When Do Applicants Search for Prior Art?

Author(s): Bhaven N. Sampat

Source: *The Journal of Law & Economics*, Vol. 53, No. 2 (May 2010), pp. 399-416

Published by: The University of Chicago Press for The Booth School of Business, University of Chicago and The University of Chicago Law School

Stable URL: https://www.jstor.org/stable/10.1086/651959

When Do Applicants Search for Prior Art?

Bhaven N. Sampat *Columbia University*

Abstract

There is concern that patent examiners lack the resources, capabilities, and incentives to properly identify the prior art against which patent applications are evaluated and that, as a result, they issue a large number of low-quality patents. In this context, the extent to which applicants have incentives to contribute prior art is an important question. This paper uses data on examiner and applicant citations in U.S. patents to examine this question. The data show that applicants contribute a surprisingly low share of citations to previous patents and routinely fail to identify even their own previous patents. However, there are also stark differences across fields. Within fields, and even within firms, there is self-sorting: applicants contribute more prior art for their more important inventions. The results suggest that incentives to search for prior art vary across industries and inventions, which reflects underlying differences in the strategic reasons for obtaining patent protection.

1. Background

In the United States, patents are to be granted only for novel and nonobvious inventions. To assess whether a patent application meets these criteria, a patent examiner compares the claimed invention with previously disclosed information, or prior art, found in references to patents and printed publications. There is growing concern that resource-constrained examiners face difficulties in identifying relevant prior art, thereby contributing to the issuance of patents of questionable validity. These concerns about patent quality have fueled calls for the most significant patent reform in more than a half century (Cohen and Merrill 2003; Federal Trade Commission 2003; Jaffe and Lerner 2004).

Although prior-art issues are central to current policy debates, little is known about the extent to which, or the conditions under which, examiners have relevant prior art before them when making determinations of patentability. Examiners conduct their own searches for prior art. However, they face severe resource constraints in doing so, and the production targets that they face create

Iain Cockburn, Rebecca Eisenberg, Bronwyn Hall, Mark Lemley, Douglas Lichtman, Jean Lanjouw, David Mowery, Matt Neidell, Richard R. Nelson, Arti Rai, Mark Schankerman, Scott Stern, and Rosemarie Ziedonis provided valuable comments on an earlier draft. I thank the Ford Foundation for financial support.

[*Journal of Law and Economics*, vol. 53 (May 2010)]
 0022-2186/2010/5302-0016$10.00

disincentives to allocating significant time to searching for prior art (Jaffe and Lerner 2004; Thomas 2001).[1] These constraints on examiners mean that the extent to which applicants supply information on relevant prior art is important to the quality of issued patents.

However, the role of patent applicants in the identification of prior art and the incentives that applicants face are poorly understood. In the United States, a duty of candor creates strong incentives to disclose known prior art that is material to patentability.[2] However, applicants are not required to conduct searches; anecdotal accounts suggest that some do and others do not (U.S. Patent and Trademark Office 1999).

This paper provides empirical data on when applicants search for prior art. Although this topic has not previously been examined empirically, it has been the subject of considerable speculation by legal scholars and economists. Some have argued that because missing prior art would result in patents that are difficult to enforce or are of questionable validity, applicants have strong incentives to conduct searches for prior art before filing patent applications. Gerald Mossinghoff, the former commissioner of the U.S. Patent and Trademark Office, has argued that "the best patent applications are written with the prior art clearly in mind starting from the beginning. . . . If you know the best prior art and you're a good attorney, you'll write a very sustainable patent" (U.S. Patent and Trademark Office 1999, p. 29). Caballero and Jaffe (1993, p. 21) suggest that "omission of important references can be grounds for invalidation of the patent, giving the applicant an incentive to make sure the citations appear."

On the other hand, some legal scholarship suggests that applicants face disincentives to searching. Applicants can obtain broader patents if the examiner does not consider prior art to be material to patentability (Kesan and Banik 2000; Wagner 2002).[3] Wagner (2002, p. 53) observes that "[t]he patentee has both the motive and intent to behave strategically. . . . It might involve declining to conduct a thorough prior-art search, thus transferring the cost to the public, as well as increasing the possibility that the PTO [Patent and Trademark Office]

[1] However, Lemley and Sampat (2008) argue that different production targets facing junior versus senior examiners create different incentives for the two groups, and they show that these differences map to differences in behavior, including the propensity to search and the likelihood that a patent application will be granted.

[2] The *Manual of Patent Examining Procedure* notes that a prior-art reference is considered material when "(1) [i]t establishes, by itself or in combination with other information, a *prima facie* case of unpatentability of a claim; or (2) [i]t refutes, or is inconsistent with, a position the applicant takes in (i) [o]pposing an argument of unpatentability relied on by the Office, or (ii) [a]sserting an argument of unpatentability" (37 C.F.R. 1.56(b)). If an applicant knowingly fails to disclose material prior art, an accused infringer can raise the defense of inequitable conduct in court, and if the court agrees, the patent will be rendered unenforceable (Allison and Lemley 1998).

[3] As an empirical matter, it is difficult to operationalize patent scope in large samples, which makes it difficult to answer the question of how much broader. Although economists traditionally have used the number of claims as a measure of patent scope, legal scholars have questioned this practice (Moore 2004). See also Lemley and Sampat (2008) for empirical evidence that the number of claims is a poor measure of scope.

will miss something and thus allow unwarranted scope."[4] These issued patents would be presumed valid,[5] and challengers (including accused infringers) must rebut this presumption with clear and convincing evidence. Given the costs of challenging validity via litigation, issued patents of dubious validity can be used to extract rents from competitors, and they can impose significant social costs (Farrell and Shapiro 2008).

Which of these competing incentives dominates in shaping the prior-art search decisions of applicants, and when? In this paper, I use a novel data set of examiner- and applicant-inserted references to U.S. patents in the 502,687 utility patents issued between January 1, 2001, and December 31, 2003, to examine these issues. The data show that applicants contribute a surprisingly low share of references to previous patents: on average, the majority of citations to previous patents come from examiners rather than applicants. Applicants routinely fail to identify even their own previous patents, which suggests that, in many cases, applicants do not conduct even cursory searches for prior art.

However, I also show that these competing incentives have different effects in different contexts. Previous empirical work shows that firms' reasons for patenting vary across industries (Cohen, Nelson, and Walsh 2000; Hall and Ziedonis 2001; Hall 2005). Although in so-called discrete-product industries (including pharmaceuticals and chemicals) patents are important for appropriating returns from research, in complex-product industries (such as electronics and computers), accumulation of patents is more for defensive purposes (for example, to use as bargaining chips in cross-licensing negotiations) or to defend against infringement claims from other firms with large patent portfolios. The empirical analyses show that these patent strategies map back to prior-art search strategies, with applicants much more likely to search for prior art in fields where individual patents are important for appropriating returns from research and development and much less likely to do so in industries where firms tend to accumulate patent portfolios for other strategic reasons. Previous research also shows that even within industries, the value of patent protection varies dramatically across inventions (Scherer and Harhoff 2000); most inventions are worthless, and, as a result, most patents will never be litigated or enforced. This too feeds back to prior-art search incentives. The analyses show that within fields and within firms, there is self-sorting, with applicants contributing more prior

[4] A second danger associated with searching stems from the doctrine of willful infringement. Because patent law imposes treble damages on applicants who willfully infringe on earlier patents (relative to those who unintentionally infringe), firms actually have incentives to not know about the patents of competitors (Lemley and Tangri 2003).

[5] The Federal Circuit has directed that the presumption of validity applies even against prior art that was not cited: the burden remains on challengers to show clear and convincing evidence to overturn this presumption (*American Hoist & Derrick Co. v. Sowa & Sons, Inc.,* 725 F.2d 1350, 1359–60 [Fed. Cir. 1984]). However, at the same time, it has stated that relevant prior art that was not cited may carry more weight than that which was cited. Allison and Lemley (1998) argue that, among patent lawyers, the conventional wisdom is that it is much easier to invalidate a patent on the basis of uncited rather than cited prior art, and their large-sample empirical analyses provide strong evidence supporting this.

art for their more important inventions. Taken together, the results suggest that incentives to search for prior art vary across different industries and inventions, reflecting cross-field and cross-invention differences in strategic reasons for patenting and the value of patent protection.

2. Examiner and Applicant Citations

2.1. An Overview

The main data set consists of the 502,687 utility patents issued by the United States Patent and Trademark Office (USPTO) between January 1, 2001, and December 31, 2003, as well as the 6,324,381 references to previous U.S. patents in these utility patents. In U.S. patents issued since the beginning of 2001, an asterisk indicates any citations resulting from examiner office actions ("examiner citations") as opposed to those originating from the disclosures of applicants ("applicant citations").

Table 1 presents numbers of examiner citations and applicant citations, both at the citation level (calculated across all citing-cited pairs) and at the citing patent level (averaged across all citing patents). Overall, 41.4 percent of all the citations in these patents come from examiners. However, on average, across the 495,713 patents citing at least one earlier U.S. patent, 63 percent of citations are inserted by examiners. Differences between the figures at the citation and patent level reflect the bimodal distribution of examiner citations. In 8 percent of the citing patents, the examiner share is zero—that is, applicants account for all citations. However, the modal examiner share is one: in 39 percent of the citing patents, all citations to previous U.S. patents come from examiners rather than applicants.

In the analyses presented in subsequent sections, I use the examiner share of references in a patent as an indicator of the extent to which applicants are searching for prior art. The numbers mentioned above are consistent with the argument that for many inventions, applicants have a weak incentive to conduct prior-art searches. However, the examiner share of references is a noisy signal of applicant incentives because it also reflects the relative search capabilities of applicants and examiners. For example, it may be that for many of these inventions, applicants are searching for prior art but are not locating it.

Although the differential effects of incentives and capabilities cannot be completely disentangled, one clearer signal of the extent of searching for prior art is the examiner-versus-applicant share of self-citations in a patent. A self-citation is a citation in a patent by assignee *i* to an earlier patent by assignee *i*. It is plausible that in a world of limited searches and search capabilities, applicants conduct prior-art searches and fail to uncover other assignees' relevant patents. However, it is less plausible that they will not discover their own patents. Thus, the examiner share of self-citations is a particularly useful indicator of an applicant's search efforts.

Table 1

Examiner Share of Citations in Utility Patents Issued between 2001 and 2003

Category	Patents with Citations	Citations	Across Citations: Examiner Citations (%)	Across Patents with Citations		
				Average Examiner Share (%)	Examiner Inserts All (%)	Applicant Inserts All (%)
All	495,713	6,324,381	41	63	39	8
Chemicals	67,196	839,663	32	51	29	16
Computers and communications	100,763	1,230,755	46	69	45	4
Drugs and medical	53,735	844,863	21	41	23	25
Electronics and electrical	105,732	1,123,372	47	68	45	4
Mechanical	84,763	1,044,873	49	67	41	5
Other	83,524	1,240,855	46	65	39	5

Of the patents in the sample, 156,702 include at least one self-citation, and these patents include 487,139 self-citations (see Table 2). Recall that patent examiners account for 41 percent of citations of previous U.S. patents. Strikingly, patent examiners also account for 41 percent of self-citations. As a result, 11 percent of examiner citations are self-citations, and 11 percent of applicant citations are also self-citations.

At the citing patent level, the average examiner share of self-citations (averaged across all patents with at least one self-citation) is 57 percent. Here, too, the distribution is bimodal. In 30 percent of patents with at least one self-citation, the examiner share is zero (that is, applicants account for all of the citations). In nearly one-half (48 percent) of patents with at least one self-citation, however, the examiner share is one (that is, examiners account for all of the self citations).

These numbers suggest that at least for some patents, applicants are not devoting much effort to searching for prior art. In the following sections, I use these data to examine how the search efforts of applicants vary by technological field and invention characteristics.

3. Differences across Fields

As discussed above, previous research suggests that both the importance of patents as a means of appropriating research results and the strategic uses of patents differ across industries. Table 1 shows differences in examiner shares of citations and self-citations across six broad industry categories.[6]

The data show that in discrete-product fields (chemicals and drugs), patent examiners account for a significantly lower share of references to U.S. patents than in complex-product industries (computers and communications, electronics and electrical, and mechanical). Although in most fields examiners account for between 45 percent and 50 percent of all patent citations, in the fields of chemicals and drugs they account for much less (32 percent and 21 percent, respectively). In other words, applicants are more active in contributing prior art in these discrete-product fields. These differences are both statistically and qualitatively significant.

Statistics calculated at the citing patent level reveal similar patterns of cross-industry differences. The average examiner share across patents is significantly lower for chemicals and drugs than for other fields. In these fields, examiners also are significantly less likely to account for all of the references to U.S. patents and are significantly more likely to account for none of the references (that is, applicants are significantly more likely to generate all of the references).

The data on self-citations in Table 2 reveal similar patterns. In the chemical and drug fields, examiners account for only 36 and 30 percent of self-citations, respectively, whereas the number is significantly higher in all other technological

[6] These categories are constructed using the patent class–to-field concordance developed by Jaffe and Trajtenberg (2002).

Table 2
Examiner Share of Self-Citations

Category	Patents with 1+ Self-Citation	Self-Citations	Across Self-Citations: Examiner Citations (%)	Across Patents with Self-Citations		
				Average Examiner Share (%)	Examiner Inserts All (%)	Applicant Inserts All (%)
All	156,702	487,139	41	57	48	30
Chemicals	24,209	83,483	36	50	41	36
Computers and communications	36,435	106,788	47	63	54	24
Drugs and medical	16,710	52,511	30	45	38	42
Electronics and electrical	35,306	110,033	42	60	51	28
Mechanical	26,363	80,150	45	60	50	28
Other	17,679	54,174	45	56	48	33

fields. In the computer and communications fields, for example, examiners account for nearly one-half of all self-citations. At the citing patent level, the average examiner share of self-citations is significantly higher in the complex-product industries than in the chemical or drug fields, as is the share of patents for which the examiner accounts for all self-citations. Strikingly, examiners account for all of the self-citations in the majority of patents in the computer and communications and the electonics and electrical fields. On the other hand, the share of patents for which the applicant accounts for all of the self-citations is greatest for chemicals and drugs.

These data suggest that applicants have stronger incentives to search for prior art in discrete-product fields than in complex-product fields. The data are consistent with the argument that applicants' commitments to obtaining quality patents may be stronger in contexts in which a specific patent may be important to exclude competitors, as opposed to contexts in which most patents are obtained for defensive purposes, as discussed in more detail in Section 5.

4. Do Patent Applicants Contribute More Prior Art for More Important Inventions?

In addition to differences across fields, there is also considerable intraindustry heterogeneity in the private value of patent protection (Scherer and Harhoff 2000). To assess whether applicants are more likely to submit prior art for more important inventions, I construct three patent-specific measures of importance for a subset of the sample patents: the 23,703 patents granted in January and February 2001 that were assigned to a firm or organization at the time of issue.[7]

The first independent variable is a count of the number of citations that each of these patents received in subsequent patents issued by December 31, 2004. Forward-citation counts are a commonly used measure of the importance of inventions, and they appear to be good predictors of other measures of an invention's importance, including whether it is licensed (Sampat and Ziedonis 2004), consumer surplus based on an invention (Trajtenberg 1990), and other measures (Lanjouw and Schankerman 2004). On average, these patents received 3.4 forward citations between 2001 and 2004.

A second measure of a patent's importance is a binary variable indicating whether it was renewed after 4 years. After a patent is issued, the patent holder must periodically pay maintenance fees to keep it in force. Currently in the United States, these fees must be paid 4, 8, and 12 years after issue. The renewal-based measure of importance assumes that, all else equal, patents that are more economically valuable are more likely to be renewed at any given point in time (Lanjouw, Pakes, and Putnam 1998). More than one-half of patents issued expire

[7] Because I am interested in within-assignee variation, I exclude the 3,605 patents that were unassigned at issue, typically those issued to independent inventors, because the assignee information for these patents is unavailable. I also exclude the small number of patents (387) that had neither applicant nor examiner citations to previous patents.

Table 3
Descriptive Statistics for Patents Issued in January and February 2001

	Mean	SD	Min	Max	*N*
Forward citations	3.400	4.952	0	91	23,703
Patent maintained after 4 years (1 = yes)	.870	.327	0	1	23,703
Family size	4.124	3.740	0	40	23,703
Examiner's share of citations	.625	.379	0	1	23,703
Examiner's share of self-citations	.583	.448	0	1	8,725
Examiner inserts all citations (1 = yes)	.394	.489	0	1	23,703
Examiner inserts all self-citations (1 = yes)	.494	.499	0	1	8,725
Applicant citations	6.81	17.29	0	626	23,703
Examiner citations	5.01	4.81	0	170	23,703
Applicant self-citations	1.63	3.28	0	58	8,725
Examiner self-citations	1.29	1.52	0	22	8,725
Total citations	11.81	17.75	1	629	23,703
Total self-citations	2.92	3.52	1	63	8,725

before their full 20-year term because the applicant chooses not to pay maintenance fees (Moore 2004). Of the patents issued in January and February 2001, 87 percent were renewed at 4 years.

The third measure of importance is the number of countries in which an applicant takes out a patent, or a patent's "family size." As Putnam (1996) suggests, because applying for patent protection in additional countries imposes additional costs, applicants are more likely to do so for inventions that are more economically valuable. The average family size of the patents in the sample from January and February 2001 is 4.1 countries.[8]

To assess how applicants' contributions of prior art varies by invention, I examine how these measures of the importance of an invention affect the examiner share of citations and self-citations in this sample and whether the examiner accounted for all citations (and thus whether the applicant submitted no patent citations) for a given patent. Because these indicators of importance are correlated (Lanjouw and Schankerman 2004), in addition to examining whether they are individually statistically significant, I report results from joint *F*-tests of the null hypothesis that they are all equal to zero.

The examiner share of references is a fraction, increasing in examiner references and decreasing in applicant references. To isolate the impact of the effects of invention characteristics on applicant-side versus examiner-side activities, I supplement these analyses by estimating analogous models with the number of applicant citations and the number of examiner citations as dependent variables. In addition, I assess how the total number of citations in patents is associated with the applicant share of references. Table 3 presents the descriptive statistics for each of the variables used in these analyses.

I begin by estimating ordinary least squares regressions of the examiner share

[8] Forward-citation and renewal data were obtained directly from the U.S. Patent and Trademark Office (USPTO); data on family size were purchased from Delphion, a private patent data vendor.

of references in patent *i* on each of the three measures of patent importance, the patent class fixed effects for each of 398 three-digit patent classes, and the assignee fixed effects for each of the 7,614 assignees in the sample. With assignee and patent class fixed effects, the estimates of the measures of importance are identified from within-assignee within-class variation. I also estimate linear probability models relating whether the examiner accounts for all citations in a patent—in other words, whether the applicant disclosed no patent citations in the patent application—as a function of the importance measures.

Table 4 presents the results. The examiner's share of citations is decreasing with each of the three proxies for importance (that is, applicants do provide more citations in their more important patents). The results suggest that a 1-standard-deviation increase in forward citations to a patent (an increase of 4.95 citations) implies a 1-percentage-point decrease in the examiner share of citations to U.S. patents in patent *i,* after controlling for assignee and technology class effects and holding the other importance measures constant. A 1-standard-deviation increase in family size (approximately 3.7 countries) implies a 4-percentage-point decrease in the examiner share. There is no statistically (or qualitatively) significant difference between renewed and nonrenewed patents, after controlling for the other measures of importance. This could, however, reflect the bluntness of renewal at 4 years as an importance variable—recall that the vast majority of patents are renewed at 4 years—or the fact that this measure provides little information about importance beyond that reflected in variation in citations and family size. Indeed, given that the three measures of importance are proxies for an underlying variable (importance), it is difficult to interpret their qualitative magnitude directly—that is, to conceptualize increasing one while keeping the others constant. Accordingly, I also report results from a joint *F*-test. This test rejects the hypothesis that the measures of importance are jointly zero, which provides evidence of statistically significant differences between applicants' more and less important patents.

The estimates from linear probability models, in the second column of Table 4, are similar. The probability that the examiner share is one is decreasing in each of the importance variables: applicants are more likely to submit applications citing no patented prior art for their more important inventions, and the results of an *F*-test show these effects to be jointly significant at the 99 percent level.[9]

I have argued that self-citations are a particularly informative signal of the extent of applicant search. Accordingly, I repeat these analyses, using as dependent variables the examiner share of self-citations and whether the examiner inserts all self-citations. These models are estimated over the subsample of 8,725 patents containing at least one self-citation. The results are broadly similar to those reported above, with the examiner share of self-citations and the probability

[9] Given the bimodal distribution of the examiner share of citations, I also explore whether applicants were more likely to contribute all references (that is, the examiner share is zero) for more important inventions, but I did not find strong relationships between the variables. This may be because this variable is more responsive to examiner characteristics than to applicant incentives.

Table 4

Ordinary Least Squares Regressions: Examiner Share of Citations versus Importance

	Citations		Self-Citations	
	Examiner Share	Examiner Inserts All (1 = Yes)	Examiner Share	Examiner Inserts All (1 = Yes)
Forward citations	−.002**	−.004***	−.003*	−.004**
	(.001)	(.001)	(.001)	(.001)
Patent renewed?	−.004	−.006	−.027	−.044
	(.010)	(.014)	(.021)	(.024)
Family size	−.011***	−.015***	−.008***	−.010***
	(.001)	(.002)	(.002)	(.002)
Constant	.926***	.527***	−.024	.606
	(.048)	(.122)	(.536)	(.399)
N	23,703	23,703	8,725	8,725
F-test on importance measures	$F(3, 15{,}783) = 30.26$ Prob > $F = .0000$	$F(3, 15{,}783) = 44.57$ Prob > $F = .0000$	$F(3, 6{,}045) = 6.75$ Prob > $F = .0002$	$F(3, 6{,}045) = 10.86$ Prob > $F = .0000$

Note. Robust standard errors are in parentheses. All regressions include patent class and assignee effects. The *F*-tests reject the hypothesis that the importance measures are jointly zero.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

that the examiner accounts for all self-citations decreasing with importance and with the measures of importance jointly significant in each model.

To investigate whether these changes are due to applicant or examiner responses to invention importance, I estimated similar models with the number of applicant references and the number of examiner references as the dependent variables. Table 5 shows the results. Although each of the importance measures is positively related to the number of applicant citations, only one of them (the number of forward citations) is positively related to the number of examiner citations. Even for this variable, the magnitude of the effect on applicant citations is only about one-quarter as large as its effect on applicant citations. Although an *F*-test shows that the importance measures are jointly significant predictors in both models, the *F*-statistic is more than six times larger for applicant citations. When the number of self-citations is the dependent variable, the *F*-statistic for the importance measures is statistically significant only when the number of applicant self-citations is the dependent variable. Taken together, these data suggest that the effect of importance on the examiner share of citations, as documented in Table 4, reflects mainly applicant-side rather than examiner-side sorting.[10]

The finding that applicants contribute a greater share of citations to previous patents (including self-citations) for more important inventions might imply that more references are considered for important inventions. However, it is also possible that by doing more thorough searches, applicants do a better job of anticipating examiners' references for their more important inventions. Under this alternative hypothesis—that applicant references are pure substitutes for examiners' references—the total number of citations in a patent

Table 6 shows the results of ordinary least squares regressions of total citations on the applicant share of citations and patent class and assignee dummies. Within classes, as the applicant share increases from zero to one, the total number of citations to U.S. patents increases by 12, and this effect is statistically significant. One potential weakness of this test is that the same factors may affect both the applicant share of citations and total citations. For example, it may be that more important or broader inventions have more predecessors and, thus, a broader universe of earlier patents and that applicants tend to devote more effort to searching for prior art for these inventions. Under this scenario, the applicant share of citations and total citations would be correlated, even if applicant citations simply anticipated examiner citations instead of providing new information. On the other hand, it could be that the universe of prior art is more limited for certain types of inventions. The potential omitted variable here is the universe of prior art, which is difficult to measure. However, for total citations, I control for variables that we can observe—the three measures of in-

[10] The data are broadly consistent with the argument that, for better or worse, there is little or no sorting across patent applications by examiners, who devote the same resources to searching for prior art for all inventions, regardless of their importance (Merges 2000).

Table 5

Ordinary Least Squares Regressions: Number of Citations versus Importance

	Applicant Citations	Examiner Citations	Applicant Self-Citations	Examiner Self-Citations
Forward citations	.168***	.047***	.031***	.008
	(.025)	(.008)	(.009)	(.004)
Patent renewed?	.503	−.046	.124	−.079
	(.394)	(.122)	(.154)	(.075)
Family size	.604***	−.002	.080***	.006
	(.044)	(.014)	(.016)	(.008)
Constant	5.568	20.357***	.213	−1.546
	(14.286)	(4.437)	(5.006)	(2.453)
N	23,703	23,703	8,725	8,725
F-test on importance measures	$F(3, 15{,}783) = 87.05$ Prob > $F = .0000$	$F(3, 15{,}783) = 12.48$ Prob > $F = .0000$	$F(3, 6{,}045) = 14.23$ Prob > $F = .0000$	$F(3, 6{,}045) = 1.77$ Prob > $F = .1506$

Note. Robust standard errors are in parentheses. All regressions include patent class and assignee effects. The *F*-tests reject the hypothesis that the importance measures are jointly zero.

*** Significant at $P < .001$.

Table 6
Ordinary Least Squares Regressions: Total Citations versus Applicant Share

	Citations		Self-Citations	
	(1)	(2)	(3)	(4)
Applicant share of citations	12.07***	11.626***		
	(.347)	(.346)		
Applicant share self-citations			1.812***	1.761***
			(.109)	(.110)
Forward citations		.191***		.034***
		(.024)		(.009)
Family size		.471***		.072***
		(.043)		(.017)
Patent renewed?		.413		−.003
		(.391)		(.163)
Constant	25.52	25.062	−1.95	−3.135
	(14.28)	(14.198)	(5.328)	(5.321)
N	23,703	23,703	8,725	8,725

Note. Robust standard errors are in parentheses. All regressions include patent class and assignee effects.
*** Significant at $P < .001$.

vention importance. Doing so has little effect on the estimated coefficient. Similar results are seen when the total number of self-citations is the dependent variable. Taken together, these results suggest that applicant citations add to the stock of prior art cited instead of simply anticipating examiner citations.

5. Discussion and Conclusions

Despite the centrality of prior art in contemporary deliberations about patent system reform, issues relating to the roles and incentives of applicants in the identification of prior art have received little empirical attention. This paper examines these issues using a novel data set of applicant and examiner citations.

On average, the applicant share of citations in previous patents is 62 percent. However, the distribution is bimodal, with examiners accounting for all of the citations in nearly 40 percent of issued patents and with applicants accounting for all of the citations in nearly 8 percent of all patents. Even for self-citations, which I argue are a stronger signal of whether applicants have conducted searches, the average examiner share is 57 percent. Here, too, the distribution is bimodal, with examiners accounting for all self-citations in 48 percent of patents and applicants accounting for all self-citations in 30 percent of patents.

The bimodal nature of the distributions suggests the presence of two search regimes, reflecting the competing incentives facing applicants. Under the first regime, an applicant does not search for prior art, leaving the job of locating relevant prior art and narrowing claims accordingly for the patent examiner. Here, the applicant can obtain a broader patent than is merited given prior art, and that patent enjoys a presumption of validity. Although this strategy relies on incomplete searching by examiners, this may not be unreasonable in light

of the resource constraints and perverse incentives facing examiners. On the other hand, missing prior art makes patents more vulnerable to validity challenges, even in the face of this presumption. If a patent holder expects to enforce that patent, incomplete searching and citation has costs. Thus, the optimal search strategy facing an applicant is a function of a number of factors, including the probability that a given patent will be litigated and its value. The empirical analyses show that these vary across fields and across inventions.

The results from the cross-industry analyses suggest that applicants are much more likely to contribute prior art in discrete- rather than complex-product fields. Previous work based on anecdotal evidence (Wagner and Parchomovsky 2005), studies of individual sectors (Hall and Ziedonis 2001), and survey evidence (Cohen, Nelson, and Walsh 2000) suggests that the strategies of firms for obtaining patents vary across these industries. The results reported above provide large-sample evidence that these differences in strategies for obtaining patents map back to patent search strategies. In discrete-product fields in which patents are reported to be important vehicles for appropriating returns from research and development (namely, to prevent rivals from producing similar or the same technology), the validity of any given patent can have significant implications. Here, firms have stronger incentives to ensure that their patents are defensible against validity challenges based on missed prior art. By contrast, in complex-product industries, many patents cover a given product, and patents are reported to be relatively unimportant relative to other means of appropriating returns from research and development. In such industries, the primary rationale for patenting is to ensure a firm's own access to the technologies of competitors (for example, to deter infringement suits or to facilitate cross licensing). The empirical result that applicants are significantly less likely to contribute prior art in such industries suggests that, in contexts of defensive patenting, there is a higher premium on obtaining a large stockpile of patents; the probability that any given patent in this stock would survive a future validity challenge (based on missing prior art) is less important.

In addition to extending previous work on cross-field heterogeneity in patent strategy, the results also show significant within-field differences. Although within-field heterogeneity in patent value has long been recognized in the empirical literature (Scherer and Harhoff 2000), within-field differences in patent strategy have received little attention. The regressions show that applicants sort across their inventions. They contribute significantly more prior art and are less likely to submit applications with no cited prior art for their more important inventions.

Previous work suggests that, ideally, the USPTO would sort across inventions (for example, by focusing examination resources on more important inventions). Because most patents are unimportant and will never be enforced or litigated, sorting would allow the patent office to focus on the minority of inventions that actually matter (Lemley 2001; Merges 1999). However, this work also suggests that sorting by examiners is not feasible, because they are unable to identify

which patents are likely to be more valuable ex ante.[11] The data here show that even if examiners do not sort, applicants do, submitting more prior art for their more important inventions. Moreover, this additional prior art appears to be adding to the stock of prior art considered rather than anticipating examiners' citations.[12]

Thus, the data suggest that current incentives to search for prior art—created by legal doctrines, including the duty of candor, the doctrine of inequitable conduct, and the presumption of validity—cut differently across industries and inventions, reflecting industry- and invention-specific differences in patent strategy.[13]

What do the data tell us about whether, in this environment, the current presumption of validity is warranted? In an environment where there are limits on examiner-side incentives, capabilities, and resources to search for prior art, the finding that, for a large share of issued patents, applicants do not have incentives to search would seem to undermine confidence in the presumption. For example, the finding that nearly 40 percent of patents are issued with no applicant citations would seem to compromise the assumption that the relevant universe of prior art for these patents has been considered.[14]

However, it is difficult to draw from these data strong inferences for the presumption of validity. That applicants systematically contribute a higher share of prior-art references in discrete-product fields and within fields for their more important inventions may be read as evidence that the probability that all relevant prior art was considered is greater in these instances and, thus, that such patents should enjoy a higher presumption of validity or higher quality. This interpretation relies on the assumption that examiners are considering this prior art

[11] Merges (1999, p. 597) also suggests that an "all applications are created equal" ethic makes sorting politically infeasible.

[12] The finding that applicants can sort, even if examiners cannot, provides support for the feasibility of policy proposals that attempt to harness applicants' knowledge to improve patent quality, in particular, Lemley, Lichtman, and Sampat (2004).

[13] One limitation in using the applicant rather than the examiner share of citations to make inferences about the incentives of the applicants is that it does not take into account the potential effect of applicants' expectations of what examiners will do. For example, if applicants believe that examiners will find missing prior art, they may decide to search for it themselves to avoid creating suspicion about the merits of the application more generally or perhaps because incomplete searching would antagonize examiners. However, Lemley and Sampat (2008) show that there is significant heterogeneity among examiners in how thoroughly they search for prior art, even within narrowly defined technical areas (the USPTO's art units). (As Cockburn, Kortum, and Stern [2003, p. 21] report on the basis of their qualitative research at the USPTO, "there are as many patent offices as there are patent examiners.") There also is considerable uncertainty about which examiner in an art unit will get an application (Lemley and Sampat 2008). This makes forming expectations about examiners' likely behavior difficult. However, to the extent that applicants form expectations based on average examiner behavior in particular fields and to the extent that these expectations affect their own search incentives, the analyses here may not capture the pure effects of field and invention characteristics on prior-art search behavior.

[14] Applicants may be citing other types of prior art in these patents, including nonpatent publications and foreign patents. However, the share of issued patents with no citations to any type of prior art is also strikingly high: nearly 25 percent for patents issued between 2001 and 2003.

properly. However, if examiners lack resources, incentives, or capabilities to inspect prior art carefully and ensure that claims are appropriately defined in its light, this need not be so. Indeed, in cases in which applicants are providing a greater volume of citations but the citations are not adequately considered by the examiner, the resulting patents are not only presumed to be valid but are less vulnerable to challenge on the basis of missing prior art. In other words, more is not necessarily better.

The result that citations reflect the strategic reasons of applicants for obtaining patents also would seem to complicate the use of patent citation data as indicators in the empirical literature in law and economics. Citation data are increasingly used to measure the economic significance of patented inventions and as proxies for spillovers and knowledge flows. Previous research suggests that citations added by examiners may create noise in citation-based measurements (Alcacer and Gittleman 2006). To the extent that applicants' strategic reasons for patenting and citing are related to other variables of interest, citation-based indicators could be systematically biased. Assessing the magnitude and extent of such biases is an important topic for future research.

References

Alcacer, Juan, and Michelle Gittelman. 2006. How Do I Know What You Know? Patent Examiners and the Generation of Patent Citations. *Review of Economics and Statistics* 88:774–79.

Allison, John, and Mark A. Lemley. 1998. Empirical Evidence on the Validity of Litigated Patents. *American Intellectual Property Law Association Quarterly Journal* 26:185–269.

Caballero, Ricardo, and A. B. Jaffe. 1993. How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth. *NBER Macroeconomics Annual* 8:15–76.

Cockburn, Iain, Sam Kortum, and Scott Stern. 2003. Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes. Pp. 17–53 in *Patents in the Knowledge-Based Economy,* edited by Wesley M. Cohen and Stephen A. Merrill. Washington, D.C.: National Academies Press.

Cohen, Wesley M., and Steve Merrill, eds. 2003. *Patents in the Knowledge-Based Economy.* Washington, D.C.: National Academies Press.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not). NBER Working Paper No. 7552. National Bureau of Economic Research, Cambridge, Mass.

Farrell, Joe, and Carl Shapiro. 2008. How Strong Are Weak Patents? *American Economic Review* 98:1347–69.

Federal Trade Commission. 2003. To Promote Innovation: The Proper Balance of Competition and Patent Law and Policy. Technical report. Federal Trade Commission, Washington, D.C.

Hall, Bronwyn. 2005. Exploring the Patent Explosion. *Journal of Technology Transfer* 30: 35–48.

Hall, Bronwyn, and Rosemarie Ziedonis. 2001. The Patent Paradox Revisited: An Empirical

Study of Patenting in the U.S. Semiconductor Industry 1979–1995. *RAND Journal of Economics* 32:101–28.

Jaffe, Adam, and Josh Lerner. 2004. *Innovation and Its Discontents: How Our Broken Patent System Is Endangering Innovation, and What to Do about It.* Princeton, N.J.: Princeton University Press.

Jaffe, Adam, and Manuel Trajtenberg. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy.* Cambridge, Mass.: MIT Press.

Kesan, Jay, and Mark Banik. 2000. Patents as Incomplete Contracts: Aligning Incentives for R&D Investment with Incentives to Disclose Prior Art. *Journal of Law and Policy* 2:23–54.

Lanjouw, Jean, Ariel Pakes, and Jon Putnam. 1998. How to Count Patents and Value Intellectual Property: Uses of Patent Renewal and Application Data. *Journal of Industrial Economics* 46:405–33.

Lanjouw, Jean, and Mark Schankerman. 2004. Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators. *Economic Journal* 114:441–65.

Lemley, Mark. 2001. Rational Ignorance at the Patent Office. *Northwestern University Law Review* 95:149–96.

Lemley, Mark, Douglas Lichtman, and Bhaven Sampat. 2005. What to Do about Bad Patent. *Regulation* 5:10–3.

Lemley, Mark, and Bhaven Sampat. 2008. Examiner Characteristics and the Patent Grant Rate. Working paper. Columbia University, Mailman School of Public Health, New York.

Lemley, Mark, and Ragesh Tangri. 2003. Ending Patent Law's Willfulness Game. Public Law Research Paper No. 142. University of California, Berkeley, Boalt Hall.

Merges, Robert. 1999. As Many as Six Impossible Patents before Breakfast: Property Rights for Business Concepts and Patent System Reform. *Berkeley Technology Law Journal* 14: 578–615.

Moore, Kimberley A. 2004. Worthless Patents. Working paper. George Mason University, School of Law, Arlington, Va.

Putnam, Jon. 1996. *The Value of Intellectual Property Rights.* Unpublished thesis. Yale University, Department of Economics, New Haven, Conn.

Sampat, Bhaven, and Arvids Ziedonis. 2004. Patent Citations and the Economic Value of Patents: A Preliminary Assessment. Pp. 277–98 in *Handbook of Quantitative Science and Technology Research,* edited by Henk F. Moed, Wolfgang Glänzel, and Ulrich Schmoch. Dordrecht: Kluwer.

Scherer, Frederic, and Dietmar Harhoff. 2000. Technology Policy for a World of Skew-Distributed Outcomes. *Research Policy* 29:559–66.

Thomas, J. R. 2001. Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties. *University of Illinois Law Review,* pp. 316–22.

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

U.S. Patent and Trademark Office (USPTO). 1999. Public Hearing on Issues Related to the Identification of Prior Art During the Examination of Patent Application. June 14. http://www.uspto.gov/web/offices/com/hearings/priorart/0714pato.doc.

Wagner, Polk. 2002. Reconsidering Estoppel: Patent Administration and the Failure of Festo. *University of Pennsylvania Law Review* 159:1–79.

Wagner, Polk, and Gideon Parchamovsky. 2005. Patent Portfolios. Public Law Working Paper No. 06-15. University of Pennsylvania Law School, Philadelphia.