HARVARD | BUSINESS | SCHOOL

Applicant and Examiner Citations in US Patents: An Overview and Analysis

Juan Alcácer
Michelle Gittelman
Bhaven Sampat

Working Paper

09-016

Applicant and Examiner Citations in US Patents: An Overview and Analysis

July 2008

Juan Alcácer
Harvard Business School
jalcacer@hbs.edu

Michelle Gittelman
Rutgers Business School
Rutgers University
michelle.gittelman@business.rutgers.edu

Bhaven Sampat
Malman School of Public Health
Columbia University
bns3@columbia.edu

Abstract

Researchers studying innovation increasingly use indicators based on patent citations. However, it is well known that not all citations originate from applicants—patent examiners contribute to citations listed in issued patents—and that this could complicate interpretation of findings in this literature. In 2001 the US Patent and Trademark Office (USPTO) began reporting examiner and applicant citations separately. In this paper, we analyze the prior art citations of all patents granted by the USPTO in 2001–2003. We show that examiner citations account for 63 per cent of all citations on the average patent, and that 40 per cent of patents have all citations added by examiners. We use multivariate regression and analysis of variance to identify the determinants of examiner shares. Examiner shares are highest for non-US applicants and in electronics, communications, and computer-related fields. However, most of the variation is explained by firm-specific variables, with the largest patent applicants having high examiner shares. Moreover, a large number of firms are granted patents that contain no applicant prior art. Taken together, our findings suggest that heterogeneity in firm-level patenting practices, in particular by high-volume applicants, has a strong influence on the data. This suggests that analysis of firm-level differences in patenting strategies is an important topic for future research.

JEL Classification Numbers:
Keywords: Technology, patents, patent examiners, prior art, citations

1. Introduction and Background

The empirical literature on technology and innovation increasingly relies on measures based on patent citations. One prominent line of research originated with the recognition that simple patent counts are noisy measures of innovative output, since there is high variance in the economic and commercial significance of individual patents. This literature thus uses the number of times a patent is cited in subsequent citations (or "forward" citation counts) as a measure of its importance or quality (Trajtenberg 1990). A second stream of research was sparked by the growing interest in spillovers and knowledge flows in the innovation literature. This literature uses citations in patent A to patent B as indicators of knowledge transmission from the inventor of B to the inventor of A.

However, some are skeptical of the use of patent citations as indicators. For example, Gregory Aharonian, editor of the electronic newsletter *PATNEWS*, remarked on the writings in economics:

> "One of the problems with many of these papers is that the economists who write about patent citation analysis have little experience with patent searching, and don't realize how worthless most patent citations are for measuring anything. For example, many of them assume that the citations that appear on the front of the patent were all used and discovered by the inventor. They then use that assumption to measure flows of information between companies and inventors ... What they don't realize is that many citations are found either by the examiner or by professional searchers ... so that such citations do not measure anything about information flow or patent importance. And without detailed analysis of many file wrappers, no one knows what percentage of citations are due to inventors, and those due to others" (posting to PIUG-L listserv, April 29, 2001).

Historically, this has been true. However, following a January 2001 change in procedure at the USPTO, the front pages of issued U.S. patents have indicated with asterisks all prior art references cited by an examiner during patent prosecution, thus distinguishing them from references taken from applicant Information Disclosure Statements. The USPTO instituted this change in hopes that indications of the sources of citations would "be helpful in compiling statistical data related to prior art submissions so that the USPTO can better consider whether changes are required to the rules governing prior art statements." In part, the change was responsive to concerns that foreign applicants are particularly lax in submitting prior art (personal communication, Karin Tyson, USPTO).

Overall, examiners account for approximately 40 percent of all citations to U.S. patents. The magnitude of examiner citations is even higher when measured on a per patent basis: for the average patent in our dataset, examiners introduced 63 percent of all patent citations to US patents. The difference in means at the citation and patent levels reflects the skewed distribution of examiner citations across patents: about 40 per cent of patents granted over the period have *all* citations inserted by examiners, i.e., applicants did not add a single citation, whereas only 8 per cent of patents had no examiner-inserted citations.

Though the overall numbers seem to support Aharonian's criticism, ultimately, the extent to which examiner citations affect inferences from citation-based measures depend on the specific application in which the citation data are used. Several recent papers have started exploring this issue in detail. For example, Alcacer and Gittelman (2006) assess whether pooling inventor and examiner citations gives unbiased estimates of knowledge flows and Sampat and Hegde (2007) assess whether examiner or applicant forward citations provide different signals of the private value of patents, as measured by applicant payment of maintenance fees.

In this paper, we describe the institutional process by which citations are generated and explore three dimensions which could plausibly affect the share of examiner versus applicant citations in a patent: the technological field of the invention, patent examiner characteristics, and applicant characteristics. Our analysis is primarily exploratory rather than causal. Our aim is to provide a set of broad stylized facts about examiner citations with the goal that these data, together with our broader descriptions of the processes through which patent citations arise, help inform the interpretation of analyses using citation based indicators in specific contexts, and stimulate future research using patents.

We find that firm-level effects explain most of the variation of examiner citation shares, followed by technological field effects and examiner-specific effects. Unpacking these firm-level effects we find surprisingly, that applicants with very high numbers of patents, whom we call the most experienced applicants, receive higher shares of examiner citations than less-experienced applicants. Additionally, non-US applicants receive far higher shares than US applicants, reflecting differences in international patenting practice. Regarding technology areas, we find strong field-specific effects that correspond to a pattern in which examiner shares are lowest in technological fields in which individual patents are more economically valuable.

Our analysis suggests the need for caution in assuming that the portion of citations that aren't knowledge flows are simply "noise" (Jaffe et al., 2000). The sheer magnitude of examiner citations (with examiners adding most or, in a large proportion of cases, all citations), means that the noise may far exceeds the signal contained in pooled citations.

More importantly, the data suggest that examiner citations are not randomly distributed: variation in the share and number of examiner citations is driven by specific characteristics of patents, applicants, and examiners. While beyond the scope of this paper, assessing exactly when and to what extent examiner citations complicate conventional interpretation of citation based data is an important task for future research.

The remainder of the paper is organized as follows. In the next section (Section 2), we briefly describe the processes involved in generating patent citations, and then discuss how various field-, assignee- and examiner-related characteristics may affect the shares of applicant and examiner citations (Section 3). In Section 4 we describe our data and measures. We present the data on patent-level characteristics associated with higher and lower examiner shares, first as descriptive statistics (Section 4) and then as coefficients from multivariate regression analyses (Section 5). In Section 6 we introduce a decomposition of variance analysis to identify the relative contributions of the different types of variables to explaining variation in examiner shares at the patent level. In Section 7 we present and analyze data on firms with disproportionately high and low shares of examiner added citations. We then proceed to analyze the incidence of examiner share in self-citations in Section 8. Section 9 presents our conclusions.

2. Where do patent citations come from?

The U.S. patent statute requires an invention to be novel and non-obvious, as well as to satisfy other technical criteria, in order to be patentable. Like the patent system itself, the novelty and non-obviousness requirements were designed with a utilitarian purpose in mind, to ensure that society only incurs the social costs of patents for inventions that are significant departures from what is already in the public domain (Barton 2001). The core of a patent is expressed in its claims, which detail aspects of the invention over which inventors and assignees may exercise exclusive ownership rights.

In order to assess whether the claims in an application are novel and non-obvious, a patent examiner compares a claimed invention to the prior art, generally embodied in previous patents and publications. If the patent examiner deems that an invention is patentable in light of these requirements, a patent is granted. Since 1947, the Patent Office has listed the prior art references against which the patentability of an invention was judged on the front-page of issued patents.[1]

In this process, prior art citations serve a number of heterogeneous functions: by anticipating the claimed invention, they may be used to limit or reject an individual claim or an entire patent. Or citations may strengthen claims, by establishing that earlier related inventions were different from or inferior to the current invention. They may be boilerplate − establishing or "teaching" facts described in the patent, with little direct relationship to the patented invention.

Though examiners are officially responsible for constructing the list of prior art references against which patentability is judged, they rely in part on applicant disclosure of the prior art submitted with the patent application via Information Disclosure Statements (IDS). In the United States, applicants (and their attorneys) have a "duty of candor" to disclose any prior art "material to patentability" of an invention, i.e. any prior patent or publication for which there is "a substantial likelihood that a reasonable

[1] Printing of prior art references on the actual patent followed from a Patent Office Order issued on December 19, 1946. The first patent to include prior art references was issued on February 4, 1947, and all patents since that date include a "references" section. This is not to say that claimed inventions were not evaluated against the prior art before this date: prior art searching became common practice with the passage of the Patent Act of 1836. Prior to 1947, however, the prior art against which the claimed invention was evaluated was available only from the "file history" of the issued patent, stored at the Patent Office.

examiner would consider it important in deciding whether to allow the application to issue as a patent" (USPTO 1998, Section 2242). If an applicant knowingly fails to disclose material prior art, an accused infringer can raise an "inequitable conduct" defense in court, and if the court agrees the patent will be rendered unenforceable. Even absent inequitable conduct, some or all of the claims of a patent can be rendered invalid in post-issuance lawsuits or re-examination, if it is subsequently shown that prior art material to patentability was not considered by the patent examiner (Allison and Lemley 1998).

Incentives to disclose prior art are complex. The following quote, by a patent attorney and former inventor, is illustrative of pressures on inventors and their attorneys for full disclosure:

> The first [time], as an inventor, I was introduced to prior art as an engineer at IBM. There, we were told to disclose and discuss all pertinent publications before they were filed. And failure, we were told by the attorneys, was punished by fraud, imprisonment, and would result in the disbarment of the attorney that was representing us. Basically, the attorneys said that we would have the time in jail to basically explain to them why they could no longer practice law, and so forth, if we didn't give them the right references. Maybe this was unique to IBM, but it's something that I've carried throughout my career in talking with inventors, and so forth, as far as how important I think the duty of disclosure is. [2]

While applicants have strong incentives to disclose known prior art, they are not legally required to search for it, and face both incentives and disincentives to doing so. On one hand, prior art searching by applicants can ensure that resulting patents are enforceable (Allison et al., 2003) as illustrated by a quote from the same attorney:

---

[2] Testimony included in "Public Hearing on Issues Related to the Identification of Prior Art During the Examination of Patent Application", June 28, 1999, before the United States Patent and Trademark Office.

> One of the worst feelings that I've even seen at a licensing table is when you're sitting there trying to license a patent and someone passes across the table a 102(b) reference [a prior art reference] that is completely out of left field, you've never seen before, that says this patent is invalid and indefensible. It's something that no one, as a practitioner, wants to face, and would rather face, have that reference come up, early on in the prosecution procedure, and be able to be discussed with the examiners, who really know what they're talking about.[3]

On the other hand, applicants may receive broader patents if the examiner fails to uncover prior art material to patentability. These patents enjoy a strong presumption of validity (Sampat 2007). Thus Wagner and Parchomovsky (2005, p. 53) suggest that "[t]he patentee has both the motive and intent to behave strategically … [i]t 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 will miss something and thus allow unwarranted scope".[4]

After applicants disclose their references, patent examiners conduct their own prior art searches. They may challenge applicants' claims based on their own prior art searches, and communicate with inventors during the examination process (see Cockburn, et al 2003, for an excellent discussion of this process). There have been considerable attempts at codification and standardization of patent examination procedures, but the search process remains largely idiosyncratic. The universe of potentially relevant prior art includes patents (over 5 million in the US alone), non-patented literature such as books and journals, and even information disclosed on the Internet. However, in a world of limited resources, the finite corps of patent examiners

---

3 Ibid.

4 In addition, since patent law imposes treble damages on applicants who willfully infringe on earlier patents (as opposed to those who unknowingly infringe), firms may have incentives to not know about competitors' patents, which also may blunt incentives for prior art searching (Lemley and Tangri 2003).

faces strict time allocation guidelines per application (Thomas, 2001; Cockburn et al. 2003). There is widespread concern that examiner time constraints have tightened in recent years, when growth in patent applications has outpaced growth in the number of examiners (Merrill and Myers, 2004), limiting the breadth and depth of examiner searches. As a result, concerns have been raised that the number of "low quality" patents issued by the USPTO, i.e. patents that would not satisfy the novelty and non-obviousness requirements if subjected to more intensive or complete examination, has increased in recent years (Jaffe and Lerner 2006).

The applicants' duty of disclosure continues *even after* the application has been submitted. Here again there is no obligation for applicants to search for prior art, but applicants are required to disclose prior art they discover (e.g. in preparing applications for related inventions). Applicants may also file new prior art to support a claim that has been rejected by the examiner.[5]

Patent examiners from foreign patent offices can also be a source of citations when patent applications are filed via the Patent Cooperation Treaty (PCT). The PCT allows an applicant to file in any treaty member state and designate other member states where she retains the option to file within 12 months of the initial application. Under the PCT process, the receiving patent office provides results of its patent searches to all patent offices designated in the original application.

It is from this complex process that patent-to-patent citation data used in analyses by social scientists emerge. Both examiner and applicant side capabilities and incentives could affect this process, and the relative shares of applicant and examiner citations could

[5] As prosecution times increased in the late 1990s and early 2000s, introduction of multiple information disclosure statements became more common In fact, starting in January 2008 new regulation was enacted to limit the number of new references introduced during prosecution.

reflect these differences. While almost all previous analyses have bundled citations originating from the applicant's disclosure with those resulting from examiner's search, the policy change discussed in the Introduction means that we can now distinguish between the two.

3. Determinants of examiner and applicant citation shares.

To identify the determinants of examiner and applicant citations shares we explore the incentives and capabilities of applicants and examiners to provide prior art. Applicant's incentives to conduct prior art searches are likely to vary by technological field. Sampat (2007) argues that, in fields where valid patents are deemed important as mechanisms for appropriating returns to R&D or to facilitate market trades in technology, applicants will expend more effort in searching for prior art. Survey research (Cohen, et al. 2000) on the value of patents across fields indicates that patents are highly valued as mechanisms for appropriating returns to R&D in only a few fields – principally chemicals and pharmaceuticals. In such fields, referred to as "discrete technologies", one or a few patents are generally sufficient to create strong monopoly rents, and we expect that on average applicants will seek to file strong individual patents, e.g., patents that would be held valid if opposed. In other fields, such as electronics and telecommunications equipment, intellectual property is more fragmented and distributed among numerous applicants, such that products may be covered by hundreds of patents, none of which has great value on its own. Cohen, et al. (2000) find that in these fields, referred to as "complex technologies", applicants are more likely to value patents as strategic bargaining chips in negotiations with rivals, rather than as mechanisms for protecting rents in product markets. Value accrues to bundles of patents rather than possession of a single patent, and the use of patents changes from protecting monopoly rents to strategic uses, e.g., hold-up or cross-licensing negotiations (Cohen et al., 2000, Hall & Ziedonis, 2001, Ziedonis, 2004). In such cases, we expect that applicants will exert less effort in prior art search, because of a lower expectation of the need to defend individual patents against rival technologies. That is, firms' decisions to search for prior art – and the resulting "quality" of issued patents – are likely to reflect field-specific factors related to the value and use of patents in those fields.

Incentives may also differ according to the type of applicant. Since the passage of the Bayh-Dole act in 1980, universities have become active patenting organizations (Mowery et al. 2001). These entities typically license their patents to firms, rather than commercialize them directly. Moreover, since under Federal law universities, rather than individual professors, take title to patents, they typically employ patent and licensing professionals to screen inventions and draft patents. Technology transfer offices may only choose to patent those inventions with a high probability of earning revenue through licensing. If strong patents are important for licensing, we would expect that patents assigned to such entities would have relatively high applicant share of references.

Firm experience and firm size may affect the examiner citation shares. Popp et al. (2003) suggest that an entity's patenting experience may affect its patenting practices. Larger firms and/or firms that file numerous patents are likely to have the resources to engage lawyers and professional patent searchers, conduct prior art searches and/or know the relevant prior art (Mossinghoff, in USPTO 1999). Greater resources and patenting capabilities would lead to the expectation that larger and/or more experienced firms would have a greater share of prior art. On the other hand, incentives may work against this expectation. Small firms may be more likely to seek to out-license patents than commercialize them in-house (Arora et al. 2001) in which case they may have stronger incentives than large firms to invest ex ante in patent quality and search for prior art. Furthermore, firms that patent heavily may economize on resources by adopting a "scattershot" approach, sending a large volume of patents for examination without extensive investment in pre-screening for patentability.[6]

Differences in country-specific patent practices can also affect applicants awareness and ability to search for prior art. In most foreign countries, applicants face no "duty of candor" similar to that in the U.S.. Thus foreign applicants may be less inclined to invest in thorough prior art searches. Michel and Bettels (2001) argue that for this and other institutional reasons, prior art searches in Europe are more selective than in the US. They argue that the duty of candor in the US leads applicants to submit *any* citation that may be relevant. Since examiners may not cull all of these applicant citations, US patents may contain many applicant citations that have little relevance to patentability.

Many applicants employ lawyers and professional prior art searchers, to draft patents and conduct prior art searches, and this could affect the share of examiner citations. Familiarity with patenting examination practices – many patent lawyers were formerly examiners – could lead to anticipation of examiners' actions, and thus reduce the examiner share of citations in a patent.

Previous research by Cockburn et al. (2003) and Lichtman (2004) suggests that examiner heterogeneity affects the prosecution and characteristics of a patent. Heterogeneity across examiners in their search capabilities can also contribute to

[6] In an analysis of Danish patents, Schneider (2007) finds that firms with large patent portfolios adopt a "trial and error" strategy, filing many applications and subsequently withdrawing those that receive signals from the patent office that they have a low probability of success.

differences in the examiner share of citations. With regard to examiner heterogeneity, the most relevant observable dimension is examiner experience. Several commentators have pointed to the difficulty the Patent Office faces in recruiting and retaining examiners (Popp et al. 2003; Thomas 2001). More generally, a high rate of examiner attrition (Bawa 2004) may limit any benefits realized from "learning by doing" in prior art searching. Thus Jennings (in USPTO, 1999, p. 34) suggests that "turnover of patent examiners" may diminish the quality of prior art searching at the USPTO, since "they leave before they gain real experience and knowledge of the pertinent art."

In the next section we present the data on examiner shares and show how they vary according to the dimensions described above. We then show that the patterns identified in the univariate analysis still hold when we control for all the variables in multivariate regressions that estimate the share and number of examiner citations on a citing patent.

4. Data and Variables

To illuminate the relative importance of these dimensions, we use a novel dataset that includes all 502,390 utility patents issued between January 1, 2001 and December 31, 2003. Our dataset is based on weekly electronic files provided by the USPTO that contain the information on the front page of utility patents as well as references to other utility patents granted in the US.

In the data, citations based on applicant Information Disclosure Statements are treated as applicant citations. (Note that this is true whether or not the citation was disclosed by the applicant in the original application, or during the patent prosecution process.) Citations from examiner searches are treated as examiner citations. One complication is that for PCT applications filed first outside the U.S (less than 10 percent of the 502,390 patents in our sample result from PCT applications) references from the foreign search report are coded as applicant citations, even though they may have resulted from searches from foreign patent offices. As a consequence, the share of examiner citations for PCT patents may be underestimated. We discuss this in greater detail (and attempt to control for it) in the analyses below.

For each citing patent, we calculate the share of examiner citations as a percentage of total patent citations. We then show how these shares vary by different dimensions of interest discussed above. Our measures capture patent-level, assignee-level, and examiner-level variables that we expect will impact the share of examiner-inserted citations.

To assess the importance of differences across technology fields, we measure technology class of a patent using 3-digit classifications at the USPTO. However, for expositional convenience, we present results using broader technology classes, as defined by the concordance between patent classes and industries provided by Jaffe & Trajtenberg (2002). Specifically, for each citing patent, we constructed dummy variables for the following broad technological categories: *Computers and Communications*, *Electrical and Electronics*, *Drugs and Medicine*, *Chemical*, *Mechanical*, and *Other*.

We create a range of variables to examine differences across types of assignees discussed above. First, we create a dummy variable *Small Entity* indicating whether the patent was assigned to a small entity. We base our coding on data from the USPTO that indicates whether an applicant qualified for reduced fees based on "small entity" status, as determined by the USPTO. [7] Analogously, the *Large Entity* dummy indicates patents assigned to firms which are not small entities.

We also created indicators for university and U.S. government assignees. The government assignee information is based on “assignee type” data from USPTO's *Cassis* database, and the university assignee data is based on institutions denoted as universities in Mowery et al. (2001). Government patents, which typically are a small share of total patents, are expected to be concentrated in fields as biomedical research (e.g., patents assigned to the National Institutes of Health) and defense technologies. A catch-all category, *Other Assignee*, includes patents by individual inventors and those which are unassigned at issue.

In addition to small entity status, we include a variable to measure an assignee’s experience with patenting. We define three categories of assignee experience based on patent volume from 1976 to 2000: *most experienced* assignees (more than 1000 patents), *experienced* assignees (more than 100 patents and less than 1000 patents) and *least experienced* assignees (less than 100 patents)[8]. These data were constructed from historical patent information found in the USPTO's *Cassis*.

As discussed above, the nationality of patent applicants also may affect citation practices and examiner citations shares. To assess these differences, we created dummies for *U.S.* and *Foreign* assignees, and also separate dummies for the following assignee countries: *Japan, Germany, Taiwan, France, United Kingdom, South Korea, Canada,*

---

[7] For more details in the definition of small entities, see http://www.uspto.gov/web/offices/pac/mpep/documents/appxr_1_27.htm

[8] A classification based on quartiles is desirable but not appropriate because the distribution of patent stocks is highly skewed (skewness index of 219.83, kurtosis index of 49,970). Most assignees have very low patent volume with a median equal to 2 patents, and the 75 percent quartile level equal to 7 patents.

*Italy, Sweden, and Switzerland..*We also include a variable, *Lawyers*, to indicate whether legal assistance was provided when drafting the patent application. Both the nationality and legal assistance data were constructed using data from the front page of the patents.

Heterogeneity in examiner experience may contribute to cross-invention differences in the examiner share of citations. As pointed out by Cockburn et al. (2003), patent examiner names are transcribed onto the front-pages of U.S. patents with considerable error. To facilitate construction of meaningful examiner experience measures, we hand-matched each of the 6,172 primary patent examiner names in the 2001-2003 data to a "standardized" patent examiner name from the 1999 and 2004 USPTO Employee Directories. We measured an examiner's experience at the patent office as the number of years since s/he examiner his/her first patent. To avoid imposing a specific parametric structure (and since our data are left-censored at 1976) we also broke this variable into quartiles, *Most experienced, Second and third experience cohorts,* and *least experienced.*

Table 1 shows the data on examiner added shares of citations, overall and disaggregated across the dimensions described above. Column 1 shows the percentage of patents without any citations by applicants or examiners; overall, this is a very low share (2%) although it ranges to as high as 10% for academic patents and patents in the Drug and Medical fields. Column 2 shows mean examiner shares for all citations; column 3 is an average of the examiner share on each patent. Overall, 41 percent of all citations on U.S. patents come from examiners (column 2), and the average examiner share at the patent level is 63 percent (column 3). The discrepancy between the means reflects the fact that 8 percent of patents, examiners insert no citations (applicants account for all citations), while in 39 percent of patents, examiners insert all citations (applicants contribute no citations). Thus the distribution of examiner citations is highly skewed, with the largest share of patents (39%) having all citations added by the examiner.

The technological field panel reveals a striking degree of field-specific variation. Patents in mechanical and computer/electronics fields have high shares of examiner-added citations, while the lowest proportions are found in chemicals and pharmaceuticals. Examiners account for *all* citations in 45 percent the patents in the computers/communications and electrical/electronic fields, and only in one quarter of drug and medical patents. The observed shares are plausibly linked to differential commitments to patent quality across fields: technologies in which individual patents are more valuable (drugs and chemicals) have higher shares of applicant prior art than in complex technologies where intellectual property rights are typically more fragmented.

Turning next to differences across types of assignees, the data show that at both the dyad and patent levels, academic institutions, and small entities have lower examiner shares than other types of assignees. This could reflect the greater reliance of these organizations on "markets for technology" to the extent that stronger patents—those more likely to survive validity challenges—are more readily out-licensed. But these patterns also could reflect other differences between university and government patents. For example, academic patents are disproportionately in the biomedical arena (Mowery et al. 2001), where examiner shares of citations are lowest. Accordingly, we re-examine these differences in multiple regressions that control for field effects.

Non-US assignees submit far fewer applicant citations than US assignees, which corresponds to our discussion of institutional practices in countries outside the US: 63

percent of citations in patents assigned to foreign assignees are added by US examiners (74 percent on the average patent), while the analogous share for US assignees is 34 percent and 55 percent, respectively. Since foreign patents made up a significant share of total granted US patents (43 percent) they have a large influence on the weight of examiner-inserted patents in total patents.

Decomposition of foreign applicants by nationality shows that the examiner share of citations is highest in East Asian countries, namely Taiwan (where examiners account for 87 percent of citations, 92 percent on the average patent) South Korea (71 percent/80percent), and Japan (69 percent/78 percent). Among European countries, Germany has the highest share of examiner citations. Since many of these foreign firms patent heavily in the electrical and electronic fields, these trends may reflect field specific differences. Accordingly, we employ multivariate regression analyses to determine the effect of nationality on examiner shares.

The next panel shows that firms that are granted a high number of patents are adding the fewest applicant citations. At the both at dyad and patent levels, the examiner share of citations is higher for the most experienced assignees: the average patent for this group has 68% of citations added by examiners, against 59% and 61% for the second and third cohort. These differences are also present in the distribution of examiner shares across patents (columns 4-7); the most experienced assignees have the highest shares of patents (43%) with all examiner citations. Again, these numbers may be driven by country and field effects: if many of the most experienced applicants are non-US and/or in computer/electronics fields, that could explain their high examiner citation shares. Again, we control for these factors in the regression analysis.

Examiner experience similarly shows that the most experienced examiners having the highest shares of examiner-added citations, followed by the least experienced examiners, although the difference between the cohorts is not as pronounced as it is for applicants, and indeed the breakdowns on the distribution of citations across patents shown in columns 4 through 7 do not show strong differences across examiner experience cohorts.

Only 7 percent of patents do not list attorneys on the front page, attesting to the importance of legal professionals in preparing patent applications. As expected, patents that do not list lawyers have higher examiner shares at both the dyad and patent levels, and have more patents without any applicant citations. This likely reflects the role of attorneys in performing prior art searches, which appear as applicant citations. The use of attorneys may also have a positive correlation with other professionals, such as patent searchers, in preparing applications. However, we cannot estimate the impact of professional searchers as they are not listed on patents.

5. Regression Model and Results

We now turn to our regression models that allow us to estimate the marginal impact each of the variables in Table 1 on the share and count of citations added by patent examiners. We estimate the following specification using data at the patent level*:*

$$Y_i = X_i\beta + \varepsilon_i \qquad (1)$$

where Y_i is either the percentage or count of examiner-imposed citations, X_i is a vector of characteristics of the patent described in section 3, and ε_i is an iid error term. When the dependent variable is a percentage, we estimate equation (1) using Ordinary Least Squares (OLS) and two-sided Tobit, since percentage values are bounded between

0 and 1.When the dependent variable are counts of citations, we estimate equation (1) using a negative binomial, and add a variable *citations* that is a count of applicant citations on a patent. Since patents granted to a given assignee may not be independent observations, in all models we specify that observations are independent across but not necessarily within assignees, and adjust the estimated standard errors accordingly, using robust standard errors clustered by assignees on the cited patents[9].

The unit of analysis in each of these regressions is any patent that contains at least one prior art citation; accordingly, we drop 11,278 patents with no citations, leaving us with 491,112 in our sample, which cite 6,235,655 patents. Table 2 provides summary statistics for each of the variables in our models. Table 3 shows results from regressing the variables above on the percent (Columns 1 and 2) and number (Column 3) of examiner citations on a given patent.10

The goal of the regression analysis is to observe whether the patterns in Table 1 are still present when we control for other variables that might affect the univariate shares. Overall the regressions provide support for the key findings in the univariate descriptive statistics, and allow us to confirm that the characteristics shown in Table 1 are associated with high shares of examiner citations. Rather than discuss in detail results that corroborate those already shown in Table 1, we limit our discussion to highlight the most notable results in the models.

---

[9] We also estimated GLM models (a la Papke and Wooldridge 1996) that explicitly recognize that the dependent variable, examiner share of citations, is a proportion . We also estimate logit models. Since the tails of the distribution are relatively thick (eg, patents with all or no citations added by examiners, which are 8 and 39% of patents, respectively) we check that our results are not being driven by these patents. We relate the suite of variables to the probability that all or none of the citations come from examiners. We also estimate models where we remove patents that fall into these two categories. Results are qualitatively similar to those reported here, and are available from the authors on request.

[10] Regressions show field effects at the level of broad technology categories. Models that use more detailed 3-digit patent classes show similar results and are available on request.

The field effects remain after controlling for nationality, assignee size, and other factors: examiner-added citations are significantly lower in the drug/medical and chemical fields, and higher in computer, electrical and electronics fields, whether measured as a percentage or number of citations. Academic institutions and small firms have significantly lower examiner shares, even after controlling for the technological fields they are patenting in. Non-US assignees receive significantly higher shares and numbers examiner citations, even after controlling for field effects. Interestingly, the results on assignee experience, which may have been driven by field and nationality, still hold in the regressions for shares of examiner citations. However, they are not significant when measured as a count of citations. It is possible that this reflects lower total citations on many of these patents, which would occur if they were of relatively minor value.

We note that the similarity between the negative binomial model, which controls for total citations on a patent, and the models that use citation shares indicates that our results are not driven by patents that contain very few (or many) applicant citations. The negative and significant coefficient on the count variable of applicant citations shows that controlling for other factors, the more applicants add citations, the fewer citations inserted by examiners.

One potential complication is the patent grant lag, and the possibility that examiners but not applicants disproportionately add citations during patent pendency. If this were true, and the grant lag were correlated with other variables in the models above, our results could be biased.

As discussed in the introduction, the applicant's duty of candor extends to the entire patent prosecution period, so if applicants become aware of material prior art after filing the initial application, they are required to disclose this art. The data show that of the 12 percent of all citations to patents that issue after the citing patent is filed, 25 percent come from applicants. While lower than the overall applicant share of citations (41 percent), this suggests that applicants are active in contributing prior art during patent prosecution, but that examiners are much more likely to do so.[11]

To account for this, we re-ran our models and explicitly controlled for the grant lag. Columns 4-6 of Table 3 show these results. The estimated coefficient on the grant variable is positive and significant, indicating that longer lags are associated with higher examiner shares of citations. However, none of the other coefficients change materially.

Another potential complication is that citations in applications filed via the PCT are coded as applicant citations, even if they are likely to emanate from foreign search reports. To see if these citations affect our results, we re-estimated all of the models after eliminating the 42,806 patents that were filed through the PCT process. Columns 7-9 of Table 3 show that results are nearly identical to those in models which included PCT-filed patents.

6. Decomposition of variance

The analysis in Section 5 shows the marginal impact of assignee, examiner, and field effects on examiner citations. In this section, we assess the importance of each of these classes of variables in explaining the total variation in the examiner share.

Following Skondral & Rabe-Hesketh (2004), we estimate a variance component model through a linear mixed model with random intercepts. We estimate the following model:

---

[11] Our analysis of 50 file wrappers for issued patents shows that applicant prior art introduced during patent prosecution is generally introduced in response to examiner rejections, or when applications are granted in other countries.

$$y_{ijkl} = \mu + u_j + v_k + w_l + r_{ijkl} \quad (2)$$

where y_{ijkl} is the percentage of examiner citations (or self-citations) in patent i, applied by assignee j, processed by examiner k and classified in technology l,

μ is a fixed intercept across observations,

$u_j, v_k, w_l$are random effects per assignee, examiner and technology groups such that $u_j \sim N(0, \sigma_u^2), vj \sim N(0, \sigma_{uv}^2)$, and $w_{\ j} \sim N(0, \sigma_w^2)$, and $r_{ijkl} \sim N(0, \sigma^2)$.

We chose this model over a more traditional analysis of variance (ANOVA) approach for several reasons. First, the large number of observations and categories in our dataset (53,022 assignees, 2,069 examiners, and 419 technology classes) make an ANOVA analysis of all the data computationally infeasible. A linear mixed model implemented through calculations that use spare variance-covariance matrices allows inclusion of groups with numerous elements. Second, the three dimensions in our model are not orthogonal: technologies, applicant, and examiner effects overlap. Accordingly, the explained variance cannot be partitioned uniquely among the independent variables, and the order in which variables are introduced will therefore affect the results. Mixed linear models incorporate the covariance between groups so the estimates are obtained controlling for variance attributed to other groups in the model.

The first column of Table 4 shows the results from estimating this model. The first panel shows the variance estimates for assignees, examiners, technologies and the model residual, and the second one shows the estimates for the fixed intercept, which can be thought of as the model average.

Overall, these three groups explain about 36 percent of the variance in the model. But the differences across the categories are striking. Of the explained variation, assignee effects explain 91 percent, examiner effects 9 percent, and technology categories only 8 percent. In other words, heterogeneity across applicants explains almost all the variation in shares of examiner citations on patents!

The magnitude of variation attributed to individual assignees could be driven by assignees with only one patent. As a robustness check we also estimated the model for assignees with between 100 and 1,000 patents (Column 2) and at least 1000 patents (Column 3)[12]. This last model captures many of the assignees in the "most experienced" cohort in our multivariate regressions. We find that the effect of assignee heterogeneity is still very high, though it falls by a few percentage points as assignee patent volume increases. Moreover, the variation explained by technologies rises from 8 percent in the full dataset to nearly 20 percent of variation for assignees with over 1000 patents. This is in line with our expectation that field effects should matter for citations, and also supports the validity of employing technology fixed effects in patent-based studies. However, field effects still account less than one-quarter of the variation in examiner citation shares as compared to individual assignees. This finding, coupled with our earlier finding that high patent-volume assignees have the highest share of citations, points to individual firm practices as an important dimension explaining high examiner shares on patents.

7. Firm-Level Analysis

Overall, these results point to individual assignee effects as a dominant factor affecting variation in the examiner share of citations. We further explore this with descriptive statistics on large-scale patent assignees. In this section, we shift our unit of analysis from individual patents to organizations. Excluding patents assigned to

[12] The sample size changes accordingly from 429,984 observations to 100,288 observations (assignees with between 100 and 1000 patents) and 180,924 (assignees with at least 1,000 patents).

individuals and unassigned patents, our sample includes 53,022 applicants. For each of these assignees, we calculated the share of citations inserted by examiners (at the dyad level) and the average examiner share of citations in the firm's patents (averaged across all of the firm's patents).

The x-axis of Figure 1 shows the average examiner share of citations for the top 10 assignees in each of the 6 broad field categories. As suggested by our earlier findings in the univariate and multivariate analyses, examiner shares are lowest for firms in the drugs/medicine and chemical fields, and highest in electronics and computers/communications.

However the chart also reveals considerable within-field variation, as our variance decomposition analyses would have predicted. Compare, for instance, Fuji versus Kodak in the chemical field, Pfizer versus Merck in the drugs/medical field, or Micron versus NEC in electronics. In each of these cases, we observe firms in similar technologies with very different examiner shares of citations.

We see further evidence of within-field firm effects in data showing firms ranked by their patent volumes and examiner shares. Table 5 shows the largest patentees ranked by (a) total patents filed in 2001-2003; (b) share of examiner citations in all patents and (c) share of examiner citations in self-citations. The first group includes firms that fall into the most experienced applicants, and we note the dominance of large, technology-intensive firms in electronics, computers, and communications. However, there is a great deal of variation by firm in the shares of examiner citations in all citations and self-citations, giving further (anecdotal) evidence that firm-level patent practices are important in assessing the level of applicant prior art.

Approximately 27 percent of the firms in our sample (14,291) have all citations in all of their patents are inserted by examiner. And of the 14,648 assignees with at least one patent with a self-citation, 43 percent (6,362) had all self-citations inserted by examiners. Table 5 shows the largest ten patentees in each of these two groups, respectively. Here again, we note the dominance of firms in electronics, computers and communications, but the firms in these categories are much smaller in terms of overall volume of patenting. These data correspond to our findings in Table 1 of a U-shaped relationship between experience and share of examiner shares, though in the regressions that control for field effects only the coefficient of the top tier of experienced assignees is significant.

These firm-level differences could reflect unobserved differences in the technological breadth of firms' patent portfolios. Firms frequently patent outside their core product areas, reflecting both firm-level R&D strategies as well as idiosyncrasies in the patent classification system. We re-estimate our models in table 3 at the firm level, and find the results are unchanged from those at the patent level. Furthermore, our decomposition of variance analysis estimates the effect of technological field separately from firm effects, so we can confirm that this within-technology variation is not driven by diversity in patent portfolios[13]. These differences more likely reflect unobserved differences in firms' patent strategies. Dimensions by which they may differ could include: the choice to patent broadly vs. only patenting valuable inventions that the firm intends to pursue internally or through licensing; efforts to invest in search ex ante versus waiting for signals from the patent office to collect additional references; and the crowdedness of the fields they select for R&D efforts.

---

13 Results for the regression analysis and the variance decomposition analysis at the firm level are available on request.

8. Evidence from self-citations

Our analysis so far provides evidence on the factors associated with high shares of patent citations added by examiners. But what is the meaning of high examiner citation shares? There are two possible interpretations. One is that the applicant share of citations (which we can observe) correlates with applicant search efforts (which we cannot observe): that, is high shares of examiner citations indicate low applicant search effort. But another interpretation is that applicants do expend effort in conducting prior art searches but fail to uncover all the relevant citations, such that examiners then find those citations that were missed by applicants. In this case, applicants do include citations they deem relevant, but they do so with error, and examiners fill the remaining gaps. If applicants lack search capabilities or resources, those gaps may be significant. In this case, variation in examiner shares correlates with applicant search capabilities, but not search efforts.

Self-citations help to identify whether applicant citations correlate with low applicant search effort or low applicant search capabilities. Alcacer and Gittelman (2006) and Sampat (2007) argue that self-citations are a more useful signal of the intensity and quality of applicant search than all applicant citations since if applicants are conducting cursory prior art searches, even if they were resource constrained, they presumably would cite their own previous patents if relevant.

Accordingly, if firms with high examiner share of citations are simply resource constrained, we would expect no relationship between the examiner share of self-citations (which are relatively easy to find, and thus less affected by resource constraints) and other citations. By contrast, if these firms are not searching for prior art, firms with systematically lower shares of all citations would also have systematically lower shares of self-citations.

We explore this issue in several ways. First, Table 6 shows the share of examiner self-citations across the same dimensions used in Table 1. Note that the distribution of examiner self-citations across technology fields, assignee types, country of origin, assignee experience, examiner experience and legal assistance is very similar to the one for share of examiner citations in Table 1. Second, we estimate regressions that are similar to those in Table 3 using self-citations instead of all citations as dependent variable: the results are very similar to those reported for all citations. At the patent level, the same factors affecting examiner shares of all citations affect self-citations. Third, Figure 1 shows a scatter plot of the examiner share of citations and self-citations for the top 50 patentees in 6 broad field categories; the X and Y axes correspond to the firm level percentage of examiner "non-self" and "self" citations respectively. Note the strong positive relationship between the two, across fields. Firms with low examiner share of self-citations tend to have low examiner share of other citations. If examiner share of self-citations are indeed a useful proxy for applicant search effort, this provides additional evidence that the variation in the examiner share of total citations correlates with differences in firms' prior art search effort, rather than capabilities . Fourth, we repeat our decomposition of variance analysis using only self-citations. The results are similar to those for all citations; firm-specific effects explain most variation on examiner self-citation shares. [14]

[14] Results for the regression analysis and the variance decomposition analysis at the firm level are available on request.

9. Discussion and conclusions

The ready availability of patent citation data has been a tremendous boon to applied research on knowledge and innovation. The role of examiners in the generation of patent citations has been thought to potentially complicate these analyses, but difficult to study. Taking advantage of a change in the way patent citation data is reported by the USPTO, starting in 2001, this paper summarized basic facts on examiner citations, and provided a descriptive analysis of factors associated with the examiner-share of citations in a patent.

The unconditional means and multivariate analyses suggest that examiner citations in patents appear to be systematically related to a number of characteristics of technological fields, types of applicants, countries of origin, and examiners.

To highlight several interesting results, our analyses suggest that examiner shares are highest in fields where intellectual property tends to be fragmented (computers and communications, electrical/electronics), and lowest in fields where patents have been shown to be more important in appropriating returns to R&D (biomedical and chemical patents). Academic institutions have lower examiner shares, even after controlling for technological field effects. Strikingly, the most prolific patentees tend to have very high shares of examiner citations. This could reflect a strategy to build patent portfolios, with relatively low value placed on any single patent (Sampat 2007). We also find that examiner citations shares are specially high for foreign firms: given that foreign firms are more likely to file PCT applications overseas and that citations added in these (by foreign examiners) are attributed to applicants in the US, the actual number of applicant citations may be overestimated. While properly testing the specific reasons for differences in citation practices across types of institutions and types of firms is beyond the scope of this paper, these simple stylized facts suggest that all citations are not created equal.

The results of the decomposition of the variance underscore the importance of firm effects. Indeed, individual firm effects account for most of the explained variance in examiner shares, far more than field effects, which are standard controls in the literature. Although work by Ziedonis (2004) and Lerner (1995), for example, explores issues related to strategic patenting, there has been little research on firm-specific patent strategies, including decisions and motives to search for prior art. The examiner citation data, including data on examiner share of self-citations, may facilitate work in this area.

The aim of this paper was primarily descriptive, and it would be premature to draw policy implications from our results. That said, we do note that applicant prior art searching, and its relationship to patent quality, are central issues in contemporary patent reform initiatives in the U.S. We interpret low applicant shares of citations as indicative of low applicant search effort, and our findings on self-citation corroborate this interpretation. Using these data to further explore prior art searching and implications for patent quality and patent system reform is another important task going forward.

We started this paper with a discussion of citation data as indicators. The data we reported suggest that examiners account for a significant share of patent citations: 41 percent at the dyad level, and 63 percent at the patent level. Moreover, the data suggest that the extent to which examiners contribute citations is not randomly distributed across patents but related to a number of field, firm, examiner, and invention specific variables. Thus an overall message that emerges from our work is that patent citations reflect the complicated interaction of applicant and examiner strategies, capabilities, and incentives. One implication of this is that research using citation data should consider whether differences in these types of factors could affect results, and introduce appropriate controls or robustness checks if possible.[15] We cannot provide definitive guidance on the right way to deal with these issues in empirical work, since this will vary across contexts, and, as noted repeatedly above, more work is needed to examine the extent and magnitude of potential biases. Rather, we hope that this paper helps raise awareness of potential complications, and stimulates further research on the interpretation of citation data.

[15] In some cases, it may make more sense to use applicant citations along, rather than pooled citations, though our analyses suggest that these too may not be pristine measures of underlying concepts like "learning" "spillovers" or "knowledge flows".

Figure 1
% examiner citations vs. % examiner self-citations

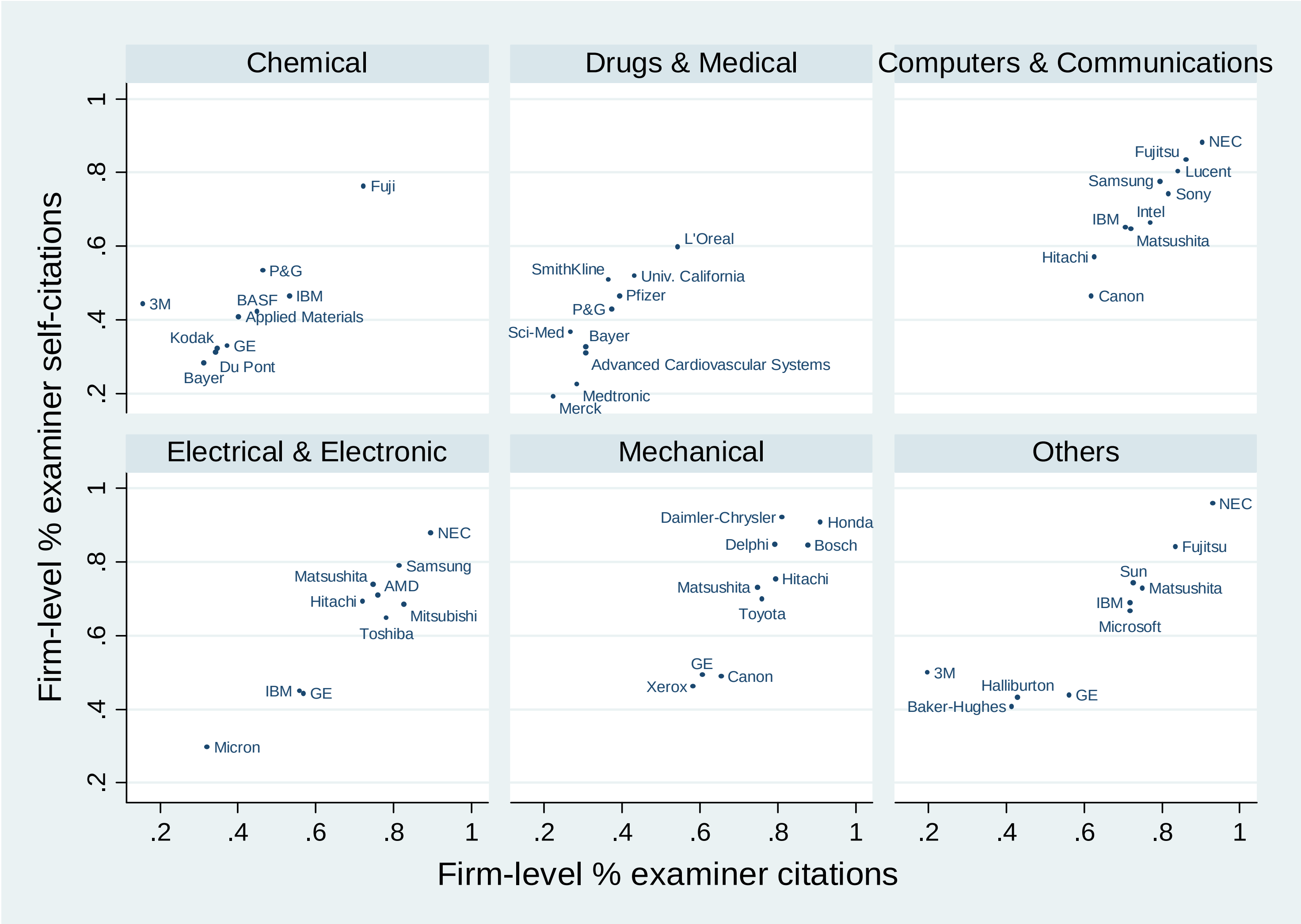

Table 1
Propensity of inventor and examiner citations by technological field, assignee type, origin, assignee nationality, assignee experience, lawyer and examiner experience

	Citations to US Patents							
	No. of Patents	Without	% examiner		Distribution at patent level			
			Dyad level	Patent level	Zero	(0%, 50%]	(50%,100%)	All
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Full sample	502,390	2%	41%	63%	8%	28%	25%	39%
Technological Field								
Chemical	69,180	4%	33%	52%	15%	32%	23%	30%
Drugs & Medical	56,651	9%	21%	43%	22%	37%	17%	25%
Computers & Communications	101,181	1%	46%	70%	4%	25%	26%	45%
Electrical & Electronic	106,323	1%	47%	69%	4%	26%	26%	45%
Mechanical	85,352	1%	49%	67%	5%	26%	28%	41%
Others	83,703	1%	46%	65%	5%	29%	27%	39%
Assignee Type								
Academia	10,521	10%	29%	45%	22%	34%	17%	26%
Small firms	61,943	2%	34%	57%	8%	35%	21%	35%
Large firms	365,047	2%	42%	64%	8%	27%	26%	39%
Government	3,038	5%	57%	68%	10%	21%	17%	51%
Others	61,841	1%	51%	68%	5%	27%	23%	46%
Origin								
American	286,824	2%	34%	55%	9%	38%	22%	32%
Foreign	215,566	3%	63%	74%	6%	15%	30%	49%
Nationality								
Japan	103,597	3%	69%	78%	5%	13%	26%	56%
Germany	33,984	3%	67%	71%	8%	14%	37%	41%
Taiwan	16,100	1%	87%	92%	1%	6%	10%	83%
France	11,945	4%	62%	68%	8%	16%	41%	35%
United Kingdom	11,428	4%	49%	64%	10%	21%	36%	34%
South Korea	11,268	2%	71%	80%	3%	13%	23%	61%
Canada	10,463	2%	42%	58%	8%	34%	25%	32%
Italy	5,182	3%	66%	72%	7%	14%	35%	43%
Sweden	4,937	2%	55%	65%	8%	18%	46%	27%
Switzerland	4,092	3%	54%	64%	12%	17%	40%	31%
Assignee Experience								
Most experienced	184,201	2%	48%	68%	6%	24%	26%	43%
Experienced	103,982	3%	37%	59%	10%	29%	26%	35%
Least experienced	214,207	2%	39%	61%	8%	31%	24%	37%
Legal assistance								
With lawyer	465,993	2%	41%	63%	8%	28%	25%	39%
Without lawyer	36,397	1%	46%	66%	6%	26%	24%	44%
Examiner Experience								
Most experienced	242,128	1%	44%	65%	6%	28%	26%	40%
2nd Tier	148,480	3%	39%	61%	9%	29%	24%	38%
3rd Tier	93,369	3%	39%	61%	9%	29%	24%	38%
Least experienced	18,413	3%	42%	63%	7%	28%	26%	38%

Table 2
Summary statistics for models at patent level

Variable	Citations				
	Obs	Mean	Std. Dev.	Min	Max
Dependent Variables					
examiner citation shares	491,112	0.631	0.373	0	1
assignee citation number	491,112	5.265	4.724753	0	299
Technology					
Chemical	491,112	0.135	0.341	0	1
Drugs & Medical	491,112	0.105	0.306	0	1
Computers & Communications	491,112	0.204	0.403	0	1
Electrical & Electronic	491,112	0.215	0.410	0	1
Mechanical	491,112	0.172	0.378	0	1
Others	491,112	0.169	0.375	0	1
Assignee characteristics					
Academia	491,112	0.019	0.137	0	1
Small firms	491,112	0.124	0.329	0	1
Large firms	491,112	0.727	0.446	0	1
Government	491,112	0.006	0.077	0	1
Others	491,112	0.124	0.330	0	1
Foreign	491,112	0.474	0.499	0	1
Most experienced	491,112	0.368	0.482	0	1
Experienced	491,112	0.204	0.403	0	1
Least experienced	491,112	0.427	0.495	0	1
Lawyer	491,112	0.927	0.260	0	1
Examiner characteristics					
Most experienced	491,112	0.593	0.491	0	1
Examiner experience: 2nd Tie	491,112	0.188	0.391	0	1
Examiner experience: 3rd Tier	491,112	0.183	0.386	0	1
Least experienced	491,112	0.036	0.187	0	1
examiner citation number	491,112	5.265	4.725	0	299

Table 3
Baseline models

	All Citing patents, all citations			Control for grant lag of citing patent			Without PCT patents		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable	% examiner citations	% examiner citations	# examiner citations	% examiner citations	% examiner citations	# examiner citations	% examiner citations	% examiner citations	# examiner citations
Estimation method	OLS	Tobit	Negative binomial	OLS	Tobit	Negative binomial	OLS	Tobit	Negative binomial
Technology									
Chemical	-0.133	-0.223	-0.213	-0.135	-0.226	-0.218	-0.136	-0.226	-0.222
	[0.007]**	[0.012]**	[0.017]**	[0.007]**	[0.012]**	[0.017]**	[0.007]**	[0.011]**	[0.018]**
Drugs & Medical	-0.195	-0.32	-0.299	-0.199	-0.327	-0.313	-0.198	-0.325	-0.307
	[0.010]**	[0.016]**	[0.031]**	[0.010]**	[0.016]**	[0.031]**	[0.010]**	[0.017]**	[0.032]**
Computers & Communications	0.049	0.095	0.028	0.041	0.083	0.00	0.049	0.095	0.03
	[0.010]**	[0.016]**	[0.012]*	[0.010]**	[0.016]**	[0.012]	[0.010]**	[0.016]**	[0.013]*
Electrical & Electronic	0.02	0.052	-0.042	0.02	0.053	-0.041	0.021	0.053	-0.042
	[0.011]	[0.019]**	[0.011]**	[0.011]	[0.019]**	[0.012]**	[0.011]	[0.019]**	[0.012]**
Mechanical	-0.002	-0.001	-0.029	0	0.001	-0.024	-0.003	-0.004	-0.032
	[0.006]	[0.010]	[0.005]**	[0.007]	[0.010]	[0.006]**	[0.007]	[0.011]	[0.006]**
Assignee characteristics									
Academia	-0.128	-0.243	-0.239	-0.134	-0.252	-0.252	-0.136	-0.255	-0.248
	[0.013]**	[0.021]**	[0.021]**	[0.013]**	[0.020]**	[0.021]**	[0.013]**	[0.022]**	[0.022]**
Small firms	-0.122	-0.228	-0.097	-0.096	-0.167	-0.074	-0.098	-0.172	-0.073
	[0.004]**	[0.006]**	[0.011]**	[0.003]**	[0.005]**	[0.010]**	[0.003]**	[0.006]**	[0.010]**
Large firms	-0.093	-0.163	-0.067	-0.125	-0.231	-0.104	-0.131	-0.241	-0.105
	[0.003]**	[0.005]**	[0.009]**	[0.004]**	[0.007]**	[0.011]**	[0.004]**	[0.007]**	[0.012]**
Government	0.024	0.035	0.005	0.018	0.029	-0.016	0.021	0.037	0.006
	[0.035]	[0.071]	[0.040]	[0.035]	[0.072]	[0.040]	[0.034]	[0.068]	[0.039]
Foreign	0.208	0.342	0.166	0.209	0.343	0.167	0.207	0.347	0.169
	[0.011]**	[0.018]**	[0.024]**	[0.011]**	[0.018]**	[0.024]**	[0.012]**	[0.021]**	[0.023]**
Most experienced	0.043	0.076	0.009	0.041	0.074	0.00	0.049	0.08	0.015
	[0.013]**	[0.024]**	[0.014]	[0.013]**	[0.024]**	[0.014]	[0.014]**	[0.025]**	[0.015]
Experienced	0.009	0.013	-0.007	0.008	0.012	-0.011	0.011	0.014	-0.005
	[0.006]	[0.010]	[0.007]	[0.006]	[0.010]	[0.007]	[0.006]	[0.011]	[0.008]
Lawyer	-0.045	-0.08	-0.034	-0.045	-0.079	-0.033	-0.047	-0.081	-0.038
	[0.010]**	[0.019]**	[0.010]**	[0.010]**	[0.019]**	[0.010]**	[0.010]**	[0.019]**	[0.011]**
Examiner characteristics									
Examiner experience: 1st Tier	0.0001	0.003	0.006	0.004	0.009	0.019	0.001	0.006	0.007
	[0.004]	[0.007]	[0.007]	[0.004]	[0.007]	[0.007]**	[0.004]	[0.007]	[0.007]
Examiner experience: 2nd Tier	-0.014	-0.02	-0.028	-0.011	-0.015	-0.019	-0.011	-0.016	-0.027
	[0.004]**	[0.007]**	[0.006]**	[0.004]**	[0.007]*	[0.006]**	[0.004]**	[0.007]*	[0.006]**
Examiner experience: 3rd Tier	-0.01	-0.013	-0.015	-0.007	-0.01	-0.008	-0.008	-0.011	-0.015
	[0.004]*	[0.007]	[0.006]**	[0.004]	[0.007]	[0.006]	[0.004]*	[0.007]	[0.006]*
Citations			-0.042			-0.042			-0.041
			[0.003]**			[0.003]**			[0.003]**
Lag				0.014	0.021	0.042			
				[0.002]**	[0.005]**	[0.004]**			
Constant	0.693	0.891	-0.224	0.66	0.843	-0.318	0.698	0.899	-0.226
	[0.011]**	[0.020]**	[0.027]**	[0.013]**	[0.025]**	[0.034]**	[0.011]**	[0.020]**	[0.025]**
Year dummy variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	491,112	491,112	491,112	491,112	491,112	491,112	448,306	448,306	448,306

Robust standard errors in brackets
* significant at 5%; ** significant at 1%

Model 1, 2, 4, 5, 7, 8: dependent variable is % of examiner citations at patent level
Model 3, 6 and 9: number of examiner citations at patent level
The number of observations for models 7, 8 and 9 is reduced because the sample used excludes patents associated to PCTs

Table 4
Variance Decomposition Analysis

		Citations		
		(1)	(2)	(3)
Random effects				
Variance	Groups name			
	Assignees	0.044	0.038	0.026
	Examiners	0.004	0.005	0.004
	Technologies	0.004	0.003	0.006
	Residual	0.093	0.097	0.093
Fixed effects				
	Intercept	0.601	0.600	0.631
		[0.004]**	[0.007]**	[0.011]**
Observations		429,984	100,288	180,934
Groups:	Assignees	53,002	1,840	275
	Examiners	2,031	1,592	1,891
	Technologies	416	402	397
% of variance explained by model		0.36	0.32	0.28
% of explained variance attributed to				
	Assignees	0.91	0.89	0.87
	Examiners	0.09	0.11	0.13
	Technologies	0.08	0.08	0.18

Table 5
Average propensity of inventor and examiner citations by assignee for all patents granted from 2001 to 2003

Ranking	Assignee name	Patents filed 01-03 (1)	Citations % examiner Dyad level (2)	Citations % examiner Patent level (3)	Self-citations % examiner Dyad level (8)	Self-citations % examiner Patent Level (9)
Top 25 assignees						
1	INTERNATIONAL BUSINESS MACHINES CORP.	10114	54%	66%	51%	59%
2	CANON KABUSHIKI KAISHA	5760	51%	63%	31%	48%
3	MICRON TECHNOLOGY, INC.	5184	20%	34%	17%	31%
4	NEC CORPORATION	4955	87%	90%	88%	88%
5	HITACHI, LTD	4765	53%	69%	53%	64%
6	MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.	4758	66%	73%	63%	70%
7	SONY CORPORATION	4108	65%	84%	53%	78%
8	SAMSUNG ELECTRONICS CO., LTD.	4087	68%	80%	73%	79%
9	MITSUBISHI DENKI KABUSHIKI KAISHA	3800	77%	82%	64%	68%
10	FUJITSU LIMITED	3679	73%	83%	72%	79%
11	GENERAL ELECTRIC COMPANY	3662	33%	54%	30%	43%
12	INTEL CORPORATION	3478	61%	77%	45%	63%
13	TOSHIBA CORPORATION	3463	68%	76%	55%	63%
14	ADVANCED MICRO DEVICES, INC.	3145	58%	73%	54%	68%
15	HEWLETT-PACKARD COMPANY	2442	55%	67%	34%	50%
16	LUCENT TECHNOLOGIES INC.	2392	76%	82%	70%	76%
17	TEXAS INSTRUMENTS, INCORPORATED	2279	68%	83%	53%	72%
18	KONINKLIJKE PHILIPS ELECTRONICS N.V.	2207	72%	77%	95%	96%
19	EASTMAN KODAK COMPANY	2161	35%	43%	30%	39%
20	ROBERT BOSCH GMBH	2132	87%	86%	85%	84%
21	SIEMENS AKTIENGESELLSCHAFT	2127	76%	75%	68%	68%
22	MOTOROLA, INC.	2100	51%	74%	50%	60%
23	FUJI PHOTO FILM CO., LTD	2074	73%	78%	65%	75%
24	XEROX CORPORATION	2022	37%	53%	24%	39%
25	SEIKO EPSON CORPORATION	1880	57%	71%	64%	76%

(continued on next page)

Table 5
Average propensity of inventor and examiner citations by assignee for all patents granted from 2001 to 2003 (continued)

Ranking	Assignee name	Patents filed 01-03 (1)	Citations: % examiner Dyad level (2)	Citations: % examiner Patent level (3)	Self-citations: % examiner Dyad level (8)	Self-citations: % examiner Patent Level (9)
Top 10 assignees with largest % of examiner citations to US patents						
631	UMAX DATA SYSTEMS INC.	83	100%	100%	100%	100%
690	FOXCONN PRECISION COMPONENTS, CO., LTD.	76	100%	100%	100%	100%
1317	BEHAVIOR TECH COMPUTER CORP.	38	100%	100%	100%	100%
1401	AMERICAN GNC CORPORATION	35	100%	100%	100%	100%
1762	FUJITSU GENERAL LIMITED	27	100%	100%	100%	100%
1797	UNITED SEMICONDUCTOR CORP.	27	100%	100%	100%	100%
1843	FARADAY TECHNOLOGY CORPORATION	26	100%	100%	100%	100%
1881	AUDEN TECHNO CORP.	25	100%	100%	100%	100%
1915	POLYMATECH CO., LTD.	25	100%	100%	100%	100%
1963	TEXAS INSTRUMENTS - ACER INCORPORATED	24	100%	100%	100%	100%
Top 10 assignees with largest % of self-citations						
129	AGERE SYSTEMS INC.	608	68%	78%	100%	100%
146	HYNIX SEMICONDUCTOR INC.	347	68%	77%	100%	100%
193	NEC ELECTRONICS CORPORATION	263	85%	87%	100%	100%
195	NOKIA CORPORATION	260	70%	74%	100%	100%
204	SAMSUNG SDI CO., LTD.	252	79%	80%	100%	100%
249	COMPAQ INFORMATION TECHNOLOGIES GROUP	210	56%	72%	100%	100%
254	FORD GLOBAL TECHNOLOGIES, L.L.C.	204	35%	51%	100%	100%
310	NOKIA NETWORKS OY	166	75%	74%	100%	100%
327	ALSTOM (SWITZERLAND) LTD	155	69%	71%	100%	100%
332	DELTA ELECTRONICS INC.	155	94%	96%	100%	100%

Table 6

Propensity of inventor and examiner self-citations by technological field, assignee type, origin, assignee nationality, assignee experience, lawyer and examiner experience

	Self-citations to US Patents							
	No. of Patents	Without	% examiner		Distribution at patent level			
			Dyad level	Patent level	Zero	(0%, 50%]	(50%,100%)	All
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Full sample	502,390	67%	41%	57%	30%	11%	11%	48%
Technological Field								
Chemical	69,180	63%	35%	50%	37%	12%	11%	40%
Drugs & Medical	56,651	67%	29%	45%	43%	11%	9%	38%
Computers & Communications	101,181	64%	46%	63%	24%	11%	11%	54%
Electrical & Electronic	106,323	66%	42%	60%	28%	10%	11%	51%
Mechanical	85,352	68%	45%	59%	28%	11%	12%	50%
Others	83,703	75%	45%	58%	31%	9%	11%	50%
Assignee Type								
Academia	10,521	70%	35%	46%	46%	7%	7%	41%
Small firms	61,943	77%	42%	56%	36%	6%	8%	50%
Large firms	365,047	61%	41%	58%	29%	11%	11%	48%
Government	3,038	62%	56%	61%	33%	5%	8%	55%
Others	61,841							
Origin								
American	286,824	67%	34%	49%	37%	13%	11%	39%
Foreign	215,566	68%	55%	69%	22%	7%	10%	61%
Nationality								
Japan	103,597	58%	53%	69%	20%	8%	11%	61%
Germany	33,984	75%	57%	65%	27%	5%	10%	58%
Taiwan	16,100	86%	79%	83%	13%	3%	6%	79%
France	11,945	75%	49%	64%	26%	7%	11%	56%
United Kingdom	11,428	78%	54%	67%	25%	6%	10%	59%
South Korea	11,268	80%	77%	82%	14%	2%	5%	79%
Canada	10,463	81%	39%	57%	32%	10%	8%	50%
Italy	5,182	81%	57%	68%	25%	5%	7%	63%
Sweden	4,937	77%	60%	66%	26%	4%	11%	58%
Switzerland	4,092	77%	48%	60%	32%	6%	9%	53%
Assignee Experience								
Most experienced	184,201	51%	42%	59%	27%	12%	12%	48%
Experienced	103,982	63%	39%	55%	33%	10%	11%	46%
Least experienced	214,207	85%	42%	56%	36%	6%	7%	50%
Legal assitance								
With lawyer	465,993	67%	41%	57%	30%	11%	11%	48%
Without lawyer	36,397	75%	39%	56%	32%	11%	10%	47%
Examiner Experience								
1st Tier	242,128	68%	42%	58%	29%	11%	11%	49%
2nd Tier	148,480	67%	39%	56%	32%	10%	11%	47%
3rd Tier	93,369	65%	41%	57%	30%	11%	11%	48%
4rd Tier	18,413	66%	44%	59%	29%	10%	11%	50%

References

Alcacer, J. and Gittelman, M. (2006) "How do I know what you know? Patent examiners and the generation of patent citations." Review of Economics and Statistics, November 2006, 88(4): 774–779.

Allison, J.R. and Lemley, M.A. (1998) "Empirical evidence on the validity of litigated patents". American Intellectual Property Law Association Quarterly Journal, 26:185–269.

Allison, J.R., Lemley, M.A., Moore, K.A., Trunkey, R.D. (2003). "Valuable Patents", George Mason Law & Economics Research Paper No. 03-31.

Arora, A, Fosfuri, A, Gambardella, A. (2001) "Markets for Technology: The Economics of Innovation and Corporate Strategy." Massachusetts Institute of Technology..

Barton, J.H. "Non-Obviousness." (2001) Mimeo, Stanford University Law School.

Bawa, R. "Nanotechnology Patenting in the US." (2004) Nanotechnology Law and Business.

Cockburn, I, Kortum, S., Stern, S. (2003) "All Are Patent Examiners Equal: Examiners, Patent Characteristics, and Litigation Outcomes" in W. Cohen and W. Merrill eds. Patents in the Knowledge Based Economy. Washington, National Academies Press.

Cohen, W., Nelson, R., and J. Walsh. (2000) "Protecting their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)" NBER Working Paper 7552

Hall, B. and Ziedonis R. (2001) "The Patent Paradox Revisted: An Empirical Study of Patenting in the U.S. Semiconductor Industry" Rand Journal of Economics: 32(1): 101-128.

Jaffe, A. B. and Trajtenberg, M. (2002). "Patents, Citations, and Innovations: A Window on the Knowledge Economy." MIT Press, Cambridge.

Jaffe, A. B. and Lerner, J. (2006) "Innovation and Its Discontents : How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It", Princeton University Press.

Jaffe, A., Trajtenberg, M. Forgarty, M. (2000) "Knowledge spillovers and patent citations: Evidence from a survey of inventors" American Economic Review 90, 2:215-

Jennings, T. (1999). Transcript of "Public Hearing on Issues Related to the Identification of Prior Art During the Examination of Patent Application." http://www.uspto.gov/web/offices/com/hearings/priorart/0714pato.doc

Lemley, M. A. and Tangri, R. K., (2003). "Ending Patent Law's Willfulness Game". UC Berkeley Public Law Research Paper No. 142.

Lerner, J, (1995) "Patenting in the Shadow of Competitors," Journal of Law & Economics, University of Chicago Press, vol. 38(2), pages 463-95.

Lichtman, D. (2004) "Rethinking prosecution history estoppel." The University of Chicago Law Review, 71(151):151–182.

Merrill, R. C. L.; Stephen A.; Myers, M. B. (2004) "A Patent System for the 21st Century." The National Academies Press, Washington, DC.

Michel, J. and B. Bettels. (2001) "Patent citation analysis: A closer look at the basic input data from patent search reports." Scientometrics, 52(1): 185-20.

Mossinghof, B. . (1999). Transcript of "Public Hearing on Issues Related to the Identification of Prior Art During the Examination of Patent Application." http://www.uspto.gov/web/offices/com/hearings/priorart/0714pato.doc

Mowery, D.C.; Nelson, R.R.; Sampat, B,N.; and Ziedonis, A.A. 2001 "The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh–Dole Act of 1980." Research Policy 30:99–119.

Papke, L. E. and Wooldridge, J. M. (1996) "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." Journal of Applied Econometrics, 11(6):619–632.

Popp, D, Juhl, T, Johnson, D. (2004) "Time In Purgatory: Examining the Grant Lag for U.S. Patent Applications," Topics in Economic Analysis & Policy, vol. 4 Iss. 1, article 29.

Sampat, B. (2007) "When Do Applicants Search for Prior Art?." Working Paper.

Sampat, B. and Hegde, D. (2007) "Examiner Citations, Applicant Citations, and the Private Value of Patents " Working Papee..

Schneider, C. (2007). "The Determinants of Patent Applications Outcomes - Does Experience Matter?" MPRA Paper No. 3359.

Skondral, A; Rabe-Hesketh, S. (2004) "Generalized Latent Variable Modeling". Chapman & Hall/CRC Ed

Thomas, J. R., (2001). Collusion and collective action in the patent system: A proposal for patent bounties. University of Illinois Law Review, pages 316-322

Trajtenberg, M. (1990) “A Penny for your Quotes: patent citations and the value of innovations”, The RAND Journal of Economics, 21.

US Patent and Trademark Office (1998). Manual of Patent Examination and Procedure.

Wagner, R. P. and Parchomovsky, G. (2005), "Patent Portfolios." U of Penn. Law School, Public Law Working Paper, 56..

Ziedonis, R. H.(2004). "Don't Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms." Management Science 50(6):4-