right)^{-a\nu}\mu^{-a(1-\nu)} \\ &= 1-\left[b\left(\tfrac{w}{A\nu}\right)^{\nu}\right]^a\mu^{-a(1-\nu)} \\ &= 1-\left[b^{\frac{1}{1-\nu}}\left(\tfrac{w}{A\nu}\right)^{\frac{\nu}{1-\nu}}\right]^{a(1-\nu)}\mu^{-a(1-\nu)} \end{aligned}$$

$$\Rightarrow \mu \sim Pareto\left[(1-\nu)\,a, b^{\frac{1}{1-\nu}}\left(\frac{w}{A\nu}\right)^{\frac{\nu}{1-\nu}}\right].$$

Likewise,

$$\Rightarrow \Pi_E \sim Pareto\left[(1-\nu)\,a, b^{\frac{1}{1-\nu}}\left(\frac{w}{A\nu}\right)^{\frac{\nu}{1-\nu}}A\right]$$

■

Obtaining $E\left[\mu_{i_{n(j)}-1}(j)\right]$

As explained, this term is of greater importance the closer the patent is to the present.

Assume $\sum_{k=1}^{n(j)-1} \frac{\mu(i_k)}{m_{i_k}}$ is known. Then:

$$E\left[\mu\left(j\right)\right] \quad = E\left[\mu_{i_{n(j)}-1}\left(j\right)\right] + \sum_{k=1}^{n(j)-1} \frac{\mu(i_k)}{m_{i_k}}.$$

We know:

$$\mu\left(j\right) > \sum_{k=1}^{n(j)-1} \frac{\mu\left(i_k\right)}{m_{i_k}} \equiv \underline{\mu},$$

that is, the initial market share of our patent cannot be less than the sum of the stolen market shares by incoming patents. Only the last patent to cite our patent may have not stolen the desired share $\frac{\mu\left(i_{n(j)}\right)}{m_{i_{n(j)}}}$.

We want to know:

$$E\left[\mu_{i_{n(j)}-1}\left(j\right)\right] + \underline{\mu} = \int_{\underline{\mu}} x g_j\left(x\right) \frac{f\left(x\right)}{F(\underline{\mu})} dx,$$

where:

$$g_j\left(x\right) = \left[\prod_{k=0}^{n(j)-1} x_k\left(1-x_k\right)^{\widehat{m}_{k,j}}\right]\left(1-x_{n(j)}\right)^{\widehat{m}_{n(j),j}},$$

and where:

$$i_0 = j,$$
$$\widehat{m}_{k,j} = \left[\sum_{t=i_k}^{i_{k+1}} m(t)\right] - 1,$$
$$x_0 = x,$$
$$x_{k+1} = x_k - \frac{\mu\left(i_{k+1}\right)}{m_{i_{k+1}}},$$
$$x_{n(j)} = \max\left\{0, x_{n(j)-1} - \frac{\mu\left(i_{n(j)}\right)}{m_{i_{n(j)}}}\right\},$$

that is, what is the value of the patent given that we know it is greater than $\underline{\mu}$. The intuition is simple. A patent can be seen as a sequence of draws, each representing the successes or failures to be cited by a new patent. Because of how the invention process was modelled, these random draws are independent from one another. However, they are not identically distributed because the likelihood of a success depends on the market share of the patent at the point in time of the randomization as well as the number of citations made by the new patent. For any patent j, it is important to know the history of successes and failures (to be cited) with respect to possible market shares at each point in time. Fortunately, since the number of successes is very limited (it is just the number of citations received), then this otherwise tedious calculation reduces to obtaining $\widehat{m}_{k,j}$, the number of failures between the k^{th} and the $k+1^{th}$ citation received by j. We then proceed to integrate over all possible initial market shares x, taking into account the evolution of market shares that is dictated by the received citations history $(x_k, \forall k)$.

References

[1] Aghion, F. and Howitt, P (1992) "A Model of Growth through Creative Destruction ". Econometrica, vol. 60, issue 2, pp. 323-351.

[2] Angell, M. (2004) "The truth about the drug companies... " The Random House Publishing, NY.

[3] Baumol, W. (2002) "The Free Market Innovation Machine" Princeton University Press.

[4] Bound, J. et.al. (1984) "Who does R&D and who Patents?". In Griliches (1984) pp. 21-54.

[5] Bryson, M. (1973) "Heavy tailed distributions: properties and tests ". Econometrica, vol. 65, issue 6, pp. 1389-1419.

[6] Eswaran, M. and Gallini, N. (1996) "Patent policy and the direction of technological change ". The RAND journal of Economics, vol. 27, no. 4, pp. 722-746.

[7] Griliches, Z. ed. (1984) "R&D, Patents and Productivity" University of Chicago Press, Chicago, IL.

[8] Griliches, Z. (1990) "Patent Statistics as Economic Indicators: A Survey". Journal of Economic Literature, vol. 28, issue 4, pp. 1661-1707.

[9] Grossman, G. and Helpman, E. (1991) "Quality Ladders in the Theory of Growth ". Review of Economic Studies, vol. 58, issue 1, pp. 43-61.

[10] Hall, B., Jaffe, A., Trajtenberg, M. (2000) "Market Value and Patent Citations: A First Look". UCBerkeley Dept. of Economics Working Paper.

[11] Hall, B., Jaffe, A., Trajtenberg, M. (2001) "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools". NBER Working Paper Series 8498.

[12] Harhoff, D., Scherer, F. and Vopel, K. (1997) "Exploring the tail of patented invention value distributions". Discussion paper no. 97-30.

[13] Harhoff, D., Scherer, F., Vopel, K. (1999) "Citations, Family Size, Opposition and the Value of Patent Rights."Research Policy, vol. 33, issue 2, pp. 363-364.

[14] Harhoff, D., Scherer, F., Narin, F., and Vopel, K. (August 1999) "Citation Frequency and the Value of Patented Inventions". Review of Economics and Statistics vol. 81, issue 3, pp. 511-515.

[15] Hopenhayn, H. and Mitchell, M. (1999) "Innovation fertility and patent design ". NBER working paper W7070.

[16] Jackson, M. (2005) "The economics of social networks (2005) ". Proceedings of the 9th World Congress of the Econometric Society (forthcoming).

[17] Jackson, M. and Rogers, B. (2005) “Search in the formation of large networks: How random are socially generated networks? ”. Working Paper.

[18] Kortum, S. (1997) “Research, Patenting and Technological Change ”. Econometrica, vol. 65, issue 6, pp. 1389-1419.

[19] Kortum, S. and Lerner, J. (1998) “Stronger protection or technological revolution: what is behind the recent surge in patenting? ”. Carnegie-Rochester Conference Series on Public policy 48, pp. 247-304.

[20] Lanjouw, J., Pakes, A. and Putnam, J. (1997) “How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data ”. The Journal of Industrial Economics, vol. 46, no. 4, pp. 405-432.

[21] Lanjouw, J. and Shankerman, M. (1999) “The Quality of Ideas: Measuring Innovation with Multiple Indicators”. NBER working paper 7345.

[22] Leiva, F. “Research vs Development ”. Working Paper.

[23] Llobet, G., Hopenhayn, H. and Mitchell, M (2000) “Rewarding Sequential Innovators: Prizes Patents and Buyouts ”. Working Paper.

[24] Pakes, A. (1985) “On Patents, R&D and the Stock Market Rate of Return”. Journal of Political Economy, vol. 93, issue 2, pp. 390-409.

[25] Pakes, A., Simpson, M., Judd, K. Mansfield, E. (1989) “Patent Renewal Data”. Brookings Papers on Economic Activity. Microeconomics, vol.1989, pp. 331-401.

[26] Romer, P. (1990) “Endogenous Technological Change ”. Journal of Political Economy, vol. 98, issue 5, pp. S71-S102.

[27] Sanders, B., Rossman, J., and Harris, L. (1958) “The Economic Impact of Patents”. The Patent and Trademark Copyright Journal vol. 2 issue 2, pp. 340-362.

[28] Schmookler, J. (1972) “Patents, Invention and Economic Change”. Eds. Griliches and Hurwicz, Harvard University Press, Cambridge, MA.

[29] Serrano, C. “The market for intellectual property: Evidence from the transfer of patents ”. Working paper.

[30] Trajtenberg, M. (1990) “A penny for your Quotes: Patent Citations and the Value of Innovations”. RAND Journal of Economics, Vol. 21, Issue 1, pp. 172-187.