ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 692 October 2010

Are Patent Citations Driven by Quality?

Christian Ghiglino and Nicole Kuschy

Are Patent Citations Driven by Quality?

Christian Ghiglino * Nicole Kuschy†

October 29, 2010

Preliminary Draft

Abstract

The present paper builds a simple model of patent citations not based on the rich-get-richer aspect of preferential attachment. In our model the dynamics of citations are driven by known heterogeneities in the applicability of existing patents and aging. The model matches closely the hazard rates of citations for the vast majority of patents in a random sample of patents granted by the USPTO between 1975 and 1999. Furthermore, we show that the long run distribution of patent citations is well fitted when the distribution of applicability across patents follows a Gamma-distribution.

We also discuss the possibility that popularity of patents might influence citation decisions if innovators are not perfectly informed about patents' applicability. We find that popularity matters but the size of the effect is very small. Finally, the possibility to distinguish between citations to patents within the same class and to different classes allows us to show that the magnitude of the influence of popularity is increasing in technological distance.

*University of Essex - Department of Economics, Colchester, CO4 3SQ, UK. `cghig@essex.ac.uk`
†University of Essex - Department of Economics, Colchester, CO4 3SQ, UK. `nkusch@essex.ac.uk`

1 Introduction

Considerable work has been devoted in recent years to the analysis of the network of patent citations. This literature has focused on a variety of issues and a relevant set of authors uses the pattern of patent citations as a proxy of knowledge spillovers and/or as a proxy for the quality of patents. Examples of this work include Hall, Jaffe, and Trajtenberg (2005) and Harhoff, Narin, Scherer, and Vopel (1999), who have found a positive correlation between the number of citations a patent has received (so-called forward citations) and its commercial value[1]. On the other hand, using a dataset constructed by the NBER and described in detail in Hall, Jaffe, and Trajtenberg (2002), evidence has been found, e.g., by Jaffe, Trajtenberg, and Henderson (2002), that knowledge spillovers are technologically localized. In fact, they find that $55 - 60\%$ of all patent citations are within the same 3-digit class[2], concluding that cited and citing patents tend to be technologically close.

Another important strand of the literature on patent citations focuses on the actual network of citations, mainly on the underlying mechanics driving the network formation. It has been found that a number of real-world networks, among them citation networks of academic papers, exhibit preferential attachment, i.e., the probability of receiving a further citation increases in the number of current citations[3]. In the case of citations, clearly a number of other factors influence the probability that a patent will be cited, such as its age, its birth-cohort and even its technological category (Marco (2007), Hall, Jaffe, and Trajtenberg (2002)). However, it appears that patents which have been cited more frequently in the past also tend to have higher probabilities to being cited in the future. If confounding factors like aging and knowledge diffusion are taken into account, it has been argued that this preferential attachment process is linear as well (see, e.g., Csárdi, Strandburg, Zalányi, Tobochnik, and Érdi (2007) or Valverde, Solé, Bedau, and Packard (2007)).

Observing preferential attachment may not mean that innovators take popularity as an indicator of quality. In fact, the cited literature on the formation of the citation network tends to be very

[1] Hall, Jaffe, and Trajtenberg (2005), e.g., have found that the stock market valuation of a firm's intangible knowledge stock is increasing in the number of citations the firm's patents have received. Harhoff, Narin, Scherer, and Vopel (1999) instead used survey data to construct private economic values of patents, and found that these estimates were positively correlated with the number of forward citations patents had received.

[2] The 3-digit classification is made by the United States Patent and Trademark Office (USPTO).

[3] See, e.g., Albert and Barabási (2002) for a very good survey of different real-world networks and their respective characteristics.

mechanical and does not provide an underlying model which would explain the observed structure or justify why preferential attachment is observed. Furthermore, the aging of patents is also modeled without a theoretic justification, it is "data-driven" (see Valverde, Solé, Bedau, and Packard (2007), Csárdi, Strandburg, Zalányi, Tobochnik, and Érdi (2007) or Mehtai, Rysman, and Simcoe (2008)). We believe that even if it is agreed that a combination of preferential attachment and aging seems to governs patent citations, it is not settled whether the number of forward citations a patent receives is related to its intrinsic quality or not. In the latter case, a patent which receives a more than average amount of forward citations by sheer luck would attract even more citations in the future - the rich will get richer. Being guided by the number of forward citations when choosing which patent to cite would then have important implications.

Attempts to model the generation of the network of patent citations using rational innovators are rare (see Ghiglino (2010)). The present paper proposes a simple model of innovation which is consistent with the observed data on patent citations, without relying on the rich-get-richer aspect of preferential attachment. The underlying structure of the model is built on a few stylized facts. Our data analysis confirms the role of technological similarity, that is, that the majority of patent citations are to technologically close patents. At the same time, those citations which are not to technologically similar patents, seem to be quite evenly distributed across the complete technological spectrum. Finally, the appeal of an existing patent decreases with its age, i.e., there is aging.

This citation behavior is something we build into our model. The basic setup assumes that each new innovation is built from a number of already existing ideas. First, in the spirit of the literature on the technology space and undiscovered ideas (see, e.g., Auerswald, Kauffman, Lobo, and Shell (2000) or Weitzman (1998)), we assume that a continuum of not yet discovered ideas exist and that an innovation is a draw from this pool of "nascent" ideas[4]. Second, we assume that each idea is related to a number of other ideas, that is, an innovation is based on an optimal combination of other ideas. However, in practice the ideas that would combine optimally will still be nascent. We assume that an innovation that is related to a particular idea j, can cite instead any existing patent

[4]Although we do not model the innovative process leading to patents, our assumptions about the arrival rate of ideas link the model to parts of the literature on endogenous growth, as, e.g., Kortum (1997).

of an idea that is compatible with idea j. This means that an existing patent can typically be used in many innovations. We define this property of being usable in many applications as " breadth of applicability". Clearly, *ceteris paribus*[5], a patent with a large breadth of applicability is more likely to be cited than a patent that is less compatible.

We view broadness of applicability as one factor that determines an idea's quality[6]. The other factor is the intrinsic productivity of the idea when used to produce a final good. When we discuss a patent's quality, we refer to a combination of both applicability and productivity. While applicability is determined by the extent to which the idea can replace other existing ideas as inputs in innovation, we assume that productivity is determined by the age of the patent. Younger patents are always assumed to be more productive than older patents when used as prior knowledge. Our model links the probability that a patent will be cited positively to its quality, i.e., the probability of being cited is increasing in the broadness of applicability of a patent and decreasing in its age. Conditional on a patent's age, this implies that broader patents are expected to have received a larger number of citations. Contrary to models built exclusively on preferential attachment, the question of possible "sunspots" or self-fulfilling prophecies does not arise in this setup. Let us note that, as citations are due to relatedness of ideas rather than knowledge spillovers between innovators, the results in the present model are not affected by issues of whether the citations are made by the innovator or added by the examiner, as discussed in detail in, e.g., Jaffe, Trajtenberg, and Fogarty (2002) and Alcácer and Gittelman (2006).

Under our assumptions, we find that the distribution of forward citations is determined by the distribution of applicability of patents. We achieve a very good fit to the data by assuming that the distribution of applicability follows a gamma distribution with a mean that is determined by the estimated mean of applicability in the data. Indeed, we match very well the citation distribution of patents that have received up to 80 citations, which make up more than 99.9% of all the patents in our data. In addition, the exponential aging arising from our model fits the data on citation rates better than any other functional form. The fit of citation rates is independent of the distribution of

[5]In particular of a given age.

[6]We assume that patent quality can be measured as the economic value of a patent.

applicability, but depends on the mean-value of applicability.

Finally, in a later part of the paper we use our model to test whether there are indications that innovators use the current number of forward citations of a patent as an indicator of patent quality. As we can distinguish between citations that are between technologically similar ideas and those that are given to technologically remote patents, we conjecture that if current forward citations are seen as signaling "quality", such signals will be more important for citations to technologically remote patents. Consequently, we look for differences in citation patterns between patents within the same category and patents from different categories. A direct inspection of the empirical patterns seems to rejects the hypothesis that there are any significant differences between the two patterns, indicating that all patent citations are built on differences in patent "quality". However, when using a more sensitive test based on the effect of citations on the hazard, we do find weak effects of popularity of patents even if we control for patent quality. Furthermore, these effects seem to increase with technological distance.

Recently, the role of popularity in the evolution of different networks, among them a patent citation network, has been investigated by Fafchamps, Goyal, and van der Leij (2010). Their work differs from ours in a number of ways. In particular, their focus is on the fat tail of the citation distribution and on statistical evidence of a snowball effect in citations. In contrast to us, they are not using a specific model of citations based on applicabilities. Using various econometric methods, they are not able to reject the role of popularity in the patent citation data, while they how that within other networks (they consider co-authorship and actor networks) the snowball effect vanishes if unobserved heterogeneity is taken into account. In this sense, our results on the importance of popularity in patent citations are consistent with their findings. With respect to our own model, we believe that there is scope for further empirical analysis of this point, particularly as patent quality is difficult to disentangle from the number of citation received.

The remainder of the paper is organized as follows: Section 2 will introduce the dataset which we will be using to compare our model to actual citation behavior. Section 3 introduces our model while section 4 compares the citation rates obtained from the theory with the rates found in the data, as well as comparing the distributions of forward citations. In section 5 we compare our simple model with a slightly altered version which allows for part of the citations relying on the signaling effect of current forward citations. Section 7 concludes.

2 The Data

As many other papers on patents citations, we use the data provided by the NBER and discussed in detail in Hall, Jaffe, and Trajtenberg (2002). In particular, we merge two datasets: 1) a random sample of 10% of all innovations having successfully applied for patent protection by the United States Patent and Trademark Office (USPTO) between January 1, 1975 and December 30, 1999 and 2) the patents that cite them within this time period. For each of the cited patents, referred to as "parents", and the according citing patents, the data provides information on their grant date, the assignee code, the technological classification, and the date of each citation received. Focusing only on those citation pairs for which all relevant information is available for both the citing and the cited patent, our dataset includes $678,363$ citing patents, $213,188$ cited patents, and a total of $1,108,468$ pairwise citation observations. The technological classification considers 412 3-digit classes from the USPTO, which were aggregated by the NBER into 36 2-digit sub-categories and into 6 1-digit categories. These categories are Chemical (excluding Drugs), Computers and Communications (C&C), Drugs and Medical (D&M), Electrical and Electronics (E&E), Mechanical, and Others. A detailed aggregation scheme can be found in the appendix.

A first pattern emerges from the analysis, as seen from table 1. Reconfirming the findings of Jaffe, Trajtenberg, and Henderson (2002), we find that the majority of citations are between patents which share the same 3-digit class, this being true for all categories considered. Interestingly, around half of those citations which are not within the same class are between patents that do not even belong to the same 1-digit category. This mix of a majority of parents within the same class and a consistent minority of parents outside the same category is persistent across all categories[7].

[7] It also does not seem to be due to a few outliers either. Around 60% of all parents in our dataset have received at least one citation from another category.

Table 1: Citations within and across categories

	% of citations		
	within class	within subcategory	within category
Overall Data	54.64%	65.70%	76.93%
Chemical	46.01%	62.46%	76.99%
C & C	51.16%	66.16%	79.62%
D & M	57.65%	71.52%	77.78%
E & E	57.14%	64.99%	77.35%
Mechanical	58.20%	65.80%	74.98%
Others	58.74%	65.73%	75.59%

Table 2: Cross-Category Citations

	Chemical	C&C	D&M	E&E	Mechanical	Others
Chemical	76.99%	1.17%	11.96%	4.01%	5.80%	8.97%
C & C	0.81%	79.62%	0.80%	7.94%	3.65%	1.50%
D & M	4.78%	0.46%	77.78%	0.80%	1.12%	1.52%
E & E	3.50%	10.79%	2.50%	77.35%	5.62%	3.33%
Mechanical	5.71%	5.60%	2.62%	6.13%	74.98%	9.08%
Others	8.21%	2.36%	4.35%	3.78%	8.84%	75.59%

It might be worthwhile mentioning that those citations across categories do not seem to be the result of patent examiner citations either. Alcácer and Gittelman (2006) found that, if at all, patent examiner citations tend to be more technologically localized than innovator citations.

Table 2 provides a general overview of "who-cites-whom". The columns indicate the category of the citing patents, and the rows the categories of the parents, i.e., the patents that are being cited. By sampling individual patents and checking the type of parents they cite, this analysis allows to better understand the process and to possibly single out the categories to which citations across categories are directed to.

In fact the data reveals an interesting regularity: upstream and downstream relationships between categories appear to be different. For example, the category Drugs and Medical, which possibly relies on a very specific and narrowly defined field of previous research, cites very often outside its own

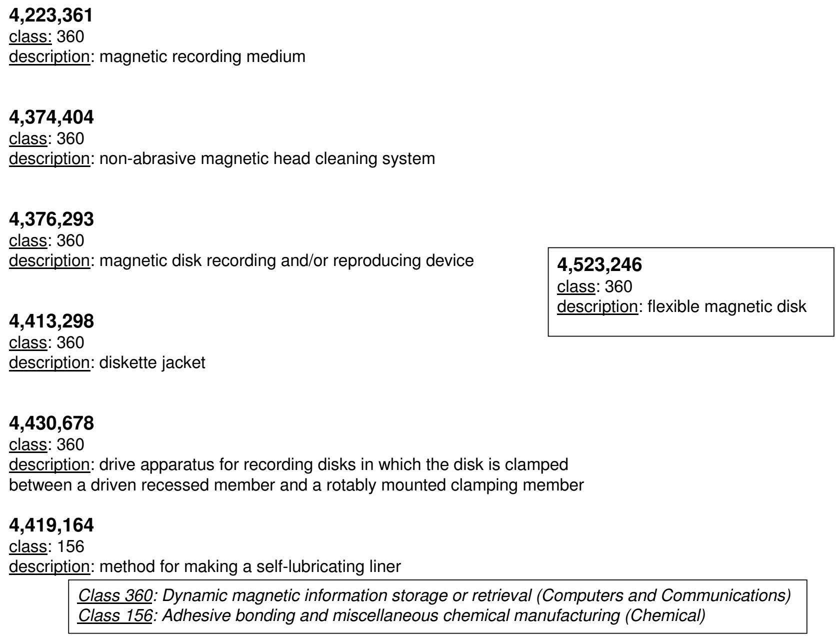


Figure 1: Computers & Communications citing Chemical

category, but this is mainly due to the fact that it cites many patents that are classified as Chemicals[8]. The link does not work in the opposite direction: D&M is not even the category to which most of the across-category citations of Chemical are directed to. This implies that many ideas in chemistry are useful when researching for drugs, and that chemical can be considered an upstream sector to D&M. A similar mechanism explains why the only other across-category in the two-digits level, is patents from C&C citing patents from E&E - a lot of ideas in the field of electronics should come in handy when researching in the field of computers. To illustrate the citation behavior, figure 1 shows an example of a citation pattern across categories. Here, a patent categorized into Computers & Communications cites the majority of its parents from the same class, and one parent from the category Chemical. From this and other examples we conclude that innovations seem to build on a mix of technologically similar and dissimilar ideas.

The citation behavior with respect to classes and categories of patents is something we are going

[8]In fact, more than 30% of *all* the citations made by the 2-digit subcategory Drugs are made to category Chemical.

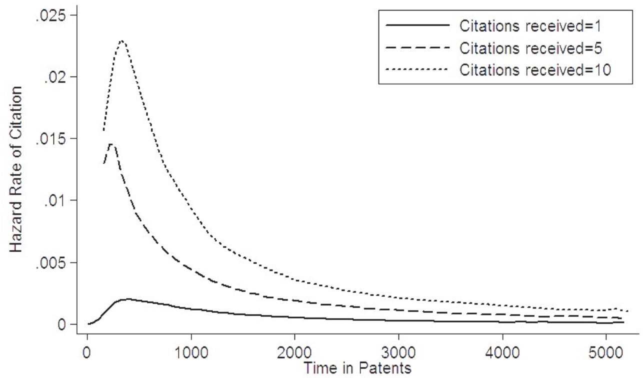


Figure 2: Non-Parametric Hazard Rates of Citation

to incorporate by assumption into our model. On the other hand, we aim to replicate endogenously the shape of the citation function and the distribution of citations across patents. Figure 2 shows the non-parametric likelihood of being cited as hazard rates for three different groups of patents, having 1, 5, and 10 forward citations respectively. The age of the patents is measured in the number of patents being granted between the grant-date of the patent and the observed date, aggregated in bins of 412 patents each.

The data shows that patents that have already received a higher number of forward citations are more likely to be cited again than their less-cited counterparts, irrespective of their age. Similarly, older patents are less likely to be cited than comparable younger patents - i.e., the older a patent is, the less likely it is that it will be cited by future research, independent of its current number of forward citations. While this decrease in the hazard of citation over time is due at least in part to a frailty effect, we will show that the decrease is also present at the individual patent level.

The data also indicate another stylized fact. The probability to receive a citation is increasing after the birth of a patent, a fact implying that either the quality of an innovation increases during the early stages after its birth, either it takes time to the innovator to become aware of its existence, a type of knowledge diffusion.

The empirical analysis pursued so far produces stylized facts that the model needs to account for.

Let us note that previous work on the network of patent citations has found that citations follow a modified preferential attachment rule that takes aging into consideration. Valverde, Solé, Bedau, and Packard (2007) find that if aging is modeled by a Weibull distribution, linear preferential attachment generates a scale-free distribution of patent citations that fits well the data. Csárdi, Strandburg, Zalányi, Tobochnik, and Érdi (2007) find preferential attachment to be super-linear if aging is not taken into account, and concede that aging apparently weakens the stratification of patents. In any case, all studies so far have found that a patent's probability of being cited is first in- and then decreasing over time and that it is increasing in the number of citations a patent has already receive.

To sum up, this literature shows that the data is consistent with a model in which the network is generated by a rule similar to linear preferential attachment corrected by age. We will see that our model, which is not based on preferential attachment, is also able to generate these features. Note that both the standard linear attachment model in the literature and the model considered here ignore the aspect of knowledge diffusion in the early stage of the patent's life.

So far we have considered the empirical hazard rates of citation. We now focus on the asymptotic, i.e., when age tends to infinity, distribution of citations. The distribution obtained from our data is represented in a log-log scale in figure 3. The highest number of citations a patent has received in our data is 605. The shape of the distribution indicates that patent citations are not driven by linear preferential attachment alone[9].

The model we construct in the next section aims at explaining both the empirical hazard rates and the distribution of citations obtained above.

3 The Model

The basic setup assumes that each new innovation is built from a number of already existing ideas. We assume that an innovation is a draw from a continuum of not yet discovered, or "nascent", ideas

[9] If they were, we would observe a distribution that appears linear on a log-log scale.

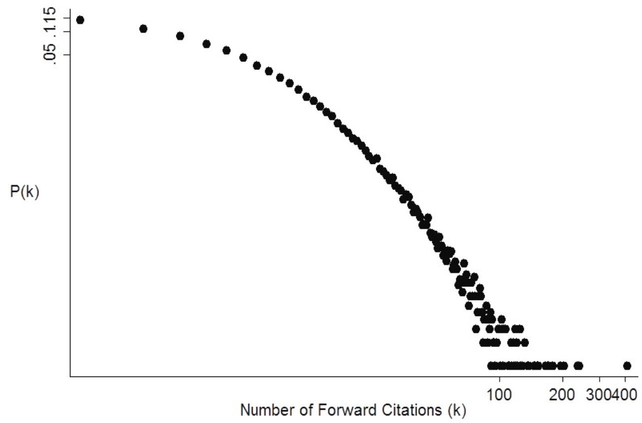


Figure 3: Observed Citation Distribution in the Data

and it is related to a number of other ideas. However, in practice these "other" ideas that would combine optimally to build the new innovation are yet undiscovered. We assume that the innovator can use and cite instead a set of different, existing, patents, that are compatible with the optimal "other" ideas. Whether an idea is used in the innovation process depends on how close is its variety from the original optimal "other" idea and on its "breadth of applicability". Finally, we introduce productivity growth in the model by assuming that, when there is a choice, the innovator chooses the youngest patent among the compatibles ones. These assumptions imply that the "value" of a patent depend on its age, its variety and its "breadth of applicability". The model is formalized in the next sub-sections.

In the model, discovered and undiscovered ideas are represented by patents and potential, or "nascent", patents. The two notions are perfectly exchangeable here. Furthermore, as the innovator only chooses citations, we may consider that innovators and their patents are indistinguishable as well.

3.1 Arrival and Variety of Patents

Time is continuous. Patents come in a continuum of potential varieties. The variety of a patent i is noted μ_i. There are S patent classes. In each class new innovations are drawn from the pool of potential ideas, where the draws follow a Poisson process with arrival rate of 1. Consequently, patents are distributed evenly across the S patent classes. Each individual class s has a support of length 1 and without loss of generality we assume that this support is represented as a circle. At time t, each class contains a large number of $N(t)$ patents, which implies that the total size of the patent network at t is $SN(t)$. We assume that patents are uniformly distributed along the support of the class they belong to and patent i's "place" on the support determines its variety, μ_i.

Define the broadness of idea i, a_i, as its broadness of applicability. For now, we do not put any restrictions on how broadness of applicability is distributed over ideas. However, we do assume that the distribution is the same in all classes and is common knowledge to the innovators. The broadness of an idea i, as well as its type, μ_i, is realized upon the arrival of the idea. Specifically, we assume that each patent's fit is defined as follows:

Assumption 1: A patent i of type μ_i and broadness a_i is assumed to have a support of

$$F_i \equiv [\underline{\mu_i}, \bar{\mu}_i]$$

where $\underline{\mu_i} \equiv \mu_i - \frac{1}{2}a_i$ and $\bar{\mu}_i \equiv \mu_i + \frac{1}{2}a_i$.

3.2 Research and Evolution of Ideas

We assume that innovation is an incremental process in which each new idea builds on existing ideas. In the model, a new patent is related to $m+1$ *potential* patents, which it uses and has to cite[10]. The set of parents to idea j is given by a correspondence $\sigma(\mu_j)$. Our assumptions on the technological identity of newly arrived patents and their related technology are summarized in Assumption 2:

[10]While in reality different patents cite different numbers of parents, for simplicity we fix the number of citations *made* by any patent.

Assumption 2: Assume a new patent j has been realized in class s. Its type μ_j is random and drawn from a uniform distribution over class s. Patent j is technologically related to m potential[11] patents from within class s and to one potential patent from outside s. The correspondence $\sigma(\mu_j) = \mu_k, k = 1, ..., m$ determines the types of the related ideas within class s.

The process of innovation as described in Assumption 2 implies that the majority of citations are between patents within the same class and a minority between patents from different classes, a pattern that matches the empirical regularities highlighted in this paper. The technological identity of these related ideas is determined by the correspondence $\sigma(\mu_j)$, which maps patent j to its parents within s uniformly. The related patent from outside s is assumed to be drawn uniformly from all other classes. Formally, we assume

Assumption 3: Let $\sigma(\mu_j) = \{\mu_1, ..., \mu_m\}$. Let $|I|$ be an open interval in $[0, 1]$. Then $\sigma(|I|)$ is the set of parents within s that correspond to the patents within $|I|$. We assume that for any open intervals $|I|$ and $|J|$ within $[0, 1]$, we have that

$$\int_{|J|} \sigma(|I|) dj = m \cdot |J|$$

The $(m+1)^{th}$ related idea is drawn uniformly over all patent classes except s.

Together, Assumptions 2 and 3 imply that citations are to all extent and purposes random draws. This also implies that it is just as likely to have the one citation outside the class occurring close to the class of the citing patent as far away from it, which is consistent with the empirical fact that those citations outside the same 3-digit class are just as likely to be within the same category as outside of it. It is easy to think of a number of reasons for why such a mix between technologically similar and different citations might occur in reality. It might be an optimal behavior for the innovator - e.g., it might be more costly to work with technologically different ideas, or it might be that their inclusion increases the probability of being patented. In this paper we assume that this regularity is due to the innovation process itself and does not depend on behavioral choices.

[11] as they may or may not already exist

3.3 Search and Innovation

We assume the innovator is fully informed about the existing patents, i.e., their type, age and applicability. The innovator has to cite the patents related to the innovation as they represent the knowledge embedded in the new patent. However, the probability that those exact $m+1$ patents with the appropriate types $\mu_{k,k=1,\dots,m+1}$ are already discovered is zero. The innovator has the possibility instead to use discovered compatible ideas and cite the associated patents. In fact, for each of the potential patents $\mu_{k,k=1,\dots,m+1}$ there will be a number of existing patents i who fit them, i.e., for whom $\mu_k \in F_i$. Any of those will be feasible to cite instead of the potential patents. Among all the compatible patents, the innovator cites the most recent, in line with the assumption of homogeneous productivity growth. Importantly, note that this assumption rules out any type of strategic consideration by the innovator. This is formalized in the following assumption on the aging of patents.

Assumption 4: Innovators use and cite the youngest patents among the set of compatible ones.

The whole citation process is then as follows:

1. In class s, a new idea j is drawn from the pool of potential ideas. Its type μ_j is drawn from a uniform distribution and broadness a_j from an unspecified distribution.
2. Idea j is optimally technologically related to m potential patents from within s (given by $\sigma(\mu_j)$) and one potential patent from outside s. However, the probability that those exact patents already exist is zero.
3. For each of the related potential patents, noted k, the innovator knows which existing patents are compatible to cite instead. These are for each citation k those patents i for which $\mu_k \in F_i$.
4. Out of all compatible patents, the innovator uses and cites the youngest.

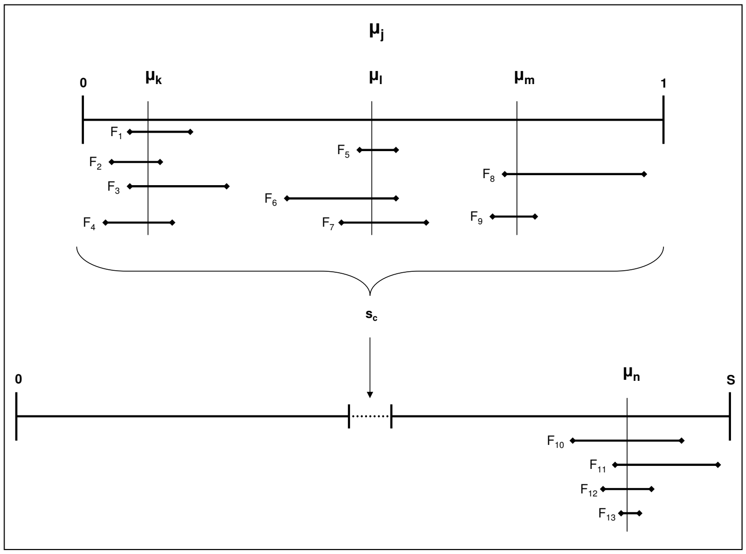


Figure 4: Citation Process for $m = 3$

Note that this process is the same for each of the $m + 1$ parents the innovator has to select.

Figure 4 illustrates the process for the case of $m = 3$. For visibility reasons, we drew the individual classes as straight lines instead of circles. In the figure, idea j of variety μ_j is technologically related to patents k, l, m from the same class as j and patent n from outside this class[12]. None of these patents exists, but for each there exist a set of other patents which are feasible to use and can be cited instead. The innovator will chose the youngest of those feasible patents (the age is not shown in the figure, so it cannot be seen which of those patents will actually end up being cited).

3.4 Citation Rates and Distribution

Within the innovation process outlined above, at time t a patent i within class s might be cited either by the patent which arrives within class s, or by any of the other $S - 1$ patents who arrive outside s. We first consider each probability separately and then combine them to obtain the total

[12]The continuous number of supports of patents from 1 to 13 is out of convenience only and does not imply any ordering of patents.

probability of being cited.

Within its own class, the probability that patent i born at time t_i is cited within t is the probability that it fits to any of the m patents which are considered for citation times the probability that patent i is the youngest patent within all patents which are feasible for citation. For each of the m citations, $\mu_{k=1,\ldots,m}$, the probability that i will fit is the probability that the type of the citation falls within F_i. For citation k this probability can be expressed as

$$\begin{aligned} Pr(\mu_k \in F_i) &= \frac{\mu_i + \frac{1}{2}a_i - (\mu_i - \frac{1}{2}a_i)}{1} \\ &= a_i \end{aligned} \tag{1}$$

Indeed, as these draws are random, it is the probability that μ_k will be within the range of applicability of patent i. And this is simply its broadness relative to the length of the support of the whole class s, which is 1.

Any patent i' such that $\mu_k \in F_{i'}$ competes for citation with i. The mean arrival rate of such patents i' competing with i is given by the expected value, over the possible values of $a_{i'}$, of $Pr(\mu_k \in F_{i'})$. This is in fact the average broadness level of patents within s, denoted $\bar{a}$. The probability that at time t patent i is the youngest patent within this interval is the probability that no other patent i' has arrived within the interval between the grant-date (birth) of patent i and t. As patents arrive as a Poisson Process with an arrival rate of 1, this is given by

$$Pr(t_i > t_{i'}) = e^{-\bar{a}(t-t_i)} \tag{2}$$

where t_i is the grant-date of patent i.

Finally, the probability of being cited within the same class is the probability that the number of forward citations increases within t. Define the number of forward citations patent i has received up to t as $k_i(t)$, and the probability that patent i will be cited as $\Pi(a_i, t_i, t)$, the probability that patent i will be cited from within its own class is

$$\Pi^{\in S}(a_i, t_i, t) = m \cdot a_i \cdot e^{-\bar{a}(t-t_i)} \tag{3}$$

The probability that patent i will be cited by any of the other $S-1$ new patents is found in a similar fashion. Of course, the probability that i fits has to take a much longer support into account:

$$\begin{aligned} Pr(\mu_{k=m+1} \in F_i) &= \frac{\mu_i + \frac{1}{2}a_i - (\mu_i - \frac{1}{2}a_i)}{S-1} \\ &= \frac{a_i}{S-1} \end{aligned} \tag{4}$$

On the other hand, the probability that patent i will be the youngest among all fitting patents is still given by equation (2). This implies that the probability of being cited from outside its own class is given by

$$\begin{aligned} \Pi^{\notin S}(a_i, t_i, t) &= (S-1) \cdot \frac{a_i}{S-1} \cdot e^{-\bar{a}(t-t_i)} \\ &= a_i \cdot e^{-\bar{a}(t-t_i)} \end{aligned} \tag{5}$$

It follows directly that the total probability that i will be cited within t is

$$\Pi(a_i, t_i, t)) = (m+1) \cdot a_i \cdot e^{-\bar{a}(t-t_i)} \tag{6}$$

The general behavior predicted by (6) seems in line with the observed citation behavior: Some patents have, for no observable reason (to us, differences in applicability) a higher probability of being cited than others, while *ceteris paribus*, older patents are less likely to be cited. Equation (6) cannot account for the observed knowledge diffusion leading to the usual hump-shaped relationship between age and probability of being cited as shown in figure 2. The reason is that our assumption on the aging of patents is too simple to take this into account. Specifically, we do not model the initial diffusion process.

We assume now that the number of citations a patent receives, k, is continuous and that the mean-field approximation is valid for our patent system. Consequently, we can express the probability that patent i will be cited as the continuous rate of change of $k_i(t)$[13]:

[13] On the validity and consequences of the mean-field approximation, see, e.g., Barabási, Albert, and Jeong (1999)

$$\frac{\partial k_i(t)}{\partial t} = (m+1) \cdot a_i \cdot e^{-\bar{a}(t-t_i)} \tag{7}$$

This allows us to calculate the number of forward citations patent i should have received by t given its broadness a_i by integrating (7) over t. Making use of the fact that $k_i(t_i) = 0$ by assumption we obtain the number of forward citations predicted by the model[14] for a patent of broadness a_i, born at t_i:

$$k_i(t) = (m+1)\frac{a_i}{\bar{a}}\left(1 - e^{-\bar{a}(t-t_i)}\right) \tag{8}$$

Equation (8) states that patents with a high relative applicability and older patents have higher numbers of (expected) forward citations. As $t - t_i$ goes to infinity, the term in brackets in (8) converges to 1. Consequently, for each individual patent i, (8) predicts that the (expected) number of citations received converges to $(m+1)\frac{a_i}{\bar{a}}$ as its age goes to infinity. Taking a_i as given and assuming that new patents are added sequentially, the grant-dates, t_i, are uniformly distributed[15]. This implies that asymptotically, as $t \to \infty$, the conditional probability distribution of forward citations will have a mass point at $(m+1)\frac{a_i}{\bar{a}}$. Formally,

$$f(k|a_i) = \delta[k - (m+1)\frac{a_i}{\bar{a}}] \tag{9}$$

where $\delta[\cdot]$ is the Dirac delta function[16]. Letting a_i now vary over patents, asymptotically the distribution of citations in our model will follow the distribution of the a_i's, i.e., the distribution of applicability, $g(a)$.

4 Fit of the Model

Section 2 has described two stylized facts associated with the patent citation network. The hazard rates of citations that are implied by the data have a very characteristic form, as does the asymptotic

or Jackson and Rogers (2007)

[14]We can drop the expectation operator, i.e., $E[k_i(t)]$, because of the mean-field approximation.

[15]This is again due to the mean-field approximation.

[16]The exact derivation is in the appendix.

distribution of forward citations. We now investigate how well the model predicts these two facts.

4.1 Citation Rates

We first focus on hazard rates for patents of a given age $t-t_i$, that is, the probability a patent issued at time t_i receives a further citation at time t. First, we need to define an appropriate and practical time scale[17]. As the model assumes that at each t, S new patents enter the system, the age of a patent is obtained by dividing the number of patents granted between patent's grant date (t_i) and t by the number of classes, which is 412 in our dataset.

Comparing the predicted probability of being cited, as given by (7), with the hazard rates in the data is problematic as the broadness of applicability of a patent is typically not reported in the data[18]. However, it is possible to eliminate broadness from the expression predicting the hazard rates. Indeed, equation (8) can be used to express broadness a_i as a function of the number of citations received by time t.

$$a_i = k_i(t) \cdot \bar{a} \cdot \frac{1}{(m+1)\left(1 - e^{-\bar{a}(t-t_i)}\right)} \tag{10}$$

The predicted hazard rate, $\frac{\partial k_i(t)}{\partial t}$, as given by equation (7), is then

$$\begin{aligned} \frac{\partial k_i(t)}{\partial t} &= (m+1) \cdot k_i(t) \cdot \bar{a} \cdot \frac{1}{(m+1)\left(1 - e^{-\bar{a}(t-t_i)}\right)} \cdot e^{-\bar{a}(t-t_i)} \\ &= k_i(t) \cdot \bar{a} \cdot \frac{e^{-\bar{a}(t-t_i)}}{1 - e^{-\bar{a}(t-t_i)}} \end{aligned} \tag{11}$$

Equation (11) is the relationship, to be tested, between the hazard rate and the number of forward citations predicted by our model for patents of age $t-t_i$. With the exception of $\bar{a}$ all the variables in equation (11) are contained in the data. Consequently, we adjust $\bar{a}$ to fit the data. For each patent, we observe its age at the time of each received citation, and we know for each citation whether it

[17] Patents are observed at each citation and at the end of the dataset.
[18] Although by assumption, it is known to the innovator.

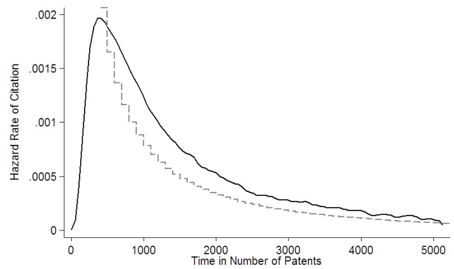
is the first citation, the second, and so on. Therefore we can calculate for each patent i the hazard rate of citation corresponding to i's observed age. Alternatively, we keep the number of citations received ($k(t)$) fixed, and have a look at a cross-section of all patents who have received this number of forward citations, irrespective of their age. We will then be able to plot the expected hazard rates of citation over different ages of patents, conditional on their level of forward citations, $k(t)$. This is the approach we opt for.

Note again that we are not aiming at replicating the upward-sloping part of the citation function, which seems to be due to knowledge diffusion. We are interested in how well we fit the downward-sloping part of the citation function. The graphs in figure 5 show the hazard rates of citation (solid black lines) estimated non-parametrically from the data, and the hazard rates of citation predicted by our model for patents which have received 1, 5, 10, and 20 citations respectively[19]. Note that the non-parametric rates, being population hazards, are decreasing over time partly due to the frailty effect, but this same effect should be present in our predicted hazard rates.

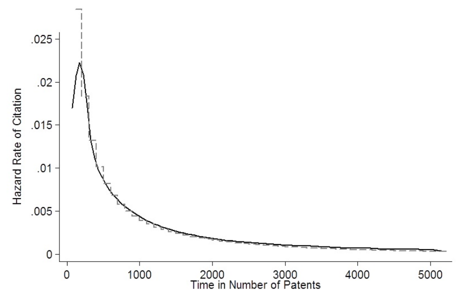
In the model, no explicit assumptions on the distribution of applicability across patents, apart from the assumption about its mean, $\bar{a}$, have been made. Furthermore, the assumption on aging is very simple: innovators prefer to cite younger patents and patents arrive as a Poisson process. The citation rates in figure 5 were obtained from our model assuming $\bar{a} = 0.00037$, implying a low average applicability, which seems reasonable, considering our very strong form of aging. The predicted hazard rates appear very close to the ones in the data, which we deem qualitative evidence in favor of our model.

It is difficult to obtain statistical evidence on how well our model fits the hazard rates of citation in the data, as these are not directly observable. Neither is the heterogeneity of applicability of patents, which is a major driving force of citations in our model. However, there are two distinct features of our predicted citation rates as given in equation (7) whose presence can be tested with a survival analysis model: (i) patents age exponentially, and (ii) citation rates should be influenced by heterogeneity of patents.

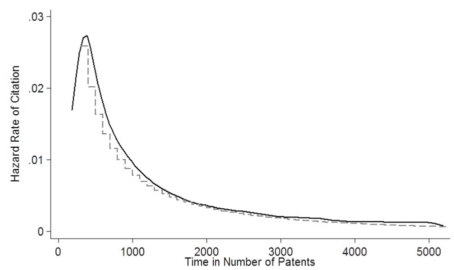
[19]The rates predicted by our model were averaged over a time-frame of 100 units of time.



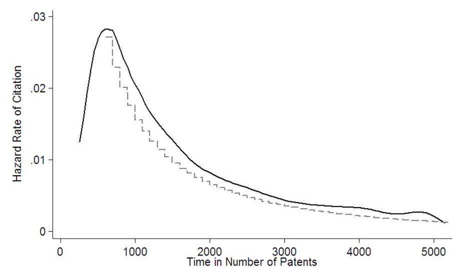
(a) 1 Citation Received



(b) 5 Citations Received



(c) 10 Citations Received



(d) 20 Citations Received

Figure 5: Probability of Citation by Data and Model

Our model predicts that the hazard rate of citation is of a proportional hazard form, i.e., the total hazard can be separated into one term depending on the age of a patent (the so-called baseline hazard), and another part depending on individual characteristics. In our case this part depends only the applicability of a patent. Readily available functional forms for estimating a proportional hazards model are the *Exponential Model*, which assumes that the hazard is independent of time, the *Weibull Model*, which models the effect of time as $\rho t^{\rho-1}$, and the *Gompertz Model*, in which time influences the hazard exponentially. Neither of these models allows for a hazard rate that is hump-shaped, as observed in the population hazard rates. However, as we only aim at fitting the downward sloping part, out of all these possibilities, clearly the Gompertz model is closest to our predicted hazard. In fact, given a set of characteristics $\mathbf{x_j}$ that possibly influence the hazard rate, the Gompertz model expresses the hazard rate of citation as

$$h(t|\mathbf{x_j}) = exp(\gamma t)exp(\beta_0 + \mathbf{x_j}\beta_\mathbf{x})$$

where γ determines the shape of the baseline hazard and t, the time since the onset of the risk of being cited, is to be interpreted as the age $(t - t_i)$ of a patent. A negative γ implies an exponential aging of patents. Furthermore, the estimated γ from a Gompertz model would coincide with the average applicability of patents in our model, $\bar{a}$. If our model correctly predicts the hazard rates of being cited, we would expect a Gompertz model to show a better fit of the data than either the Exponential or the Weibull model. We also would expect a negative value for γ. Table 3 shows the log likelihoods of all three models, together with the shape parameters (ρ and γ) influencing the baseline hazard where applicable, and the Akaike Information Criterion (AIC). The AIC is defined as

$$AIC = -2lnL + 2(k + c)$$

where k is the number of covariates and c the number of model-specific distributional parameters. When running the regression, in each model we include category-dummies, to control for different citation practices across categories. As citation practices have changed over the years (there appears to be a trend to cite more patents), we also control for the grant-year of the patents. As we categorize patents into six categories, one of which is our reference group[20], this implies that for us, $k = 6$. The exponential model only has one specific distributional parameter, while both the Gompertz and the Weibull have two, which implies that $c = 1$ for the exponential model and $c = 2$ for both the Weibull and Gompertz models. To focus on the downward-sloping part of the hazard function, we only consider patents that are at least 2 years old.

The Gompertz model performs slightly better than the Weibull model both in terms of the lower AIC. In addition, the value of γ estimated by the Gompertz model is very close to the value of $\bar{a} = 0.00037$ which we used to construct the hazard rates in figure 5.

The regressions summarized in table 3 do not take unobserved heterogeneity of patents into

[20]The reference group is category "Others".

Table 3: Aging of the Patent Population

	Shape Parameter	log likelihood	AIC	Observations
Exponential	n/a	363,046	-726,078	914,414
Weibull	0.33382	397,609	-795,202	914,414
Gompertz	-0.00036	396,454	-792,893	914,414

Both shape parameters are significant at the 1%-level.

Table 4: Aging of Individual Patents

	Shape Parameter	θ	log likelihood	AIC
Exponential	n/a	0.90423	587,080	-1,174,145
Weibull	0.44724	0.86125	609,761	-1,219,504
Gompertz	-0.00030	0.86523	610,250	-1,220,482

All parameters are significant at the 1%-level.

The regressions were run over $149,975$ patents and $914,414$ observations.

account, which may bias the results. However, all three models can be easily adjusted to include it, using a shared frailty model, which assumes that unobserved characteristics of a patent (i.e., its applicability) are present. All three models find significant evidence for unobserved heterogeneity. It is assumed that the heterogeneity of patents follows a Gamma distribution, and θ is the estimated variance of this distribution. As can be seen from table 4, all three models predict similar variances of unobserved heterogeneity. The Gompertz model is now preferred to the Weibull by both its log likelihood value and the AIC criterion, while the estimate of γ remains very close to the average quality value used to construct the hazard rates in figure 5. Indeed, these results suggest that *individual* hazard rates are decreasing over time exponentially, as predicted by our model.

4.2 Distribution of Citations

The observable outcome of the citation process is the distribution of forward citations, which we wish to match. From equation (9), we know that in our model the conditional probability distribution of forward citations as $t \to \infty$ is given by the Dirac delta "function":

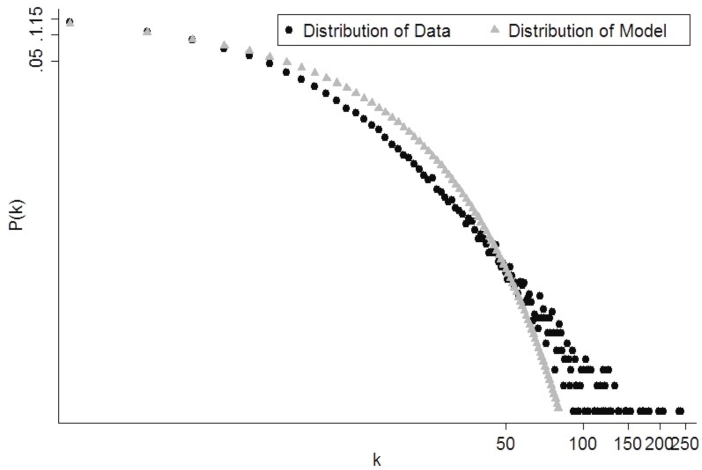


Figure 6: Citation Distributions: Data vs. Model

$$f(k|a_i) = \delta[k - (m+1)\frac{a_i}{\bar{a}}].$$

To arrive at the unconditional probability distribution, we integrate $f(k|a_i)$ over the probability distribution of applicabilities, $g(a)$, and find that

$$f(k) = \frac{\bar{a}}{m+1} \cdot g\left(\frac{k\bar{a}}{m+1}\right). \tag{12}$$

We can now perform a sort of calibration exercise. Set $m+1$ to be the average number of citations made in the data, which is 7.7. The distribution of applicabilities of patents will determine the form of the probability distribution of citations received in the data, and from the data we know that we are looking for a $g(a)$ with a mean-value of around 0.00037. We need to optimize on the set of distributions with such a mean-value in order to match the empirical distribution of patents. Among typical distributions, the Gamma-distribution with a shape-parameter of 0.85 and a scale-parameter of 0.0004, is consistent with the mean of 0.00037. A comparison between the distribution of forward citations in the data and the one predicted by the model is shown in figure 6.

Figure 6 was obtained assuming an average number of citations made of 7.7. For different values of m, the shape- and scale-parameters have to be adjusted accordingly to match the distribution of citations as well. We do not make any claim that this distribution of applicability would be the only distribution leading to the observed distribution of forward citations. In particular, it does not fit the curvature of the distribution of the data perfectly, and it is not able to reproduce the occurrence of patents having received more than 100 citations. However, these cases are rare. In fact, the correlation between the probability of a specific number of forward citations from the data and the probability estimated using this particular Gamma distribution is 0.9935, and the mean absolute error between the two distributions in figure 6 is only 0.13%-points. In particular, although figure 6 shows that we are unable to fit the fat tail of the distribution, there are very few patents observed with such high numbers of citations. Indeed, out of our more than $200,000$ patents, only 57 have received 100 citations or more. Overall, more than 99.9% of all the patents in our data have received no more than 80 citations. It might therefore be argued that the failure to correctly predict the distribution of those patents with very high numbers of forward citations can be disregarded.

As we mentioned in the introduction, there exists a considerable literature in economics that argues that the number of citations received by a patent is indicative of its underlying quality. To the best to our knowledge, this is the first paper that proposes an attachment process resting on the difference in the expected value of the patents for future research - captured by its applicability. In this way, the distribution of forward citations is linked to the underlying distribution of intrinsic differences in applicability.

5 Citations Within / Across Classes

So far we have assumed that a unique applicability parameter governs both the citations made to patents within the same class and the citations to different classes. In the current section we relax this assumption. Indeed, it could be argued that some patents, related to general purpose technologies, might be more broadly useful outside their own class than other patents. Our model can be modified to account for this possibility.

5.1 Theory

Our modified model assumes that the citation process driving citations across classes is the same as described in section 3, albeit allowing for two different types of applicability:

$$\left.\frac{\partial k_i^I(t)}{\partial t}\right|_{j\in s} = m \cdot a_i^I \cdot e^{-\bar{a}^I(t-t_i)} \tag{13}$$

for the hazard rate of citation from within the same class, and

$$\left.\frac{\partial k_i^O(t)}{\partial t}\right|_{j\notin s} = a_i^O \cdot e^{-\bar{a}^O(t-t_i)} \tag{14}$$

which is the hazard rate of citation from other classes. Equations (13) and (14) imply that intra and infra-class citations follow

$$k_i^I(t) = m\frac{a_i^I}{\bar{a}^I}(1 - e^{-\bar{a}^I(t-t_i)}) \tag{15}$$

and

$$k_i^O(t) = \frac{a_i^O}{\bar{a}^O}(1 - e^{-\bar{a}^O(t-t_i)}) \tag{16}$$

respectively. Therefore, following the same steps as before, we find the probability distributions of citations received:

$$f^I(k) = \frac{\bar{a}^I}{m} \cdot g^I\left(\frac{k\bar{a}^I}{m}\right), \tag{17}$$

$$f^O(k) = \bar{a}^O \cdot g^O(k\bar{a}^O). \tag{18}$$

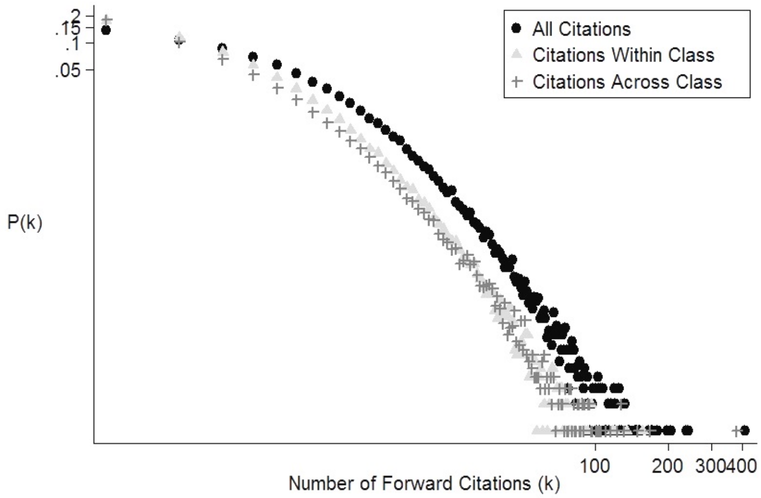


Figure 7: Citation Distributions: Within vs. Across Classes

5.2 Empirical Evidence

If each patent is characterized by two applicability levels that are distributed independently of each other, we would expect to observe that the distribution of citations from within a class and the distribution of citations across classes are also different. Figure 7 compares the two distributions to the overall distribution of all citations made. At first sight, the two distributions appear very similar. To "test" this observation we tried to reproduce the distribution with the same gamma function as the one used for the whole data set (controlling for the fact that there are more citations within a class than across classes), an approach that is equivalent to assuming that $a_i^I = a_i^O$ and $g^I(\cdot) = g^O(\cdot)$. We find that our fit is best for the whole dataset. The mean absolute errors increase from 0.13%-points for all citations to 0.26%-points for within class citations and to 0.27%-points for across class citations. We consider this as evidence that there is scope for improving the predictive power of the model by allowing $a_i^I \neq a_i^O$ and/or $g^I(\cdot) \neq g^O(\cdot)$.

As an additional test of the equality between a_i^I and a_i^O, we computed the implied applicabilities within and across classes from the data, making use of equations (15) and (16):

Table 5: Fit of Applicabilities

	AIC for a_i^I	AIC for a_i^O	Number of citations made
Intra-Class Citations	-10,424.67	75,369.38	404,826
Infra-Class Citations	142,216.2	-74,228.21	363,323

$$a_i^I = k_i^I(t) \cdot \bar{a}^I \cdot \frac{1}{m(1 - e^{-\bar{a}^I(t-t_i)})}, \tag{19}$$

$$a_i^O = k_i^O(t) \cdot \bar{a}^O \cdot \frac{1}{1 - e^{-\bar{a}^I(t-t_i)}}. \tag{20}$$

Under the hypothesis that each patent is characterized by one unique applicability parameter, we should find that both (19) and (20) are equally good in estimating citation hazard rates. In particular, the value of a_i^I from (19) should provide a valid regressor for estimating the hazard rate of infra-class citations, and vice versa. This hypothesis is rejected by the data. We ran a number of Gompertz regressions to estimate the intra- and infra-class hazard rates of citations separately. When we considered the intra-class citation process individually, the fit of our model, as measured by the AIC criterion, improved substantially when the computed value of a_i^I was included as a regressor, as opposed to a_i^O. The same holds true for infra-class citations, where the model's fit is improved considerably if a_i^O is used as a regressor instead of a_i^I. We measure the fit of the model as before with the AIC-criterion, and the results are briefly summarised in table 5.

6 The Role of Popularity

In this section we both allow for different levels of applicabilities and relax the assumption that innovators are perfectly informed about the applicabilities of all existing patents. In lieu of the assumption of perfect information, we assume that innovators have no information on the applicabilities of patents they cite from other classes. For intra-class citations we retain the assumption of perfect information. Under these circumstances, it appears natural to assume that innovators "estimate" the relevant infra-class applicabilities making use of their knowledge of within-class applicabilities and possibly the number of infra-class forward citations.

We investigate whether we find evidence in the data that popularity is used as an indicator of applicability. The possibility of this being the case is raised in part by the distribution of infra-class citations in figure 7. On the one hand, it is obvious that neither of the citation processes is of a linear preferential attachment form, that is, a process where the probability of citing a patent would be proportional to its number of forward citations. If it was, the corresponding distribution should be a straight line. Nevertheless, the distribution of infra-class citations in figure 7 has a lower degree of curvature than the distribution of intra-class citations. This leaves the possibility that preferential attachment plays a role in the citation process, and that this role is more important for citations across classes than for citations within classes.

6.1 Theory

If preferential attachment is supposed to play a role in a model in which citations are based on quality, it has to be the case that innovators are not perfectly informed about the applicability of all patents. We introduce uninformed innovators with the following assumption:

Assumption 5: While innovators are perfectly informed about the intra-class applicability of patents, they have no knowledge about the broadness of applicability of patents across classes. They have perfect information on the types and age of all patents in existence.

The first implication of assumption 5 is that the citation process within classes remains as described in section 3, and patents face a hazard rate of being cited from within their class given by (13). The infra-class citation hazard rates will follow (14), in which a_i^O is unknown to the innovator. The best he can do is to "estimate" a_i^O, using his knowledge of a_i^I and k_i^O. Applying the relationship between a_i^O and k_i^O from (20), the innovator forms his expectation according to

$$E[a_i^O] = \epsilon a_i^I + (1-\epsilon) k_i^O(t) \cdot \bar{a}^O \cdot \frac{1}{\left(1 - e^{-\bar{a}^O (t - t_i)}\right)}. \tag{21}$$

where $\epsilon \in [0, 1]$ is a parameter.

With $E[a_i^O]$ known, the innovator forms an expectation over which infra-class patents are feasible for him to cite. From this set of (possibly) feasible patents, he will cite the youngest. Innovators become only aware of the actual applicability of the cited patent after they patent the idea. As they based their citation decision on $E[a_i^O]$ and not on the actual a_i^O, it is possible that $\mu_{m+1} \notin F_i^O$. I.e., in such a case the cited patent is not compatible with the related idea μ_{m+1}, as described in section 3. If such a citation is included, the patented idea will incur a loss in its productivity.

6.2 Direct Empirical Evidence

If innovators do use forward citations as indicators of patents applicability, we would expect to find

1. a positive correlation between the number of forward citations and the proportion of citations from another class (as forward citations beget forward citations), and
2. that hazard rates of citations within classes do not depend on the number of previous citations, while the hazard rates across classes do.

In particular we want to explore whether our poor fit of the patent citation distribution for very highly cited patents might be induced by popularity playing a role in the citation process. If popularity does play a role for the infra-class citation process only, we might observe a snowball effect in infra-class citations. If the highest numbers of forward citations in our data are the outcome of such a snowball effect, we would expect that they are infra-class citations.

Figure 7 is a first indication against the conjecture that high numbers of citations received are overwhelmingly infra-class citations. The largest numbers of citations seem to be just as likely to be from within classes than across classes. A more detailed analysis is pursued in the appendix. Indeed,

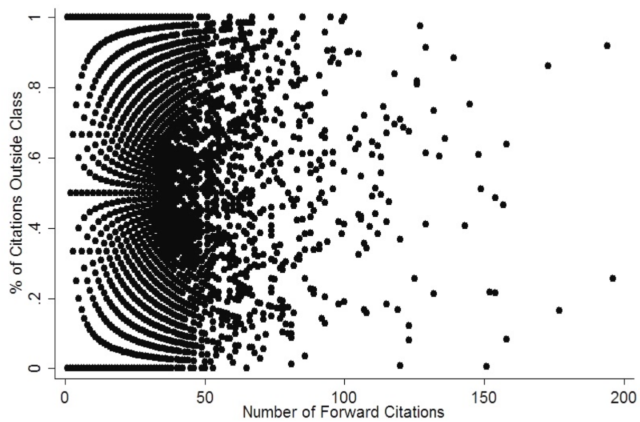


Figure 8: Forward Citations vs. Percentage Outside Own Class

from the citation distributions of selected individual classes shown in appendix C, it can be seen that the effect is not due to the fact that there is too much aggregation over classes when considering the total data set.

Statistical evidence of the existence of a snowball effect in infra-class citations can be sought from a correlation between the number of received citations and the percentage of received citations that are infra-class. If a snowball effect is present in infra-class citations, we expect positive correlations of these two variables. Contrary to this, in the data we find that the correlation coefficient between the total number of forward citations and the percentage of infra-class citations are not significantly different than zero. The exact coefficient is 0.0391. To better visualize the lack of correlation, figure 8 shows the scatter plot of the number of forward citations against the percentage of infra-class citations received.

6.3 Hazard Rates and Popularity

The second implied effect is that hazard rates across classes should depend on the number of previous citations, even if we control for their applicability. The role of popularity in a patent citation network has been investigated empirically also by Fafchamps, Goyal, and van der Leij (2010), who test for statistical evidence of a snowball effect. Indeed, they cannot reject the hypothesis that popularity

Table 6: Importance of Preferential Attachment

	All Citations			Within-Class Citations			Across-Class Citation		
citations	1.039 ($8.9\cdot10^{-5}$)	1.036 ($1.2\cdot10^{-4}$)	1.038 ($9.8\cdot10^{-5}$)						
overall applic.		$7.86\cdot10^{53}$ ($3.15\cdot10^{54}$)							
citations in class				1.055 ($1.6\cdot10^{-4}$)	1.052 ($2.7\cdot10^{-4}$)	1.055 ($1.6\cdot10^{-4}$)			
applic. in class			$3.48\cdot10^{23}$ ($8.51\cdot10^{23}$)		$2.19\cdot10^{25}$ ($9.25\cdot10^{25}$)				$1.10\cdot10^{50}$ ($4.64\cdot10^{50}$)
citations across class							1.062 ($1.6\cdot10^{-4}$)	1.056 ($2.4\cdot10^{-4}$)	1.062 ($1.6\cdot10^{-4}$)
applic. across class						$4.61\cdot10^{19}$ ($2.02\cdot10^{20}$)		$7.13\cdot10^{47}$ ($2.25\cdot10^{48}$)	

The number of forward citations used in the regressions is always including the current citation. The qualitative results would be the same if the current citation was excluded.

All values are significant at the 1%-level, where the critical values have been adjusted to take the large sample size into account, according to Leamer (1978), p.114.

does play a role in the patent citation data. Notwithstanding their results, we wish to test whether popularity matters within our own model, i.e., if we directly control for applicability of patents and exponential aging.

We calculate the implied applicabilities of patents within a well-defined period (which we deem a pre-estimation period), using equations (10), (19) and (20), based on the assumption that all citations are made by informed innovators. We then estimate the hazard rates of citation of the same patents during a later period (our estimation period) with a Gompertz regression. We sampled all cited patents that have been granted between 1975 and 1985, and our pre-estimation period runs from 1975 to 1990. Our estimation period runs from 1991 to 1999 and consists of $50,286$ cited patents and a total of $211,866$ citation observations.

Controlling for categories of the cited patents[21] and their grant-years, we also add as regressors the number of current forward citations and the implied applicability values from the pre-estimation period. We differentiate between forward citations received from within and across classes. We expect the relative importance of these regressors to give an indication of whether intra-class citations are more likely to be made by informed innovators than infra-class citations. If all citations are driven only by intrinsic applicability differences, we expect that neither infra- nor intra-class citations will be significant once we control for applicability. Our basic results are summarized in table 6

[21] As before, we use the category "Others" as a reference category.

The results from table 6 reject our hypothesis that the likelihood of citation depends solely on applicability differences, and not on popularity. No matter which citation process (overall, only within classes, only across classes) we consider, controlling for patent applicability never renders the effect of the number of forward citations insignificant. However, there are interesting differences between both the magnitudes of the reported hazard ratios as well as their standard errors, which would be consistent with the idea of two separate citation processes governing intra- and infra-class citations. The number of forward citations has a much bigger impact (in terms of hazard ratio) on infra-class citations than it does on total or intra-class citations. Our earlier findings that $a_i^I \neq a_i^O$ are confirmed in so far that the impact of applicability on the hazard rate is always biggest if the "correct" applicability parameter is used.

We ran a number of further regressions as robustness checks for the results in table 6, and these gave us qualitatively the same results. In particular, we find that as we increase the technological distance between patents (e.g., if we consider only the citations made to patents in a different 1-digit category, instead of a different class), the impact of popularity on the hazard of being cited increases.

Overall, our results indicate that there are subtle differences between the citation processes within and across classes. Popularity of patents appears to influence the likelihood of citation no matter which process is considered, even if we explicitly control for applicability. However, it also appears that this influence is increasing in the technological distance between citation pairs. These findings are generally in line with assuming that both intra- and infra-class citations are sometimes used as indicators for quality, and that information about patent applicability is decreasing in technological distance.

7 Conclusions

Existing research on patterns of patent citations either focuses on using the citation network as a signal of geographic or technological spillover, or analyzes the evolution of the network itself. In the first case the network is exogenously given, while in the second case patent citations are assumed to be governed by an automatic rule, as for example preferential attachment, without providing an underlying microeconomic model to such a process. A stylized fact of patent citations is that patents with high current numbers of forward citations are also more likely to be cited again in the future. Without a microfounded model, however, it is not possible to determine whether patents with a high number of forward citations are intrinsically better or whether their citations are the result of a rich-get-richer effect driven by popularity only. Clearly, the distinction is important as in the later case the number of citations does not reflect quality and the existence of these "sunspots" may introduce inefficiencies in the process of innovation.

In the present paper we first build a simple theoretical model of innovation in which popularity does not matter. This model explains most of the observed pattern of patent citations. The model assumes that new ideas arrive to innovators as a Poisson process. The new ideas build on existing ones and a successful new idea needs to cite these parental ideas. The decision to use and cite a parent is determined by two factors. First, the parent needs to be useful as a building block of the new idea. In fact, typically many existing ideas can be parent. Indeed, ideas are characterized by their broadness of applicability. This is a measure not of the productivity of a patent, but of their compatibility with other ideas. Second, among the compatible ideas, the innovator chooses to cite the most recent one. This assumption is a way to model that the most productive parent is chosen, and it presumes a widespread productivity growth in the economy. It also implies that patents age.

Statistically, there is strong evidence for heterogeneity in patents affecting the citation process. The citation process outlined above is able to closely match observed hazard rates of citations, assuming a mean for applicability of patents of 0.00037. In our model, the functional form of the citation distribution is determined by the distribution of patent applicabilities. Adjusting this distribution to take a Gamma-form with a mean of applicability of 0.00037, we are able to match the observed distribution of citations closely, with a mean absolute error of only 0.13%-age points.

Our model predicts exponential aging of patents and that the strength of aging is determined by the mean value of applicability. The data support the hypothesis that patents age. First, we find that exponential aging provides a better statistical fit to the data than no, linear, or power law aging.

Second, we find that the parameter governing the strength of aging in the data is between 0.0003 and 0.00036. I.e., it is very close to the mean value of applicability of 0.00037 which we assumed to match the citation hazard rates.

In the second part of the paper, we investigate whether giving a role to popularity per se in the willingness to cite improves the fit of the model. Of course, in a model driven by applicability differences, the use of popularity as an indicator of applicability only makes sense if innovators are uninformed about the applicability of patents. We modify the baseline model assuming that innovators do not know the applicability of patents outside their own 3-digit class. The implication of this assumption is that innovators use the current number of infra-class forward citations as well as the intra-class applicability to estimate infra-class applicability of patents. This opens up the possibility of a rich-get-richer effect of popularity in which earlier citations might beget future citations.

We use various methods to test the importance of popularity. The direct empirical evidence is not conclusive. On the one hand, the distributions of infra- and intra-class citations appear very similar, which is in favor of a single citation process. On the other hand, the fit of the citation distribution predicted by our model is the lowest for the distribution of infra-class citations. This may indicate the presence of another factor than applicability in the citation process. We finally estimate the hazard rates of citations considering infra- and intra-class citations as two individual processes. We find that current numbers of forward citations matter for the hazard of citation within both processes, even if we control for expected patent applicability. This impact seems to increase with technological distance, while the impact of our computed applicability indicators decreases. This evidence is consistent with the idea that forward citations are used as applicability-indicators by a certain fraction of less informed innovators, and that this fraction increases with technological distance.

An obvious issue that has not yet been addressed in the present paper is the hump-shaped form of citation hazard rates. We believe the initial increase in the hazard rate of being cited to derive from knowledge diffusion about patents' applicabilities. The derivation of a model that incorporates knowledge diffusion and its empirical validity are issues left at present for future research.

References

Albert, R., and A.-L. Barabási (2002): "Statistical mechanics of complex networks," *Reviews of Modern Physics*, 74, 47–97.

Alcácer, J., and M. Gittelman (2006): "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations," *The Review of Economics and Statistics*, 88(4), 774–779.

Auerswald, P., S. Kauffman, J. Lobo, and K. Shell (2000): "The production recipes approach to modeling technological innocation: An application to learning by doing," *Journal of Economic Dynamics and Control*, 24, 389–450.

Barabási, A.-L., R. Albert, and H. Jeong (1999): "Mean-field theory for scale-free random networks," *Physica A*, 272, 173–187.

Csárdi, G., K. J. Strandburg, L. Zalányi, J. Tobochnik, and P. Érdi (2007): "Modeling innovation by a kinetic description of the patent citation system," *Physica A: Statistical Mechanics and its Aplications*, 374(2), 783–793.

Fafchamps, M., S. Goyal, and M. van der Leij (2010): "Preferential Attachment and Individual Heterogeneity," mimeo.

Ghiglino, C. (2010): "Random Walk to Innovation: Why Productivity Follows a Power Law," mimeo.

Hall, B. H., A. Jaffe, and M. Trajtenberg (2005): "Market Value and Patent Citations," *The RAND Journal of Economics*, 36(1), 16–38.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2002): "The NBER Patent-Citations Data File: Lessons, Insights, and Methodological Tools," in *Patents, Citations & Innovations: A Window on the Knowledge Economy*, ed. by A. B. Jaffe, and M. Trajtenberg, chap. 13, pp. 403–459. The MIT Press.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999): "Citation Frequency and the Value of Patented Inventions," *The Review of Economics and Statistics*, 81(3), 511–515.

Jackson, M. O., and B. W. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random are Social Networks?," *American Economic Review*, 97(3), 890–915.

Jaffe, A. B., M. Trajtenberg, and M. S. Fogarty (2002): "The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees," in *Patents, Citations & Innovations: A Window on the Knowledge Economy*, ed. by A. B. Jaffe, and M. Trajtenberg, chap. 12, pp. 379–401. The MIT Press.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (2002): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," in *Patents, Citations & Innovations: A Window on the Knowledge Economy*, ed. by A. B. Jaffe, and M. Trajtenberg, chap. 5, pp. 155–177. The MIT Press.

Kortum, S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65(6), 1389–1419.

LEAMER, E. E. (1978): *Specification Searches: Ad Hoc Inferences with Nonexperimental Data*. John Wiley & Sons, Inc.

MARCO, A. C. (2007): "The dynamics of patent citations," *Economics Letters*, 94, 290–296.

MEHTAI, A., M. RYSMAN, AND T. SIMCOE (2008): "Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion," .

VALVERDE, S., R. V. SOLÉ, M. A. BEDAU, AND N. PACKARD (2007): "Topology and Evolution of Technology Innovation Networks," *Physical Review E*, 76(5).

WEITZMAN, M. L. (1998): "Recombinant Growth," *The Quarterly Journal of Economics*, CXIII(2), 331–360.

A Aggregation of Patent Classes

Table 7: Classification of Patent Classes

Cat. code	Category name	Subcat. code	Subcategory name	Patent classes
1	Chemical	11	Agriculture, Food, Textiles	8, 19, 71, 127, 442, 504
		12	Coating	106, 118, 401, 427
		13	Gas	48, 55, 95, 96
		14	Organic Compounds	534, 536, 540, 544, 546, 548, 549, 552, 554, 556, 558, 560, 562, 564, 568, 570
		15	Resins	520-528, 530
		19	Miscellaneous - Chemical	23, 34, 44, 102, 117, 149, 156, 159, 162, 196, 201-205, 208, 210, 216, 222, 252, 260, 261, 349, 366, 416, 422, 423, 430, 436, 494, 501, 502, 510, 512, 516, 518, 585, 588
2	Computers & Communications	21	Communications	178, 333, 340, 342, 343, 358, 367, 370, 375, 379, 385, 455
		22	Computer Hard- & Software	341, 380, 382, 395, 700-702, 704-710, 712-714
		23	Computer Peripherals	345, 347
		24	Information Storage	360, 365, 369, 711
3	Drugs & Medical	31	Drugs	424, 514
		32	Surgery & Medical Equipment	128, 600-602, 604, 606, 607
		33	Biotechnology	435, 800
		39	Miscellaneous - Drugs& Medical	351, 433, 623

4	Electrical & Electronics	41	Electrical Devices	174, 200, 327, 329-332, 334-338, 392, 439
		42	Electrical Lightning	313-315, 362, 372, 445
		43	Measuring and Testing	73, 324, 356, 374
		44	Nuclear and X-rays	250, 376, 378
		45	Power Systems	60, 136, 290, 310, 318, 320, 322, 323, 361, 363, 388, 429
		46	Semiconductor Devices	257, 326, 438, 505
		49	Miscellaneous - Elec.	191, 218, 219, 307, 346, 348, 377, 381, 386
5	Mechanical	51	Materials Processing & Handling	65, 82, 83, 125, 141, 142, 144, 173, 209, 221, 225, 226, 234, 241, 242, 264, 271, 407-409, 414, 425, 451, 493
		52	Metal Working	29, 72, 75, 76, 140, 147, 148, 163, 164, 228, 266, 270, 413, 419, 420
		53	Motors, Engines & Parts	91, 92, 123, 185, 188, 192, 251, 303, 415, 417, 418, 464, 474-477
		54	Optics	352, 353, 355, 359, 396, 399
		55	Transportation	104, 105, 114, 152, 180, 187, 213, 238, 244, 246, 258, 280, 293, 295, 296, 298, 301, 305, 410, 440
		59	Miscellaneous - Mechanical	7, 16, 42, 49, 51, 74, 81, 86, 89, 100, 124, 157, 184, 193, 194, 198, 212, 227, 235, 239, 254, 267, 291, 294, 384, 400, 402, 406, 411, 453, 454, 470, 482, 483, 492, 508

6	Others	61	Agriculture, Husbandry, Food	43, 47, 56, 99, 111, 119, 131, 426, 449, 452, 460
		62	Amusement Devices	273, 446, 463, 472, 473
		63	Apparel & Textile	2, 12, 24, 26, 28, 36, 38, 57, 66, 68, 69, 79, 87, 112, 139, 223, 450
		64	Earth Working & Wells	37, 166, 171, 172, 175, 299, 405, 507
		65	Furniture, House, Fixtures	4, 5, 30, 70, 132, 182, 211, 256, 297, 312
		66	Heating	110, 122, 126, 165, 237, 373, 431, 432
		67	Pipes and Joints	138, 277, 285, 403
		68	Receptacles	53, 206, 215, 217, 220, 224, 229, 232, 383
		69	Miscellaneous - Others	1, 14, 15, 27, 33, 40, 52, 54, 59, 62, 63, 84, 101, 108, 109, 116, 134, 135, 137, 150, 160, 168, 169, 177, 181, 186, 190, 199, 231, 236, 245, 248, 249, 269, 276, 278, 279, 281, 283, 289, 292, 300, 368, 404, 412, 428, 434, 441, 462, 503

Source: Hall, Jaffe, and Trajtenberg (2002), p.452-454, adapted by the authors.

B Derivation of the probability distribution of citations

We follow the argument of Albert and Barabási (2002), starting from the equation of the number of citations received:

$$k_i(t) = (m+1)\frac{a_i}{\bar{a}}\left(1 - e^{-\bar{a}(t-t_i)}\right) \tag{22}$$

Assuming that a_i is the same for all patents, the only possibility that a patent has received less than a given number of citations, k, is that it has been born after a certain time, i.e., t_i is bigger than a certain value. In particular:

$$\begin{aligned} k &= (m+1)\frac{a_i}{\bar{a}}\left(1 - e^{-\bar{a}(t-t_i)}\right) \\ k\frac{1}{m+1}\frac{\bar{a}}{a_i} &= 1 - e^{-\bar{a}(t-t_i)} \\ -\bar{a}(t-t_i) &= ln\left[1 - k\frac{1}{m+1}\frac{\bar{a}}{a_i}\right] \\ t_i &= t + \frac{1}{\bar{a}}ln\left[1 - k\frac{1}{m+1}\frac{\bar{a}}{a_i}\right] \end{aligned} \tag{23}$$

which implies for the probability that $k_i(t) < k$, conditional on a_i:

$$F\left(k_i(t) < k|\, a_i\right) = F\left(t_i > t + \frac{1}{\bar{a}}ln\left[1 - k\frac{1}{m+1}\frac{\bar{a}}{a_i}\right]\right) \tag{24}$$

Note that $ln(\cdot)$ is only defined for values of $k\frac{1}{m+1}\frac{\bar{a}}{a_i} < 1$.

Applying the mean-field approximation, we assume that new patents are added sequentially, one patent in each of the S patent classes per t. Therefore the probability density of t_i is:

$$F(t_i) = \frac{S}{m_0 + t} \tag{25}$$

Here, m_0 is the number of patents we assume were present at the beginning of time, which we leave unspecified.

Substituting equation (25) into equation (24), we get:

$$\begin{aligned} F\left(t_i > t + \frac{1}{\bar{a}}ln\left[1 - k\frac{1}{m+1}\frac{\bar{a}}{a_i}\right]\Big|\, a_i\right) &= 1 - F\left(t_i \leq t + \frac{1}{\bar{a}}ln\left[1 - k\frac{1}{m+1}\frac{\bar{a}}{a_i}\right]\right) \\ &= 1 - \frac{S\cdot t}{m_0+t} - \frac{S}{m_0+t}\left(\frac{1}{\bar{a}}ln\left[1 - k\frac{1}{m+1}\frac{\bar{a}}{a_i}\right]\right) \end{aligned} \tag{26}$$

The conditional probability density for the number of citation received, $f(k|a_i)$ can be obtained using:

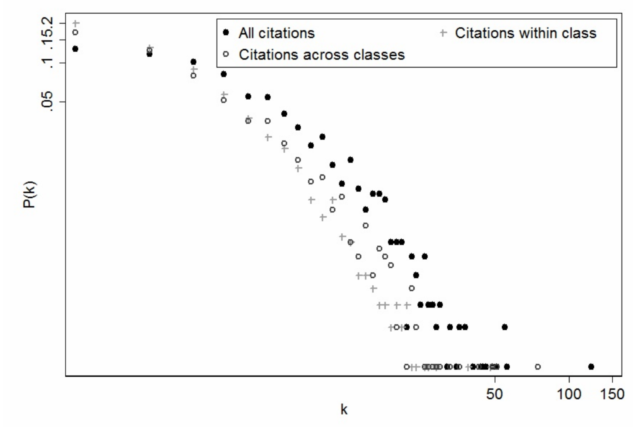
$$\begin{aligned} f(k|a_i) &= \frac{\partial F(k_i(t) < k|a_i)}{\partial k} \\ &= \frac{S}{(m_0+t)(m+1)}\cdot\frac{1}{a_i}\cdot\frac{1}{1 - k\frac{1}{m+1}\frac{\bar{a}}{a_i}} \end{aligned} \tag{27}$$

This expression goes to 0 as $t \to \infty$ for any value of k except $k = \frac{a_i}{\bar{a}}(m+1)$, for which the denominator of the final term goes to 0, offsetting the fact that the first term goes to 0. This is the reason for why we observe that the conditional probability density for $f(k|a_i)$ is determined by the dirac delta function, as stated in the main text:

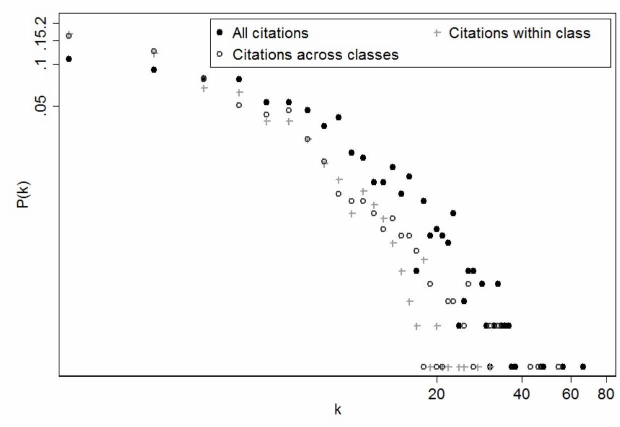
$$f(k|a_i) = \delta[k - (m+1)\frac{a_i}{\bar{a}}] \tag{28}$$

C Distribution of Citations by Classes

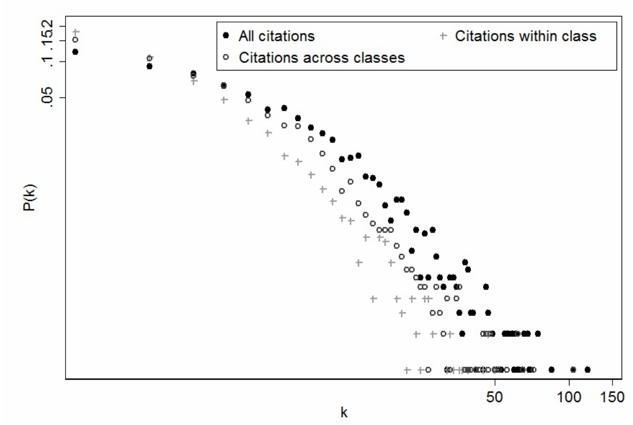
Below are citation distributions for four random patent classes. The class labeled “Radiant Energy” belongs to the category Electrical & Electronics and a total of 2,124 parents in our dataset belong to this class. The class “Communications: Electrical” belongs to Computers & Communications and 1,573 cited patents are classified as it. “Stock Material or Miscellaneous Articles” is categorized as Others and encompasses 3,893 parents, while “Drug, Bio-Affecting and Body Treating Compositions” belongs to Drugs & Medical and is the largest class out of the four shown here, with 5,161 cited patents.



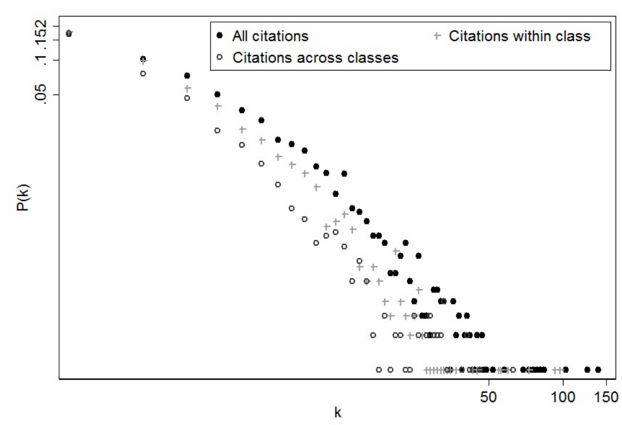
(a) Radiant Engergy



(b) Communications: Electrical



(c) Stock Material or Miscellaneous Articles



(d) Drug, Bio-Affecting and Body Treating Components

Figure 9: Citation Distributions by Class

Overall, the individual citation distributions mirror the picture given by the citation distribution of the dataset as a whole. There is by and large no sign of snowball-effects affecting individual classes.