"How do I know what you know? The role of inventors and examiners in the generation of patent citations"

Juan Alcacer
jalcacer@stern.nyu.edu
Michelle Gittelman
mgittelm@stern.nyu.edu

Department of Management and Organizations
NYU Stern School of Business

March 17, 2004

ABSTRACT

There is a large and growing literature in the technology and innovation literature that relies on patent data to study patterns of technological evolution, knowledge creation and diffusion, and firm technology strategy. Analysis of prior art – citations to patents by other patents -- has been a core methodology in the literature, in which citations are the "paper trails" tracking social, organizational, and geographic pathways of knowledge flows. However, in many instances researchers have been limited in their interpretations of their findings because citations made by patent examiners have not been separated out from citations made by inventors. We leverage a recent (2001) change in the reporting of patent data that indicates whether prior art citations are made by inventors or examiners. Our data consist of citing-cited pairs of patents generated from a large, random sample of patents issued over the period 2001-2003. Legal rules favor a "natural division" of citations, in which inventors reveal what they know and examiner fill in gaps. However, the data do not support these patterns across a number of dimensions. We expect that inventors would be more likely than examiners to cite their own prior patents (individual self-citation) and their firm's prior patents (firm self-citation). However, we find that there is no statistical difference between inventor and examiner citations for self-citations along either of these dimensions. We also hypothesize that inventors are more likely than examiners to cite technologies that are near to the citing patent in space, technology class, and time. We find this to be the case for geography. However, the magnitude of the difference in geographic citing patterns is so small as to be potentially economically insignificant. Regarding technology and vintage effects, examiners are more likely to proximate citations than inventors, reversing the expected pattern. Overall, our results do not change the presumption that patents trace out knowledge flows: inventors face strong legal pressures to reveal all they know, and citations do contain a signal of knowledge flows. However, our results indicate that the "invisible hand" of administration and the law is strong in generating citation streams. We would interpret our results to mean that citations might indeed track knowledge flows, but that the potential for both Type 1 and Type 2 error in making this inference is high.

"A person shall be entitled to a patent unless -

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States."

35 U.S.C. 102 Conditions for patentability; novelty and loss of right to patent.

Introduction

There is a large and growing literature in the technology and innovation literature that relies on patent data to study patterns of technological evolution, knowledge creation and diffusion, and firm technology strategy. Key areas of research that use patent data include: the geographic localization of knowledge flows (Almeida and Kogut, 1999; Jaffe, Trajtenberg, and Henderson 1993); knowledge diffusion across and within firm boundaries (Rosenkopf and Nerkar, 2002; Song, Almeida and Wu, 2003); technological positioning of firms (Podolny, Stuart, and Hannan, 1996); factors associated with the production of important innovations (Cockburn and Henderson, 1998; Gittelman and Kogut, 2003); university-firm technology transfer and universities as a source of important innovations (Henderson, Jaffe, Trajtenberg, 1998, Mowery, Sampat, and Ziedonis, 2003); and the impact of the structure of knowledge on knowledge diffusion and firm strategy and (Sorenson, Rivkin and Fleming, 2002, Ziedonis, 2003). The availability of data in electronic format has vastly increased the size of the datasets being used in the literature. Whereas limited sample sizes (a few hundred patents) characterized early papers, it is now common to find papers analyzing tens or hundreds of thousands of patents[1].

Analysis of prior art – citations to patents by other patents -- has been a core methodology in the literature. Two assumptions frequently made are that citations indicate a transfer of knowledge from the cited to the citing patent, and that highly-cited patents proxy for important inventions. The assumption of knowledge transfer characterizes much research on knowledge diffusion, in which citations are the "paper trails" tracking social, organizational, and geographic pathways of knowledge flows. However, in many instances researchers have been limited in their interpretations of their findings because citations made by patent examiners have not been separated out from citations made by inventors.[2] Notwithstanding a few important attempts to understand whether citations capture spillovers and quantify the impact of examiner citations (Jaffe, Trajtenberg and Fogarty, 2000; Cockburn, Kortum and Stern, 2003), little is known about whether and how examiner citations differ from inventor citations, and researchers have been forced to treat examiner citations as "noise" in the signal they are trying to measure.

Our paper takes patents themselves as the object of analysis and asks: what is the role of the key actors involved in the patent creation process (inventors, firms, lawyers, examiners) in assembling knowledge contained in lists of prior art? We leverage a recent (2001) change in the reporting of patent data that indicates whether prior art citations are made by inventors or examiners. Our data consist of citing-cited pairs of patents

[1] To illustrate the dramatic increase in the size of the datasets, a seminal paper published in 1993 (Jaffe, Henderson and Trajtenberg) had a sample size of 2,400 citing patents and 9,950 cited patents, while a 2003 working paper by a doctoral student (Singh, 2003) analyzed a sample of 500,000 citing patents and 1.3 million cited-citing dyads. Our dataset comprises 1,500 citing patents and 16,089 citing-cited dyads.
[2] Unless we specify otherwise, we use the term "inventor citations" to mean prior art citations not made by patent examiners. These might be made by individual inventors or by a firm's attorneys.

generated from a large, random sample of patents issued over the period 2001-2003. We complement the statistical analysis with reading of examiner manuals, and discussions with examiners, patent attorneys, and other patent professionals.

This paper provides the first comprehensive analysis of patent citations that accounts for differences between inventor-added citations and examiner-added citations. Thompson (2003) has studied inventor and examiner citation streams in an analysis of geographic citation patterns; we go beyond that first effort to include many other dimensions that capture inventor and examiner behavior. We hypothesize that inventor-added citations represent knowledge that is bounded along several dimensions: personal and organizational boundaries, geographic and technological distance, and vintage effects. We also hypothesize that examiners are relatively free of these limitations, but are subject to different ones that result from the processes and work practices they adopt in examining patents. We contrast inventor citations with those made by examiners to show differences along these dimensions. We hypothesize that examiner-added citations are not randomly-generated noise but systematic inputs into the patenting process that change key patterns in the data.

One methodological problem we face is that we are not able to separate citations specified by inventors from those added by inventors' lawyers. Insofar as lawyers are likely to be cognitively and behaviorally much closer to examiners than to inventors this is a limitation of our analysis. However, we are able to run separate analyses for patents in which lawyers are indicated and those in which lawyers were absent to broadly show how these additional and important actors influence patterns in the data.

What is the practical meaning of prior art?

A granted patent is a novel, non-obvious and manmade invention: an addition to the world's stock of technological knowledge and a stepping stone for future inventions (the latter a primary intention of the patent system). A patent consists of several components that define the invention, assign rights to individuals and organizations, and delineate the scope of those rights. The description discloses the invention so that it can be understood by others "skilled in the art". The core value of a patent is expressed in its claims, which detail aspects of the invention over which inventors and assignees may exercise ownership rights. Claims cover intellectual property that is not already anticipated by existing patents or public knowledge. To make the case that claims are valid (new and non-obvious), patents contain prior art. Prior art may consist of patented and nonpatented information; in fields where patenting is relatively new, much of the prior art may be in the public domain, eg, published in trade journals[3].

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; prior art may strengthen claims, by establishing that earlier versions of the inventions were different from or inferior to the current invention; or they may be "boilerplate" that establish facts described in the patent. Claims and prior art thus operate together to describe the invention, show its novelty over existing knowledge, and

[3] Hence, patents in biotechnology contain prior art lists that rely much more extensively on the published literature than patents in more mature technologies. To some degree, this reflects the reliance of biotech innovations on "open science", but also reflects the large amount of knowledge in that field that is published. Several studies have found that patents whose prior art is weighted towards published (rather than) patented citations subsequently receive more forward citations by other patents. While this finding has been interpreted as indicating that science or published knowledge is associated with greater inventiveness, it may more realistically represent a pattern in which these patents become heavily cited because they encapsulate a package of prior art that is dispersed widely in the literature.

delineate the scope and the corresponding strength of the intellectual property covered by the claims.

The role of prior art in proving (or disproving) claims underscores that the contents of a patent are not just codified knowledge but legal tools that embody and reflect the strategies of a variety of actors: individuals, firms, competitors, government employees. While some portion of the prior art contained in patents traces out knowledge flows, citations also reflect the heterogeneous objectives and interests of these different actors. Firms (and their attorneys) use patents not only to bring innovations to a market but also to create bargaining positions vis-à-vis other firms in the industry (Hall and Ham Ziedonis, 2001; Somaya, 2003; Ziedonis, 2003). The language used in claims and the selection of prior art are carefully crafted to minimize the risk of costly litigation; patents may also be written so as to create intellectual property positions that might invalidate other patents. The role of patent examiners is to certify the validity of a patent's claims to novelty and nonobviousness: examiners check the lists of prior art submitted by inventors and attorneys and make changes based on their own search of prior art and reading of the claims. Examiners are often treated in the literature as objective, independent arbiters of prior art, but in fact they communicate with inventors during the examination process, and there is a great deal of heterogeneity in the practices of individual examiners (Cockburn et al, 2003). Furthermore, examiners are subject to administrative pressures that limit the degree to which they search comprehensively. In this paper, we explicitly include the predicted behaviors of inventors and examiners in developing our hypotheses about the generation of prior art lists and their utility as measures of knowledge spillovers.

What is the meaning of patent citations for studies of technology and knowledge transfer?

A central premise of the technology literature utilizing patent citation data is that citations reveal evolutionary pathways across innovations, organizations, and time. We may identify two basic streams in the literature. The first uses patent citations to trace out pathways of knowledge diffusion across and within organizations, geographic space, and populations of inventors. The second category, which might be broadly termed productivity studies, is fundamentally concerned with determinants of technological performance. Here, citation data are used both as a measure of the impact of patents (as captured by forward citation counts) as well as to develop variables that measure different structural characteristics of inventions: breadth, originality, vintage, complexity, and fragmentation of the knowledge base.

Each of these types of studies rely on rather broad assumptions about the meaning of citations and citing-cited patent dyads. We focus our hypotheses on studies in the first category, which make the strongest assumption, namely that citations indicate a knowledge flow between inventors listed on citing and cited patents. It is generally accepted that prior art citations indicate a technological relationship between patents A and B, which may be quite strong (patent B wouldn't be possible without patent A) or relatively trivial (patent B belongs to a class of patents of which A is representative). However, patents do not cite *all* possible technological antecedents. By studying which patents were *actually* cited, compared to patents that could *potentially* have been cited, researchers use prior art lists to trace out knowledge diffusion patterns across individuals and firms, and within firms over time[4]. The interpretation of prior art citations is that, despite noise added by examiner citations, they reveal what is “inside the heads of “ inventors prior to an invention.

The legal rules work in favor of this interpretation. Inventors are required to submit patent applications that contain all published art they are aware of (patented and in the public domain) that is relevant to the claims they are presenting. Ultimately, if a patent is approved that does not list prior art (whether by the examiner or inventor), it might be vulnerable to infringement litigation and be found invalid. If, during litigation, it can be shown that inventors *knew* of prior art and failed to disclose it, they are subject to even greater penalties: inventors’ premises may be searched to prove failure to disclose prior art that was known to the inventors[5]. The following quote is illustrative of pressures on inventors and their attorneys for full disclosure:

“The first, as an inventor [sic], I was introduced to prior art as a 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 basically 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.

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 reference not cited in the patent] that that completely

4 This methodology was introduced in the seminal paper that used patent citations to measure localization of R&D spillovers, Jaffe, Trajtenberg and Henderson (1993) and subsequently employed by Almeida and Kogut (1999). A different methodology was employed by Jaffe and Trajtenberg (1998) in considering international spillovers.

5 This risk raises the interesting possibility that inventors have incentives *not to know* about related inventions. Where patents are at risk, communication with competitors (for instance, attending conferences or social gatherings) may represent a negative externality for firms not only because engineers could *reveal* too much information to competitors, but because engineers may *learn* too much information from competitors!

is 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".[6].

Given strong pressures for inventors to reveal what they know, we may make some simple assumptions about the kinds of knowledge inventors "ought to" be revealing in their patents. We hypothesize that there is natural division between inventor and examiner citations: that examiner-added citations are *less* likely to occur for knowledge that inventors are more likely to know, and *more* likely to occur for prior art that expect inventors to be unaware of. We assume that inventors' knowledge of technological antecedents is cognitively bounded, specifically that inventors' knowledge of prior art is likely to be strongest for technologies that are closest to them along a number of dimensions: social, organizational, technological, geographic, and chronological. Our hypothesis is that increasing distance along each of these dimensions will decrease inventor awareness, thereby increasing the chances that examiners will add citations. While cognitive awareness of technologies is clearly more complex than these simple univariate assumptions, our "natural division" hypotheses allow us to test the degree to which the data deviate from the simplest assumptions about inventor and examiner knowledge. If we find that these simple hypotheses are rejected, we might question the appropriateness of making more complex assumptions about the data without further validation.

First, we hypothesize that inventors would be *most* aware of their own prior inventions, meaning that overall inventors would be more likely (examiners less likely) to

---

[6] 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.

cite their own work. We may define several degrees of self-citation. First, there is self-citation by the same inventor, for which we expect the strongest prior knowledge by the inventor. Even if inventors move across firms, they can be expected to know of their own prior work:

> H1: Examiner-added citations are *less* likely to occur in citing-cited patent pairs that include self-citation by the same inventor. Conversely, examiner-added citations are *more* likely to occur in citing-cited patent pairs that do not share any of the same inventors.

Self-citation may also occur for prior inventions by the same firm. This is somewhat weaker, insofar as personnel turnover may mean that inventors are not aware of past patents at their firms. However, insofar as they work on projects that relate to past inventions, they should be aware of those patents:

> H2: Examiner-added citations are *less* likely to occur in citing-cited patent pairs that include self-citation at the firm level. Conversely, examiner-added citations are *more* likely to occur in citing-cited patent pairs that include patents by different firms.

Finally, we expect that lawyers should know about patents they worked on in the past, and will be likely to cite their own patents in prior art:

> H3: Examiner-added citations are *less* likely to occur in citing-cited patent pairs that include self-citation at the lawyer level. Conversely, examiner-added citations are *more* likely to occur in citing-cited patent pairs that are include patents by different lawyers.

Another dimension of knowledge about prior art is knowledge about pertinent art across different technology classes. Consistent with the natural division hypothesis, we expect that inventors will be more aware of inventions in similar technological domains than they are of more technologically distant patents:

> H4: Examiner-added citations are *less* likely to occur in citing-cited patent pairs that are in the same technological field. Conversely, examiner-added citations are *more* likely to occur in citing-cited patent pairs that include patents in different technological fields.

Geography is another dimension by which knowledge may weaken with distance. Indeed, the core theoretical insight of the spillover literature is that despite the intangible nature of ideas, knowledge does not diffuse readily across space but flows more readily within regions. Following this logic, we hypothesize inventors are more likely to be aware of other technologies that are developed nearby than they are technologies developed in distant locations:

> H5: Examiner-added citations are *less* likely to occur in citing-cited patent pairs that are within the same geographic region. Conversely, examiner-added citations are *more* likely to occur in citing-cited patent pairs that are include patents in different geographic regions.

Finally, we hypothesize that inventors' knowledge of other technologies is more likely to include recent inventions that overlap in time with their own inventive activities, and that they are less likely to know about antecedents that occurred in the distant past:

> H6: Examiner-added citations are *less* likely to occur in citing-cited patent pairs that are close in time. Conversely, examiner-added citations are *more* likely to occur in citing-cited patent pairs that are far apart in time.

Statistical support for the hypothesized "natural division" patterns would support the contention that citations are a noisy signal of knowledge flows, and that the primary source of noise is added by examiner citations. If however, the patterns are rejected by the tests (both univariate and multivariate), which would present a more puzzling and complex picture of the process by which citations are generated. Why should examiner citations be more likely to fall in areas where we would expect *inventors* to be most knowledgeable? Similarly, why would inventors know more about technologies that are distant from their own activities than they know about those that are most related?

Several answers present themselves, which we group under the headings of "administrative process hypotheses" and "strategic citation hypotheses". One is the patenting process itself, in which inventors and lawyers craft lists of prior art for reasons that are unrelated to knowledge flows but are instead driven by legal and strategic considerations. Many prior art citations may be added that do not represent a knowledge transfer or evolutionary relationship between inventions, but rather are added ex post to strengthen specific claims. In this regard, lawyers play a critical role. Attorneys operate much more like examiners than they do inventors -- many patent attorneys were formerly US patent examiners: their job is to anticipate information that an examiner would likely insist upon reviewing, and include that information in the patent application. In so doing, they are likely to turn up prior art that was not previously known to inventors. The following quote, from the same attorney quoted above, illustrates this:

> So, as an attorney. . .I ask the inventors to err on the side of giving the information to me, or to the people I'm working with, to make sure that we consider it. And I paint the same picture of fraud and disbarment, and all these solid, you know, strong feelings, that I don't want to share a jail cell with these guys, and I want to make sure and get the right art in. But, more importantly, I also paint some stories, war stories, from licensing and litigation efforts that went completely south because of *some references that we weren't even -- that we were not aware of, that probably should have been turned up if the right kind of search tools had been used earlier on in the game.*

In a study to investigate this type of measurement error, Jaffe, Trajtenberg and Fogarty (2000) surveyed patent inventors to identify their familiarity with prior art cited in their patents. The authors report "a large amount of noise in citation data; it appears that something like one-half of all citations do not correspond to any perceived communication, or even necessarily to a perceptible technological relationship between the inventions." Only 28 per cent of inventors reported a high degree of familiarity with patents cited in their own patents. In about 40 per cent of cases, inventors learned about prior art cited in their patents *during* the process of writing their patent applications, indicating that the patenting process itself is important in generating citing-cited patent pairs.

Overall, the administrative process hypothesis implies that when lawyers are involved, inventor-generated citations would not be very different from examiner-generated citations. Our data do not allow us to distinguish between individual citations added by inventors and those put in by attorneys. However, they do allow us to test for whether attorneys and deep patenting experience at the firm level align examiner and inventor citation patterns. First, we expect that firms with a great deal of experience in patenting will have learned how to craft prior art in such as a way as to minimize missing citations, thereby minimizing examiner citations. Insofar as patent attorneys are cognitively and behaviorally much closer to examiners than they are to inventors when finding prior art, we expect that the difference between inventor citations and examiner citations to be small for firms employing counsel:

H7: There will be no significant differences between inventor-added citations and examiner-added for citing-cited pairs on which lawyers are listed on the citing patent.

The administrative process hypothesizes that examiner and inventor citations streams will be similar, however, examiners may differ from inventors by citing patents they previously examined. Examiners are civil servants who face strong administrative pressures to process as many patents as possible in as short a time as possible; indeed, their bonus pay is tied to yearly quotas of patents they examine. Faced with these pressures, examiners might frequently refer patents that they are familiar with and have used in the past. Cockburn et al (2003) note “patent examiners identify and frequently refer to ‘favorite’ examples of prior art that usefully describe (“teach”) the technology area and the bounds of prior art in a way that facilitates the examination of a wide range of subsequent applications.” In addition, since examiners are narrowly specialized within their technological area, they may be responsible for much of the relevant prior art in that area. We therefore hypothesize self-citatoin patterns at the examiner level:

H8: Citing-cited patent pairs that share the same examiner are more likely to be added by the examiner.

Another problem with the “natural division” assumption is that incentives for inventors to add citations are complex; it cannot be assumed that inventors will always disclose all that they know. Since prior art limits the scope of claims, inventors and attorneys may desire to keep prior art to a minimum. Against this objective is the risk of examiner rejection or, worse, infringement suits by competitors if the patent is approved without all the relevant prior art. To the extent that inventors face a greater risk of lawsuits from competitors than from inventors within their own firm, we may expect that incentives to minimize prior art would be greater, on balance, for self-citations than for competitors citations:

H9: Inventor citations are less likely to occur than examiner-added citations for self-citation at the firm level.

At the individual level, inventors may not cite their own prior work if it was performed at another firm. Employers in high-technology industries frequently require inventors to sign nondisclosure agreements to prevent leakage of valuable knowledge stemming from job mobility, but given the difficulty of separating knowledge at the individual level, these agreements are likely inadequate to protect against knowledge transfer of this sort. Inventors may avoid citing their own prior patents at former employers in order to minimize the risk of litigation related to breach of confidentiality. We group patterns consistent with these behaviors into the "strategic citations" hypothesis:

H10: Inventors are less likely than examiners to self-cite when inventors change employers.

Data

Starting in January 2001, the USPTO has indicated on the front page of patent images prior art citations that were added by examiners. We collected the front page images of all utility patents granted between January 1st 2001 and August 26th 2003. This yielded a group of 442,839 patents citing back to 5,434,883 patents in their prior art. Table 1 provides summary information for the dataset. Patent examiners are an important source of citations representing about 40 per cent of all citing-cited dyads in each of the three years. The magnitude of examiner citations is even higher when measured on a per patent basis: for the average patent in our dataset, examiners imposed 67 per cent of all prior art citations. Moreover, between 38 and 40 per cent of patents granted over the period have *all* citations imposed by examiners; in contrast, only 8 per cent of patents had no examiner-added citations. About 70 per cent of the patents have at least half or more of their prior art citations introduced by examiners[7].

Tests of our main hypotheses requires that we match individual elements of citing and cited patents; to reduce this to a manageable task, we create a sample of 1500

[7] We do not consider non-patent references or non-US patents in our analysis. All references to prior art involve US patent citations.

citing patents from the full dataset. We randomly select 500 patents from each year in the data, resulting in a sample of 1500 citing patents. Patents are selected at random within each year rather than within specified periods to avoid seasonal variations that would bias our samples, e.g., monthly variations in examination schedules across Art Units or groupings of patent approvals by company.

The 1,500 citing patents generate 17,866 prior art citations; only 26 patents (2.4 per cent of the sample) have no citations at all. Detailed data are not available for patents granted before 1976, leading us to remove 1,767 of the cited patents granted before that date, yielding a final sample of 16,089 citing-cited patent pairs. We perform Mann-Whitney and Kolmogorov-Smirnov tests to compare the distribution of the dataset with that of the samples. Table 2 shows basic statistics and results of these tests for three variables: citations per patent, percentage of examiner citations per patent and application year. For all variables and tests used, we cannot reject the hypotheses that the distribution of our sample is similar to the distribution of the full sample. To explore whether the distribution across technology classes in our samples is similar to that of the full datasets, we compare the top 30 most frequent classes in both groups and find an overlap of over 80 per cent for each year.

Our statistical tests depend on identification of common elements linking the citing and cited patents: inventors, assignees, geographic locations, technological domains, and vintage. Additionally, we wish to control for common examiners and patent attorneys in understanding the allocation of citations to examiners and inventors. However, except for classifications, patent data are not standardized, resulting in a great deal of variation in data formats across common elements. To correct for this, we perform a number of operations on the data, in order to identify common assignees, geographic locations, and individuals.

Regarding assignees, we perform three steps. We first standardize names by correcting for differences in spelling and format (for example: Sam Sung Electronics/Samsung Electronics; German letter ö typed as o or oe; Minnesota Mining and Manufacturing Co./Minnesota Mining & Manufacturing Company /3M). In the second step, we group assignees with different names (e.g., Nokia Finland and Nokia USA) that are subsidiaries of the same corporate parent. We identify the ultimate parent for each assignee using the Directory of Corporate Affiliations based on their parent in the year of patent application, going back to 1991. Assignees on patent applications before 1991(27 per cent) were matched to the 1991 directory. We further correct for mergers, acquisitions, and name changes since 1976 using The Directory of Corporate Affiliations. Taken together, these changes reduce the number of unique assignee names by 28 per cent, from 5933 to 4239: 1002 assignee names are eliminated through corrections for name variations; 1694 unique assignee names are removed in the second step accounting for corporation affiliations and mergers.

We assigned the geographic location of a patent based on the locations of inventors. We identify locations for all inventors listed on citing and citing patents, not just first inventors. Our sample generated 40,797 inventors (58 per cent located in the United States) and 8,474 different locations (51 per cent in the USA). Similar to assignee data, locations also present problems and require significant cleaning and checking. We first perform a manual cleaning of city, state, and country names[8]. Second, we identify

[8] To our surprise, state and country data was far from perfect. Unfortunately, the USPTO uses the same abbreviations for countries and states. For example, CA can be California or Canada, IL can be Illinois or

longitude and latitude point data using the United States Postal Office for locations within the USA and GEOnet name server of the National Imaginary and Mapping Agency for all other locations. These steps allow us to identify 73 per cent of locations, leaving 2,301 locations, mainly Chinese, Japanese, and Taiwanese locations, unidentified. To identify these locations, country natives checked each list to match place names to those given in the GEOnet and USPT databases. Because of these efforts, we are able to identify at least one inventor location for all but four patents.

Prior studies have used both administrative units as well as actual distance in miles to measure common geographic location. We adopt both methods, as each captures different aspects of the relationship between geography and knowledge flows. We first measure administrative boundaries at the country, city, state, economic area and county level. [9] The last three variables are defined only for citations where both patents have an American inventor. Finally, latitude and longitude coordinates enable us to calculate distance in miles between patent locations using the great circle distance equation.

Identifying a common inventor across patents presents a challenge since it is reasonable to expect that two names, although identical, can correspond to two different individuals. Although in many cases, inventor names contain initials for middle names, reducing the chances of mistakenly identifying two names as the same, we follow different principles for matching inventors between citing and cited patents to develop more stringent tests. First, we identify whether the full name is the same: first name,

Israel. This problem is also present in the NBER dataset. Although this problem does not seem to affect a great number of patents, researchers should be aware of it.

[9] In an effort to identify geographic areas that mimic economic activity and not state or administrative boundaries, the Bureau of Economic Analysis (BEA) defined 171 economic areas that span the US. Each economic area consists of one or more nodes – metropolitan or similar areas that serve as centers of economic activity – and the surrounding counties that are economically related to the nodes. The main factor used in determining the economic relationships among counties is commuting patterns, so each economic area includes, as far as possible, the place of work and the place of residence of its labor force.

middle, and last names are all the same for this to result in a match. Second, we identify if full names are the same *and* company assignee is the same. Third, we identify whether full names are the same *and* the locations for citing and cited patent are no more than 100 miles apart. The first matching principle is the most flexible because it recognizes that inventors can move to other locations or companies. However, it is also the most likely to generate false matches leading to Type 2 measurement error (inferring a common link when no link exists). The last two matching rules are more restrictive, in that they are less likely to assume common inventors in cases where the same names belong to different people.

We also match the names of examiners and lawyers on citing and cited patents. Name matching of examiners is more reliable because we are able to make use of a roster with the names of all examiners, produced by the USPTO, reducing ambiguous cases. The USPTO also provides a list of all the lawyers and law firms that are registered to introduce patent applications. This roster not only allows us to match names of lawyers on citing-cited patent pairs but also provides information on whether the lawyer is an in-house counselor or not. Companies that introduce numerous applications per year, such as Intel, IBM, Procter & Gamble, have a group of internal lawyers that deal with the applications for that company. In some cases, in-house counselors and external law firms are both involved in a patent application.

Finally, we identify whether the citing-cited pairs belong to the same technology class. We use the International Patent Classification (IPC) instead of the United States Classification (USC) for this purpose. A number of reasons drive this choice. First, the IPC system follows a nested hierarchical structure, allowing us to look at different levels of aggregation in the technology domain. Second, the IPC system is more similar to a traditional industry end-use classification system than the US system, which classifies patents by function. One problem with the IPC is that older patents are not reclassified when classification codes change (which happens infrequently). This would make our matching test more conservative, insofar as patents that belong to the same class are coded differently because the more recent patent was subject to a newer classification code. To account for this, we update IPC codes based on USPTO-IPC concordance tables.

Table 3 shows definitions for the variables used to measure linkages between citing and cited patents for all of the elements discussed above.

Univariate tests

We first conduct univariate tests of means of inventor and examiner citations to observe patterns in the data that correspond to our natural division or strategic citation hypotheses. Our null hypothesis is the "administrative process" hypothesis, that inventor citations are generated by a process similar to the process generating examiner citations, rather than a process of revealing only prior knowledge. Our natural division hypotheses are supported when examiners have a higher propensity than inventors to add distant citations. The "strategic citations" hypotheses are supported when inventors have a greater propensity than examiners to add more "distant" prior art. Results are shown in Table 4.

Two key findings stand out: first, the statistical results provide at best mixed support for the natural division hypothesis. Second, even if statistical differences between examiners and inventors are significant, the magnitude of the difference is so small for most measures as to be economically insignificant. Concerning statistical differences, examiners have a higher propensity to cite distant prior art (supporting the natural division hypothesis) when adding citations that belong to different corporate parents (H2), and when adding geographically distant citations for several of the geography measures (H5): citations that are foreign; out-of-state; across different economic areas; and distance measured in miles. The greatest difference in inventor and examiner proportions occurs for foreign citations, possibly as a result of examiners having better access to search tools and databases for foreign patents. From these tests we reject H8, that firms do not cite their own prior art for strategic reasons.

There is also support for both the administrative and strategic citations hypotheses. In particular, the natural division hypothesis is rejected for H1, which in which we have the *strongest* presumption of inventor knowledge: inventors are not more likely to add self-citations at the individual level than examiners. For the more restricted inventor measures, in which we have higher confidence of identifying the same individuals (same inventor/same company; same inventor/same city), examiners are *more* likely to include self-citations than inventors themselves. Since we have difficulty accepting that inventors "forget" about their own past patents that are relevant to current patents, we either assume that, consistent with H10, they omit self-citations for strategic or legal reasons, and examiners are adding back what inventors "should have" included in their lists. We probe more deeply into these patterns in our multivariate tests.

The univariate tests regarding technology also violate the natural division hypothesis (H4): inventors are more likely to add prior art from different technological classes than examiners (4 digit IPC codes). Two possible explanations present themselves. The first is that inventors have a greater breadth of knowledge about patented technologies than examiners, who are narrowly specialized by technological field. The second (and we feel, more plausible) explanation stems from the patenting process itself. Patents are not classified until they go through the examination process, so inventors are adding citations without knowledge of the ultimate detailed classification code of the invention. In the process of examining patents, examiners develop classification codes based on individual claims, at the same time searching for prior art that is relevant to those claims. This endogenous process in which classifications and prior art are simultaneously generated by examiners would be consistent with a pattern in which examiners match citing and cited patents on technology to a greater degree than inventors. Finally, we reject the natural division hypothesis for H6 on vintage effects. Examiners are more likely to add recent citations than inventors, with a mean difference in years of 7 versus 9.8 years. The long time lags for both inventors and examiners (7 and 10 years) mean the difference is probably not due to administrative delays, such that examiners know about more recently-issued patents than inventors. While we do not have a strong theoretical explanation for this finding, we note that it is consistent with a strategic citations hypothesis insofar as inventors and lawyers may choose to cite older patents whose owners are less likely to litigate than owners of recently issued patents.

Statistically then, we find weak support for the natural division hypothesis predicting that prior art lists reveal "what inventors know", with examiner citations adding more distant prior art. More important than the lack of statistical support is the fact that the magnitude of the differences is so small as to make inventor and examiner citation patterns virtually the same. We find no economically significant differences in the means between examiner and inventor citing patterns, except for international citations and technology class. That is, examiners are not adding “random noise” to inventor citations – inventor and examiner citations track eachother quite closely. Citations that would be presumed to represent a knowledge flow would be just as likely to be generated by examiners than by inventors.

To test this, we turn to multivariate logit regressions that estimate the odds of a given citation dyad being generated by an examiner or an inventor, conditional on the dimensions measured in the univariate means tests.

Multivariate analysis

To tests our hypotheses, we estimate the following empirical specification:

$$Prob(examiner\ citation = 1 \mid X) = F(\beta X) \qquad (1)$$

where $F(Z)=e^z/(1+e^z)$ is the cumulative logistic distribution, our dependent variable is binary and equal to 1 if the citation was imposed by the examiner and 0 if it comes from the inventor, with X a set of variables that indicate similarities between cited and citing patent along the dimensions shown in table 3: self-citation by individuals, assignees, corporate parents, lawyers and examiners; geographic location; technology classes and time. Since citation pairs for a given citing patent are not necessarily independent, we estimate the β coefficients in equation 1 using robust maximum-likelihood estimation to correct for heteroscedasticity. Table 5 shows the summary statistics and table 6 the correlation values for all variables.

Table 7 contains results in two sets, the first (models 1-2) for all dyads, the second (models 3-5) for dyads where the citing patents had law firms associated with their application process (models 3 and 4), or in which no law firm was associated with the citing patent (model 5). Coefficients for all variables are expressed as odds ratios. An odds ratio greater than one indicates that the citation is more likely to be originated by examiners than by inventors; odds ratios less than one indicate that examiners are less likely than inventors to add a citation. Statistical significance for a given coefficient indicates that examiners and inventors differ in the propensity to add a citation.

We first consider self-citation at the firm level. In Model 1, self-citation is measured for the company level: the coefficient is positive, (inventors self-cite more than examiners) but statistically insignificant. For self-citation at the level of the corporate parent (model 1), the coefficient is again positive but barely significant at the 5 per cent level. These results do not support the "natural division" hypothesis at the firm level, whether for intra-company (where we would expect knowledge flows to be strongest) or intra-parent citations.

The coefficient for self-citation at the individual inventor level in Model 1 is close to one but not statistically significant. Hence, we reject the natural division hypothesis and conclude that there is no difference between examiner and inventor citations at the individual level. Self-citation at the examiner level does appear to influence citation patterns: citing-cited pairs that share the same examiner are more likely to be added by the examiner, and this result is significant at the 5 per cent level.

Turning to the other variables, the coefficient on distance is equal to one and highly significant, indicating an equal probability of a citation being generated by inventors or examiners across distance. We illustrate this pattern in figures 1 and 2, which shows the distribution of inventor- and examiner generated dyads in the United States and between US and non-US dyads, which show that they closely track one another. The graphs do not reveal a pattern consistent with the natural division hypothesis. We examine the effects of geography in more detail in models that follow.

Technology and vintage effects are consistent with the univariate tests, leading us to reject the natural division hypothesis in both cases.Examiners are less likely to add citations when the cited and citing patents differ in technological class across all models. We estimate models with technology specified at the 4-digit classification level. As shown in model 1 an inventor is 43% (1/0.70) more likely to cite patents from different 4-digit technological classes than from the same classes of the citing patent, suggesting that inventors are adding more breadth of prior art than examiners. We also find a similar pattern for the variable *years,* which indicates that examiners are more likely to cite more recent patents than inventors, however, the magnitude of the difference is small.

To test whether these results change when lawyers are involved in the process we re-estimate model 1 for citing patents that list lawyers (models 3 and 4) and those without lawyers (model 5). Model 4 adds lawyer self-citations, which are not significant, so we reject H3 that inventor-added citations are more likely to include self-citation at the lawyer level. Comparing model 5 with models 3 and 4 provides an opportunity to test for administrative effects, where we hypothesize that lawyers act as proxies for examiners and dilute the true differences between examiner and inventor citations. If the administrative hypothesis is correct, the sample in which no lawyers are involved should exhibit greater evidence of a "natural division", such that nearer (more distant) citations are more likely to be added by inventors (examiners). The test provides no support for this hypothesis: model 5 does not show significant changes from models 3 and 4. One concern is that the sample of patents in which lawyers are not involved is so small (5 per cent of dyads) that this is not an adequate test for the administrative effect. Analysis of lawyer experience and social position may be called for, specifically to test whether very experienced lawyers, and lawyers who have close connections to the patent office (for instance, were formerly patent examiners), approximate examiner citation patterns more closely than less experienced lawyers. We note that overall, our results do support administrative process effects, since for many variables we find no statistical difference between inventor and examiner citations, even when we expect to find strong differences between them (in particular, self-citations at the firm and individual level).

We explore further the role of geographic location in table 8. Acknowledging the non-linear relationship between distance and communication patterns, we re-estimate model 1 from table 7, but instead of measuring distance in continuous miles between citing and cited patent we define distance with binary variables that indicate whether there is at least one pair of inventors in the citing-cited pair that is in the same country, state, county, economic area, or city. Note that only 7,632 observations are used to estimate models 2 through 5 since the geographic definitions used in these models require at least one inventor in the US listed on both citing and cited patents. This underscores the very high proportion of dyads (52%) that include at least one non-US patent.

We focus our attention on the geographic component (results for the other independent variables remain similar to those in table 7). Examiners are 87% more likely to cite patents that are in other countries than inventors, confirming their role as connectors of knowledge across national boundaries. This lends strong support for Jaffe and Trajtenberg's (1998) prior finding that knowledge spillovers as evidenced by patent citations are strongly national in character. Within the United States, examiners are more likely to cite patents originating in different states, counties, and economic areas. These results provide strong support for the natural division hypothesis for geographic knowledge flows. However, the result of city is not statistically significant. Cities may be too small to be economically meaningful units in measuring knowledge flows, particularly knowledge flows related to local patterns of employment, communication, and transactions. The economic area is designed to overcome these limitations, and is statistically significant. Overall, we find strongest support for the natural division hypotheses in geographic citing patterns. At the same time, the magnitude of the difference in geographic citing patterns is very low, as shown in figures 1 and 2. This raises the possibility that examiner citations are potentially inflating localization patterns that are being attributed to knowledge spillovers.

Discussion and conclusion

Knowledge is difficult to measure, and researchers have understandably been eager to apply patent citation data to test macro-level theories of knowledge creation and diffusion. However, the question has always remained as to the extent to which these data actually do track knowledge flows. In particular, the addition of examiner citations and the process by which firms and attorneys craft patents would seem to add significant noise – and possibly distortions to – the assumed patterns of knowledge flows. Apart from a few studies that show the importance of examiner citations and potential weakness in the knowledge transfer assumption (Cockburn et al, 2003; Jaffe et al, 2000) ours is among the first to compare the generation of inventor citations to examiner citations to test for differences that would violate knowledge flow hypotheses.

Our "natural division" hypotheses most approximates a world in which inventors reveal what is in their heads, and examiners fill in the missing pieces. We find the greatest statistical support for this view of the world in our analysis of the geography of citing-cited pairs. However, the closeness with which examiner and inventor citations track eachother across distance raises the question of the economic significance of these effects. Regarding self-citation, we reject the natural division hypothesis as there is no statistical difference between inventor and examiner citation. Instead, we find support for the "administrative process" hypothesis, in which the process of writing patents means inventor citations anticipate and mirror those of examiners. In our univariate tests, we found some support for the "strategic citations" hypothesis for self-citation at the personal level. The fact that self-citations at the individual level (measured three different ways) show either no difference between inventors and examiners (for two of the measures) or are more likely to come from examiners than inventors (for the same-city measure) provides, we feel, the strongest support for strategic effects in citation patterns. We intend to explore these effects further with more detailed analysis of inventor mobility and interviews geared to understanding incentives to add or suppress self-citations from prior employers.

Overall, our results do not change the presumption that patents trace out knowledge flows: inventors face strong legal pressures to reveal all they know, and therefore citations do contain a signal of knowledge flows. Furthermore, studies have consistently found patterns in the data that would likely not be found if no spillovers were present. However, while researchers have argued that examiner citations represent added noise to the signal of inventor citations, our analysis indicates that changes introduced by the patenting process are not random noise: we find that the “invisible hand” of administration and the legal system is strong in generating citation streams. We would interpret our results to mean that citations might indeed track knowledge flows, but that the potential for both Type 1 and Type 2 error in making this inference is high. Where administrative effects are strong, Type 1 error may occur: the hypothesis of knowledge transfer would be rejected when in fact it is present if distant citations added by attorneys and inventors are “washing out” the true signal of knowledge flows. On the other hand, since examiners are adding citations that are “near” to inventors, researchers also risk Type 2 error, in which inferences of knowledge transfer are made when in fact many such citations are added by examiners.

Our results point to interesting processes by which citations are generated that have not received as much attention in the literature. Prior art citations are not only lists of knowledge held by inventors, but legal and strategic tools in which the interests of inventors, attorneys, examiners, and competitors come into play. We intend to further explore these strategic effects with more analysis of competitive citations, self-citations and inventor mobility. Support for the strategic citations hypothesis suggests that citations should not be viewed as a noisy signal of knowledge flows, but as a multi-dimensional signal involving heterogeneous processes and actors -- knowledge flows among inventors, strategic and legal positioning by firms and individuals, and a complex administrative process of codification by lawyers and examiners-- that intersect to create and shape technology fields.

References

Adams, James and Adam Jaffe (1996), "Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data", *Rand Journal of Economics,* 27 (4):700-721.

Almeida, Paul and Bruce Kogut (1999), "Localization of Knowledge and the Mobility of Engineers in Regional Networks", *Management Science*, 45(7):905-918.

Cockburn, Iain and Rebecca Henderson, (1998) "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery", The Journal of Industrial Economics, 46(2):157-182.

Cockburn, I. S. Kortum and S. Stern (2004) "Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes" in Wes Cohen and Stephen Merrill, *Patents in the Knowledge-based Economy,* Washington, DC: National Academies Press.

Gittelman, Michelle and Bruce Kogut. 2003. Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns. *Management Science*. 49(4):366-382.

Jaffe, A, M. Trajtenberg and M. Fogarty (2000)The meaning or patent citations: Report on the NBER/Case Western Research survey of patentees" NBER Working paper 7631.

Jaffe, A., Trajtenberg, M. and Henderson R (1993) "Geographic location of knowledge spillovers as evidenced by patent citations", *Quarterly Journal of Economics* 108:577-598.

Jaffe, Adam and Manuel Trajtenberg, (1998), "International Knowledge Flows: Evidence from Patent Citations", NBER Working Paper 6507 (April).

Hall, B, and R. Ham Ziedonis (2001), "The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979-1995", *The Rand Journal of Economics*. 32 1: 101-129.

Henderson, R., A.B. Jaffe, and M. Trajtenberg (1998). "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-88," *Review of Economics & Statistics*, 119-127.
Lamoreaux and Sokoloff, 2003

Mowery, Sampat, and Ziedonis (2003) "Learning to Patent: Institutional Experience, Learning, and the Characteristics of U.S. University Patents After the Bayh-Dole Act, 1981–1992", *Management Science,* 48(1):73–89

Podolny, J, Stuart, T, Hannan, M. 1996. Networks, knowledge, and niches: Competition in the worldwide semiconductor industry, 1984-1991 *American Journal of Sociology,* 102 3: 659-690.

Sorenson, Rivkin and Fleming (2002), Complexity, Networks and Knowledge Flow (working paper).

Rosenkopf, Lori and Atul Nerkar (2001), "Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry", *Strategic Management Journal* 22(4):287-306

Singh, Jasjit (2003), "Social Networks as Drivers of Knowledge Flows", Working Paper, Harvard University.

Somaya, Deepak, (2003) "Strategic determinants of decisions not to settle patent Litigation." *Strategic Management Journal*. 24 1:17-

Song, Jaeyong, Almeida, Paul, Wu, Geraldine (2003). Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Management Science* 49(4): 351-365.

Thompson, Peter (2003) "Patent Citations and the Geography of Knowledge Spillovers: What do Patent Examiners Know?" Working paper, Carnegie Mellon University.

United States Department of Commerce, 1999 "Public Hearing on Issues Related to the Identification of Prior Art During the Examination of Patent Application", June 28.

Ziedonis, Rosemarie (2003). "Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms", Working paper.

Table 1
Summary statistics of full dataset

	2001	2002	2003 *	Total
Citing patents	166,064	167,424	109,351	442,839
Cited patents	1,960,448	2,040,345	1,433,690	5,434,483
Inventor	57%	59%	60%	59%
Examiner	43%	42%	40%	42%
% examiner citation x patent				
Average	63%	63%	63%	
0%	8%	8%	7%	
10%	5%	5%	6%	
20%	6%	6%	7%	
30%	6%	6%	6%	
40%	5%	5%	6%	
50%	8%	8%	9%	
60%	3%	3%	3%	
70%	6%	6%	7%	
80%	4%	4%	5%	
90%	6%	6%	6%	
100%	42%	41%	39%	
All citations by examiner	40%	39%	38%	

* From January 1 to August 26 2003

Table 2
Comparison of full dataset and 3 year sample

	2001		2002		2003	
	Universe	Sample	Universe	Sample	Universe	Sample
Citin patents						
Total	166,064	500	167,424	500	109,351	500
With no citations	2.4%	2.0%	2.3%	2.4%	1.9%	2.8%
Cited patents						
Total	1,960,448	5,668	2,040,345	5,902	1,433,690	6,296
Citation/patent						
Mean	11.84	11.33	12.23	11.80	13.18	12.62
Std. Dev	17.74	13.63	18.47	21.00	20.75	18.41
% Examiner citations						
Mean	0.63	0.63	0.63	0.62	0.63	0.64
Std Dev	0.37	0.37	0.37	0.37	0.37	0.38
Application year						
Mean	1,998.69	1,998.68	1999.699	1999.7	2,000.40	2,000.32
Std. Dev	1.25	1.19	1.22	1.18	1.24	1.13
Mann-Whitney Test	z	Prob >|z|	z	Prob >|z|	z	Prob >|z|
Citation/patent	0.606	0.5442	0.861	0.389	0.549	0.5828
% Examiner citations	0.38	0.70	0.746	0.455	(0.85)	0.40
Application year	0.317	0.751	0.045	0.9369	0.932	0.3514
Kolmogorov-Smirnov Test	D	p-value				
Citation/patent	0.0301	0.769	0.0305	0.755	0.0313	0.73
% Examiner citations	0.0246	0.929	0.0305	0.755	-0.0377	0.499
Application year	0.0099	1	0.007	1	0.0178	0.998

H0: Sample= Universe, H1: Sample~= Universe

H0: Distribution of Sample= Distribution of Universe, H1: Distriution of Sample~= Distribution of Universe

Table 3
Variable definitions and measurement

Dimension	Variable	Definition
Dependent variable	Examiner	1 if examiner citation, 0 if inventor citation
Self Citation, by:		
Inventors	same_inventor	1if citing and cited patents have the same first inventor, 0 otherwise
	same_inventor_all	1 if citing and cited patents have at least 1 inventor in common, 0 otherwise
Firms	dif_company1	0 if citing and cited patents have the same first assignee, 1 otherwise
	dif_company_all	0 if citing and cited patents have at least 1 assignee in common, 1 otherwise
Corporate parent	dif_parent1	0 if citing and cited patents have the same first ultimate parent 1 otherwise
	dif_parent_all	0 if citing and cited patents have at least 1 ultimate parent in common, 1 otherwise
Lawyers	same_law_firm	1 if citing and cited patents have the law firm, 1 otherwise
	same_law_firm_all	1 if citing and cited patents have either at least 1 law firm in common, 0 otherwise
Examiners	same_examiner	1 if citing and cited patents have the same primary examiner, 0 otherwise
	same_examiner_all	1 if citing and cited patents have either the same primary or assitant examiner, 0 otherwise
Location, by:		
Country	dif_country1	0 if the first inventors in the citing and cited patents are in the same country, 1 otherwise
	dif_county_all	0 if at least 1 inventor in the citing and cited patents is in the same country, 1 otherwise
State (US)	dif_state1	0 if the first inventors in the citing and cited patents are in the same state, 1 otherwise
	dif_state_all	0 if at least 1 inventor in the citing and cited patents is in the same state, 1 otherwise
Economic area (US)	dif_ea1	0 if the first inventors in the citing and cited patents are in the same economic area, 1 otherwise
	dif_ea_all	0 if at least 1 inventor in the citing and cited patents is in the same economic area 1 otherwise
County (US)	dif_county1	0 if the first inventors in the citing and cited patents are in the same county, 1 otherwise
	dif_county1_all	0 if at least 1 inventor in the citing and cited patents is in the same county 1 otherwise
City (all)	dif_city1	0 if the first inventors in the citing and cited patents are in the same city, 1 otherwise
	dif_city1_all	0 if at least 1 inventor in the citing and cited patents is in the same city 1 otherwise
Miles	distance1	Distance in miles between the location of first inventors for citing and cited patents
Technology	dif_technology4	0 if citing and cited patents have same primary IPC technology classficiation, 1 otherwise
	dif_technology4_all	0 if citing and cited patents have at least 1 IPC technology classficiation in common, 1 otherwise
Time	years	application year citing-application year cited

Table 4.
Comparison of Means, Inventor and Examiner Citations

	Inventor Citations (n=9370)	Examiner Citations (n=6725)	T test	H_o: Examiners=Inventors ("administrative process" hypothesis) H_{A1}: Examiners>Inventors with increasing distance ("natural division" hypothesis) H_{A2} :Inventors>Examiners with increasing distance ("strategic citations" hypothesis)
Self Citation:				
Same inventor	0.063	0.062	0.20	Accept Ho
Same inventor, same company	0.039	0.046	-4.12**	Accept HA2
Same inventor, same city	0.022	0.033	-2.29**	Accept HA2
Different company	0.89	0.90	-1.55	Accept Ho
Different parent	0.87	0.88	-3.2**	Accept HA1
Same law firm [a]	0.09	0.08	1.5	Accept Ho
Technology Class:				
Different Technology, 4 digit IPC code	0.49	0.38	13.1**	Accept H A2
Geographic Distance:				
Different country, all inventors	0.34	0.47	-0.17**	Accept HA1
Different state, all inventors, US only[b]	0.70	0.73	-3.04**	Accept HA1
Different city, first inventor only	0.96	0.95	2.75**	Accept HA2
Different city, all inventors	0.9	0.91	-1.05	Accept Ho
Different economic area, all inventors, US only[b]	0.74	0.77	-3.02**	Accept HA1
Distance, miles	2197	2605	-11.0**	Accept HA1
Vintage:				
Years	9.8	7.1	28.3**	Accept HA2

* $p<0.05$ ** $p<0.01$

a. $N_{inventors}$=6986; $N_{examiners}$=4988

b. $N_{inventors}$=5253; $N_{examiners}$=2379

Table 5

Variable	Obs	Mean	Std. Dev.	Min	Max
distance	16,089	2367.6870	2321.8880	0	10780.76
dif_country_all	16,089	0.3951	0.4889	0	1
difstateall	7,632	0.7087	0.4544	0	1
dif_county_all	7,632	0.8147	0.3885	0	1
dif_city_all	16,089	0.9025	0.2966	0	1
dif_company_all	16,089	0.8923	0.3100	0	1
dif_parent_all	16,089	0.8736	0.3323	0	1
same_inventor_all	16,089	0.0623	0.2417	0	1
same_inventor_city_all	16,089	0.0272	0.1627	0	1
same_inventor_compar	16,089	0.0422	0.2012	0	1
same_examiner_all	16,089	0.0697	0.2547	0	1
same_lawyer	1,998	0.1161	0.3204	0	1
same_law_firm	11,974	0.0866	0.2813	0	1
years	16,089	8.7048	6.1553	0	27
dif_technology4 _all	16,089	0.4452	0.4970	0	1

Table 6
Correlation table

	distance	dif_country_all	dif_state_all	dif_county_all	dif_city_all	dif_company_all	dif_parent_all	same_inventor_all	same_inventor_city_all	same_inventor_company_all	same_examiner_all	same_layer_all	same_firm_all	dif_technology4 _all	years
distance	1														
dif_country_all	0.787	1													
	0														
dif_state_all	0.5637	.	1												
	0	1													
dif_county_all	0.4596	.	0.7439	1											
	0	1	0												
dif_city_all	0.2968	0.2655	0.5761	0.7755	1										
	0	0	0	0											
dif_company_all	0.2908	0.2519	0.4526	0.5429	0.5515	1									
	0	0	0	0	0										
dif_parent_all	0.3084	0.2636	0.4859	0.5738	0.5521	0.9133	1								
	0	0	0	0	0	0									
same_inventor_all	-0.2292	-0.2025	-0.4196	-0.5403	-0.6154	-0.4743	-0.4775	1							
	0	0	0	0	0	0	0								
same_inventor_city_all	-0.1613	-0.1351	-0.2884	-0.388	-0.5089	-0.3472	-0.3363	0.6488	1						
	0	0	0	0	0	0	0	0							
same_inventor_company_all	-0.192	-0.1659	-0.3363	-0.4389	-0.5381	-0.6046	-0.5522	0.8147	0.5894	1					
	0	0	0	0	0	0	0	0	0						
same_examiner_all	-0.0296	-0.0145	-0.0264	-0.033	-0.054	-0.0671	-0.0523	0.0677	0.0802	0.0783	1				
	0.0002	0.0655	0.0213	0.004	0	0	0	0	0	0					
same_layer_all	-0.2042	-0.1618	-0.4053	-0.426	-0.4121	-0.4886	-0.4798	0.4435	0.2872	0.3965	0.1082	1			
	0	0	0	0	0	0	0	0	0	0	0				
same_firm_all	-0.262	-0.2313	-0.4257	-0.4945	-0.4792	-0.6004	-0.6255	0.4674	0.3324	0.4496	0.05	0.7049	1		
	0	0	0	0	0	0	0	0	0	0	0	0			
dif_technology4 _all	0.0372	0.0509	0.1277	0.1439	0.134	0.1572	0.1466	-0.1347	-0.113	-0.1487	-0.1405	-0.2471	-0.1464	1	
	0	0	0	0	0	0	0	0	0	0	0	0	0		
years	0.0041	0.0061	0.0278	0.0378	0.0445	0.0584	0.0518	-0.0645	-0.0446	-0.0589	-0.127	-0.0569	-0.0512	0.0878	1
	0.599	0.4417	0.015	0.001	0	0	0	0	0	0	0	0.011	0	0	

Table 7
Results of logit regressions
Coefficients expressed as odd ratios

	(1)	(2)	(3)	(4)	(5)
	All dyads	All dyads	Cited patents have lawyers	Cited patents have lawyers	Cited patents do not have lawyers
dif_company_all	1.18		1.046	1.166	1.521
	[0.174]		[0.752]	[0.224]	[0.327]
dif_parent_all		1.256			
		[0.051]			
same_inventor_all	0.944	0.986	1.028	0.941	1.136
	[0.655]	[0.914]	[0.849]	[0.641]	[0.816]
same_inventorcity_all					
same_inventorcompany_all					
same_examiner_all	1.267	1.265	1.266	1.272	1.219
	[0.041]*	[0.043]*	[0.063]	[0.044]*	[0.656]
distance	1	1	1	1	1
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.002]**
dif_technology4_all	0.7	0.7	0.662	0.686	1.149
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.651]
years	0.924	0.924	0.916	0.924	0.93
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.002]**
same_law_firm_all			0.783		
			[0.115]		
Observations	16,089	16,089	11,970	15,336	753
Log Likelihood	-10385	-10381	-7694	-9873	-486

Robust p values in brackets
* significant at 5%; ** significant at 1%

Table 8

Geographic location

	(1)	(2)	(3)	(4)	(5)
	All dyads	US only	US only	US only	US only
same_inventor_all	0.986	1.191	1.253	1.222	1.143
	[0.914]	[0.263]	[0.155]	[0.202]	[0.422]
same_examiner_all	1.253	1.285	1.283	1.288	1.3
	[0.053]	[0.084]	[0.083]	[0.082]	[0.070]
dif_company_all	1.116	1.226	1.208	1.196	1.338
	[0.365]	[0.229]	[0.264]	[0.303]	[0.070]
years	0.923	0.918	0.918	0.918	0.919
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
dif_technology4_all	0.699	0.74	0.74	0.74	0.742
	[0.000]**	[0.005]**	[0.005]**	[0.005]**	[0.006]**
dif_country_all	1.874				
	[0.000]**				
dif_state_all		1.354			
		[0.010]*			
dif_county_all			1.418		
			[0.007]**		
dif_ea_all				1.404	
				[0.009]**	
dif_city_all					1.175
					[0.263]
Log Likelihood	-10288	-4520	-4522		-4529
Observations	16089	7632	7632	7632	7632

Robust p values in brackets

* significant at 5%; ** significant at 1%

Figure 1
K-density graph for distance in miles
Examiner vs. Inventor citations
Citing and cited patents in Continental USA

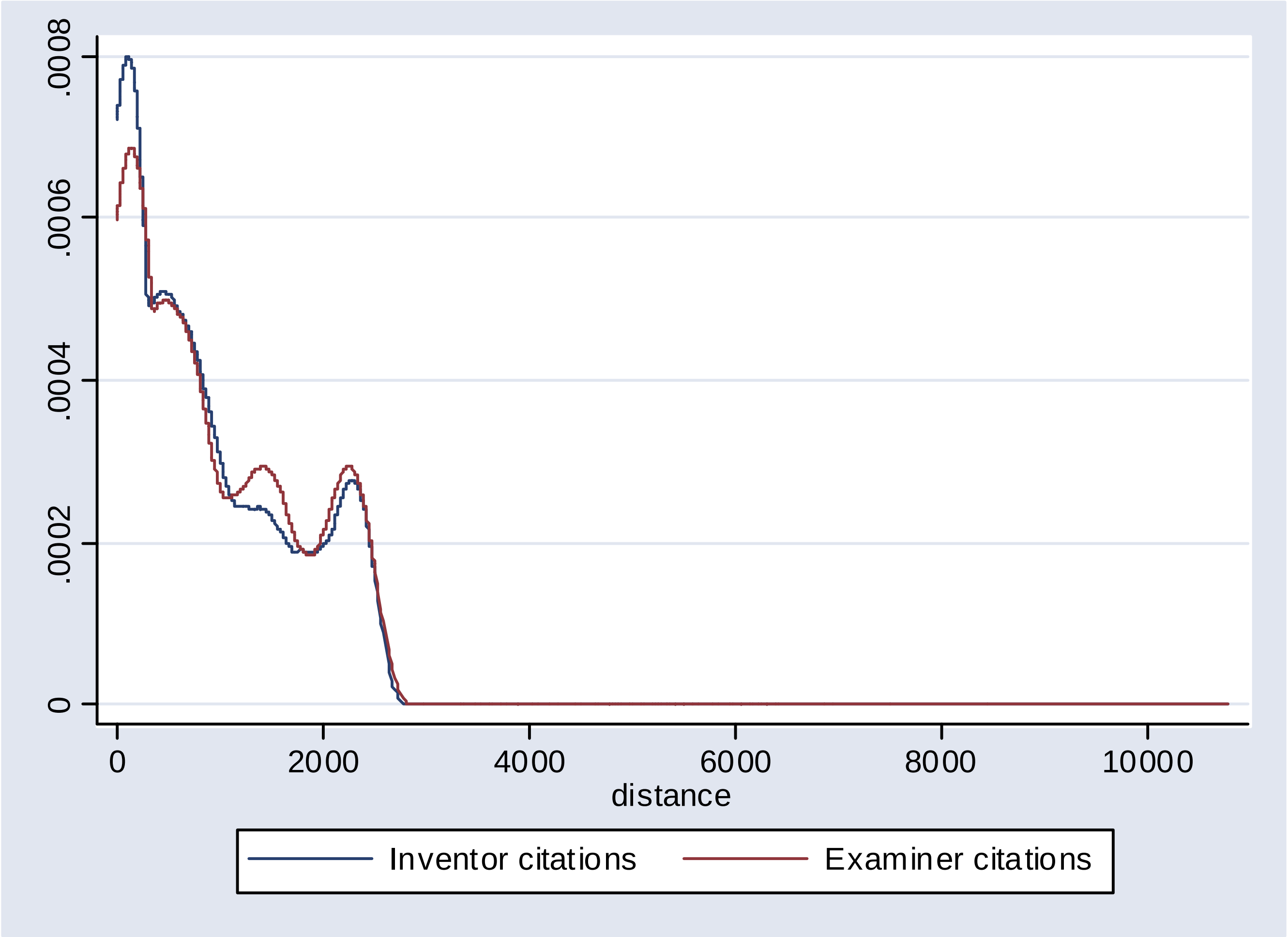

Figure 2

K-density graph for distance in miles

Examiner vs. Inventor citations

Citing patents in Continental USA, foreign cited patent