DISCUSSION PAPER SERIES

DP13982

OUT OF SIGHT: A STUDY OF UNCITED PATENTS

Neil Gandal, Michal Shur-Ofry, Michael Crystal and Royee Shilony

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

OUT OF SIGHT: A STUDY OF UNCITED PATENTS

Neil Gandal, Michal Shur-Ofry, Michael Crystal and Royee Shilony

Discussion Paper DP13982
Published 03 September 2019
Submitted 01 September 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

OUT OF SIGHT: A STUDY OF UNCITED PATENTS

Abstract

Scientific understanding of innovation processes and of the patent system increasingly relies on big data analyses of patent citations. Much of that research focuses on highly cited patents. This study, conversely, offers the first systematic exploration of uncited patents—patents that receive no citations. Analyzing data on all US patents issued between 1976 and 2008, we focus on the ratio of uncited patents out of all patents granted each year. We track the changes in the percentage of uncited patents during that period, and across technological fields, controlling for patents' age. We also investigate traits of uncited patents by examining the association between lack of citations and various factors including the number of inventors, number of technological subclasses, number of backward citations, and number of claims in the patent. We find a robust pattern whereby the percentage of uncited patents declined between 1976 and the mid 1990s, but has been significantly increasing since then. These findings are consistent across technological fields and hold after controlling for patent characteristics. We discuss these and additional findings, and propose possible explanations. We suggest that the trend of increase in uncited patents raises, and reinforces, concerns regarding patent quality and "patent explosion". More broadly, our focus on "negative information" embedded in patent data opens up a new avenue for further research that can deepen our understanding of the patent system.

JEL Classification: N/A

Keywords: Uncited Patents, Patent Citations, networks, Big Data, Negative Knowledge, Innovation

Neil Gandal - gandal@post.tau.ac.il
Tel Aviv University and CEPR

Michal Shur-Ofry - michalshur@mail.huji.ac.il
Hebrew University

Michael Crystal - crystal1@mail.tau.ac.il
Berlglas School of Economics

Royee Shilony - royee.shilony@mail.huji.ac.il
Hebrew University

Out of Sight: A Study of Uncited Patents

Neil Gandal,[±] Michal Shur-Ofry,[±] Michael Crystal,[ϒ] Royee Shilony[Ψ]

Abstract

Scientific understanding of innovation processes and of the patent system increasingly relies on big data analyses of patent citations. Much of that research focuses on highly cited patents. This study, conversely, offers the first systematic exploration of uncited patents—patents that receive no citations. Analyzing data on all US patents issued between 1976 and 2008, we focus on the ratio of uncited patents out of all patents granted each year. We track the changes in the percentage of uncited patents during that period, and across technological fields, controlling for patents' age. We also investigate traits of uncited patents by examining the association between lack of citations and various factors including the number of inventors, number of technological subclasses, number of backward citations, and number of claims in the patent.

We find a robust pattern whereby the percentage of uncited patents declined between 1976 and the mid 1990s, but has been significantly increasing since then. These findings are consistent across technological fields and hold after controlling for patent characteristics. We discuss these and additional findings, and propose possible explanations. We suggest that the trend of increase in uncited patents raises, and reinforces, concerns regarding patent quality and "patent explosion". More broadly, our focus on "negative information" embedded in patent data opens up a new avenue for further research that can deepen our understanding of the patent system.

Keywords: Uncited Patents; Patents; Patent Citations; Patent Quality; Networks; Big Data; Negative Knowledge; Innovation

[±] Tel Aviv University, Berglas School of Economics; gandal@tauex.tau.ac.il
[±] Hebrew University of Jerusalem Law Faculty; michalshur@huji.ac.il.
[ϒ] Tel Aviv University, Berglas School of Economics and CEPR
[Ψ] Hebrew University of Jerusalem Law Faculty
We are grateful to Reut Ginat for excellent research assistance.

Table of Contents

1. Introduction

Patent data is increasingly recognized as an important source of knowledge about innovation. A growing body of literature in economics, business management, network science, and the legal field, suggests that analyses of patent data can provide ample information about innovation processes, about the traits of specific inventions, and about technological domains (for pioneering studies see, e.g., Schmookler, 1966; Griliches, 1984; Trajtenberg, 1990). This growing awareness, together with developments in big data analyses have led to an upsurge in the exploration of patent databases in the recent decades (e.g., Lanjouw & Schankerman, 2001; Allison et al., 2004; Erdi at al., 2013).

A significant part of the research in this area is devoted to the study of patent citations, and specifically to the study of highly cited patents (e.g. Trajtenberg: 1990; Harhoff et al., 1999; Hall et al., 2005). The perception underlying this line of scholarship is that patent citations reflect technological relations between the citing and the cited patents (Jaffe et al, 2003). Therefore, a large number of follow-on citations received by a patent is perceived as an indication of the impact, quality, social value, and even breakthrough nature, of the patented technology (e.g., Fleming, 2001; Dahlin & Behrens 2005; Erdi at al., 2013; Arts & Veugelers 2015; Poncheck 2015; Risch 2015).

Yet, there is another side to patents citations that can provide important information about innovation and about the patent system, but has been essentially ignored in the literature to date: patents that do not receive any citations. This paper provides the first systematic exploration of uncited patents. We analyze all U.S. patents with complete data granted between 1976 and 2008 (3,079,587 observations). We focus on the ratio of uncited patents out of all patents granted each year, and track the changes in this percentage during a period of more than three decades.

We also inquire whether there are differences in uncited patents across technological fields, comparing, primarily, the fields of pharmaceuticals and software-related patents. We then perform an initial inquiry as to the association between an "uncited status" and various factors including the number of backward citations, the number of inventors, the number of subclasses, the number of claims in the patent, and the degree of similarity between the uncited patents and the prior art which they cite.

Our findings reveal a robust pattern in which the ratio of uncited patents out of all granted patents decreased between 1976 and the mid-1990s, but has been significantly increasing since then. Graphically, then, the ratio of uncited patents across the period of our study forms a rough "U" shape.[1] These findings are robust after controlling for patents' age, and are consistent across technological fields. While we cannot offer a decisive explanation for this trend, because citations are perceived as an indication of technological quality and impact, these findings are concerning.

Interestingly, we observe differences among technological fields in the ratio of uncited patents. Prominently, the percentage of uncited patents is significantly higher for pharmaceutical drug patents, in comparison to patents in the field of software and communication.[2]

We also find that uncited patents are *negatively* associated with the number of backward citations, the number of subclasses, the number of claims in the patent, the number of inventors,[3] and the "degree of similarity" between the patent and the cited prior art.[4] In other words, large numbers of backward citations, subclasses, claims, and inventors, and a higher "degree of similarity" between the patent and the cited prior art increase a patent's chances of being cited.

Since numbers of inventors, claims, subclasses and backward citations have been recognized in the literature as positively related to highly cited

[1] See the graphs in Section 3 *infra.*

[2] For the purpose of our study we broadly define this field to include patents in information and communications technology (ICT), fin-tech and med-tech — see Section 3.2 *infra*.

[3] The finding pertaining to inventors only applies when we include self-citations as citations, but not when we exclude self-citations from the analysis – see the discussion in Section 3.3, *infra*..

[4] As explained below, the associations with number of inventors and "degree of similarity" is weak.

patents, these latter findings are not surprising. Yet, they provide an additional indication that the quality of uncited patents may indeed be lower, in comparison to cited patents. More importantly for our study, controlling for these characteristics, the "U" shaped pattern described above continues to hold.

The general picture emerging from our study in concerning. Out of all patents granted by the USPTO, the percentage of uncited patents across domains has been increasing significantly since 1996 and until 2008. Since patent citations are an indications of impact and quality, and given the patent system's mission to serve as a vehicle for promoting valuable innovation, the trend we identify is troubling. We discuss possible explanations for this trend, and locate our findings in the context of the literature concerning "patent explosion" and patent quality.

More broadly, our focus on the negative information embedded in patent repositories opens up a new research avenue in the analysis of patent data. While our study constitutes a first step in this exploration, it certainly does not exhaust all possible applications of this approach. We therefore call for follow-on studies that would shed further light on this area, and complement the research on the positive aspects of patent citations. Such studies could provide us with a deeper and more nuanced understanding of the patent system, as well as the innovation processes it seeks to promote.

This Article proceeds as follows: Section 2 provides the theoretical background for the ensuing examination. It begins by reviewing the primary research on patent citations and citation patterns, and then proceeds to briefly review the legal literature on patent quality that is relevant for the following analysis. Section 3 describes our dataset, methodology, and findings, which are also graphically presented in a series of figures and detailed in technical appendices. Section 4 discusses the potential significance of these findings, and sketches, in broad strokes, potential directions for further research.

2. Theoretical Background

Our study of uncited patents draws on several strands of literature. The first is the large body of economic, network science, and legal scholarship that explores patent data, and specifically patent citations. These studies are based on the understanding that mining the data recorded in patent repositories—which include, *inter alia*, information about patents' inventors, classification into technological subclasses, number of claims and additional factors—can provide ample information about innovation processes and about the traits of certain inventions (e.g., Trajtenberg, 1990; Jaffe at al. 2003; Erdi at al., 2013).

A prominent thread within this body of research is devoted to patent citations, and particularly to highly cited patents. Patent citations are citations of prior art pertaining to the invention. These citations are commonly comprised of previous patents, and infrequently also of scientific literature. Because prior art plays a crucial legal role in the decision whether the invention deserves patent protection, the citation of relevant prior art is required as part of submitting a patent application (e.g., Erdi et al., 2013). Those citations are reviewed by the Patent Office examiners, who often contribute additional citations (e.g., Alcacer et al., 2009). Citations, therefore, reflect relations between inventions: broadly speaking, *backward citations*—citations made by an invention—reflect the previous technologies related, or providing building blocks to the invention, while *forward citations*—citations received by an invention—reflect its impact on subsequent technologies (Trajtenberg et al., 2003b; Lanjouw et al., 2001).

Thus, forward citations have become an acceptable, if noisy, indication for the technical quality of inventions (e.g., Strandburg et al., 2006). The underlying assumption is that if the technology embedded in the patent is valuable for technological progress, future patents relying on that technology would cite the original patent (cf. Trajtenberg, 1991, Fleming, 2001; Erdi et al., 2013; Arts & Veugelers 2015).

Indeed, numerous studies found positive correlations between large numbers of forward citations and various external indications of value, including expert evaluations of the patented inventions, payment of patent renewals fees, amounts of R&D investment, and the likelihood of the patent being involved in litigation (see, respectively, Alberta et al., 1990; Harhoff et al., 1999; Trajtenberg, 1990; Allison et al., 2004). Others investigated various traits of highly cited patents, in an effort to understand the factors associated with successful inventions (e.g. Wuchty et al. 2007; Fleming, 2001; Arts & Veugelers 2015 and the discussion *infra*).

Additional studies, of particular relevance to our research, concentrate on patent citation *patterns*. These studies examine temporal changes that occur in patent citations over the years in order to draw broader insights about the patent system and about innovation processes. Several studies indicate that patents reach the peak of their citations during the early years after patent grant, with certain variations among technological domains (Hall at al., 2002; Mehta et al., 2010).

Kuhn et al. (2018) detect a recent change in patent citation patterns whereby a small minority of patent applications are generating a large majority of patent citations, and argue that the technological similarity between citing and cited patents has significantly weakened in recent years. Strandburg et al. observed an increasing gap between the least and most cited patents since the late 1980s, and suggested that this increase may result from issuing patents on more trivial advances, of lesser technical value. A subsequent work showed that this trend has leveled around 2000 (Strandburg et al., 2006; Csardi et al., 2009).

Several studies have shown that newer patents have more backward citations (i.e., cite more patents) in comparison to older patents (e.g., Strandburg et al., 2009; Mehta et al., 2010). However, despite this trend, the average number of citations per patent has been declining over the years, which may suggest a decline in patent quality (Mehta et al., 2010). Concomitantly, numerous legal studies expressed concerns about the decline in patent quality. This scholarship observes that many patents do not reach commercialization, and are not actually used in practice (e.g., Keiff, 2001; Lemley, 2001; Sichelman, 2010). It argues that the grant of low-quality patents contributes to a sharp increase in the numbers of patent application and patents granted over the past decades, often referred to as a "patent explosion". It further submits that this phenomenon buttresses non-practicing entities (commonly known as patent trolls), and produces other negative externalities that overall hinder, rather than promote, innovation (e.g., Cotropia, 2014; Lemley, 2016).

The rich literature on patent citations largely focuses on information about cited patents, especially those that receive a large number of citations. Conversely, uncited patents, patents that receive no citations, have hardly received scholarly attention. This state of affairs is not surprising, given the general tendency to concentrate on positive aspects of knowledge and ignore negative information. However, negative information—in our case, information about patents with no citations—is essential for getting a full picture of the patent system, and more broadly of our innovation ecosystem (Shur-Ofry, 2016). While, several studies of uncited *scientific papers* were conducted in recent years (e.g., Larviere et al., 2009; Van Noorden, 2017) uncited patents have not been systematically explored. This article seeks to fill in this gap, by providing a first systematic inquiry of uncited patents. What percentage of patents receive no citations at all? Are there differences in these percentage across technological domains? Does the pattern change over time? And are there specific traits which can be associated with uncited patents? These are the questions which we seek to explore in the following analysis.

3. Data, Methods and Findings

In order to shed light on these questions, we analyzed data extracted from the USPTO database, on all US patents granted between 1976 and 2008.[5] Our data includes a total number of 3,079,587 patents.

For each year during in our study period, we identified the absolute numbers and percentage of patents granted that have not been cited. As we explain below, we control for patents' age and include citations for 10 years following patent grant.

We examined the temporal pattern of non-citation, namely the changes in the percentage of uncited patents across the period of our study, variations across technological fields, and associations between an "uncited" status and several factors that were identified in the literature as related to patents' impact and quality.

3.1 Temporal Pattern of Non-Citation

Previous studies indicate that patents usually reach their citation peak within approximately four years after issuance (Hall at al., 2002; Strandburg et al., 2006; Mehta et al., 2010). Nevertheless, an "uncited" status is never final. The exploratory and combinatory nature of innovation, together with the legal requirement to cite any relevant prior art, imply that a patent can receive citations at any point during its lifetime, and also after its expiry (Strandburg et al., 2006; cf. Fleming & Sorenson, 1992). Thus, older patents have more opportunities to gain citations. As a result, in order to accurately compare the percentages of uncited patents across the period of our study we needed to control for patents' age. We therefore considered citations received in the first ten years following the grant of the patent. Consequently, our data includes patents issued no later than 2008, so as to allow all patents in our database an equal ten-year period to gain citations.

Figure 1 shows the percentage of uncited patents during the period of our study, after controlling for patents' age

[5] http://www.patentsview.org/download/. A small number of patents (two percent) did not have complete data needed for the analysis; hence, they were not included in our data. This small amount of missing data is typical in such analyses and does not affect our results.

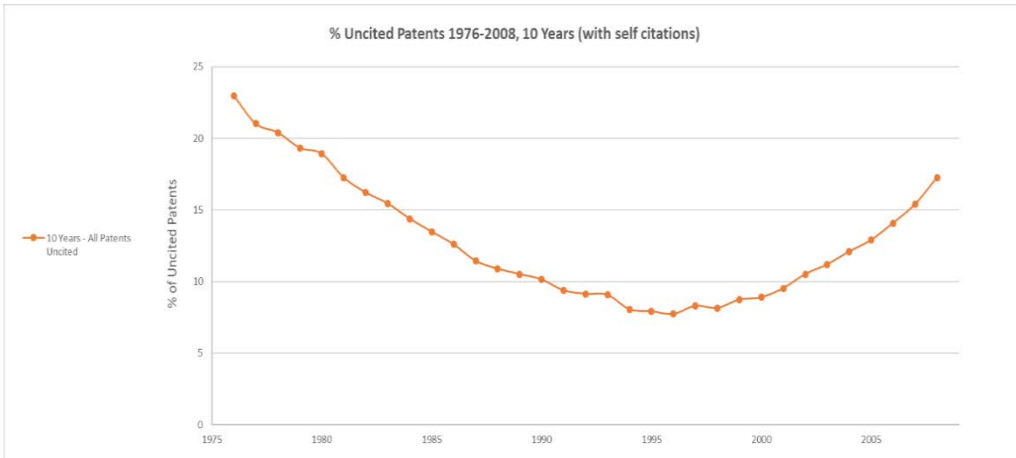

Fig. 1: Uncited Patents 1976-2008 (10 years)

The red curve in Figure 1 shows the ratio of uncited patents to total patents issued each year during our study period. The graph is U-shaped: From 1976 to (roughly) 1996, the percentage of patents that were uncited decreased over time from essentially 22 percent in 1976 to less than 8 percent in 1996. Since 1996, the percentage of uncited patents increased steadily over time, reaching 17 percent in 2008.[6] In section 4 we discuss the significance of this pattern and suggest possible explanations

3.2 Non-Citation Patterns across Technological Fields

Does this temporal pattern subsist across technological fields? Are there differences in the percentages of uncited patents among the different fields? We performed an initial inquiry of this question, by distinguishing between three categories of patents: (1) pharmaceutical drug patents (2) software-related patents, which we defined in a broad manner, as including patents in information and communications technology (ICT), fin-tech and med-tech, and (3) all other patents, i.e., patents not included in categories (1) and (2). In order to identify the patents which belong to each of these groups, we used the patent classification system, which assigns each patent into subclasses, in

[6] Interestingly, this trend is contrary to the trend observed in studies of scientific *papers*, which indicate that the percentage of uncited papers is consistently decreasing (see Larviere et. al., 2009, referring to the period 1900-2007).

accordance with the invention's technological features.[7] Appendix 1 details the subclasses included in our first and second categories for the purpose of our analysis.[8]

Figure 2 shows the percentage of uncited patents for each of the three categories, together with the general, all-patents data.

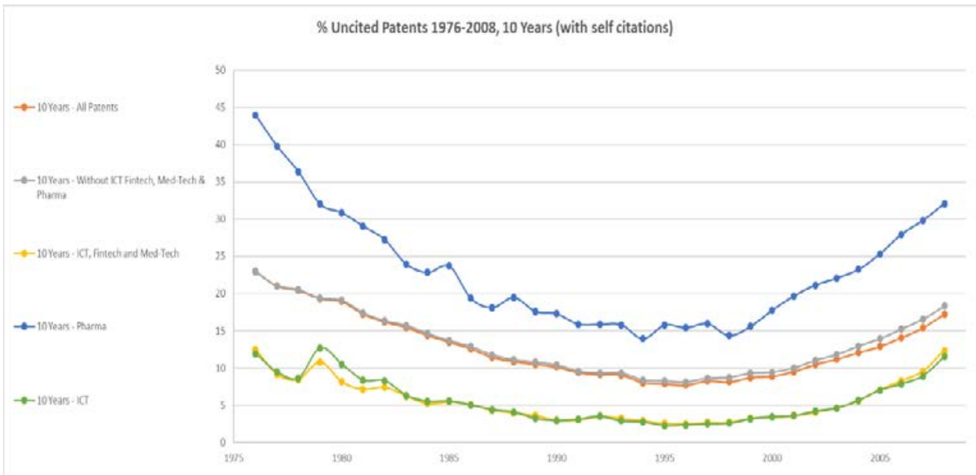

Fig. 2: Uncited patents by categories (10 years)

Similar to our overall data, the pattern of uncited patents in all the three categories—software-related (yellow curve) pharmaceuticals (blue curve), and others (grey curve)—exhibits a U-shape. The percentage of uncited patents in each of these categories decreased between 1976 and (roughly) 1996, and increased since 1996 and until the end of our period.

Despite the general similarity in the pattern, there are striking differences among the three groups in the percentages of uncited patents. The percentage of uncited patents in the pharmaceutical drugs field is significantly higher across the entire period, in comparison to software related patents. The

[7] For identifying subclasses relevant for pharmaceutical drug patents we used the U.S. PATENT & TRADEMARK OFFICE, OFFICE OF PATENT CLASSIFICATION, *available at* http://www.uspto.gov/patents-application-process/patent-search/classification-standards-and-development. For identifying subclasses relevant for med-tech, fin-tech, and information and communications technology patents we relied on previous compilations by Gandal et al., (2018) and Gandal & Cohen (2019.)

[8] The third category is a residual group and includes patents that are not included in the other two groups (for example, certain mechanical patents). Notably, because inventions can be classified into more than a single subclass our categories are not completely exclusive.

ratio of uncited patents in our third, residual, category lies somewhere between the two other categories, and closely tracks the overall non-citation ratio. The fact that a higher percentage of pharmaceutical patents remain uncited, in comparison to software related patents, is somewhat counter intuitive, given that patent protection in the pharmaceutical field is often perceived in the literature as more necessary and justified, in comparison to the field of software. In section 4 we suggest possible explanations for this result.

In the analysis, we included self-citations as citations. However, in order to check the robustness of our findings we relaxed this assumption and re-analyzed our data with self-citations excluded. Our econometric results are robust to excluding self-citations. Figure 3 below shows the percentages of uncited patents by categories, with self-citations excluded.

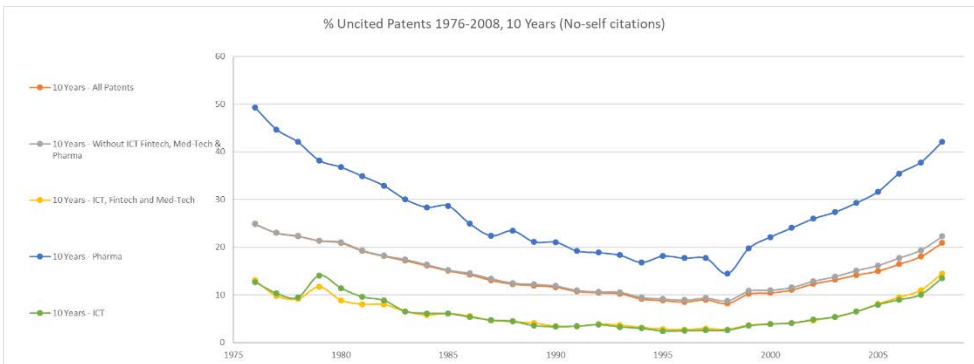

Fig. 3: Uncited patents by categories (10 years), no self-citations

As is apparent from Figure 3, while excluding self-citations somewhat raises the percentages of uncited patents, the temporal U-shape pattern for all categories remains unchanged.

3.3 Lack of Citations and Patent Traits

We further examined possible association between lack of citations and the following five traits, which we include as explanatory variables in our study:

1) Number of backward citations – the number of citations made by a patent to preexisting patents;

2) Number of subclasses –the number of subclasses to which the patent was assigned by the patent office examiners, in accordance with the invention's technological traits;
3) Number of claims listed on the patent, which is determined by the applicant;
4) Number of inventors appearing on the patent; and
5) Backwards similarity – a variable which reflects the degree of similarity between the patent and its backward citations.

The information about the first four factors appears on the patent. In order to calculate backwards similarity we used the similarity index developed in Lanjouw & Schankerman (2001). Simply put, under this index similarity is determined by the fraction of backward citations that belong to the same subclass as the citing patent, out of the total number of backward citations. Thus, the value of this variable is between 0 and 1. A similarity index that equals 1 implies that all cited patents belong to the same subclasses as the citing patent, while a value of 0 implies that all backward citations belong to subclasses that are different from the subclasses of the citing patent.

Intuitively, higher values of the first four factors would increase the likelihood of a patent to be cited. Indeed, these factors are perceived as indications of the invention's technological breadth, quality, and impact. Thus, studies found positive association between high numbers of forward citations received by a patent and its number of sub-classes (e.g., Schoenmakers, 2010; Yoshikane et al., 2012; Kelley et al., 2013), backward citations (e.g., Yoshikane et al., 2012; Kelley et al., 2013), and claims (e.g., Tong & Frame, 1994; cf. Lanjouw & Schankerman, 2004). Additional research indicates that patents produced by more than one inventor are more highly cited than those produced by a single inventor (Wuchty et al., 2007). The relations between backward similarity and patent's quality and impact are more nuanced according to existing studies (see, e.g. Phene et al., 2006; Nemet et al. 2012).

The following two tables provide summary data for these variables. Table 1 described the cited patents, while table 2 describes the uncited patents:

	# of Obs.	Mean	Std. Dev	Min	Max
Backward Citations	2,754,321	11.75095	20.36809	1	1328
Number of Inventors	2,754,321	2.242832	1.611063	1	51
Number of Subclasses	2,754,321	4.369612	3.207579	1	260
Number of Claims	2,754,321	15.65761	13.24943	1	887
Backward Similarity	2,754,321	0.6145774	0.3541722	0	1

Table 1: Descriptive Statistics: 1976-2008: Cited Patents

	# of Obs.	Mean	Std. Dev	Min	Max
Backward Citations	325,266	7.853425	10.57848	1	773
Number of Inventors	325,266	2.209788	1.611371	1	32
Number of Subclasses	325,266	3.981225	3.165645	1	164
Number of Claims	325,266	12.21976	9.965565	1	706
Backward Similarity	325,266	0.6070431	0.3809577	0	1

Table 2: Descriptive Statistics: 1976-2008: Uncited Patents

Tables 1 and 2 taken together show that overall 89 percent of the patents in this period received at least one citation in the 10 years following its issuance.

The data in Tables 1 and 2 further indicate that uncited patents have, on average, fewer backwards citations and less claims than cited patents, and also belong to fewer technological subclasses. Backward similarity values and the number of inventors are roughly the same for cited and uncited patents.

We then conduct the regression analysis. The details of our regression and formal econometric analysis appear in Appendix 2. The dependent variable is CITED, where CITED equals one if the patent receives one or more citations in the first ten years following its issuance. If the patent receives no citations during the first ten years following its issuance, CITED equals zero.

As mentioned above, the independent variables in the regressions are backward citations, inventors, subclasses, claims, and backwards similarity.

Before we run the regressions, we show the correlation among the variables used in our analysis: As can be seen from Table 3, in the raw data, there is a positive correlation between whether a patent is cited and the numbers of backward citations, subclasses, and claims.

	Cited	b. cites	inventors	subclasses	claims	b. similarity
Cited	1					
Backward citations	0.1072	1				
Number of inventors	0.0084	0.013	1			
Number of subclasses	0.0526	0.0649	0.0758	1		
Number of claims	0.0914	0.1904	0.1	0.0748	1	
Backward similarity	0.0065	-0.1143	-0.0098	-0.0621	-0.0396	1

Table 3: Correlations among variables

We then run the regression analysis (see Appendix 2.) From the regression analysis, we find a negative association between “uncited status” and all independent variables. Table 4 in the Appendix 2 shows that all of these associations are statistically significant.

Since numbers of claims, subclasses, inventors and backward citations have been previously recognized as positively related to highly cited patents, these findings are not entirely surprising. Yet, they provide an initial indication that the quality of uncited patents is indeed lower.

More importantly, the regression results in Appendix 2 demonstrate that our results concerning the “U” shape temporal pattern of uncited patterns continue to hold, even after we control for all these factors, regardless of whether we use the first or second dependent variable in the analysis. In other words: after controlling for a series of prevalent factors that might affect citations, the likelihood of not being cited decreases from 1976 to 1996, and increases consistently from 1996 to 2008.

3.4 A Long Tail of Patent Citations?

Finally, we take a look at our raw data, without controlling for patents’ age. As we explain above, patents may continue to receive citations at any point. While our 10-year limit is necessary in order to compare “apples and apples”, looking at the raw data without controlling for age can provide us with a different insight regarding the actual existence of citations beyond the 10-year period.

Figure 4 shows our results for patents that did not receive any citations, by categories, without controlling for patent's age.

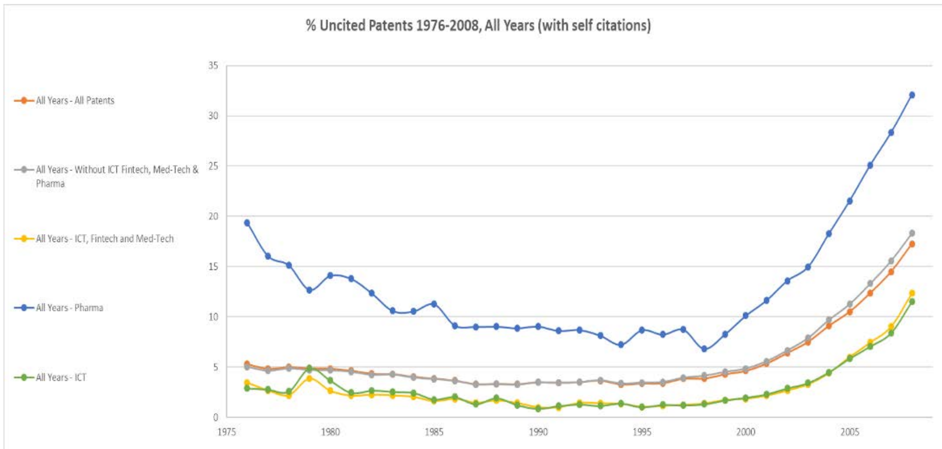

Fig. 4: Uncited patents by categories (all years, self-citations included)

Indeed, this figure compares "apples and oranges", in the sense that the oldest patents in our database had more than 40 years to gain citations, in comparison to the youngest patents, that had only 10 years. Despite this distorting factor, this figure still demonstrates a decrease in the percentage of uncited patents between 1976 and 1996, and a change of this trend from 1996 onwards. Interestingly, however, the percentage of "old" uncited patents drops significantly in all categories when we lift the 10-year limitation. For example, 22 percent of all patents issued in 1976 were uncited after 10 years (Figure 1), but only 4 percent remained uncited in 2018, 42 years from issuance. For pharmaceutical patents issued in 1976 the percentage of uncited patents drops from 44 percent after 10 years to roughly 20 percent after 42 years. Likewise, out of all patents issued in 1996, 8 percent were uncited after 10 years, but only 3.5 percent remained uncited in 2018, after the lapse of 22 years.[9]

The comparison between Figures 2 and 4 implies that even if patents remain uncited for ten years after grant, many of them can still receive at least *one* citation after that period. In other words, the patent citation tail is a "long tail". While we cannot offer a definitive explanation, these findings seem consistent with a recent stream of research, that regards innovation not as a

[9] Obviously, with respect to younger patents issued after 1996 the gap between uncited after "10 years" and uncited after "all years" narrows.

strictly cumulative and linear process, but rather as a combinatory, exploratory and networked process, whereby old technologies may gain new significance at a later stage (see, e.g., Fleming & Sorenson, 1992; Fleming, 2001; Strandburg et al., 2006; Shur-Ofry, 2017). This line of inquiry certainly warrants further research.

4. Discussion

What are the possible interpretations of our findings, and what is their significance for innovation policy?

Several factors may explain the U-shape temporal pattern. A plausible explanation for the left side of the U, namely the decline in the ratio of uncited patents between 1976 and 1996, may be the substantial improvement in patent search tools. During the first years of our period patents were published on paper, and had to be searched manually (Grigg, 2003). This state of affairs gradually changed with the introduction of a system that allowed a computerized CD-ROM-based search in 1988. In 1996 the USPTO launched a website that allowed internet-based search, which was first limited to bibliographic patent information, but since 2000 allowed to search the full-text of all patents issued from 1976 onwards (and some more limited search of earlier patents) (Grigg, 2003). This timeline is largely consistent with our findings that the percentage of uncited patents reached its lowest value around 1996.

The increase in the ratio of uncited patents since 1996 is possibly connected to the sharp increase in the overall number of patents issued by the USPTO during that period. In 1976, the first year of our study, the cumulative number of patents in the USPTO registry was approximately 4 million patents. In 2008 the number of patents exceeded 7 million.

Figure 5 shows the percentage of uncited patents against the *cumulative* number of US patents during the period of our study.

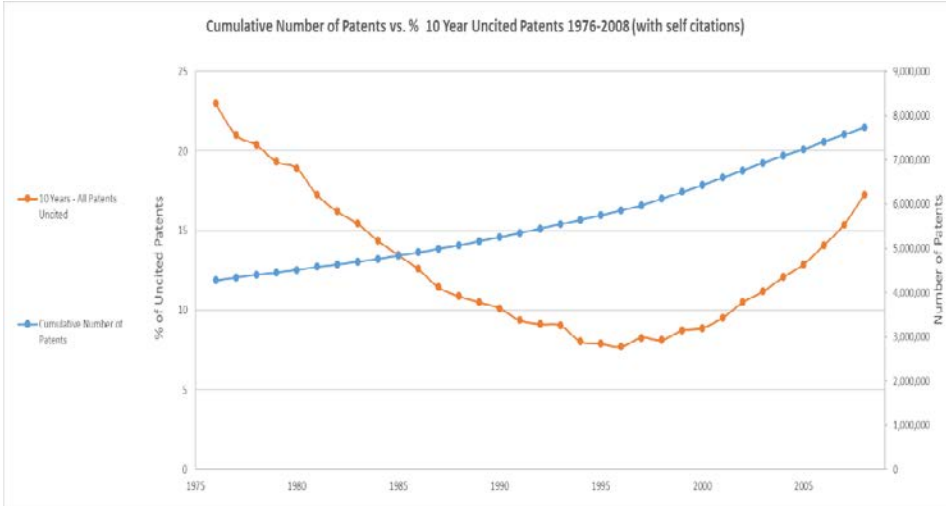

Fig. 5 —Cumulative number of US patents vs. Percentage of Uncited Patents 1976-2008

A closer look at the number of patents issued *per year* during our study period reveals a sharp rise in issued patents during the second half of our period, from the mid-1990s. Figure 6 shows the percentage of uncited patents against the yearly numbers of US patents issued during the period of our study.

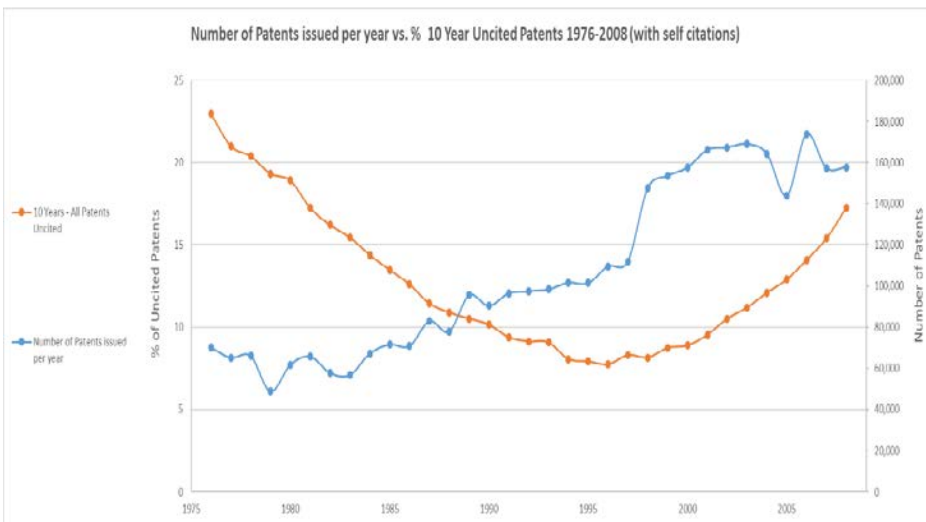

Fig. 6 —Yearly Numbers of Patents Issued vs. Percentage of Uncited Patents, 1976-2008

This significant increase in patents issued since the mid-1990s is largely consistent with the increase we find in the percentage of uncited patents during the same period. Hence, one possible interpretation of our results is that despite the improvement in search capabilities, the search for prior art becomes more difficult the more patents there are in the registry. Therefore, in an era of "patent explosion" more patents are left uncited.

A second interpretation, which is not contradictory to the first one, concerns patent quality. Numerous legal scholars expressed concerns about a decline in patent quality (e.g., Hemphill & Sampat, 2012; Cotropia, 2014; Lemley, 2016; Chien, 2018; Frakes & Wasserman, 2019). This literature generally maintains that the issuance of low quality patents is a major cause for the recent patent explosion. It argues that low quality patents often find their way to the hands of non-practicing entities (commonly known as patent trolls), create "patent thickets" that hinder technology commercialization, and produce other negative externalities that overall impede, rather than promote, innovation. The USPTO too declared patent quality as a priority, and has been considering initiatives to improve it.[10]

To a certain extent, the picture emerging from our study reinforces patent-quality concerns. Because citations are perceived as an indication of technological quality and impact, the increase in the percentage of uncited patents may imply an increase in low-quality patents that do not serve as building blocks for subsequent innovation. Moreover, our analysis reveals negative relations between uncited patents and several traits that are considered signals of patent quality, namely numbers of claims, backward citations, subclasses, and inventors.[11]

The differences we find in the percentage of uncited patents between categories, particularly between drug patents and software related patents are counter-intuitive. The pharmaceutical field is considered the "poster child" of the patent system, and the area in which strong patent protection is most justified (e.g., Oullette, 2010). Conversely, there is a longstanding policy debate that casts doubt on the justifications and necessity of software patents (e.g., González,, 2006; Bessen & Meurer, 2008; Bessen & Maskin, 2009). However, our analysis demonstrates that the percentages of uncited patents are consistently higher for drug patents in comparison to software-related patents.

[10] See https://www.uspto.gov/patent/patent-quality; https://www.uspto.gov/about-us/organizational-offices/office-commissioner-patents/office-deputy-commissioner-patent-19.

[11] Yet, we note that these factors are not exhaustive, and further studies could shed light on additional traits of uncited patents. We outline below several directions for such research.

One possible explanation may relate to the tendency of pharmaceutical companies to protect drugs by a series of secondary patents that do not involve new active ingredients, in order to prolong their life-cycle. Studies indicate that this phenomenon, also known as "ever-greening", has been growing over time. For example, Oullette (2010) observes that the average number of patents per drug increased from 2.5 in the late 1980s to nearly 3.5 in 2005. Kapczynski et al., (2012) find that around 50% of drugs are protected by secondary patents, while Feldman (2018) maintains that 78% of the drugs associated with new patents in the FDA's records between the years 2005 and 2015 were not new drugs coming on the market, but existing drugs that are "recycled" by their owners. These secondary patents are often considered low-quality patents (e.g. Hemphill & Sampat, 2012), which may explain the relatively high percentage of uncited drug patents. Another possible explanation could be that innovation in the area of drugs has less cumulative traits than in the software-related domains.[12] This hypothesis, however, requires further exploration.

The finding that the percentage of uncited patents in software-related fields is *lower* than the average for all technological fields is also surprising, in light of prominent criticisms that software patents are often trivial, and have limited use as sources for subsequent developments (e.g., Bessen & Meurer, 2009). One should note that our definition of software-related patents is broad and includes patents in the fields of information and communications technology, fin-tech and med-tech. Subject to this broad definition, our analysis indicates that software patents are related to subsequent technologies, no less (and even more) than other categories of patents. Notably, because the youngest patents in our database were issued in 2008, our results do not reflect the potential impact of the recent U.S. Supreme Court cases that raised the threshold of patentability in this field.[13]

Overall, although our study indicates that a substantial majority of patents receive at least one citation within 10 years from grant, the trend we identify is disturbing. If innovation is a networked process, the fact that more and more patents remain outside the network is a cause for concern.

Nevertheless, we cannot offer a single conclusive explanation for our findings, and one should be cautious in their interpretation. Our analysis is merely a first step, and our focus on uncited patents opens up myriad research questions, the exploration of which would allow to better understand the association between uncited-ness on the one hand, and lack of social or private

[12] Cf. Van Noordern (2017) who discusses uncited science and suggests that certain domains may be less "cumulative" than others.

[13] See, prominently, Bilski v. Kappos, 561 U.S. 593 (2010); Alice Corp. Pty. v. CLS Bank Int'l, 134 S. Ct. 2347 (2014). See also the analysis in Chien, 2018.

value, on the other. Future research could examine the links between uncited patents and a series of additional factors associated with *highly* cited patents. These factors include, for example, external expert evaluations (cf. Alberta et el., 1990), R&D investments in the underlying technologies (cf. Trajtenberg, 1990), payment of patent renewal fees (cf. Harhoff et al. 1999), or a high level of recombinations in the patent's backward citations and subclasses (e.g. Fleming, 2001; Arts & Veugelers 2015).

Similarly, our study examined citations of patents by subsequent patents. Yet patents can also be cited in scientific literature. Although such citations are quite rare (Glänzel et al., 2003), they indicate that the patent constitutes a source of knowledge, and therefore imply social value. Therefore, another research direction would be to examine whether, and to which extent, patents that receive no citations whatsoever from subsequent patents receive citations from scientific literature.

Finally, inventions could potentially have commercial value despite lack of patent citations. Thus, another direction for future exploration would be to cross the data on uncited patents with data concerning patent assignments and licenses, and with data about patents involved in litigation (in itself a proxy for commercial value—cf. Allison et al., 2004).

These directions for future research are of course non-exhaustive, but illustrate the broad potential of the approach we choose in this study.

5. Conclusion

Our systematic study of patents that receive no citations yields three principal insights. First, we find a robust U-shape pattern, whereby the percentage of uncited patents decreased between 1976 and 1996, but has been constantly increasing since then. Second, we find counter-intuitive differences in the rates of uncited patents between different technological fields, primarily drug patents and software-related patents. Third, our analysis reveals that uncited patents are negatively associated with several indications for patent quality. From the perspective of innovation policy, these findings are troubling. They raise, and reinforce, concerns regarding patent quality and "patent explosion".

On a more general note, this study's systematic focus on patents that are "out of sight" opens up new avenues for future research, and demonstrates how exploration of negative information embedded in patent data can provide us with important knowledge and a deeper, more nuanced, understanding of our ecosystem of innovation.

Appendices

Appendix 1: Categorization of Patents according to Classes

a. Drug-Related: USPC Classes 424, 514
b. Software Related Classes: we relied on a the list of classes previously compiled in Gandal et. al (2018) and Gandal & Cohen (2019), which includes ICT/Information Security (USPC), Fin-Tech (IPC) and Med-Tech (IPC)

Appendix 2: Details of Formal Econometric Analysis

The dependent variable is CITED. This variable takes on the value one if a patent receives one or more citations in the first ten years following its issuance. Since the dependent variable is a binary variable, we run a Logistic regression. The same qualitative results are obtained using a Probit regression. Because all independent variable except for backwards similarity are highly skewed, we enter these variables in logarithms. The independent variables included in the regression are

l_back_cites – the natural logarithm of the number of citations the patent made to preexisting patents
l_inventors – the natural logarithm of the number of inventors on the patent.
l_subclasses – the natural logarithm of the number of subclasses listed on the patent.
l_claims – the natural logarithm of the number of claims
b_similarity – the backward similarity as defined above.

Finally, we include dummy variables for the grant year. These are the primary variables of interest. We include data from 1976-2008; therefore we have dummy variables for each year from 1977-2008.

Regression Analysis

The estimation equation is as follows, where for compactness we do not list the dummy variables for year.

(I) $\text{cited}_j = \beta_0 + \beta_1 * \text{l\_back\_cites}_j + \beta_2 * \text{l\_inventors}_j * \text{l\_subclasses}_j + \beta_3 * \text{l\_claims}_j + \beta_4 * \text{b\_similarity}_j + \epsilon_j$

The results of the logistic regression are shown in Table 4:

Dependent variable: CITED	**Coefficient**	**Standard error**	**Z-Statistic**
Independent Variables			
l_back_cites	0.3778102	0.0021927	172.31***
l_inventors	0.0292338	0.0031838	9.18***
l_subclasses	0.2001326	0.00298	67.16***
1_claims	0.3211272	0.0022841	140.59***
Backward similarity	0.1530326	0.0051152	29.92***
3,079,587 observations	Pseudo R squared	0.04	

Table 4: Regression Results: (all "p-values less than 0.001)

All of the estimated parameters are highly significant (at the 99% level of confidence.)

The estimated coefficients on the yearly dummy (binary) variables from the regression in equation I are shown in Table 5 below.

Table 5: Coefficients on the yearly dummy (binary) variables

year	coefficient
1977	0.0805028
1978	0.0054316
1979	0.0325794
1980	0.0230138
1981	0.0641844
1982	0.1468927
1983	0.1651518
1984	0.2410426
1985	0.2755763
1986	0.3478947
1987	0.4455987
1988	0.4704984
1989	0.4827714
1990	0.5273918
1991	0.5905338
1992	0.6024416
1993	0.5846586
1994	0.6786498
1995	0.6754624
1996	0.6925536
1997	0.6064685
1998	0.6123309
1999	0.5263189
2000	0.4824541
2001	0.3941743
2002	0.2503819
2003	0.1428378
2004	0.0374031
2005	-0.0714987
2006	-0.1657945
2007	-0.2564237
2008	-0.3784648

Thus after controlling for patent characteristics, Table 5 shows that the pattern is exactly as in the raw data. From 1976-1996, more patents are being cited over time. This is because the estimated coefficients on the yearly dummy variables increase (essentially monotonically) over that period. From 1996 through 2008, fewer patents are cited over time and this decline is also essentially monotonic.

Robustness Analysis

We ran the following robustness regressions:

- We excluded self-citations, that is, citations to patents held by the same assignee.
- We include dummy variables for the eight IPC classes.
- We include dummy variables for software and pharmaceutical patents

The results are qualitatively unchanged.[14] In particular, the graph is still U-shaped. These results are available from the authors upon request.

[14] In the case where we exclude self-citations, the coefficient associated with the number of inventors is negative and significant, rather than positive and significant. But our main result (the U-shaped graph) continues to hold.

References

Aditi, Mehta, Rysman, Marc, and Simcoe, Tim, *Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion*, 25.7 JOURNAL OF APPLIED ECONOMETRICS 1179-1204 (2010).

Alberta, M.B., Averyb, D. , Narina, F. & McAllister, P., *Direct Validation of Citation Counts as Indicators of Industrially Important Patents*, 20 RESEARCH POL'Y 251 (1991)

Allison, John R., Lemley, Mark A., Moore, Kimberly A., & Trunkey, R. Derek, *Valuable Patents*, 92 GEO. L. J. 435 (2004)

Allison, John R., *Patent Value* in Peter Menell & David Schwartz eds., RESEARCH HANDBOOK ON THE ECONOMICS OF INTELLECTUAL PROPERTY LAW, VOLUME II (2018)

Arts, Sam and Veugelers, Reinhilde, *Technology Familiarity, Recombinant Novelty, and Breakthrough Inventions*, 24 INDUS. & CORP. CHANGE, 1215 (2015)

Beck, Roger L. Competition for Patent Monopolies, 3 RES. L. & ECON. 91 (1981)

Bessen, J. & Maskin, E., *Sequential Innovation, Patents, and Imitation* 40(4) RAND JOURNAL OF ECONOMICS, 611 (2009)

Bessen, James & Meurer, Michael J., PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK (Princeton University Press, 2008)

Chein, Colleen V., *Comparative Patent Quality*50 ARIZONA ST. L. J. 71 (2018)

Csardi, Gabor & Tobochnik, Jan & Erdi, Peter & Zalányi, László & Strandburg, Katherine. *Patent Citation Networks Revisited: Signs of a Twenty-First Century Change?* 87 NORTH CAROLINA L. REV. (2009)

Dahlin K.B. & Behrens, D.M. *When Is an Invention Really Radical?: Defining and Measuring Technological Radicalness*, 34(5) RESEARCH POL'Y 717 (2005)

Feldman, Robin, *May Your Drug Price Be Evergreen*, 5 J. L. & BIOSCI. 590 (2018)

Fleming, Lee & Sorenson, Olav, *Technology as a Complex Adaptive System: Evidence from Patent Data*, 30 RESEARCH POL'Y 1019 (2001)

Fleming, Lee, *Recombinant Uncertainty in Technological Search*, 47 MANAGEMENT SCI. 117 (2001)

Frakes, Michael & Wasserman, Melissa F., *Irrational Ignorance at the Patent Office*, 72 VANDERBILT L. REV. (2019)

Gambardella, A., P. Giuri, and M. Mariani, *The Value of European Patents. Evidence from a Survey of European Inventors*, Final report of the PatVal-EU Project. Pisa: Laboratory of Economics and Management (LEM), Sant' Anna School of Advanced Studies. Available at: www.alfonsogambardella. it/patvalfinalreport.pdf.

Gandal, N., Kuniesky, N., and L. Branstetter, 2018, *Network-Mediated Knowledge Spillovers: A Cross-Country Comparative Analysis of Information Security Innovations,* NBER Working Paper No. 23808.

Gandal, N., & Cohen, 2019, S. *Networks and Spillovers in Software in Israeli Hi-Tech*, CEPR discussion paper #13467, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3319788

Glänzel, W. & Meyer, M., *Patents Cited in the Scientific Literature: An Exploratory Study of Ereverseí Citation Relations*, 58 SCIENTOMETRICS 415 (2003)

Golden, J.M., *Innovation Dynamics, Patents, and Dynamic-Elasticity Tests for the Promotion of Progress* 24 HARV. J. L. & TECH. 47 (2010)

González, A.G.,. The Software Patent Debate, 1(3) JOURNAL OF INTELLECTUAL PROPERTY LAW & PRACTICE, 196 (2006)

Grigg, K., *The Evolution Of The Patent And Trademark Depository Library And The Role Of The Patent And Trademark Librarian In The Digital Age* (2003), available at https://peer.asee.org/12462

GRILICHES ZVI ED., R&D, PATENTS, AND PRODUCTIVITY (1984)

Hall, B., *Exploring the Patent Explosion*, 30 JOURNAL OF TECHNOLOGY TRANSFER, (2005)

Hall, B., Z. Griliches, and J. Hausman, *Econometric Models for Count Data with an Application to the Patents-R&D Relationship*, 52 ECONOMETRICA, 909 (1984)

Hall, B., Jaffe, A. Jaffe, and Trajtenberg, M., *Market Value and Patent Citations: A First Look*, 36 RAND JOURNAL OF ECONOMICS (2005)

Hall, B., Jaffe, A. Jaffe, and Trajtenberg, M., *The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools*, in JAFFE. ADAM B & TRAJTENBERG, MANUEL EDS., PATENTS, CITATIONS & INNOVATIONS: A WINDOW ON THE KNOWLEDGE ECONOMY (2003)

Harhoff, D., F., Narin, F. Scherer, and K. Vopel, *Citation Frequency and the Value of Patented Inventions*, 81(3) REVIEW OF ECONOMICS AND STATISTICS, 511-515 (1999)

Hemphill, C. Scott and Sampat, Bhaven N., *Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals* 31 J. OF HEALTH ECONOMICS (2012)

JAFFE. ADAM B & TRAJTENBERG, MANUEL EDS., PATENTS, CITATIONS & INNOVATIONS: A WINDOW ON THE KNOWLEDGE ECONOMY (2003)

Kapczynski A, Park C, Sampat B. *Polymorphs and Prodrugs and Salts (Oh My!): An Empirical Analysis of "Secondary" Pharmaceutical Patents* 7 PLOS ONE (2012)

Kelley, D. J., Ali, A., & Zahra, S. A. *Where Do Breakthroughs Come From? Characteristics of High-Potential Inventions* 30(6) JOURNAL OF PRODUCT INNOVATION MANAGEMENT (2013)

Kieff, Scott F., *Property Rights and Property Rules for Commercializing Inventions*, 85 MINN. L. REV. 697, 698 (2001)

Kuhn, Jeffrey M., Younge, Kenneth A., and Marco, Alan, *Patent Citations Reexamined: New Data and Methods*, available at http://dx.doi.org/10.2139/ssrn.2714954 (2018)

Kuhn & Kenneth A. Younge, *Patent Citations: An Examination of the Data Generating Process* ACADEMY OF MANAGEMENT PROCEEDINGS 2016(1):18173

Lanjouw, Jean O. & Schankerman, Mark, *Characteristics of Patent Litigation*, 32 RAND J. ECO. 129 (2001)

Lanjouw, Jean O. & Schankerman, Mark, *Patent Quality and Research Productivity: Measuring Innovation with Multiple* Indicators, 114 THE ECONOMIC JOURNAL 495 (2004)

Larviere, V., Gingras, Y. & Archmbault, E., *The Decline in the Concentration of Citations, 1900-2007*, 60(4) J. ASS'N OF INFORMATION SCIENCE AND TECHNOLOGY, 858 (2009)

Lemley, Mark A., *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495 (2001)

Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341 (2010)

Motohashi, Kazuyuki, *Licensing or Not Licensing? An Empirical Analysis of the Strategic Use of Patents by Japanese Firms*, 37 RES. POL'Y 1548, 1550 (2008)

Nemet, G.F., Johnson, E., *Do Important Inventions Benefit from Knowledge Originating in Other Technological Domains?*, 41(1) RESEARCH POL'Y (2012)

Ouellette, Lisa. L. *How Many Patents Does It Take To Make a Drug? Follow-On Pharmaceutical Patents and University Licensing*, 17 MICHIGAN TELECOMMUNICATIONS AND TECHNOLOGY L. REV. 299 (2010)

Risch, Michael, *Patent Troll Myths*, 42 SETON HALL L. REV, 457, 478 (2012)

Phene A, Fladmoe-Lindquist K, Marsh L. *Break Through Innovations in the U.S. Biotechnology Industry: The Effects of Technological Space and Geographic Origin* 27(4) STRATEGIC MANAGEMENT J. (2006)

Ponchek, Talya, *Does the Patent System Promote Scientific Innovation—Empirical Analysis of Patent Forward Citations*, 25 ALB. L.J. SCI. & TECH. 289, 320 (2015)

Saunders, Kurt M., *Patent Nonuse and the Role of Public Interest As A Deterrant to Technology Supression*, 15 HARV. J.L. & TECH. 389, 390 (2002)

Schmookler, Jacob Invention and Economic Growths (1966)

Schoenmakers, Wilfred & Duysters, Geert, *The Technological Origins of Radical Inventions*, 39 RES. POL'Y. 1051 (2010).

Shur-Ofry, Michal, *Access to Error*, 34 CARDOZO ARTS & ENTERTAINMENT LAW REVIEW 357 (2016)

Shur-Ofry, Michal, *Connect the Dots: Patents and Interdisciplinarity*, 51 MICHIGAN JOURNAL OF LAW REFORM 55 (2017)

Strandburg, Katherine J., Csárdi, Gábor, Tobochnik, Jan, Érdi, Péter & Zalányi, László, *Law and the Science of Networks: An Overview and an Application to the "Patent Explosion"*, 21 BERKELEY TECH. L.J. 1293 (2006)

Tong, X & Davidson, J., *Measuring National Technological Performance with Patent Claims Data*, 23 RES. POL'Y, 133 (1994)

Trajtenberg Manuel, *A Penny for Your Quotes: Patent Citations and the Value of Innovation* 21 RAND J. ECO. 172 (1990)

Trajtenberg, Manuel, Henderson, Rebecca & Jaffe, Adam B., *University versus Corporate Patents: A Window on the Basicness of Invention*, in JAFFE. ADAM B &

TRAJTENBERG, MANUEL & JAFFES, ADAM B., EDS., PATENTS, CITATIONS & INNOVATIONS: A WINDOW ON THE KNOWLEDGE ECONOMY (2003c)

Van Noorden, Richard, *The Science that's Never Been Cited*, 552 NATURE 163 (2017)

Yoshikane, F., Suzuki, Y., & Tsuji, K. *Analysis of the Relationship between Citation Frequency of Patents and Diversity of their Backward Citations for Japanese Patents* SCIENTOMETRICS. 92(3) (2012)

Wuchty, Stefan, Jones, Benjamin F. & Uzzi, Brian, *The Increasing Dominance of Teams in Production of Knowledge*, 316 SCIENCE 1036 (2007)