Determinants of Patent Quality: An Empirical Analysis

Bhaven N. Sampat*
Assistant Professor
Mailman School of Public Health
and
School of International and Public Affairs
Columbia University

bns3@columbia.edu

September 2005

*This research was supported by funding from the Robert Wood Johnson Foundation. Iain Cockburn, Rebecca Eisenberg, Bronwyn Hall, Mark Lemley, Douglas Lichtman, Jean Lanjouw, David Mowery, Richard R. Nelson, Arti Rai, Mark Schankerman, Scott Stern, and Rosemarie Ziedonis provided helpful comments on an earlier draft of this paper. I also benefitted from feedback from participants at the 2004 NBER Summer Institute and the Berkeley "Innovation Seminar."

1 Introduction

In an 1813 letter discussing the originality of Oliver Evans' patent on "Elevators, Conveyers, and Hopper-boys" in light of previously known inventions and ideas, Thomas Jefferson, the first Commissioner of the U.S. Patent Office, pointed to "the difficulty of drawing a line between the things which ares worth to the public the embarrassment of an exclusive patent, and those which are not" (The Writings of Thomas Jefferson, p. 355).

Though following Jefferson's tenure the patent office operated primarily as a registration system, today patent examiners draw the line between what is and is not patentable by comparing applications to related information already in the public domain, or the *prior art*. And there is growing concern that difficulties examiners face obtaining access to the relevant prior art are contributing to the issuance of patents of questionable validity. These concerns about diminished patent "quality" have prompted a range of calls for patent system reform (NAS 2004; FTC 2003).

While central to contemporary debates about patent quality and patent system reform, issues relating to the identification of prior art have received little empirical attention from economists. Primarily, this is because scholars have lacked useful data on the identification of prior art, and the role of patent examiners and applicants in generating prior art against which patents are evaluated.

In this paper, I use a novel dataset of examiner and applicant inserted references to the prior art in the 502,687 utility patents issued between January 1, 2001 and December 31, 2003 to shed light on these issues. In particular, I argue that the examiner share of references in a patent provides a useful lens on differences in applicants' and examiners' search capabilities, and how they vary across types of prior art, and on how incentives facing applicants vary across industries and inventions.

Using these data, I find strong evidence that patent examiners have a comparative disadvantage in searching for non-patent prior art and foreign patents, suggesting that all else equal, patents are likely to be of lower quality for technological areas for which most prior art is not embodied in U.S. patents. In addition, the data suggest that patent applicants are less likely to search for prior art in fields where patents are useful primarily for "strategic" purposes, rather than to appropriate returns from R&D. Patents in these fields are less plausibly valid: patent quality problems are likely most pronounced in these fields. Finally, I show that within fields, patent applicants devote more effort to identifying prior art for more technologically and commercially valuable inventions.

The remainder of the paper is structured as follows. Section 2 discusses the importance of "prior art" in the U.S. Patent System, the roles of patent applicants and examiners in identifying prior art, and the relationships between prior art and patent "quality." Section 3 provides an overview of the data, and assesses whether patent examiners are particularly advantaged (or disadvantaged) relative to applicants in identifying particular types of prior art. Section 4 examines differences across technological fields, and argues that the examiner share of citations provides a window on cross-field differences in firms' patent strategies. Section 5 assesses whether patent applicants are more likely to conduct prior art searches for more "important" patents. Section 6 concludes.

2 Prior Art and Patent Quality

To assess whether an invention disclosed in a patent application satisfies the "novelty" and "non-obviousness' criteria for patentability, a patent examiner compares it to *prior art* embodied in references to patents and printed publications. If the patent examiner deems that the invention is novel and non-obvious in light of the prior art (and satisfies various other criteria for patentability), a patent is granted, and the prior art references are listed on the front page of the issued patent.

Though examiners officially are 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. 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. prior patents and publications for which there is "a substantial likelihood that a reasonable examiner would consider it important in deciding whether to allow the application to issue as a patent" (MPEP). 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).

While applicants have strong incentives to disclose known prior art, little is know about the extent to which they actually search for prior art. Some observers argue that since missing prior art will result in patents that are difficult to enforce or of questionable validity, applicants have strong incentives to conduct prior art searches before filing patent applications.[1] Thus Ger-

[1]Even absent inequitable conduct, some or all of the claims of a patent can be rendered

ald Mossinghoff, the former commissioner of the USPTO, recently suggested 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" (USPTO 1999a). Caballero and Jaffe (1993) suggest "omission of important references can be grounds for invalidation of the patent, giving the applicant an incentive to make sure the citations appear" (21). And Allison et al. (2004) suggest, "[c]iting more prior art will make a patent more valuable in litigation, as it much harder to prove a patent is invalid if the PTO has already considered it and rejected the relevant prior art" (139).

On the other hand, applicants may also face disincentives to searching. Applicants can get broader patents if the examiner does not consider prior art material to patentability (Kesan and Banik, 2000; Kesan 2000; Wagner 2002). Thus Wagner (2002) suggests 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"(53). A second reason applicants may not search for prior art stems from the doctrine of willful infringement. Since 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 competitors' patents (Lemley and Tangri 2003).

Finally, prior art searches are costly and such searches may not always

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).

warrant the costs.[2] In particular, there is growing evidence that while patents are important mechanisms for appropriating returns from R&D in chemicals and pharmaceuticals, in most other industries firms rely primarily on mechanisms other than patents (e.g, secrecy) to exclude competitors from using their technologies (Cohen et al. 2000). Nevertheless, there has been a growth of patenting in so-called "complex product" industries (including electronics, computers, and communication technologies) over the past decade (Hall 2005).[3] A growing body of empirical research suggests that in these fields much patenting is for "strategic" purposes, e.g. patenting to preserve freedom to practice and to use as bargaining chips in cross-licensing negotiations (Hall and Ziedonis 2001; Cohen et al. 2000; Hall 2005). Some scholars suggest that in these cases, firms care more about the quantity of patents–and in particular building large patent portfolios–than about obtaining "quality" patents: patents that would be deemed valid if actually tested in court (Lemley and Shapiro 2005; Wagner and Parchomovsky 2004).

Notwithstanding disagreement about the extent of prior art disclosure by applicants and the factors affecting it, there is widespread agreement that applicants' disclosures are unlikely to identify the universe of relevant prior art. In recognition of this, patent examiners conduct their own prior art searches, primarily via full-text or bibliographic databases of patent documents and the scientific and technical journal literature (see Cockburn et al. 2003 for an excellent discussion of this process).

[2]Statutory bars prohibiting patentability of applications filed more than one year after an invention has been on sale, in use, or described in a printed publication also may create time constraints making prior art searching costly, or in cases, infeasible.

[3]As opposed to "discrete product" industries, in "complex product" industries any individual product typically is covered by many patents (Cohen et al., 2000).

If these databases spanned the universe of relevant prior art, and if examiners could search them (and evaluate the results of these searches) without cost, in principle no anticipatory prior art references would be missed. Of course, in a world of limited resources, the finite corps of patent examiners faces strict time allocation guidelines per application (Thomas, 2001; Cockburn et al. 2004). And searching prior art databases is difficult. Some observers believe that these difficulties are most pronounced in searching for non-patent prior art. For example, Thomas (2001) argues that"[i]n comparison to much of the secondary literature [non-patent art], patents are readily accessible conveniently classified, and printed in a common format. Identification of a promising secondary reference, and full comprehension of its contents, often prove to be more difficult tasks" (318).

There is growing commentary that these various constraints on effective prior art searching are increasingly binding, and that the USPTO is issuing more and more "low quality" patents, i.e., patents that would not have been issued had the examiner considered the entire universe of relevant prior art. Issuance of such patents can impose a range of social costs, including *inter alia* standard deadweight losses from monopoly pricing, encouragement of predatory and/or rent- seeking behavior, and, in the context of cumulative invention, taxes on subsequent innovators. In response to these concerns, a number of reforms to the US Patent System have been proposed by groups as diverse as the National Academy of Sciences (2004) and the Federal Trade Commission (2004).

However, these issues have been subject to very little systematic empirical research. In particular, little is known about three issues central in these debates:

1. Do examiners face particular difficulties in identifying certain types of prior art?

2. Do applicants' incentives to search for prior art vary across industries? and

3. Do applicants' incentives to search for prior art vary across inventions?

In the following sections, I use data on examiner and applicant citations to prior art to examine these questions.

3 Differences Across Types of Prior Art

The main dataset consists of the 502,687 utility patents issued by the USPTO between January 2001 and December 2003, and all prior art references in these patents, including the 6,324,381 references to earlier U.S. patents, the 1,382,430 references to foreign patents, and the 1,655,166 references to non-patent literature.

Table 1 shows the basic numbers on examiner and applicant inserted references in these 502,687 patents. Patent examiners account for 41 percent of the citations to previous U.S. patents, but only 10 percent of references to non-patent prior art and 12.3 percent of references to foreign patents.

The stark differences in examiner share of citations across types of prior art support the argument that patent examiners have a comparative disadvantage, as compared to applicants, at searching for prior art not embodied in

U.S. patents.[4] However, these figures could also reflect cross- field differences in (relative) search capabilities which are correlated with cross-field differences in types of prior art referenced. For example, if examiners faced particular difficulties at searching for *all* types of prior art in fields where most prior art is embodied in the non-patent literature (or conversely, that examiners are better at searching for all types of prior art in fields where most of the prior art is embodied in U.S. patents) the aggregate numbers would show the examiner share of references to non-patent prior art to be lower than the examiner share of references to other types of prior art. But in this scenario, the difference would reflect not differences in search capabilities across types of prior art, but rather differences in search capabilities across fields.

In an attempt to isolate potential differences in capabilities across types of prior art from cross-field differences in search capabilities, I estimated linear probability regressions of whether a citation was inserted by the examiner on dummy variables indicating type of prior art, and dummies variables for each 3-digit patent class. Column 1 of Table 2 reports these results. Even after controlling for technology class, references to non-patent prior art and foreign patents are 26 and 27 percent less likely to be inserted by examiners than references to U.S. patents (the left-out category).

The likelihood that a citation comes from the examiner may reflect differences in applicants' search incentives across inventions. For example, even if applicants' and examiners' search capabilities were identical across types of prior art, if applicants had greater incentives to conduct prior art searches for

[4] Note that the data do not tell us anything about absolute search capabilities, but rather differentials across types of prior art in the in relative search capabilities of applicants and examiners.

inventions where most of the relevant prior art was embodied in non-patent prior art, the examiner share of references to non-patent prior art would be lower than that for other types of prior art.

In order to control for cross-invention differences in incentives, I added fixed effects for each citing patent to the model. Accordingly, the estimates are identified based on differences *within* individual patents in the share of examiner-inserted references to different of prior art.

The results, reported in Column 2, suggest that the examiner share of references to foreign patents and to non-patent literature are, respectively, 26 percentage points and 19 percentage points lower than the examiner share of references to U.S. patents (the left-out category).

The statistics and regression results above are calculated at the citation level. Table 3 shows data on the examiner share of citations at the patent level, calculated across the 502,687 patents in the sample. Column 1 shows that the average examiner share of references to U.S. patents is 63 percent. Examiners account for an average of 18 percent of references to non-patent literature, and an average 21 percent of references to foreign patents.

However, like the figures in Table 1, the examiner share of citations at the patent level reflects not only differences in search capabilities across types of prior, art, but also incentives across inventions and fields. To control for these differences, I also examined differences *within patents* in the examiner share of references to U.S. prior art versus foreign and non-patent prior art. In the 179,139 patents citing both U.S. patents and non-patent prior art, on average the examiner share of references to the former was greater by 29 percentage points ($p<.001$). Similarly, in the 262,943 patents citing both U.S. patents and foreign patents, on average the examiner share of references to the former was greater by 36 percentage points ($p<.001$).

The differences between the examiner shares of all citations and the average examiner share of citations across patents suggests there is significant variation across patents in the share of citations inserted by examiners. Indeed, for all types of prior art, the distribution across patents of the share of examiner references is bimodal, peaking at zero and one. Columns 2 and 3 of Table 3 show the data for the tails of this distribution. Examiners insert all of the references to U.S. patents for 39 percent of the citing patents in the sample. However, they insert none of the references to U.S. patents for 8 percent of the citing patents; here, all of the references are from the applicants' information disclosure statement. By contrast, examiners account for all of the references to non-patent literature in 12 percent of the patents, and all references to foreign patents in 16 percent of the patents. Strikingly, examiners account for none of the references to non-patent prior art (and thus applicants all of the references) in 69 percent of the patents citing non-patent prior art; the analogous figure for references to foreign patents is 68 percent.

Taken together, the results in this section provide strong evidence that patent examiners have a strong comparative disadvantage in searching for prior art not embodied in U.S. patents. All else equal, one implication of this is that patent applications in fields where most prior art is embodied in the non-patent literature or foreign patents are likely to be of lower quality. However, the examiner share of citations reflects not only relative search capabilities, but also applicants' incentives across fields and inventions, which I explore below.

4 Differences Across Technological Fields

If search capabilities and/or applicant incentives varied across industries, we would also expect to see cross-field differences in the likelihood that a reference is examiner-inserted, and in the examiner share of citations to prior art. Table 4 shows the share of citations inserted by examiners, by category of prior art and broad technological field, using the broad technological field categories constructed by Jaffe and Trajtenberg (2001).

Previous empirical literature suggests that while patents are important for appropriating returns to R&D in "discrete product" industries like chemicals and pharmaceuticals, in other "complex product" industries patents are useful primarily for "strategic" purposes. The data show that in Chemicals and Drugs and Medicine, patent examiners (applicants) account for a significantly lower (higher) share of references to U.S. patents than in complex product industries i.e. in the Computers and Communications, Electric and Electronics, and Mechanical fields. These differences are both statistically and qualitatively significant. Whereas in most fields applicants contribute approximately 50 percent of references to U.S. patents, for Drugs and Medical patents applicants contribute 79 percent, and for Chemical patents applicants contribute 67 percent.

These data provide suggestive evidence that applicants are more likely to conduct prior-art searches in fields where patents are more important for appropriability (vis-a-vis fields where patents are employed for other strategic purposes). However, Table 4 also shows that while the examiner share of references to non-patent and foreign prior art is lower in pharmaceuticals and chemicals than the average across other fields, these differences are not nearly as stark as the cross-field differences in references to U.S. patents.

Table 5 shows the data at the level of citing patents, rather than citations, and reveals similar trends. For references to U.S. patents, the average examiner share across patents is significantly lower in chemicals and pharmaceuticals than other fields. In these fields, examiners also are significantly less likely to account for all of the references to U.S. patents, and significantly more likely to account for none of the references to U.S. patents (i.e., applicants are significantly more likely to generate all of the references). Table 5 also shows that at the patent level there is little evidence of systematic differences across the discrete and complex product fields in the average examiner share of non-patent references or citations to foreign patents.

However, the analysis in the previous section suggested that patent examiners are less capable at identifying relevant prior art embodied in the non-patent literature and foreign patents than at identifying U.S. patented prior art. Accordingly, the examiner share of references to U.S. prior art provides a better signal of whether applicants searched for prior art than references to other types of prior art, since examiners are much more likely to discover relevant prior art that the applicant did not report if that prior art was embodied in U.S. patents. In other words, if an applicant does not search for prior art and thereby does not report a piece of relevant prior art on his/her information disclosure statement, the examiner is less likely to discover it if it is codified in the non-patent literature or a foreign patent than if it is codified in a U.S. patent, since examiner capabilities for searching for U.S. patents exceed their capabilities for searching other sources of prior art.

Since the share of examiner-inserted references in a patent is in part a function of applicants' and examiners' search capabilities, it is not a perfect indicator of differences across fields in applicants' incentives to search for prior art, and thus differences across in applicant commitment to patent quality In the following section, I examine trends across fields in examiner share of references to "self" citations, a context where the examiner share of references is likely to be a more precise signal of applicants' search activities.

5 Self-Citations

A "self" citation is a citation in a patent by assignee i to an earlier (U.S.) patent by assignee i. Table 6 tabulates the share of assignee-assignee self citations in patents issued between 2001 and 2003.[5] Consistent with previous research (Jaffe and Tratenberg, 2003) approximately 11 percent of all citations in this sample are self-citations.

Recall that across all fields, patent examiners account for 41 percent of citations to previous U.S. patents. Strikingly, patent examiners also account for 41 percent of self-citations. At the patent level, the average share of self-citations inserted by the examiner is 57 percent. Examiners account for all self-citations in 48 percent of patents with self- citations, while applicants account for all self-citations in only 30 percent of patents with self-citations.

These figures provide strong evidence that in many cases, patent applicants are not devoting much effort to searching for prior art. It is plausible that applicants might search for prior art but fail to uncover relevant references; after all, no search can be exhaustive. However, it is less plausible that applicants

[5]Note that the total number of patent-patent citations is smaller than that reported in Table 1, since I had to drop unassigned citing and cited patents, as well as pre- 1976 cited patents for which assignee and applicant information is unavailable.

are searching for prior art and missing so much of *their own* prior art. Moreover, examiners may be better able to find relevant prior art missing from the applicant's disclosure when the search domain is well-defined. That is, identifying relevant prior art within a firm i's patents is easier than identifying relevant prior art across the universe of all patents. As a result, differences across fields in examiner "self" citations provide a particularly informative signals of when applicants are conducting prior art searches.

Table 7 shows the examiner share of self citations across fields, calculated at the citation level. In chemicals and pharmaceuticals examiners account for only 36 and 29 percent of self-citations, respectively, while the number is significantly higher in other technological fields. In computers and communications, for example, examiners account for nearly half of all self citations.

Table 8 shows similar trends calculated at the level of the citing patent. The average examiner share of self citations is significantly higher in the complex product industries than in pharmaceuticals or chemicals, as is the share of patents where the examiner accounts for all self citations. Strikingly, examiners account for *all* of the "self" citations in the majority of patents in the Computers/Communications and Electonic/Electrical fields. On the other hand, the share of patents for which the applicant accounts for all of the self-citations is greatest in Pharmaceuticals and Chemicals.

Consistent with the data on all references to U.S. patents, patent examiners account for a smaller share of references to self-citations (and conversely, applicants account a larger share) in fields where patents are important for appropriating returns to R&D, and a higher share in fields where patents are used more for strategic purposes. This provides strong *prima facie* evidence of cross-field differences in applicants' commitments to patent quality. I explore the implications of this in more detail below.

6 Do Patent Applicants Contribute More Prior Art for More Important Inventions?

It also is possible that applicants treat different inventions differently, i.e., even within fields, are more committed to getting "quality" patents for particular inventions. It is well known that the lion's share of inventions are economically and technologically useless (Lemley, 2000). As discussed in the Introduction, there is little sorting among patent applications by examiners. However, applicants themselves may be better informed about the commercial and/or technological importance of patents. Accordingly, it is plausible that applicants would devote more effort to obtaining quality patents, i.e., search for prior art, for relatively important inventions.

To examine this, I constructed three patent- specific measures of importance for a subset of the sample patents: the 28,131 patents granted in January and February 2001. The first is a count of the number of citations each of these patents received in subsequent patents issued by January 1, 2005. Forward citations counts are a commonly used measure of the "importance" of inventions, and appear to good predictors of other measures of an invention's importance, including whether it is licensed (Sampat and Ziedonis 2001), consumer surplus based on an invention (Trajtenberg 1999), and others (Lanjouw and Schankerman 2004). On average, these patents received 3.2 forward citations between 2001 and 2004.

The second measure of a patent's importance is a binary variable: whether the patent was renewed after four years. After a patent is issued patent holders must periodically pay maintenance fees to keep it in force. Currently, in the United States, these fees must be paid four, eight, and twelve years after issue. The renewal based measure of importance assumes that, all else equal, more economically valuable patents are more likely to be renewed at any given point in time (Lanjouw et al. 1998). Over half of patents issued expire before their full twenty year term because the applicant chose not to pay maintenance fees (Moore, 2004). Of the patents issued in January and February 2001, 15 percent expired because the applicant chose not to pay 4-year maintenance fees.

The third measure of importance is the number of countries in which an applicant takes out a patent, or patent "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 more economically valuable inventions. The average family size of the patents in the January-February 2001 sample is 3.96 countries.

To assess whether applicants contribute more prior art more "important" inventions, I examined the two types of prior art where examiner versus applicant citations provide the clearest picture of applicants' incentives: the examiner share of references to U.S. patents, and the examiner share of "self" citations. As discussed above, for these types of prior art (as opposed to non-patent references and foreign patents), examiners are more likely to find relevant references that applicants do not submit with their applicants. Hence, these references provide the clearest signals of whether (and when) applicants search for prior art.

The examiner share of references is a share variable, 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 supplemented these analyses with analogous regressions with the *number* of applicant references and the *number* of examiner references as dependent variables. Table 9 shows the descriptive statistics for each of the variables used in these analyses.

I began by estimating ordinary least squares regressions of the examiner share of references in patent i on each of the three measures of patent importance, separately, and patent class fixed effects for each 3-digit patent class.[6]. Columns 1 through 3 of Table 10 show the results, with examiner share of references to U.S. patents as the dependent variable. For all three measures of patent importance, the examiner share (applicant share) of references in patent i is lower (higher) for more important inventions, and the differences are both statistically and qualitatively significant. A one standard deviation increase in the number of forward citations to patent i (an increase by 4.7 citations) implies a 3 percentage point decrease in the examiner share of citations to U.S. patents in patent i, after controlling for technology class effects. Similarly, the examiner share of citations is four percentage points greater for patents which would not be renewed four years after issue. And a standard deviation increase in family size implies a 3 percentage point decrease in the

[6]Given the bounded nature of the dependent variable, some have argued that OLS is inappropriate. Accordingly, I estimated similar regressions using the GLM-based quasi-likelihood method proposed by Papke and Wooldridge (1996). In addition, since the distribution of the examiner share of references is bimodal, I estimated probit models of how invention specific characteristics affect the probability that an examiner inserts *all* or *none* of the references in a patent. The results of these additional analyses were qualitatively similar to those reported below, and are available by request from the author.

examiner share of citations to U.S. patents in patent i.

These results suggest that after controlling for technology class, across all invention applicants account for a greater share of references for inventions that are more important (using any measure of importance).

However, this relationship could also reflect systematic (unobserved) relationships between types of assignees and types of patents. For example, if U.S. assignees were more likely to cite U.S. patents, and on average their patents also tended to be more "important," we would see similar trends, even if there were no sorting across applications by patent applicants. To assess this, I re-estimated these regressions with assignee-specific fixed effects for each of the 7897 assignees. With the inclusion of assignee fixed effects, the coefficients on the importance variables are identified based on within-assignee variation. In other words, the estimates measure the extent to which an examiner citations differ between an assignees' more and less important inventions (within a class).

Columns 4 through 6 report the results. Here too, the effects of importance are unambiguous: patent applicants account for a greater share of references to U.S. patents in their more important inventions, for each of the three measures of importance.

To investigate whether these changes are due to applicant or examiners' responses to invention "importance," I estimated similar models with the *number* of applicant references and the *number* of examiner references as the dependent variables. Tables 11 and 12, respectively, report the results.[7] Table 11

[7]These models were estimated via OLS. I also estimated negative binomial models, which explicitly account for the integer nature of the dependent variable. The results are similar, and are available from the author by request.

shows that after controlling for patent class, patent applicants insert a greater number of references in patents which will generate more "forward" citations in the four years after issue, which will be renewed after four years, and for which applications are filed in multiple jurisdictions. This result holds in both the baseline specification and the models with assignee fixed effects.

Table 12 shows analogous results for the number of examiner-inserted references, after controlling for the number of applicant references. [8] The coefficients on the number of applicant references suggests that, not surprisingly, examiner and applicant references are substitutes. All else equal, the greater the number of relevant references contributed by the applicant, the fewer the examiner will insert.[9] For two of the importance measures, whether the patent was renewed and family size, there is no statistically significant effect of invention importance on the number of examiner references. This is consistent with the notion that (for better or worse) there is little or no "sorting" across patent applications by examiners: they devote the same resources to searching for prior art for all inventions, regardless of their importance.

However, examiners do insert a greater number of references in patents which will receive more "forward" citations, in the models with and without assignee fixed effects. While this could provide some evidence for examiner-level sorting, note that the magnitude of the impact is extremely small.[10] The effect of this measure of importance on applicant citations is 8 times larger

[8] Similar results hold if the number of applicant references is omitted from the right hand side.

[9] By definition, all of the applicant references were deemed "material to patentability" by the examiner.

[10] It is also possible that this result reflects crowdedness of technological fields which the 3 digit class dummies are too blunt to absorb.

than its effect on examiner citations in the baseline model, and 4 times larger in the model with assignee fixed effects.

Taken together, these results suggest that applicants contribute a higher number of prior art references and a greater share of citations to previous patents (including self-citations) for more important inventions. If, for more important inventions, applicants are contributing prior art that otherwise would have been missed by the examiner, this would suggest that patent "quality" problems are likely to be least pronounced for more such inventions. Another possibility, however, is that by doing more thorough searches, applicants do a better job of *anticipating* examiner references for more important inventions. There are several reasons why this could be in an applicant's interest, including to earn the goodwill of the examiner or to reduce patent approval times. In addition, applicants may wish to avoid prior art induced claim changes, which could limit extension of patent scope via prosecution history estoppel (see Lichtman 2004).

Note however, that under this alternative hypothesis–that applicant references are pure substitutes for examiner references–the *total* number of citations in a patent would be unrelated to the applicant share of references. Table 16 shows results of OLS regressions of total references on the applicant share of references and patent class dummies. Model 1 shows within classes, as the applicant share increases from zero to one, the total number of citations to U.S. patents increases by seventeen, and this effect is highly statistically significant. One potential weakness of this test is that the same factors may affect both the applicant share of references and total references: 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 references would be correlated, even if applicant references simply anticipated examiner references rather than provided new information. Accordingly, in Model 2, I control for each of the three measures of invention importance discussed above, and also for the the number of claims in the patent. Doing so did had little effect on the estimated coeffiecient, increasing confidence in the conclusion that applicant citations are in fact adding new information rather than simply anticipating examiner citations. Similar results are seen in Models 3 and 4, with the total number of self-citations as the dependent variable.

These results suggest that more applicant citations actually add to the universe of cited prior art, rather than displace examiner citations. Together with the result that the applicant share is higher for more important inventions, this suggests a possible self-sorting mechanism in the U.S. patent system, with a larger share of the relevant prior art being considered for more important inventions.[11] This is especially interesting in light of the widespread recognition that examiner-side sorting at the USPTO is infeasible (Lemley 2001).

Irrespective of whether one interprets these results as evidence that applicants provide more information for more important inventions, or instead simply as evidence that applicants better anticipate examiners for more important inventions, they do suggest that unlike patent examiners, applicants have information about which inventions are likely to be commercially and

[11] On the other hand, it is possible that patent applications with too much applicant inserted prior art will get less rather than more attention from overworked patent examiners, in which case more applicant citations would not translate into more rigorous review or quality (Allison et al. 2003).

economically important, and act on this information in preparing their patent applications (cf. Allison et al. 2003). Lemley et al. (2000)

7 Discussion and Conclusions

Despite the centrality of issues relating to the identification of prior art in contemporary discussions of patent quality and patent system reform, these issues have been subject to limited empirical scrutiny. In this paper, I explored factors affecting the identification of prior art using a novel dataset of applicant and examiners citations.

To review the main findings, I found strong differences across types of prior art in the likelihood that a reference is examiner-inserted, and the share of examiner inserted references, even within patents. It is difficult to reconcile these data with any explanation other than that patent examiners have a comparative disadvantage at searching for prior art not embodied in U.S. patents. All else equal, this suggests that patent quality will be worse in fields where most prior art is not embodied in U.S. patents. Note that this typically means new fields, perhaps those where excessively broad patents impose the greatest social costs (Merges and Nelson 1990).

I also found stark differences across fields in the best indicators of the extent of applicant search: the examiner share of U.S. patent citations and the examiner share of self-citations. Using these data, I found evidence that applicants are more likely to search for prior art in fields where patents are important as mechanisms for appropriating returns from R&D, like pharmaceuticals and chemicals, and less likely to search for prior art in "complex product" industries, like electronics, communications technologies, and computing, where patents primarily are used for other strategic purposes. The data thus provide the first large-sample evidence (albeit indirect evidence) that patent quality is likely to be most problematic in these "strategic" patenting industries.

I also found robust evidence that even within fields, patent applicants are more likely to contribute prior art for more important inventions. One interpretation of this is that for"important" patents, prior art searching is incentive compatible, and helps to "bullet proof" their patents. If this were true, patent quality problems would be least pronounced for the most important patents. A weaker interpretation is that applicants take more effort to anticipate examiner citations for more important inventions. Either way, the data suggest that applicants know, when they file their applications, which of their applications are likely to be most valuable. Lemley et al. (2005) suggest that harnessing that knowledge ought to be an important part of current efforts at patent system reform.

Finally, this paper shows that citations in patents result from the complex interaction of examiner and applicant activities, reflecting constraints and incentives facing each. Patent citation data–in particular data on U.S. patents citing previous U.S. patent–are increasingly used by economists to measure the economic significance of patented inventions, and as proxies for spillovers and knowledge flows. The data presented in this paper suggest reason to be cautious in interpreting the results of such analyses. For example, it is unclear whether we can treat citations as indicators of knowledge flows, when examiners account for 41 percent of all citations. Similarly, treating all citing patents as the same may be problematic when, for over nearly two-thirds of all issued patents, either all or none of the citations to previous patents come from examiners. And the fact that "self" citations are as likely to come from examiners as applicants would seem to complicate their interpretation. The impact of examiner-inserted references on inferences from citation based indicators is an important topic for future research: Thompson (2005) and Alcacer and Gittleman (2004) are excellent first steps in this direction.

Table 1: Examiner and applicant references in patents issued between 2001-2003 by type of prior art (at citation level)

	Type of Prior Art							
	U.S. Patent		Non-Patent Literature		Foreign Patent		Total	
Examiner Citation	Num	%	Num	%	Num	%	Num	%
No	3,707,597	58.6	1,490,330	90.0	1,212,023	87.7	6,409,950	68.5
Yes	2,616,784	41.4	164,836	10.0	170,407	12.3	2,952,027	31.5
Total	6,324,381	100.0	1,655,166	100.0	1,382,430	100.0	9,361,977	100.0

Based on references in the 502,687 utility patents issued from January 1, 2001 to December 31, 2003

Table 2: OLS Regressions of whether a citation is examiner-inserted on type of prior art

	(1)	(2)
Reference to Foreign Patent	-.268***	-.263***
	(.0003)	(.0004)
Reference to Non-Patent Literature	-.259***	-.190***
	(.0003)	(.0004)
Const.	.401***	.388***
	(.0002)	(.0002)
Obs.	9361977	9361977
Class Effects	Yes	No
Patent Effects	No	Yes
R^2	.14	.54

Left-out category is citations to U.S. Patents. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

Table 3: Examiner share of references by type of prior art (at patent level)

Type of Prior Art	Average Share Across Patents	Share of Patents with Examiner Share=1	Share of Patents with Examiner Share=0
U.S. Patent	0.63	0.39	0.08
Non-Patent Literature	0.18	0.12	0.69
Foreign Patent	0.21	0.16	0.68
Total	0.42	0.27	0.37

Table 4: Examiner and applicant references in patents issued between 2001-2003 by type of prior art and field (at citation level)

	Type of Prior Art			
Technological Category	U.S. Patent	Non-Patent Literature	Foreign Patent	Total
Chemical	0.33	0.10	0.11	0.24
Computers and Communications	0.46	0.13	0.10	0.37
Drugs and Medical	0.21	0.08	0.06	0.14
Electric and Electronic	0.47	0.08	0.13	0.37
Mechanical	0.49	0.08	0.20	0.42
Other	0.46	0.15	0.18	0.40
Total	0.41	0.10	0.12	0.32

Based on references in the 502,687 utility patents issued from January 1, 2001 to December 31, 2003

Table 5: Examiner share of references by type of prior art and field (at patent level)

Citations to U.S. Patents

		Share of Patents With	
Category	*Avg Examiner Share*	*Examiner Share=1*	*Examiner Share=0*
Chemical	0.51	0.29	0.16
Computers, Communications	0.69	0.45	0.04
Drugs, Medical	0.41	0.23	0.25
Electrical, Electronic	0.68	0.45	0.04
Mechanical	0.67	0.41	0.05
Other	0.65	0.39	0.05

Citations to Non-Patent Literature

		Share of Patents With	
Category	*Avg Examiner Share*	*Examiner Share=1*	*Examiner Share=0*
Chemical	0.19	0.13	0.67
Computers, Communications	0.21	0.15	0.69
Drugs, Medical	0.18	0.09	0.56
Electrical, Electronic	0.14	0.10	0.80
Mechanical	0.14	0.11	0.81
Other	0.22	0.17	0.70

Citations to Foreign Patents

		Share of Patents With	
Category	*Avg Examiner Share*	*Examiner Share=1*	*Examiner Share=0*
Chemical	0.19	0.13	0.68
Computers, Communications	0.16	0.12	0.76
Drugs, Medical	0.13	0.09	0.76
Electrical, Electronic	0.20	0.15	0.71
Mechanical	0.28	0.21	0.59
Other	0.29	0.23	0.61

Table 6: Examiner and applicant self-citations (at citation level)

	Self-Citation					
	No		Yes		Total	
Examiner Citation	Num	%	Num	%	Num	%
No	2,301,284	59.1	285,679	58.6	2,586,963	59.1
Yes	1,589,857	40.9	201,460	41.4	1,791,317	40.9
Total	3,891,141	100.0	487,139	100.0	4,378,280	100.0

Based on references in the utility patents issued from January 1, 2001 to December 31, 2003

Table 7: Examiner share of self-citations by field of citing patent (at citation level)

Cat	Examiner Citation
Chemical	0.36
Computers, Communications	0.46
Drugs, Medical	0.29
Electrical, Electronic	0.42
Mechanical	0.45
Other	0.43
Total	0.41

Based on references in the utility patents issued from January 1, 2001 to December 31, 2003

Table 8: Examiner share of self-citations by field of citing patent (at patent level)

Cat	Average	Examiner Share=1	Examiner Share=0
Chemical	0.50	0.41	0.36
Computers, Communications	0.63	0.54	0.24
Drugs, Medical	0.45	0.38	0.43
Electrical, Electronic	0.60	0.51	0.28
Mechanical	0.60	0.50	0.28
Other	0.56	0.48	0.33
Total	0.57	0.48	0.30

Calculated over the utility patents issued from January 1, 2001 to December 31, 2003 with at least one self-citation

Table 9: Summary statistics for patents issued in January and February, 2001

Variable	Mean	Std. Dev.	N
Examiner Share of Patent Citations	0.633	0.377	27734
Examiner Share of Self Citations	0.584	0.448	8856
Examiner Patent Citations	5.107	4.836	28131
Applicant Patent Citations	6.549	16.567	28131
Examiner Self Citations	1.296	1.511	8856
Number of Applicant Self Citations	1.62	3.281	8856
Number of Forward Citations	3.232	4.785	28131
Patent Expired After 4 Years?	0.147	0.354	28131
Family Size	3.96	3.716	28131

Table 10: OLS Regressions of the examiner share of references to U.S. patents in $PATENT_i$ on the "importance" of $PATENT_i$

	(1)	(2)	(3)	(4)	(5)	(6)
Number of Forward Citations	-.007***			-.003***		
	(.0005)			(.0006)		
Patent Expired After 4 Years?		.041***			.032***	
		(.006)			(.008)	
Family Size			-.008***			-.010***
			(.0007)			(.001)
Const.	.656***	.627***	.665***	.878***	.608***	.657***
	(.003)	(.002)	(.003)	(.034)	(.045)	(.044)
Class Effects	Yes	Yes	Yes	Yes	Yes	Yes
Assignee Effects	No	No	No	Yes	Yes	Yes
Obs.	27734	27734	27734	27734	27734	27734
R^2	.105	.099	.103	.48	.479	.482

Based on patents issued in January and February 2001. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

Table 11: OLS Regressions of the number of applicant references to U.S. patents in $PATENT_i$ on the "importance" of $PATENT_i$

	(1)	(2)	(3)	(4)	(5)	(6)
Number of Forward Citations	.499***			.223***		
	(.042)			(.037)		
Patent Expired After 4 Years?		-1.763***			-1.163***	
		(.215)			(.266)	
Family Size			.507***			.551***
			(.049)			(.074)
Const.	4.936***	6.807***	4.543***	16.623***	15.851***	16.230***
	(.137)	(.110)	(.173)	(.939)	(.924)	(.913)
Class Effects	Yes	Yes	Yes	Yes	Yes	Yes
Assignee Effects	No	No	No	Yes	Yes	Yes
Obs.	28131	28131	28131	28131	28131	28131
R^2	.063	.046	.055	.527	.525	.53

Based on patents issued in January and February 2001. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

Table 12: OLS Regressions of the number of examiner references to U.S. patents in $PATENT_i$ on the "importance" of $PATENT_i$

	(1)	(2)	(3)	(4)	(5)	(6)
Applicant Patent Citations	-.018***	-.016***	-.015***	-.023***	-.021***	-.021***
	(.003)	(.003)	(.003)	(.007)	(.007)	(.007)
Number of Forward Citations	.061***			.054***		
	(.007)			(.009)		
Patent Expired After 4 Years?		.016			.023	
		(.072)			(.102)	
Family Size			-.016			.002
			(.011)			(.018)
Const.	5.028***	5.207***	5.273***	11.791***	11.548***	11.554***
	(.035)	(.034)	(.046)	(.505)	(.506)	(.509)
Class Effects	Yes	Yes	Yes	Yes	Yes	Yes
Assignee Effects	No	No	No	Yes	Yes	Yes
Obs.	28131	28131	28131	28131	28131	28131
R^2	.161	.157	.157	.438	.436	.436

Based on patents issued in January and February 2001. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

Table 13: OLS Regressions of the examiner share of self-citations in $PATENT_i$ on the "importance" of $PATENT_i$

	(1)	(2)	(3)	(4)	(5)	(6)
Number of Forward Citations	-.006***			-.003**		
	(.0009)			(.001)		
Patent Expired After 4 Years?		.039**			.035*	
		(.017)			(.021)	
Family Size			-.003*			-.008***
			(.001)			(.002)
Const.	.607***	.580***	.595***	-.147	-.179	-.074
	(.006)	(.005)	(.008)	(.536)	(.533)	(.536)
Class Effects	Yes	Yes	Yes	Yes	Yes	Yes
Assignee Effects	No	No	No	Yes	Yes	Yes
Obs.	8856	8856	8856	8856	8856	8856
R^2	.096	.092	.092	.466	.466	.467

Based on patents issued in January and February 2001. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

Table 14: OLS Regressions of the number of applicant self-citations in $PATENT_i$ on the "importance" of $PATENT_i$

	(1)	(2)	(3)	(4)	(5)	(6)
Number of Forward Citations	.050***			.034***		
	(.010)			(.012)		
Patent Expired After 4 Years?		-.208*			-.201	
		(.113)			(.175)	
Family Size			.058***			.084***
			(.012)			(.019)
Const.	1.431***	1.641***	1.374***	1.318***	1.649***	.562
	(.044)	(.036)	(.055)	(.374)	(.354)	(.438)
Class Effects	Yes	Yes	Yes	Yes	Yes	Yes
Assignee Effects	No	No	No	Yes	Yes	Yes
Obs.	8856	8856	8856	8856	8856	8856
R^2	.091	.085	.089	.36	.359	.362

Based on patents issued in January and February 2001. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

Table 15: OLS Regressions of the number of examiner self-citations in $PATENT_i$ on the "importance" of $PATENT_i$

	(1)	(2)	(3)	(4)	(5)	(6)
Number of Applicant Self Citations	-.030***	-.030***	-.031***	-.031***	-.030***	-.031***
	(.010)	(.010)	(.010)	(.011)	(.011)	(.011)
Number of Forward Citations	.003			.009**		
	(.003)			(.004)		
Patent Expired After 4 Years?		.044			.070	
		(.053)			(.083)	
Family Size			.010*			.010
			(.005)			(.011)
Const.	1.332***	1.340***	1.303***	-1.527	-1.453	-1.564
	(.022)	(.020)	(.029)	(1.285)	(1.286)	(1.295)
Class Effects	Yes	Yes	Yes	Yes	Yes	Yes
Assignee Effects	No	No	No	Yes	Yes	Yes
Obs.	8856	8856	8856	8856	8856	8856
R^2	.079	.078	.079	.283	.282	.282

Based on patents issued in January and February 2001. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

Table 16: OLS Regressions of total citations in $PATENT_i$ on applicant share of references, controls

	Total Cites			Total Self-Cites		
	(1)	(2)	(3)	(4)	(5)	(6)
Applicant Share of Cites	17.594***	16.862***	16.195***			
	(.260)	(.259)	(.260)			
Applicant Share Self-Cites				1.978***	1.939***	1.879***
				(.083)	(.083)	(.083)
Family Size		.334***	.312***		.058***	.053***
		(.028)	(.028)		(.011)	(.011)
Patent Expired After 4 Years?		-.450*	-.266		-.018	.010
		(.268)	(.267)		(.123)	(.122)
Number of Claims			.131***			.020***
			(.007)			(.003)
Class Effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	27734	27734	27734	8856	8856	8856
R^2	.192	.209	.218	.138	.144	.149

Based on patents issued in January and February 2001. Robust standard errors in parentheses. *** denotes p<.001, ** denotes p<.01, and * denotes p<.05.

References

[1] J. Alcacer and M. Gittelman. How do I know what you know? Patent examiners and the generation of patent citations. *Working Paper*, page 54, May 2004 2004.

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

[3] J. Allison, M. Lemley, K. Moore, and R. D. Trunkey. Valuable patents. *UC Berkeley School of Law Working Paper*, 2003.

[4] J. Allison and M. Lemley. The growing complexity of the United States patent system. *UC Berkeley Public Law and Legal Theory Working Paper2*, 2001.

[5] I. Cockburn, S. Kortum, and S. Stern. Are all patent examiners equal? the impact of examiner characteristics. *NBER Working Paper w8980*, 2002.

[6] W. M. Cohen and S. A. Merrill, editors. *Patents in the Knowledge-Based Economy.* National Academies Press, Washington, DC, 2003.

[7] W. Cohen, R. R. Nelson, and J. Walsh. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not),". *NBER Working Paper 7552*, 2000.

[8] Federal Trade Commission. To promote innovation: The proper balance of competition and patent law and policy. Technical report, Federal Trade Commission, 2003.

[9] B. Hall and R. Ziedonis. The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry 1979-1995. *RAND Journal of Economics*, 32(1):101–128, 2001.

[10] B. Hall. Exploring the patent explosion. *Journal of Technology Transfer*, 30(2), 2005.

[11] A. B. Jaffe and M. Trajtenberg. *Patents, Citations, and Innovations: A Window on the Knowledge Economy.* MIT Press, Cambridge, 2002.

[12] J. P. Kesan and M. Banik. Patents as incomplete contracts: Aligning incentives for r&d investment with incentives to disclose prior art. *Journal of Law and Policy*, 2:23–54, 2000.

[13] J. P. Kesan. Carrots and sticks to create a better patent system. *Berkeley Technology Law Journal*, 17(2):138–173, 2002.

[14] J. Lanjouw, A. Pakes, and J. Putnam. How to count patents and value intellectual property: Uses of patent renewal and application data. *Journal of Industrial Economics*, XLVI(4):405–433, 1998.

[15] J. Lanjouw and M. Schankerman. Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114, 2004.

[16] M. A. Lemley and R. K. Tangri. Ending patent law's willfulness game. *UC Berkeley Public Law Research Paper*, 2003.

[17] M. A. Lemley. Rational ignorance at the patent office. *Northwestern University Law Review*, 95:1495, 1496, 2001.

[18] M. Lemley and C. Shapiro. Probabilistic patents. *Journal of Economic Literature*, 2005.

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

[20] A. A. Lipscomb and A. E. Bergh, editors. *The Writings of Thomas Jefferson*. The Writings of Thomas Jefferson. Thomas Jefferson Memorial Association, Washington, 1905.

[21] R. P. Merges. As many as six impossible patents before breakfast: Property rights for business concepts and patent system reform. *Berkeley Technology Law Journal*, 14(2), 1999.

[22] R. Merges and R. R. Nelson. On the complex economics of patent scope. *Columbia Law Review*, pages 839–916, 1990.

[23] R. C. L. Merrill, Stephen A. and M. B. Myers, editors. *A Patent System for the 21st Century*. The National Academies Press, Washington, DC, 2004.

[24] K. Moore. Worthless patents. *Working Paper, George Mason Law School*, 2004.

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

[26] J. Putnam. *The Value of Intellectual Property Rights*. Unpublished Thesis, 1996.

[27] C. Ricardo and A. B. Jaffe. 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, 1993.

[28] B. N. Sampat and A. Ziedonis. Cite seeing: Patent citations and economic value. *Mimeo*, 2001.

[29] J. R. Thomas. Collusion and collective action in the patent system: A proposal for patent bounties. *University of Illinois Law Review*, pages 316–322, 2001.

[30] P. Thompson. Patent citations and the geography of knowledge spillovers: What do patent examiners know? *Mimeo*, 2004.

[31] M. Trajtenberg. A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics*, 21, 1990.

[32] P. Wagner and G. Parchomovsky. Patent portfolios. *U of Penn. Law School, Public Law Working Paper*, 2005.