Financial Patent Quality: Finance Patents After State Street

Citation
Lerner, Josh, Andrew Speen, Mark Baker, and Ann Leamon. "Financial Patent Quality: Finance Patents After State Street." Harvard Business School Working Paper, No. 16-068, December 2015.

HARVARD | BUSINESS | SCHOOL

Financial Patent Quality: Finance Patents After *State Street*

Josh Lerner
Mark Baker
Andrew Speen
Ann Leamon

Working Paper 16-068

Financial Patent Quality: Finance Patents After *State Street*

Josh Lerner
Mark Baker
Andrew Speen
Ann Leamon

Working Paper 16-068

Financial Patent Quality:
Finance Patents After *State Street*

Josh Lerner, Andrew Speen, Mark Baker, Ann Leamon **

November 2015

*** The report was sponsored by Askeladden L.L.C. as part of its Patent Quality Initiative. All errors or omissions are our own.*

Abstract

In the past two decades, patents of inventions related to financial services (“finance patents”), as well as litigation around these patents, have surged. One of the repeated concerns voiced by academics and practitioners alike has been about the quality of these patents in particular, and business method patents more generally. In particular, because so much of the prior work in these areas has not been patented, concerns have been expressed as to the extent to which the awards reflect this knowledge.

Inspired by these issues, this paper empirically examines the quality of finance patents in the years after the landmark litigation between State Street Bank and Signature Financial Group. We show that relative to two sets of comparison groups, finance patents in aggregate cite fewer non-patent publications and especially fewer academic publications. This finding holds across the major assignee groups. In addition, it appears that patents assigned to individuals and associated with non-practicing entities (NPEs) cite less academic work than those assigned to non-NPE corporations. While not statistically significant due to the small number of academic citations in finance patents, we observe qualitatively similar patterns of under-citation when we restrict our analysis to finance patents held by individuals and NPEs, as opposed to non-NPE corporations. These findings raise questions about the quality of finance patents.

We also explore litigated finance patents and discuss how the results here may reflect differences in the quality of finance patents relative to other areas. We find that, as earlier work has suggested, finance patents are more likely to be litigated than non-finance patents, but increased academic citations appear to reduce that possibility relative to others. Collectively, these findings raise important questions about the quality of finance patents and the proliferation of litigation in this domain.

1. Introduction

The patenting of financial techniques surged in the years following the landmark 1998 *State Street* decision by the U.S. Court of Appeals for the Federal Circuit. *State Street Bank v. Signature Financial Group* eliminated the so-called “business method exception,” a strong bias in courts that business methods were inherently un-patentable subject matter. Under *State Street,* the court affirmed the patentability of a piece of financial software that valued mutual funds since it produced a “useful, concrete, and tangible result.”[1] In doing so, *State Street* established

[1] In particular, the court held “…that the transformation of data, representing discrete dollar amounts, by a machine through a series of mathematical calculations into a final share price, constitutes a practical application of a mathematical algorithm, formula, or calculation, because it produces "a useful, concrete and tangible result"--a final share price momentarily fixed for recording and reporting purposes and even accepted and relied upon by regulatory authorities and in subsequent trades.” See *State Street Bank and Trust v. Signature Financial Group,* 927 F. Supp. 502, 38 U.S.P.Q.2d 1530 [D. Mass. 1996]; 149 F.3d 1368 [Fed. Cir. 1998]). We note that while the “business method exception” can be traced back to a 1908 court decision (*Hotel Security Checking Co. v. Lorraine Co.,* 160 F. 467 [2d Cir. 1908]), the USPTO has indeed issued patents for what many consider to be financial service business methods (e.g., U.S. patent 4,346,442 was granted due to the program’s key role in manufacturing (Allen 1984)).

that business methods were statutory subject matter on an equal playing field with more traditional technologies.[2]

Patent protection of business methods, however, did not come without controversy. As summarized in Hunter (2004, Table 1), critics questioned (a) the capabilities of the USPTO to process applications, (b) the validity of such patents in terms of obviousness and novelty, and (c) their overall impact on innovation and competition. Indeed, the validity of business method patents has been debated in subsequent Supreme Court cases. Most recently, in June 2014, the Supreme Court ruled in *Alice Corp. v. CLS Bank* that the patent on Alice's computerized trading program, which mitigated settlement risk and facilitated the exchange of financial obligations, was invalid. The Court found the patent to be merely an abstract idea and thus ineligible for patent protection. While the Court made no categorical rejection of business methods or software, *Alice* has amplified concerns over the extent of financial-related software patentability.[3]

In this report, we empirically evaluate the quality of finance patents that were awarded subsequent to the *State Street* decision.[4] We note at the outset that this report *does not* evaluate the theoretical foundations for business method patentability; rather, we use a number of econometric techniques to compare the characteristics of finance patents relative to non-finance patents.

It is important to distinguish the different ways in which patent quality can be measured. First, patent quality can be broken into two parts: "strength" and "value." Patent *strength* reflects the validity, or defensibility, of a patent. In other words, does the patent sufficiently adhere to prerequisites established under Title 35 of the U.S. Code? Patent *value* can further be broken down into two distinct types: social and private value. Social value reflects the extent of innovativeness and technological significance of the underlying invention. Has the patent disseminated information and/or broken down barriers critical to the creation of other inventions? Has the underlying invention made a significant contribution to social welfare? In contrast, private value reflects the rents that the patent owner receives from the patent. What financial gains can the patent owner reap due to the possession of monopoly rights obtained from

[2] The USPTO defines business methods mainly through patent class 705: "Data Processing: Financial, Business Practice, Management, or Cost/Price Determination." The class definition is as follows:

> This is the generic class for apparatus and corresponding methods for performing data processing operations, in which there is a significant change in the data or for performing calculation operations wherein the apparatus or method is uniquely designed for or utilized in the practice, administration, or management of an enterprise, or in the processing of financial data.
>
> This class also provides for apparatus and corresponding methods for performing data processing or calculating operations in which a charge for goods or services is determined.

For more details, see http://www.uspto.gov/web/offices/ac/ido/oeip/taf/def/705.htm.

[3] Another notable case was *Bilski v. Kappos*. While the *Bilski* decision affirmed a ruling by the Board of Patent Appeal and Interferences that rejected the patentability of a method of hedging against price risk in commodities trading, the Supreme Court also rejected a *per se* exclusion of business methods. In particular, Bilski established that the "useful, concrete, and tangible result" was insufficient to determine patentable processes and thereby raised further uncertainties regarding the validity of business method patents.

[4] For relevant discussions on whether financial patent inhibit competition or incentivize innovation, or how they fit into patentability definitions as described by the 35 U.S.C., see Raskind (1999) and Spulber (2011).

patent protection? What would be the opportunity cost of losing patent protection? We highlight that there are no necessary connections between “strength” and “value.” A patent highly valuable to the patent owner—for instance, one that yields large settlements from other firms—need not possess robust strength. For instance, the fear of an injunction, with its potentially catastrophic effects, may lead a firm to pay a substantial settlement to the holder of a dubious patent.

To explore the quality of finance patents, we focus mainly on the extent to which the patents drew on the prior art. One of the frequent critiques of business method patents is that they often do not fully identify relevant previous inventions, particularly non-patented ones.[5] We seek to understand whether these awards did indeed pay less attention to prior art, particularly non-patented discoveries, than comparable inventions.

We break our analysis down into three parts:

- In Part I, we evaluate financial services patent quality overall, comparing finance patents to two sets of non-finance patents with grant years between 2001 and 2010. We identify 2,799 finance patents from the subclasses used in Lerner (2002): 705/35, 705/36R, 705/37, 705/38, and some patents from 705/4 that deal with finance. We then compare this set of finance patents to samples from two other groups—the overall most popular classes (“Comparison Group 1”) and then those patents at the cutting edge of innovation (“Comparison Group 2”).
- In Part II, we explore financial services patent quality for different assignee types, namely (a) corporations (excluding NPEs) and (b) individuals and NPEs. The methodology is otherwise the same as the Part I methodology described above.
- In Part III, we examine whether patents with more academic citations are more likely to be litigated. We examine the propensity for patents to be litigated as a function of the number of citations to prior non-patent art for finance patents and the comparison group.

In summary, our analysis generates several intriguing results, suggesting that finance patents are problematic with respect to the citation of prior academic research. Overall, the finance patents in our study cite less non-patent prior art of any type—even though examiners add more of these citations to finance patents. Finance patents tend to cite fewer academic publications than do non-finance patents, and this difference grows as we consider the leading journals. Even after controlling for a number of variables and separating patents into assignee groups, finance patents still appear weaker across the board with respect to non-patent prior art (and especially academic citations) relative to the comparison groups. These findings raise substantial questions regarding finance patents.

Moreover, this effect is not observed evenly across awardees. Patents awarded to corporations generally cite more prior art, particularly academic research; while those awarded to individuals and associated with non-practicing entities (NPEs) generally have fewer academic citations.

A natural question is whether these differences in academic citations really matter. In the final analysis, we examine the litigation of finance patents relative to the matching groups. As shown

[5] For relevant discussions, see Merges (1999, 589-90), Raskind (1999, 85), and Thomas (2001, 318-19)

in earlier studies, the finance patents are litigated more often and more intensely, a fact that has drawn concern from many observers. We find that for finance patents—but not for other fields—academic citations are critically related to litigation. Finance patents with more academic citations had less litigation. The same effect was not seen for non-finance patents. The results suggest that the absence of links to academic knowledge—a problem that is particularly dramatic for finance patents, especially those awarded to individuals and associated with NPEs—is directly associated with the proliferation of litigation in this area.

The remainder of this paper is structured as follows. Section 3 describes the construction of the dataset. Section 4 briefly describes the main variables we use in our analyses. We explain the rationale behind these variables and synthesize the literature evaluating their abilities to measure patent qualities. Section 5 gives an overview of our dataset with descriptive statistics. Section 6 examines the quality of finance patents relative to selected comparison groups of non-finance patents, which are divided into the three parts described above. Section 7 includes a number of robustness tests. Section 8 explores implications. Finally, in Section 9, we include three appendices. Appendix A offers highly detailed information on the programmatic methods employed and assumptions made to construct our dataset. Appendix B provides additional information on variables used throughout the analysis. Appendix C describes data limitations to consider when interpreting our findings.

3. Constructing the Dataset

This section describes how we constructed the dataset used to investigate our hypotheses.

Patent Awards. We identify patent awards from publicly available patent data assembled by Ronald Lai, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming in Harvard's Patent Network Dataverse ("Lai et al. dataset").[6] Lai et al. primarily aggregate their data from two primary sources. First, they incorporate the National Bureau of Economic Research's U.S. Patent Citations Data File.[7] This database has been extensively used in patent literature in its own right and includes, in its original version, comprehensive data on utility patents and citations from 1975 to 1999.[8] Next, they incorporate patent data from the U.S. Patent and Trademark Office's (USPTO) weekly publications. Importantly, the dataset "disambiguates" inventor and assignee

[6] Ronald Lai, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming, 2011, "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)", http://hdl.handle.net/1902.1/15705, Harvard Dataverse, V5.

[7] Many patent characteristics, such as citations made to these patents and the number of claims, start as of 1975. For a description of these data, see B. H. Hall, A. B. Jaffe, and M. Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498, 2001.

[8] For a partial list of studies, see Iain Cockburn, Bronwyn H. Hall, Woody Powell, and Manuel Trajtenberg, "Patent Data Project – NSF Proposal," February 2005. The NBER patent database includes U.S. patents granted from 1963 and citations made to these patents from 1975.

names to ensure consistency across the universe of patents, as the USPTO does not require unique identifiers for either inventors or assignees.[9]

Given the proliferation of finance patents after *State Street*, we exclusively look at patents in the post- *State Street* era from 2001 to 2010. We note that there are, in essence, three eras of finance patenting: pre-Second Pair of Eyes Review (SPER), post-SPER up to *Alice*, and post-*Alice*. Given major changes in the examination process for finance patents that started in March 2000 under SPER, as well as the availability of "examiner-added" data starting in January 2001, we begin our dataset at 2001. While we do not capture awards between 2011 and 2014 due to data limitations, we cover most of the second era. Pre-2001 finance patents may not accurately reflect the current state of quality in these patents, and in fact, earlier studies have shown them to be problematic in quality (Lerner 2002). Our data extends to 2010, which is the last year of available data in the Lai et al. dataset, the most contemporaneous database available for academic patent research.

Characteristics of Assignees and Inventors. The Lai et al. database contains key information on patentees, including assignee names and types (such as U.S./foreign corporations or U.S./foreign individuals) and inventor nationality. We employ assignee and inventor data throughout to ensure that our finance and non-finance samples contain equal distributions of different assignees.

Features of Patents. We compile patent characteristics using both the Lai et al. dataset and Google Patents. In particular, we collect counts of citations made (backward citations), citations received (forward citations), and claims from the Lai et al. dataset. We supplement the Lai et al. dataset with information from Google Patents, which contains the full text of all USPTO patents.[10] Using Google Patents, we collect counts for the number of non-patent and examiner-added citations for all patents in our dataset. We supplement this information with the journal database SCImago Journal and Country Rank[11] in order to determine counts for the subset of non-patent citations to academic literature. In addition to counts of references to academic literature overall, we also explore citations to *leading* academic journals. We accomplish this by employing SCImago's ranking system, a measure of a journal's "impact, influence, or prestige," to determine which journals are in the top 1, 10, and 25 percentiles within their respective disciplines as of 2010. For more details on this process and the programmatic methods used to conduct this search, see Appendix A.

---

[9] For more information on the disambiguation process, see Guan-Cheng Li, Ronald Lai, Alexander D'Amour, David M. Doolin, Ye Sune,Vetle I. Torvik, Amy Z. Yu, Lee Fleming, "Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010)," *Research Policy* 43, 2014: 941-955.

[10] The Google Patents database contains all patent applications and grants from the USPTO, European Patent Office (EPO), World Intellectual Property Organization, German Patent and Trade Mark Office (DPMA), Canadian Intellectual Property Office (CIPO) and the Chinese Patent Office (SIPO).

[11] For a review of SCImago and a comparison to Web of Science, see Borja González-Pereiraa (2010), Falagas et al. (2008), and Barnett and Lascar (2012). In addition, Jorge Mañana-Rodríguez (2015) offers a critical analysis of the SCIMago database. We find that SCImago's wide coverage and ranking system offer a suitable source for patent citation data. We also note SCImago's use in a number of peer-reviewed publications (e.g., Chinchilla-Rodríguez 2015; Cimini et al. 2014).

Litigation. We employ the Derwent LitAlert patent litigation database (via WestLaw) to identify the occurrence and intensity of litigation for each patent in our full samples. Litigated patents include patents involved in any type of suit from 2001 until 2010. Intensity is defined very broadly and is a composite measure of the total number of actions (e.g., complaints filed, memorandum opinions, final orders, etc.), which can come from a single or multiple suits.

Below, we specify each part of our analysis.

Part I

We first identified all finance patents in Lai et al.'s dataset. We then created the two comparison sets ("Comparison Group 1" and "Comparison Group 2"). Comparison Group 1 is composed of 2,799 patents from the five most popular three-digit patent classes in the Lai et al. database, which runs from 1975-2010. Comparison Group 2 is composed of 2,799 patents from the five most popular three-digit classes in the post-*State Street* era (1998-2010) among universities with heavy R&D spending. This set proxies for patents that are on the frontier of technological and scientific development.

To construct these comparison groups, we match each finance patent (n=2,799) to a randomly selected non-finance patent with the same assignee type and grant year. We did this for each comparison set separately, which results in the two samples of 2,799 patents each that we use in the analysis. Both groups include class 514 (Drug, Bio-Affecting and Body Treating Compositions).

In accordance with Lerner (2002), finance patents include the following classes:

- 705/35: Finance (e.g., banking, investment or credit);
- 705/36R[12]: Portfolio selection, planning or analysis;
- 705/37: Trading, matching, or bidding;
- 705/38: Credit (risk) processing or loan processing (e.g., mortgage).
- *Some patents within* 705/4: Insurance (e.g., computer implemented system or method for writing insurance policy, processing insurance claims, etc.).[13]

We compare finance patent characteristics (focusing on prior art references) with two sets of comparison groups of leading patent classes. Non-finance patents in our comparison groups whose prior art references we examine include:

[12] Class 705/36R refers to Class 705/36 described in Lerner (2002), namely, "Portfolio selection, planning or analysis." We add this class to distinguish from class 705/36T, which is titled "Tax strategies."

[13] These patents were determined in the following way. First, we pulled in the "Patent Description" file from the Lai et al. dataset. We then reviewed a number of patents within the 705/4 category that were indeed finance focused to generate a list of keywords to search this file. We identified five patents that were related to each of the following topics, which were also used in Lerner (2002): (a) the calculation of annuity rates, (b) the investment of insurance company assets, and (c) the management of risk through related topics. After reviewing these patents, we generated a list that has both "positive" keywords and negative keywords (namely, EFT, ATM, and remote). We exclude all patents with either no matches or at least one negative match. Finally, all patents with one or more positive matches are integrated into our main dataset.

Table 1: Patent Classes Contained in Comparison Groups[14]

Comparison Group 1*		Comparison Group 2**	
Class	**Title**	**Class**	**Title**
257	Active Solid-State Devices (e.g., Transistors, Solid-State Diodes)	424	Drug, Bio-Affecting and Body Treating Compositions[15]
428	Stock Material or Misc. Articles	435	Chemistry: Molecular Biology and Microbiology
435	Chemistry, Molecular Bio and Biology	514	Drug, Bio-Affecting and Body Treating Compositions
438	Semiconductor Device Manufacturing Process	530	Chemistry: Natural Resins or Derivatives; Peptides or Proteins; Lignins or Reaction Products Thereof
514	Drug, Bio-Affecting and Body Treating Compositions	600	Surgery
***** Representative sample of patent classes overall. These classes are the five most popular in the Lai et al. dataset across all years (1975-2010).		** Representative sample of cutting-edge technologies. These classes are the five most popular from the post *State Street* era (1998-2010) among R&D intensive universities.[16]	

We take the sample of 2,799 finance patents and match each patent to one in each comparison group with the same grant year and assignee type.[17] Matching datasets have been used in dozens of patent analyses. This approach is well-regarded in the literature as a way to help control for "artificial" differences—such as temporal variations in patenting trends or general differences in patenting styles among different assignees—that could distort the findings (see, e.g., Lanjouw and Schankerman 2002; Lerner 2002).

[14] We note that while there is an overlap in classes (namely 435 and 514), the actual patents composing our samples are unique to the sample (i.e., no patent is included in both samples).

[15] Patent class 424 was split into two classes, class 424 and class 514, in the mid-1980s (Narin and Olivastro 1988, 471). For a more technical description of the types of formulations/compositions that fall into the different classes, see http://pharmexcil.org/uploadfile/ufiles/1899013087_Current_Status_of_Pharmaceutical_Patenting_at_USPTO.pdf.

[16] "R&D intensive universities" include the top 20 university in terms of total R&D expenditure from 2004-2009. To align with assignee data, we aggregate all University of California campuses to one entity. We exclude 2010 data due to changes in collection methodologies starting this year. For data and discussion of methodology changes starting FY 2010, see National Science Foundation, National Center for Science and Engineering Statistics, Higher Education R&D Survey, at http://ncsesdata.nsf.gov/herd/2013/html/HERD2013_DST_17.html.

[17] Assignee types include: U.S. company/corporation, foreign company/corporation, U.S. individual, foreign individual, U.S. federal government, foreign government, U.S. county government, U.S. state government. If the assignee type was missing or listed as 0, we checked the "country" variable to see the inventor's country of origin. As discussed in the technical appendix, we have reason to believe these patents are actually assigned to their inventors, so if the inventor was listed as American we recoded the patent as "U.S. individual." Otherwise, it was recoded as "foreign individual."

Part II

In Part II, we analyze the characteristics of finance patents across sub-populations of assignees. First, we break down our data from Part I by assignee categories. Because prior research has identified that a high frequency of NPE activity is related to the litigation of financial services patents, we pay special attention to the subset of assignees we can identify as NPEs. We analyze patents either (a) assigned to NPEs or (b) re-assigned to NPEs after issue, as well as those assigned to individuals.[18] To identify re-assignments to NPEs we used a computer program to "match" names contained in a list of over 3,000 uniquely identified NPEs to re-assignees for each patent as reported by in Google Patents. For more details, see Appendix A.

Part III

To assess the quality of litigated patents, we identify finance patents within our sample that have been litigated through August 2015. In addition to *whether* a patent has been litigated, we also provide a rough measure of the intensity of litigation. Intensity is defined very broadly as a composite measure of the total number of actions (e.g., complaints filed, memorandum opinions, final orders, etc.) which can come from a single or multiple suits, as recorded in the LitAlert database. We then examine whether litigated patents exhibit differences in non-patent prior art, and whether the relationship between litigation and quality holds equally in our finance and non-finance samples.

A number of data limitations related to this methodology are described in Appendix C.

4. Variable Descriptions

In this section, we describe in more detail the primary "quality" variables we will use to investigate the three parts of our analysis. To measure quality, we examine a number of patent characteristics that have been used in prior academic literature. For a more expansive list and discussion of variables used in this analysis, see Appendix B.

1. Non-patent prior art references

There are two different types of citations contained in patent documents: Citations to other patents and citations to non-patent literature (for example, academic publications, trade journals, software documentation, and company documents). Both types of citations facilitate the USPTO's Patentability Search, which determines whether the patent meets the requirements of novelty and non-obviousness. Prior art references are therefore a key indicator of patent validity in economics literature. A lack of cited prior art in patents, as has been frequently suggested, increases the probability that the discovery of uncited prior art would render a patent invalid (e.g., Allison and Tiller 2003, 1037).

[18] NPEs include all unique entities identified from two sources: PlainSite and IPCheckups NPE Tracker List (http://www.ipcheckups.com/npe-tracker/npe-tracker-list/) In the IPCheckups list, we exclude Academic/Research Entities, as well as Allied Security Trust (a defensive patent aggregator) from our list. A few entities were deleted because they had very generic names and were therefore likely to create false matches.

Consistent with much patent literature, we distinguish between patent and non-patent prior art references in our analysis. The number of *non-patent* citations (e.g., academic and trade journals, company publications, government reports, software documentation) is often used as indicator of a patent's quality. Non-patent prior art has specifically been found to reveal the "proximity" of the invention to "technological and scientific developments" (Callaert et al. 2006).[19] Indeed, the OECD Patent Statistics Manual (2009, 116) states that "[t]he more scientific references [that] are found in patents, the closer the technology is considered to be to basic research [i.e., pure, or fundamental research as opposed to applied science]." Many authors, however, have noted that all types of non-patent prior art are not of equal quality (Allison and Tiller 2003, Allison and Hunter 2006). Allison and Hunter (2006, 742) note that academic publications are likely to be the "most objective and reliable [references]". In addition, some non-patent publications, such as those in peer-reviewed academic journals, are likely to be more "cutting-edge," while others—such as popular press—may be less indicative of scientific developments. As a result, we are particularly interested in academic references, as they are likely to provide the clearest signal of innovations at the frontier of research.

Some researchers (e.g., Cohen 1995, 1178) argue that scholarly publications apply less to many new developments in computer programming—a key component of finance patents—since the knowledge is often embedded in alternative sources, or have more to do with the nature of the technology itself (Duffy and Squires 2008). Others (e.g., Lerner 2002), however, find substantial reason to think differences arise more from reasons inherent to the patent examiner system, such as examiner familiarity with subject matter and the challenges associated with searching the non-patent prior art. We therefore look closely at academic citations in two ways. First, we count the total number of academic references for each patent. To do so, we start by including any reference to an article in the SCImago Journal & Country Ranking. This index contains more than 19,000 journals that are included in the Scopus database of literature.[20] Because applicants often reference journals not by their full tiles, but by an abbreviated version of the title, we tracked common forms of abbreviations using a list from Web of Science that we supplemented by other sources such as the National Library of Medicine (NLM) catalog.[21] We also look specifically at academic references to leading journals to examine patents at the forefront of academic research. We accomplish this by employing SCImago's ranking system, a measure of a journal's "impact, influence, or prestige," to determine which journals are in the top 1, 10, and 25 percentiles within their respective disciplines as of 2010.[22] We note that because we aimed to

[19] For a relevant discussion, see also Roach and Cohen (2013).

[20] Given its prevalence in scientific patents, we added "Chemical Abstracts" to our list of overall academic journals. Chemical Abstracts is a periodical index that summarizes recently published scientific documents.

[21] While at least one common abbreviated version of each journal was included in our list for nearly all (96%) journals contained in the top 25%, roughly 44% of journals outside of this range included an abbreviated version. As a result, the *proportion* of top 25% journals contained in academic prior art references is likely over-estimated.

[22] We look at percentile rankings *within* each of the 27 subject areas available on SCImago. As a result, the *number* of journals considered "leading" within each discipline varies. For example, there are 716 journals listed under the "Chemistry" subject area and 8 journals with an SJR ranking at the 99th percentile or greater. By way of comparison, there are 686 journals in the "Economics, Econometrics and Finance" subject category and 7 journals with an SJR ranking at the 99th percentile or greater. If we had used a pre-set number of journals within each ranking, the journals' relative prestige would not have been on an equal playing field. For journals that appear in more than one

limit "false matches" between journal titles in our journal list and *similar* (but not identical) titles in Google patents, our estimates are likely on the conservative side, as we no doubt miss certain academic references that do not meet our criteria. For comparative purposes, however, this methodology provides the fairest look at the data as "missed" references should not exhibit a bias to our finance or non-finance samples.

For the purposes of this paper, we take a higher number of academic references, and especially academic references to leading journals, to reflect technologies (a) on the scientific and technological frontier and (b) of higher quality with respect to their examination process and validity.

2. Examiner-added citations to academic publications

There are two primary routes for a citation to appear on a patent. The first way is for an applicant or his/her lawyer to add the citation out of a "duty of candor" to reveal relevant subject matter. Examiners, however, are officially responsible for the identification of prior art that ought to be included in the list (Alcacer et al. 2009). These citations are extremely important, as the invention must be novel and non-obvious in light of the state of current knowledge.

The typical background of patent examiners, however, may be better suited to some patent classes than others—for instance, an examiner with an advanced degree in materials science would likely be more acquainted with prior art in that field. This situation may not be the case for finance patents. Lerner (2002) found that examiners of finance patents reviewed fewer patents overall and fewer patents in the same patent class than those in other categories, indicating less experience in the field. Moreover, they were less likely to have a doctorate in a related field, further reducing the likelihood that they would be able to supply prior art references that had not been included by the applicant. This finding suggests that systematic differences exist across patent classes with respect to the quality of examiner reviews.[23] To better understand this knowledge gap, we look at differences between finance and non-finance patents in the number of examiner-added academic citations. [24]

subject area in 2010, we define its percentile as the top percentile among all categories in which the journal appears. We also note that a small percentage of "journals" in the SCImago database are not "academic," such as *Federal Register* (a journal of the U.S. government). We exclude the following newspaper/magazine publications, which are considered "journals" in SCImago and could bias results, but are not academic journals: *New York Times; New Yorker*; *Washington Post*; *Newsweek*; and *The Economist*.

[23] Lerner (2002) offers two illustrative examples of patents granted in 1999 to demonstrate obvious omissions of relevant finance academic publications from prior art analyses that appear to be "obvious" extensions of the literature.

[24] A natural question is whether examiners typically play an important role in non-patent prior art searches. Research suggests that examiners are far less adept in identifying non-patent prior art relative to patent prior art. Lemley and Sampat (2012, 820), for example, found that examiners accounted for under 10 percent of citations to non-patent prior art in their issued patents, but provided over 40 percent of the citations for patent prior art. This discrepancy, especially with respect to the recently established patentability status of business methods in the State Street decision, is likely due to the far less formalized approach that must be taken to identify references in printed publications relative to patents (Thomas 2001, 319). Cotropia et al. (2013) further found that examiner-added citations are overwhelmingly used in rejections to narrow claims. That is, examiners tend to rely mainly on the prior

For the purposes of this paper, we interpret a difference in examiner-added prior art citations to academic publications as consistent with weaker patents in terms of defensibility.

3. Other patent measures

To put the above measures in perspective and give a more complete discussion, we also gathered a number of other measures. These include: the number of backward U.S. patent citations; the number of claims in the patent; the number of forward citations from U.S. patents; and a measure of "patent originality." For a more thorough examination of these variables and how they have been used in economics literature, see Appendix B.

We use these same measures in our Part 2 analysis, but examine quality in our finance and non-finance datasets within the two groups of our dataset: (1) corporations and (2) individuals and non-practicing entities (NPEs).

4. Litigation

There is an established literature lending empirical support to the claim that litigated patents, in aggregate, are generally more valuable. The litigation-value connection is relatively straightforward; the very fact that a party is willing to engage in litigation suggests that the expected value of the patent in dispute is greater than the expected cost of litigation. Given that nearly two-thirds of patents expire prematurely because the owners do not pay the maintenance fees required by the USPTO, which are many magnitudes lower than litigation-related fees, litigated patents theoretically represent a subset of the most valuable patents. Research suggests that overall patent litigation rates are around 1%-2% (Lanjouw and Schankerman 2001, 131-35; Lemley 2001, 12).

A large amount of empirical literature supports this theory (Allison et al. 2004; Allison et al., 2009; Ball and Kesan 2009; Bessen 2008; Lanjouw and Schankerman 1997). It is important to emphasize, however, that this literature connects litigated patents to expected *private value*, and generally does not make assertions about the social value of an invention. Further, litigated patents may not be a *representative* subset of valuable patents. Patents whose value is driven by licensing fees, defensive use, and signaling, for example, may largely be omitted from litigation data (Moore 2005, 1548).

Industry-specific effects further cloud how litigation rates ought to be interpreted across classes. As explained by Lerner (2010, 811), the probability of resorting to litigation to resolve legal harms rises as a result of four considerations: "(1) the likelihood that the offense is detected by the potential plaintiff, (2) the size of the stakes under dispute, (3) the uncertainty about the outcome of the controversy between the two parties, and (4) the cost of settlement relative to that of trial." As a result, while the litigation-value connection may hold for patents in aggregate, it may not hold for specific industries. This is particularly true for finance patents, given their relative immaturity. The litigation rates of finance-related patents may, for example, reflect the

---

art that they themselves identify when determining validity. Taken together, these facts suggest that finance patents may be fundamentally weaker due to the examiners' lack of knowledge regarding the relevant academic literature.

high levels of uncertainty about the scope of patent protection following the *State Street* decision, and the rents to be garnered from initiating sham litigation.

For the purposes of this paper, we interpret a difference in quality measures among litigated and non-litigated patents as reflecting differences regarding the nature of litigation itself. In other words, if litigated finance patents are not of higher quality, litigation may be more about rent-seeking or overall uncertainty than about innovation.

A summary of our analysis and the associated economic metrics in the above discussion is presented in Table 1.

Table 1: Summary of Key Variables

Variable	Used in which parts?	Source(s)	Finance/Business Method-Related Reference Material	Primary Interpretation
Backward Non-Patent Citations	I, II, and III	Google Patents	Allison and Tiller (2003); Hunter (2004); Allison and Hunter (2006)	Validity/innovativeness, with significant heterogeneity among publication types.
Backward Non-Patent Citations to *Academic Journals*	I, II, and III	Google Patents; SCImago	Lerner (2002) Duffy and Squires (2008)	Validity/Innovativeness, with possibility that certain industries have less academic material to cite. May also reflect nature of underlying technologies.
Backward Non-Patent Citations to *Leading* Academic Journals.	I, II, and III	Google Patents; SCImago	Lerner (2002) Duffy and Squires (2008)	Same as above
Examiner Added Citations to Academic Journals	I, II, and III	Google Patents; SCImago	Lerner (2002)	Validity/quality of examination, with possibility that certain industries have less academic material to cite.
Litigation (Binary)	III	Derwent LitAlert (via Westlaw)	Lerner (2010)	A discrepancy in relationship between litigated patents and value indicators would reflect difference in nature of litigated patents (e.g., rent-seeking or innovation).
Litigation (Intensity)	III	Derwent LitAlert (via Westlaw)	Lerner (2010) Allison et al.(2009)	Same as above

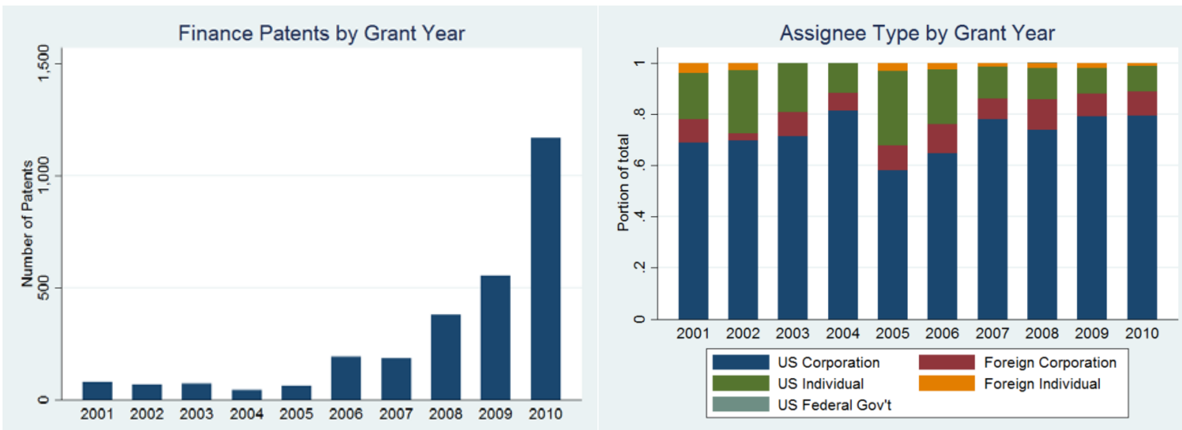
5. Summary Statistics

Figure 1.A provides an overview of the distribution of patent counts in our database. Our finance patent sample is skewed toward the later years. In Figure 1.B, we show the distribution of assignee types.[25] Figure 1.B is consistent with data presented in Lerner (2010, Table 1).

Figure 1: Finance Patents Dataset

A. Finance Patent Count by Grant Year

B. Assignee Types by Grant Year



Note: The finance sample consists of 2,799 finance patents issued between 2001 and 2010. Not all patents are explicitly assigned; a substantial number within our sample period were either missing a value or labeled as "type 0," which is not found in the codebook accompanying the data. Following extensive investigation of the data (see technical appendix), we recoded type 0 and valueless patents to be assigned to individuals. If the variable *country* was present and equal to "United States," they were considered American individuals. Otherwise, they are coded as being foreign.

6. Results

Our analysis proceeds in three parts. In Part I we look at our measures of patent quality among finance and non-finance classes in aggregate. We focus mostly on the state of non-patent academic prior art to determine systematic differences between finance and non-finance patents in aggregate. Due to the possibility that different assignee types exhibit different patenting behaviors, Part II examines patenting behavior *within* two main groups: (1) non-NPE corporations and (2) individuals/NPEs. This gives a more refined look at finance patent quality and can identify if patent quality varies among assignees of finance patents relative to our non-finance samples. Finally, Part III examines the quality of litigated finance and non-finance patents. Variation in patent quality among litigated patents may reflect fundamental differences in the nature of such litigation.

---

[25] Certain patents in our dataset had an "assignee type" labelled as 0 (starting at 2005). Upon inspection it appears that the majority of these patents were assigned to their original inventors. We used this fact and the data we had on the inventor's country of origin to classify these ambiguous assignees as either type 4 or 5, respectively American and foreign individuals. In addition, a number of patents in our dataset were missing data for assignee type, assignee name, and country. We believe these are overwhelmingly awards to individuals, and code them accordingly.

Part I

We first compare summary statistics for each of our indicator variables. We organize the variables in three categories: overall prior art, academic prior art (a subset of all non-patent prior art citations), and other. The results are summarized below.

Prior Art – Overall

These results look at the state of prior art for finance patents. Table 1 reports that relative to both comparison groups, finance patents cite fewer non-patent publications and roughly equal numbers of patents on average. Normalizing by number of claims, backward citation counts favor comparison groups for both types of prior art. Interestingly, finance patents, on average, have more examiner-added citations for both types of prior art. That is, while finance patents cite fewer patents and less non-patent literature, examiners have added more of these citations to these patents.

Prior Art – Academic Publications

We also consider the state of academic prior art and examiner additions in this domain. Interestingly, Table 1 shows that relative to Comparison Groups 1 and 2, finance patents cite substantially fewer academic publications, which, as shown in Lerner (2002), can reflect the quality of the patent both in terms of validity and technological significance. In fact, the ratio of academic citations between non-finance and finance patents becomes higher when we compare citations to leading publications versus overall publications, and markedly so for the most innovative non-finance patents. In other words, finance patents tend to cite fewer academic publications than non-finance patents, and this difference generally grows when we move to the leading journals. The same trend generally holds for the number of examiner-added citations to academic publications.

Other Variables

Finally, we assess patent originality, claims per patent, and forward citations per patent. Table 1 shows that average originality ratings are higher for the non-finance samples. This suggests that finance patents draw from a narrower field of technologies. The average finance patent has more claims than the average non-finance patent in comparison groups. As noted in Appendix B, while the difference in claims is interesting, further investigation is required—perhaps looking at the nature of the claims themselves—to understand what exactly is driving this difference. Strikingly, the average finance patent receives considerably more forward citations than comparison groups, which may reflect a field effect: the thin pool of patent awards would lead to many citations to earlier awards almost automatically.

Table 1: Finance versus Non-Finance Patent Characteristics Overall

This finance sample consists of 2,799 patents awarded by the USPTO between 2001 and 2010. We include awards assigned to any type of entity: corporations, individuals, and governments. Comparison Group 1 is composed of 2,799 patents from the top five classes overall in the Lai et al. dataset (1975-2010). Comparison Group 2 is composed of 2,799 patents from the top five classes registered by R&D-intensive universities in the post-*State Street* era (1998-2010). Comparison groups are matched by assignee type and grant year to the finance sample.

	Mean			**Standard Deviation**		
Prior Art – Overall	Finance	Non-Finance Group 1	Non-Finance Group 2	Finance	Non-Finance Group 1	Non-Finance Group 2
Backward patent citations (U.S. only)	36.86	32.83**	41.06*	61.37	58.77	105.5
Backward citations per claim (U.S. only)	2.44	2.66	3.55***	6.20	6.058	10.90
Backward non-patent citations	16.92	19.35**	33.94***	41.45	45.28	61.49
Backward non-patent citations per claim	1.01	1.87***	3.61***	2.89	6.161	10.52
Examiner added patent citations (U.S. only)	7.32	3.90***	2.88***	6.82	5.130	4.17
Examiner added non-patent citations	2.49	0.96***	2.11***	3.94	2.717	4.89
Prior Art – Academic Publications						
Non-patent academic citations - **Overall**	0.36	4.58***	8.57***	1.29	24.71	22.56
Non-patent academic examiner-added citations - **Overall**	0.09	0.24***	0.62***	0.37	1.14	2.05
Non-patent academic citations - **Top 25%**	0.21	4.00***	7.41***	1.03	21.15	19.71
Non-patent academic examiner-added citations - **Top 25%**	0.04	0.20***	0.53***	0.25	1.02	1.82
Non-patent academic citations - **Top 10%**	0.13	3.08***	5.76***	0.68	15.94	15.33
Non-patent academic examiner-added citations - **Top 10%**	0.02	0.15***	0.40***	0.16	0.75	1.44
Non-patent academic citations - **Top 1%**	0.06	1.01***	2.04***	0.37	4.57	6.08
Non-patent academic examiner-added citations - **Top 1%**	0.01	0.04***	0.13***	0.11	0.29	0.58
Other Variables						
Originality	0.27	0.47***	0.43***	0.26	0.28	0.28
Claims	25.41	18.46***	18.68***	22.06	15.37	16.31
Number of forward cites (U.S. only; patents granted up to 2004; n = **262** per group)	26.51	6.56***	5.25***	30.43	14.72	10.34
Observations	2,799	2,799	2,799	2,799	2,799	2,799

Mean significance at: * 10% level; ** 5% level, *** 1% level
Mean significance tested by two-sample t test assuming unequal variance (two-tailed)
Significance marked in non-finance portion only, but is relative to the finance patents

Next, we built a regression model to provide a more robust analysis. Table 2 controls for a number of variables that may confound the results. In our regressions, we look at patent quality in three respects to give a broad overview of the state of non-patent prior art: non-patent citations overall; non-patent citations to academic journals; and non-patent citations to academic journals ranked in the top 10% of their respective fields (i.e., leading journals).[26] For our explanatory (independent) variables, in addition to looking at whether the patents fall into a finance or non-finance class, we add a number of controls to adjust for "artificial" effects stemming from characteristics extraneous to our analysis. In the first specification, we control for inventor nationality, grant years, and an interaction between grant year and assignee type. These variables control for the possibility that inventors of various origins might have different propensities to cite prior art, as well as for different patterns in citing patents across time for different assignee types. In a second specification, we a control for the total number of claims in the patent, which accounts for the possibility that patents with more claims may cite more prior art.

For each of these regression specifications, we run two different types of regressions. First, we run a standard ordinary least squares (OLS) regression. Second, we run negative binomial regressions, which are better able to "fit" the skewed nature of the variable we are explaining and thus provide a clearer measure of the relationship between our dependent and independent variables. In other words, because of the ordinal, nonnegative, and "over-dispersed" (i.e., a variance that exceeds the mean) nature of the count of prior art citations, negative binomial regressions can provide more accurate models relative to OLS.[27] In Panel A, we look at finance patents relative to a sample of non-finance patents in the top five classes overall from 1975-2010 in the Lai et al. dataset ("Comparison Group 1"). Panel B replicates these regressions for classes popular among R&D intensive universities ("Comparison Group 2").

Table 2 is consistent with our summary statistics from Table 1. In particular, even after controlling for other patent characteristics that may play into the non-patent prior art citations, we find that finance patents cite less non-patent prior art. These results are statistically significant for all specifications at the 5% level, and usually at the 1% level.

---

[26] Journal rankings are calculated using the SJR rating from SCImago Journal and Country Ranks database of journals, which measures a journal's "impact, influence or prestige." For more details, see http://www.scimagojr.com/files/SJR2.pdf.

[27] Negative binomial regressions are typically used instead of Poisson regressions when the data for the dependent variables are "over-dispersed," or where the conditional variance is larger than the conditional mean. Negative binomial regressions are common in patent literature, see, e.g., Lerner (2010) and Hunter (2004).

Table 2: The State of Non-Patent Prior Art in Finance and Non-Finance Patents

The finance sample consists of 2,799 patents awarded by the USPTO between 2001 and 2010 in classes 705/35, 705/36R, 705/37, 705/38 and part of 705/4. See Table 1 for details. Each dependent variable looks at a different type of non-patent prior art: all types (A); all academic (B); and leading academic (C), defined as non-patent citations to journals ranked in the top 10% of their respective disciplines from SCImago Journal & Country Rank. Independent variables include various controls: Year*Assignee fixed effects to control for differences in assignee patenting behavior among the different grant years; grant year fixed effects to control for differences in patenting behavior across years; foreign inventor effects (U.S., Japan, and EPO); and, in Specification (2), number of claims. Within each column, we run two different types of regressions: OLS and negative binomial.

	(Column A) Dependent Variable: Non-Patent Prior Art – *All Types*				**(Column B) Dependent Variable: Non-Patent Prior Art – *Academic Journals***				**(Column C) Dependent Variable: Non-Patent Prior Art – *Top 10% of Academic Journals***			
	OLS Specification		Negative Binomial		OLS		Negative Binomial		OLS Specification		Negative Binomial	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Panel A: Finance versus Comparison Group 1												
Finance	-2.79 (1.16)**	-4.28 (1.20)***	-0.13 (0.05)**	-0.19 (0.05)***	-4.32 (0.48)***	-4.39 (0.49)***	-2.48 (0.10)***	-2.52 (0.10)***	-3.02 (0.31)***	-3.05 (0.31)***	-3.21 (0.12)***	-3.27 (0.12)***
Grant Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.03	0.04	-	-	0.02	0.02	-	-	0.02	0.02	-	-
Pseudo R^2	-	-	0.01	0.01	-	-	0.06	0.06	-	-	0.08	0.08
Observations	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598
Panel B: Finance versus Comparison Group 2												
Finance	-17.31 (1.39)***	-18.87 (1.43)***	-0.70 (0..05)***	-0.76 (0.05)***	-8.31 (0.43)***	-8.40 (0.44)***	-3.15 (0.08)***	-3.18 (0.08)***	-5.71 (0.29)***	-5.77 (0.30)***	-3.85 (0.10)***	-3.89 (0.10)***
Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.06	0.07	-	-	0.08	0.08	-	-	0.08	0.08	-	-
Pseudo R^2	-	-	0.02	0.02	-	-	0.09	0.09	-	-	0.11	0.11
Observations	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598	5,598

Notes: * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level. Robust standard errors in parentheses. We report R^2 for OLS regressions and Pseudo R^2 for negative binomial regressions; however, these measures should be looked at independently as they are not directly comparable.

Part II

A. Distribution of Patent Awards Among Assignees

We first explore the distribution of finance patents across assignee types relative to our comparison groups. Panel A of Table 3 breaks down all assigned patents by type of entity and shows that a disproportionate number of finance patents is awarded to U.S. corporations and individuals. This is consistent with Lerner (2002). While 9% of finance patents were awarded to non-U.S.-based corporations from 2001 to 2010, 50% and 34% of patents, respectively, were awarded to this group in Comparison Groups 1 and 2.

Panel B looks at the proportion of patents either assigned or re-assigned to an NPE. While no statistically significant difference was found between finance (2.1%) and Comparison Group One (2.6%), finance patents are more frequently assigned or re-assigned to NPEs relative to Comparison Group 2 (0.4%). [28]We limit our analyses in Part II to corporations and individuals, so we drop the sole government observation and its match in each comparison group. The number of observations is therefore 2,798 for each group when NPEs are involved.

Table 3: Breakdown of Patent Awards by Assignee Type

Panel A data include all patents awarded between 2001 and 2010 to corporations, individuals, and governments. Panel B data consist of matched patents from our finance sample, which are matched by assignee type and grant year. NPEs were gathered from various sources, including PlainSite and IPCheckups.

Panel A: All Patents Assigned to Corporations, Individuals, and Governments from 2001-2010			
	Finance	Comparison Group 1	Comparison Group 2
Corporations	**86%**	**96%**	**92%**
U.S.	76%	46%	58%
Foreign	9%	50%	34%
Individual	**14%**	**3%**	**7%**
Government	0%	**1%**	**1%**
Total Number of Identified Patents	2,799	170,354	97,720

Panel B: Breakdown of NPE Awards Within Matched Samples			
NPE	**61**	**74**	**10*** **
Total Number of Identified Patents	2,798	2,798	2,798

Notes: Panel A does not show significance tests but simply the breakdown of patents across assignee types. Panel B tests for significance and finds it only in the difference between finance patents and non-finance patents from Comparison Group 2 assigned to NPEs. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level. Significance determined using a two-sample test of proportions.

[28] Our matching algorithm leaves open the possibility for false matches, as well as missed NPE patents. While efforts were taken to limit the number of false matches, variation in assignee names in many cases will lead to missed results. See Appendix C for more details.

B. Prior Art by Assignee Groups

Next, we replicate our Part I analysis within different assignee groups. Table 4 looks at all patents assigned to individuals or assigned/re-assigned (as of August 2015) to NPEs. Our list of over 3,000 NPEs was gathered from multiple sources, but primarily came from PlainSite and IPCheckups.[29]

The results are qualitatively consistent with Table 1. On average, finance patents awarded to individuals and NPEs cite fewer non-patent publications than do non-finance patents awarded to individuals and NPEs, and examiners add *more* non-patent citations. The average finance patent from individuals/NPEs cites especially less academic literature. Again, the difference generally widens as we look at citations to higher ranked journals. In addition, while examiners add *more* non-patent citations overall, they add far fewer citations to academic journals.

Table 5 paints a similar picture for non-NPE corporations. Of particular importance, finance patents have fewer non-patent citations and cite less academic literature. Examiners still appear to add more prior art (patent and non-patent overall) to the average patent, though significantly less *academic* prior art.

Similar to Part I, we employ a set of regressions to look more rigorously at the state of non-patent prior art within these assignee groups. Table 6 shows that differences in counts of non-patent prior art for patents awarded to individuals or associated with NPEs are not statistically significant for the Ordinary Least Squares (OLS) specification. But for the rest of Table 6, including counts of non-patent prior art for individuals/NPEs in the second half of Column A (when fitted to a negative binomial), the differences are significant. In general, these finance patents cite less academic prior art when controlling for other factors.

Similarly, Table 7 shows that finance patents awarded to non-NPE corporations, after controlling for a number of variables and being sorted by assignee group, still appear weaker across the board with respect to non-patent (and especially academic) prior art relative to both comparison groups. We ultimately find that overall, finance patents cite less non-patent prior art.

In Table 8, we look at patents awarded to individuals and NPEs in contrast to corporations (excluding NPEs). We see that once again, finance patents are less likely to cite non-patent prior art of all types. We also see that patents awarded to individuals and associated with NPEs

[29] For the PlainSite list, see http://www.plainsite.org/tags/patent-trolls/. For the partial list of NPEs featured in the NPE Tracker, see http://www.ipcheckups.com/npe-tracker/npe-tracker-list/. In the IPCheckups list, we exclude Academic/Research Entities and Allied Security Trust (a defensive patent aggregator) from our list. We note that NPE tracking—especially shell companies of large NPEs—is extremely complex. As a result, certain entities in the list may have been incorrectly labeled as an NPE in the respective sources. An article quotes a PlainSite representatives explaining its list of NPEs affiliated with Intellectual Ventures: "'We combed through 15GB of this data and linked up every patent assignment with the PlainSite entity, law firm and attorney databases to create an improved version of the USPTO assignment database, which we've made available for free. Then we tagged all of the companies that have links to attorneys and mailing addresses frequently used by Intellectual Ventures. The resulting list is about 2,000 companies. We have not verified that each and every company is definitely a shell corporation for Intellectual Ventures (doing so would be prohibitively expensive), but some obvious overlaps are apparent: managing corporations, telephone numbers, and other factors.'" See Kim (2012).

generally cite significantly less academic prior art in the two OLS regressions (Columns 3 and 5). When we confine the examination to finance patents in an unreported analysis, we find similarly that finance patents awarded to non-NPE corporations cite more academic prior art than those awarded to individuals and associated with NPEs. Because the overall number of academic citations in the finance patents is so low, however, the differences are not statistically significant.

Overall, this analysis indicates that finance patents under-report the total amount of academic prior art relative to their peers in other patent classes. It might be thought that examiners would correct this pattern during the examination process for finance patents. Indeed, in March 2000, after a flood of patent applications to Class 705, the USPTO established the “second pair of eyes review” (SPER) for these business method patents—a “quality control” mechanism that had two examiners review each application. This likely led to a more expansive prior art search. Allison and Hunter (2006) lend empirical support to this interpretation based on a sample of main class 705 patents. In particular, they found that SPER “produced a significant improvement in the quantity of all types of prior art cited in main class 705 patents,” as well as the quantity of non-patent prior art added by examiners (p. 749). But even with the SPER, finance patents lag the comparison groups in terms of citing non-patent, and especially academic, prior art.

Table 4: *Summary Statistics* of Patent Characteristics: Awarded to Individuals or Associated with NPEs

The finance sample consists of 440 patents assigned by the USPTO between 2001 and 2010 in classes 705/35, 705/36R, 705/37, 705/38 and part of 705/4 to an individual or assigned/re-assigned to an NPE. Comparison Group 1 is composed of 467 patents from the top five classes overall in the Lai et al. dataset (1975-2010). Comparison Group 2 is composed of 409 patents from the top five classes from R&D intensive universities in the post-*State Street* era (1998-2010).

	Mean			**Standard Deviation**		
Prior Art – Overall	Finance	Non-Finance Group 1	Non-Finance Group 2	Finance	Non-Finance Group 1	Non-Finance Group 2
Backward patent citations (U.S. only)	21.49	26.15	18.38	28.29	48.79	28.22
Backward citations per claim (U.S. only)	1.62	1.90	2.03***	2.622	3.44	5.49
Backward non-patent citations	10.72	13.07*	15.89**	19.64	26.43	30.06
Backward non-patent citations per claim	0.79	1.47**	2.06***	2.19	5.14	8.87
Examiner added patent citations (U.S. only)	7.36	4.28***	3.33***	6.68	4.76	4.13
Examiner added non-patent citations	3.04	1.29***	1.76***	4.65	3.71	4.03
Prior Art – Academic Publications						
Non-patent academic citations – **Overall**	0.36	2.91***	3.82***	1.55	9.12	10.47
Non-patent academic examiner-added citations - **Overall**	0.11	0.31	0.54**	0.44	1.90	1.82
Non-patent academic citations - **Top 25%**	0.25	2.48***	3.28***	1.46	8.05	9.33
Non-patent academic examiner-added citations - **Top 25%**	0.04	0.267***	0.43***	0.26	1.73	1.60
Non-patent academic citations - **Top 10%**	0.09	1.90***	2.64***	0.64	6.9	7.95
Non-patent academic examiner-added citations - **Top 10%**	0.02	0.18***	0.32***	0.18	1.14	1.26
Non-patent academic citations - **Top 1%**	0.03	0.60***	1.03***	0.27	2.71	3.80
Non-patent academic examiner-added citations - **Top 1%**	0.01	0.05*	0.10***	0.15	0.35	0.45
Other Variables						
Originality	0.22	0.49***	0.408***	0.239	0.28	0.28
Claims	23.17	18.31***	15.94***	21.85	14.85	13.21
Number of forward cites (U.S. only; patents granted up to 2004; n = 63, 60, 56 respectively)	21.59	7.37***	3.91***	30.77	19.20	6.87
Observations	440	467	409	440	467	409

Mean significance at: * 10% level; ** 5% level, *** 1% level
Mean significance tested by two-sample t test assuming unequal variance (two-tailed)
Significance marked in non-finance portion only, but is relative to the finance patents

Table 5: *Summary Statistics* of Patent Characteristics: Corporations (excluding NPEs)

The finance sample consists of 2,358 patents assigned by the USPTO between 2001 and 2010 in classes 705/35, 705/36R, 705/37, 705/38 and part of 705/4 to a corporation that was not identified as an NPE. Comparison Group 1 is composed of 2,331 patents from the top five classes overall in the Lai et al. dataset (1975-2010). Comparison Group 2 is composed of 2,389 patents from the top five classes from R&D intensive universities in the post-*State Street* era (1998-2010).

	Mean			**Standard Deviation**		
Prior Art – Overall	Finance	Non-Finance Group 1	Non-Finance Group 2	Finance	Non-Finance Group 1	Non-Finance Group 2
Backward patent citations (U.S. only)	39.74	34.18***	44.96*	65.34	60.50	113.10
Backward citations per claim (U.S. only)	2.60	2.81	3.81***	6.65	6.45	11.56
Backward non-patent citations	18.09	20.60*	37.04***	44.26	48.08	64.88
Backward non-patent citations per claim	1.05	1.94***	3.88***	3.01	6.34	10.76
Examiner added patent citations (U.S. only)	7.32	3.93***	2.80***	6.85	5.20	4.17
Examiner added non-patent citations	2.38	0.90***	2.17**	3.79	2.46	5.02
Prior Art – Academic Publications						
Non-patent academic citations – **Overall**	0.37	4.92***	9.39***	1.23	26.76	23.94
Non-patent academic examiner-added citations - **Overall**	0.09	0.23***	0.63***	0.36	0.92	2.08
Non-patent academic citations - **Top 25%**	0.20	4.30***	8.12***	0.93	22.88	20.90
Non-patent academic examiner-added citations - **Top 25%**	0.04	0.19***	0.54***	0.24	0.81	1.86
Non-patent academic citations - **Top 10%**	0.13	3.32***	6.30***	0.69	17.23	16.20
Non-patent academic examiner-added citations - **Top 10%**	0.02	0.14***	0.41***	0.16	0.64	1.47
Non-patent academic citations - **Top 1%**	0.06	1.09***	2.22***	0.38	4.85	6.37
Non-patent academic examiner-added citations - **Top 1%**	0.01	0.04***	0.14***	0.10	0.26	0.595
Other Variables						
Originality	0.28	0.47***	0.44***	0.27	0.27	0.28
Claims	25.83	18.50***	19.15***	22.08	15.47	16.74
Number of forward cites (U.S. only; patents granted up to 2004; n = 199, 202, 206 respectively)	28.07	6.32***	5.61***	30.23	13.15	11.09
Observations	2,358	2,331	2,389	2,358	2,331	2,389

Mean significance at: * 10% level; ** 5% level, *** 1% level

Mean significance tested by two-sample t test assuming unequal variance (two-tailed)

Significance marked in non-finance portion only, but is relative to the finance patents

Table 6: *Regressions*: Non-Patent Prior Art Contained in Patents Awarded to Individuals or Associated with NPEs

The finance sample consists of **907** patents awarded to individuals or awarded/re-assigned to NPEs by the USPTO between 2001 and 2010 in classes 705/35, 705/36R, 705/37, 705/38 and part of 705/4. See Table 4 for details. Each dependent variable looks at a different type of non-patent prior art: all types (A); all academic (B); and leading academic (C), defined as non-patent citations to journals ranked in the top 10% of their respective disciplines from SCImago Journal & Country Rank. Independent variables include various controls: Year*Assignee fixed effects control for differences in assignee patenting behavior among the different grant years; grant year fixed effects control for differences in patenting behavior across years; foreign inventor effects (U.S.; Japan; and EPO); and, in Specification (2), number of claims. Within each column we run two types of regressions: OLS and negative binomial.

	Column A				Column B				Column C			
	Dependent Variable: Non-Patent Prior Art – Overall				Dependent Variable: Non-Patent Prior Art – Academic				Dependent Variable: Non-Patent Prior Art – Top 10% Academic			
Regression Type	OLS		Negative Binomial		OLS		Negative Binomial		OLS		Negative Binomial	
Specification	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Panel A: Finance versus Comparison Group 1												
Finance	-2.34 (1.59)	-2.54 (1.63)	-0.27 (0.12)**	-0.28 (0.12)**	-2.70 (0.46)***	-2.71 (0.46)***	-2.23 (0.22)***	-2.26 (0.24)***	-1.90 (0.31)***	-1.90 (0.30)***	-3.50 (0.26)***	-3.54 (0.27)***
Grant Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.04	0.04	-	-	0.05	0.05	-	-	0.06	0.06	-	-
Pseudo R^2	-	-	0.01	.01	-	-	0.06	0.06	-	-	0.09	0.09
Observations	907	907	907	907	907	907	907	907	907	907	907	907
Panel B: Finance versus Comparison Group 2												
Finance	-6.61 (1.73)***	-7.03 (1.83)***	-0.50 (0.12)***	-0.50 (0.12)***	-3.48 (0.52)***	-3.57 (0.54)***	-2.54 (0.22)***	-2.55 (0.24)***	-2.56 (0.39)***	-2.63 (0.41)***	-3.81 (0.25)***	-3.49 (0.26)***
Grant Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.12	0.13	-	-	0.07	0.07	-	-	0.06	0.06	-	-
Pseudo R^2	-	-	0.02	0.02	-	-	0.07	0.08	-	-	0.12	0.12
Observations	907	907	907	907	907	907	907	907	907	907	907	907

Notes: * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level. Robust standard errors in parentheses. We report R^2 for OLS regressions and Pseudo R^2 for negative binomial regressions; however, these measures should be looked at independently as they are not directly comparable.

Table 7: *Regressions*: Non-Patent Prior Art Contained in Patents Awarded to Corporations (excluding NPEs)

See Table 5 for details of the various patent groups. Each dependent variable looks at a different type of non-patent prior art: all types (A); all academic (B); and leading academic (C), defined as non-patent citations to journals ranked in the top 10% of their respective disciplines from SCImago Journal & Country Rank. Independent variables include various controls: Year*Assignee fixed effects control for differences in assignee patenting behavior among the different grant years; grant year fixed effects control for differences in patenting behavior across years; foreign inventor effects (U.S.; Japan; and EPO); and, in Specifications (2), number of claims. Within each column we run two different types of regressions: OLS and negative binomial.

	Column A				Column B				Column C			
	Dependent Variable: Non-Patent Prior Art – Overall				Dependent Variable: Non-Patent Prior Art – Academic				Dependent Variable: Non-Patent Prior Art – Top 10% Academic			
Regression Type	OLS		Negative Binomial		OLS		Negative Binomial		OLS		Negative Binomial	
Specification	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Panel A: Finance versus Comparison Group 1												
Finance	-2.98 (1.36)**	-4.79 (1.40)***	-0.11 (0.06)**	-0.18 (0.06)***	-4.69 (0.57)***	-4.77 (0.58)***	-2.54 (0.11)***	-2.58 (0.11)***	-3.27 (0.37)***	-3.31 (0.37)***	-3.22 (0.13)***	-3.29 (0.13)***
Grant Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.03	0.04	-	-	0.02	0.02	-	-	0.02	0.02	-	-
Pseudo R^2	-	-	0.01	0.01	-	-	0.06	0.06	-	-	0.08	0.08
Observations	4,689	4,689	4,689	4,689	4,689	4,689	4,689	4,689	4,689	4,689	4,689	4,689
Panel B: Finance versus Comparison Group 2												
Finance	-19.38 (1.60)***	-21.08 (1.65)***	-0.75 (0.05)***	-0.81 (0.05)***	-9.16 (0.50)***	-9.24 (0.51)***	-3.26 (0.08)***	-3.28 (0.08)***	-6.26 (0.34)***	-6.32 (0.34)***	-3.89 (0.11)***	-3.93 (0.11)***
Grant Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.06	0.06	-	-	0.08	0.08	-	-	0.08	0.08	-	-
Pseudo R^2	-	-	0.02	0.02	-	-	0.09	0.09	-	-	0.11	0.11
Observations	4,747	4,747	4,747	4,747	4,747	4,747	4,747	4,747	4,747	4,747	4,747	4,747

Notes: * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level. Robust standard errors in parentheses. We report R^2 for OLS regressions and Pseudo R^2 for negative binomial regressions; however, these measures should be looked at independently as they are not directly comparable.

Table 8: Non-Patent Prior Art Contained in Patents Awarded to Individuals/NPEs versus Corporations (excluding NPEs)

This set of regressions combines the observations described in both tables; see Tables 4 and 5 for details of the various patent groups. Each dependent variable looks at a different type of non-patent prior art: all types (A); all academic (B); and leading academic (C), defined as non-patent citations to journals ranked in the top 10% of their respective disciplines from SCImago Journal & Country Rank. Independent variables include various controls: Year*Assignee fixed effects control for differences in assignee patenting behavior among the different grant years; grant year fixed effects control for differences in patenting behavior across years; foreign inventor effects (U.S.; Japan; and EPO); and, in Specification (2), number of claims. Within each column, we run two different types of regressions: OLS and negative binomial.

	Column A				Column B				Column C			
	Dependent Variable: Non-Patent Prior Art – Overall				Dependent Variable: Non-Patent Prior Art – Academic				Dependent Variable: Non-Patent Prior Art – Top 10% Academic			
Regression Type	OLS		Negative Binomial		OLS		Negative Binomial		OLS		Negative Binomial	
Specification	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Panel A: Individual/NPE versus non-NPE Corporations (Comparison Group 1)												
Finance	-2.80 (1.17)**	-4.29 (1.20)***	-0.13 (0.05)**	-0.19 (0.05)***	-4.35 (0.48)***	-4.42 (0.49)***	-2.47 (0.10)***	-2.51 (0.10)***	-3.04 (0.31)***	-3.07 (0.31)***	-3.20 (0.12)***	-3.26 (0.12)***
Individuals and NPEs	-2.39 (2.81)	-2.86 (2.80)	-0.02 (0.15)	-0.03 (0.15)	-2.54 (0.46)***	-2.57 (0.46)***	-0.52 (0.53)	0.53 (0.54)	-1.74 (0.31)***	-1.74 (0.31)***	-0.50 (0.78)	0.50 (0.78)
Grant Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.03	0.04	-	-	0.02	0.02	-	-	0.02	0.02	-	-
Pseudo R^2	-	-	0.01	0.01			0.06	0.06	-	-	0.08	0.08
Observations	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596

Notes: * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level. Robust standard errors in parentheses. We report R^2 for OLS regressions and Pseudo R^2 for negative binomial regressions; however, these measures should be looked at independently as they are not directly comparable.

Table 8 Continued: Non-Patent Prior Art Contained in Patents Awarded to Individuals/NPEs versus Corporations (excluding NPEs)

	Column A				Column B				Column C			
	Dependent Variable: Non-Patent Prior Art – Overall				Dependent Variable: Non-Patent Prior Art – Academic				Dependent Variable: Non-Patent Prior Art – Top 10% Academic			
Regression Type	OLS		Negative Binomial		OLS		Negative Binomial		OLS		Negative Binomial	
Specification	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Panel B: Individual/NPE versus non-NPE Corporations (Comparison Group 2)												
Finance	-17.46 (1.39)***	-19.01 (1.43)***	-0.71 (0.05)***	-0.76 (0.50)***	-8.31 (0.43)***	-8.39 (0.44)***	-3.16 (0.08)***	-3.18 (0.08)***	-5.70 (0.29)***	-5.76 (0.30)***	-3.87 (0.10)***	-3.91 (0.10)***
Individuals and NPEs	13.18 (5.31)**	12.45 (5.31)**	0.54 (0.14)***	0.54 (0.14)***	-0.82 (0.55)	-0.86 (0.56)	0.42 (0.56)	0.43 (0.57)	-0.61 (0.39)	-0.64 (0.40)	0.67 (0.76)	0.70 (0.77)
Grant Year* Assignee	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S. Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Japan Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EPO Inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Claims	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.06	0.06	-	-	0.08	0.08	-	-	0.08	0.08	-	-
Pseudo R^2	-	-	0.02	0.02	-	-	0.09	0.09	-	-	0.11	0.11
Observations	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596	5,596

Notes: * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level. Robust standard errors in parentheses. We report R^2 for OLS regressions and Pseudo R^2 for negative binomial regressions; however, these measures should be looked at independently as they are not directly comparable.

Part III

In this section, we examine whether the quality of litigated finance patents differs from that of non-litigated finance patents, and compare these findings with our non-finance comparison groups. Given previous findings that litigated patents cite more prior art than non-litigated patents (Allison et al. 2004; Allison et al. 2009), we investigate whether this finding holds true for our sample of litigated finance patents. If these relationships do not hold for finance patents, it may suggest key differences in the nature of the litigated patents.

We begin with some descriptive statistics on the frequency of litigation and note comparisons to previous literature. We follow with summary statistics and regression analysis comparing litigated and non-litigated patents. We look at this relationship separately for finance and non-finance groups.

A. Descriptive Statistics

Table 9 compares the frequency of a litigation event between finance and non-finance patents. Consistent with Lerner (2010), finance patents are disproportionately litigated. Our litigation rates are low compared to the literature, which may reflect that many of the patents in our sample have had relatively little time to be litigated yet.

Litigation intensity is a composite measure that looks at the total number of litigation "events" in the LitAlert database. Events include any formal matter relating to the litigation, including complaints filed, memorandum opinions, and final orders. We find that among litigated patents, finance patents are more intensely litigated than comparison groups. This finding is consistent with Allison et al. (2009, 18), who found that software business method patents are overrepresented in their group of "most-litigated" patents.

Table 9: Descriptive Statistics for Litigated Patents

This finance sample consists of 61 patents awarded by the USPTO between 2001 and 2010 in classes 705/35, 705/36R, 705/37, 705/38 and part of 705/4 that have engaged in litigation. We include awards assigned to any type of entity: corporations, individuals, and governments. We define litigation as any event in the Derwent LitAlert database accessed via Thomson Reuters Westlaw. Litigation intensity is the number of litigation events for a given patent in the LitAlert database.

	Count	Percent of Patents Litigated from Sample	Average Litigation Intensity		Median Litigation Intensity	
			Overall	Among Litigated	Overall	Among Litigated
Finance Patents	59	2.11	1.29	61.41	0	30
Comparison Group 1	15	0.54	0.27	50.33	0	30
Comparison Group 2	13	0.46	0.21	46.08	0	15

B. Summary Statistics

Table 10 compares our measures within the finance (Panel A) and non-finance (Panel B) groups. Given the infrequency of litigated patents, we combine comparison groups 1 and 2 to increase our sample of litigated patents. Consistent with prior literature, we find that in the non-finance sample, litigated patents make more reference to prior art. These findings are significant at the 5% level for both patent and non-patent citations. Because our sample of litigated non-finance patents is small (n=28), these figures must be interpreted cautiously. Among finance patents, litigated and non-litigated patents have roughly the same number of backward patent citations (not statistically significant) but vastly more non-patent citations (significant at the 1% level). In addition, litigated patents have more claims (significant at the 10% level).

A major difference between litigated and non-litigated patents, however, is their emphasis on academic prior art. Panel A shows that litigated finance patents have fewer academic citations on average than non-litigated patents. These findings are significant at the 1% level for top 25% journals, and at the 10% level for top 10%. In contrast, the average litigated patent in our non-finance sample has in some cases *more* academic citations than non-litigated non-finance patents. This finding is significant at the 1% level for the top 1% journals.

B. Regression Analysis

Table 11 examines how the probability of litigation (Panel A) and the intensity of litigation (Panel B) varies with respect to the type of patent (finance or non-finance) and the number of non-patent academic and non-patent, non-academic citations. We run two regressions each in Panel A and Panel B, where we look at how an additional non-patent prior art citation affects litigation for finance and non-finance patents. We break the non-patent prior art into two groups: academic citations and non-patent non-academic citations, which would include trade journals, books, and so forth. In the probit regressions, we control for the grant year of the application, nationality of the inventor, and the number of claims contained in the patent. Certain controls in the negative binomial regression (claims in Equation 1 and inventor nationality in Equation 2) made it too complex to run, so we excluded them as necessary to simplify the model. Although the negative binomial did not converge after 300 iterations, we believe its results are illustrative and report them as of that point.

Both panels show that finance patents are more likely to be litigated (Panel A), and to have increased litigation intensity (Panel B) relative to non-finance patents. The differences are significant at the 1% level for each specification, whether finance only or finance with additional terms and controls.

Equation 2 in Panel A shows that although finance patents are more likely to be litigated, each additional academic citation reduces this possibility. While we must interpret this relationship with caution, given our small number of litigated patents, the interaction term of finance and non-patent academic citations (italicized) is negative and significant at the 5% level. This finding thus suggests that the probability of litigation for a finance patent relative to a non-finance patent falls for every additional academic citation in the patent. In Panel B's Equation 2, the finance and non-patent academic citations interaction term (italicized) is also negative and, in this case, significant at the

10% level. This result suggests that additional academic citations tend to reduce the litigation intensity for finance patents relative to non-finance patents. The other significant variable in both versions of Equation 2, the non-patent non-academic term, is extremely small, which means it has no meaningful impact. Our findings, therefore, suggest that additional academic citations may significantly reduce the probability of a finance patent's litigation and the intensity of that litigation, relative to a non-finance patent.

Table 10: Summary Statistics: The Quality of Litigated and Non-Litigated Finance Patents

This finance sample consists of **59** patents awarded by the USPTO between 2001 and 2010 in classes 705/35, 705/36R, 705/37, 705/38 and part of 705/4 that have engaged in litigation and **2,538** that have not been litigated. Non-finance patents are inclusive of Comparison Groups 1 and 2. We include awards assigned to any type of entity: corporations, individuals, or any level of government. We define litigation as any event in the Derwent LitAlert database accessed via Thomson Reuters Westlaw.

	Panel A Litigated? – Finance				**Panel B Litigated? – Non-Finance (Comparison Groups 1 & 2 combined)**			
	Mean		**Standard Deviation**		**Mean**		**Standard Deviation**	
Prior Art – Overall	Lit	Not-Lit	Lit	Not-Lit	Lit	Not-Lit	Lit	Not-Lit
Backward patent citations (U.S. only)	34.36	36.92	28.85	61.99	69.96	36.74**	66.32	85.76
Backward citations per claim (U.S. only)	2.09	2.45	3.34	6.25	4.35	3.10	4.72	8.90
Backward non-patent citations	42.29	16.38***	74.04	40.31	53.68	26.20**	62.14	53.51
Backward non-patent citations per claim	1.69	1.00*	3.04	2.89	5.48	2.69	14.08	8.57
Examiner added patent citations (U.S. only)	7.24	7.32	6.94	6.82	2.75	3.46	3.46	4.73
Examiner added non-patent citations	3.54	2.46	3.90	5.59	1.14	1.52	1.80	3.99
Prior Art – Academic Publications								
Non-patent academic citations - **Overall**	0.27	0.37	0.64	1.30	7.86	6.52	15.74	23.83
Non-patent academic examiner-added citations - **Overall**	0.07	0.09	0.25	0.37	0.18	0.42**	0.48	1.65
Non-patent academic citations - **Top 25%**	0.07	0.21***	0.03	1.04	6.21	5.65	13.75	20.57
Non-patent academic examiner-added citations - **Top 25%**	0.02	0.04	0.13	0.25	0.11	0.36***	0.32	1.47
Non-patent academic citations - **Top 10%**	0.07	0.13*	0.25	0.69	4.29	4.38	10.69	15.72
Non-patent academic examiner-added citations - **Top 10%**	0.02	0.02	0.13	0.16	0.07	0.27***	0.26	1.14
Non-patent academic citations - **Top 1%**	0.03	0.06	0.18	0.37	0.57	1.52***	1.29	5.39
Non-patent academic examiner-added citations - **Top 1%**	0.00	0.09***	0.00	0.11	0.04	0.09	0.19	0.45
Other Variables								
Originality	0.27	0.26	0.03	0.26	0.51	0.45	0.30	0.28
Claims	38.15	25.13*	56.32	20.65	24.43	18.56*	17.84	15.86
Number of forward cites (U.S. only; patents granted up to 2004**; n = 61** for lit and **2,538** for not-lit for finance. Non-finance excluded due to too few observations)	31.67	26.13	30.17	30.47	2.00	5.94***	1.00	12.77
Observations	59	2,740	59	2,740	28	5,471	28	5,471

Mean significance at: * 10% level; ** 5% level, *** 1% level

Mean significance tested by two-sample t test assuming unequal variance (two-tailed)

Significance marked in non-litigated portion only, but is relative to the litigated patents.

Table 11: Summary Statistics: Citation Trends of Litigated and Non-Litigated Finance Patents

Here, the non-finance sample consists of litigated patents awarded during the same period within Comparison Group 1 and Comparison Group 2, as well as awards that were not litigated. We include awards assigned to any type of entity: corporations, individuals, and governments. See Tables 4 and 5 for details of the various patent groups. We define litigation as any event in the Derwent LitAlert database accessed via Thomson Reuters Westlaw. Intensity is defined very broadly and is a composite measure of the total number of actions (e.g., complaints filed, memorandum opinions, final orders, etc.) in the LitAlert database, which can come from single or multiple suits.

	Panel A		**Panel B**	
	Regression Type: Probit		**Regression Type: Negative Binomial**	
	Dependent Variable: Litigated (Yes = 1)		**Dependent Variable: Litigation Intensity**	
	(1)	(2)	(1)	(2)
Finance	0.52***	0.53***	1.41***	1.44***
Non-Patent Academic Citations		-0.00		-0.00
		(0.00)		(0.00)
Finance * Non-Patent Academic Citations		-0.16**		-0.37*
		(0.08)		(0.19)
Non-Patent Non-Academic Citations		0.00***		0.01***
		(0.00)		(0.00)
Finance * Non-Patent Non-Academic Citations		0.00		-0.00
		(0.00)		(0.00)
Grant Year Fixed Effects	Yes	Yes	Yes	Yes
American Inventor	Yes	No	Yes	No
Japanese Inventor	Yes	No	Yes	No
European Inventor	Yes	No	Yes	No
Claims	No	Yes	No	Yes
Pseudo R^2	0.07	0.01	0.07	0.10
Observations	8,298	8,298	8,298	8,298

Notes: Significance at: * 10% level; ** 5% level, *** 1% level. Robust Standard Errors in parentheses. All numbers rounded to two decimal places, allowing for "significant zeros." The negative binomial models failed to converge after 300 iterations so results were reported at that point.

7. Robustness Checks

We employ a number of robustness checks to determine the sensitivity of our results to the particular specifications we employed in our base results. We adjust our methodology in the following ways and observe the changes in output:

1. We run non-parametric tests, as opposed to t-tests, to measure if there are statistically significant differences in medians (as opposed to means) between the finance and comparison groups with respect to our key variables of interest.
2. We re-run each OLS specification in Part I and II with a log-transformed dependent variable.
3. We re-run each set of regressions using the “top 1%” journals instead of “top 10%” as the dependent variable, our standard for “high-quality” journal citations.
4. Instead of comparing the *count* of academic references, we compare the *proportion* of non-patent prior art references that are from academic references.

Robustness Test 1: Non-Parametric Tests

We know the distribution of patents is skewed, and hence t-tests, which assume a normal distribution, may be problematic. Using median tests instead of two-sample t-tests returns similar, but not identical, significance results for all three parts of the analysis. In general, the difference is that the median test produces more conservative results, i.e., lower levels of statistical significance. Still, the main results are largely consistent—the number of overall academic citations and those from the top 10% of journals have similar significance levels for all three parts.

Robustness Test 2: Log-Normal Regressions

The idea behind logarithmic translations is that proportionality matters. If one were, say, comparing pay raises as a proxy for performance, then a doctor would always appear to be performing better than a mechanic: the fact that the doctor received a 2% raise and the mechanic a 10% raise would be lost in the data. Similarly, the difference in citations described in Section 6 could be a function of the size of the field; if the classes in our comparison groups have far more patents available, it is unsurprising that they would cite more of them. The logarithms make the difference between 1 and 2 as dramatic as that between 10 and 100.

Using logarithmic transformations of dependent variables—backward citations, forward citations, non-patent academic citations, and academic citations from the top 1%, 10%, and 25% of journals—generally does not change the direction of our regression coefficients for linear, negative binomial, or probit models. Transforming the same variables in Part III, where they are independent variables in regressions examining the propensity for and intensity of litigation, actually tends to yield higher levels of significance for either the finance coefficient or the transformed independent variable’s coefficient. This suggests that the size of a field’s literature base does not substantially impact our results.

Robustness Test 3: Regressions with Top 1% Journals

We previously reported results for regressions where the dependent variables are the total number of academic non-patent citations and the number of academic citations from journals in the top 10% of that field, according to the SCImago ranking. We run the same regressions using journals in the top 1% of their field. Unsurprisingly, the regression coefficients were smaller, but they were directionally identical and significance levels were similar. Using the top 25% of journals tended to provide different levels of significance, but this could be due to the fact that we have more coverage of journal abbreviations for journals in the top 1% and 10%. If non-standard abbreviations were used when listing references (and they very often are), we would miss some of the academic citations for the top 25%. Consistency of results when using the top 1% suggests that our results hold for elite journals.

Robustness Test 4: Comparing Academic Reference Proportions

A final robustness check compares the *proportion* of academic references instead of the raw number. This is necessary because given fields' relationships to academia may differ. For example, the initial phases of pharmaceutical research are often done at universities through public grant money. Once there are early results to build on, private companies develop a specific drug and patent it, often citing the relevant academic papers. This may not be the case for financial innovation. Similarly, the outlet for research performed by academics may differ. Academic finance innovation would likely come from professors at business schools; these professors may be more heavily involved in industry than their counterparts in other fields. Thus, research on, for example, evolutionary game theory is almost certainly going to be submitted for publication in a journal, but financial developments may be sent to any number of other entities.

To test this, we created a new variable equal to the ratio of academic citations to overall non-patent citations. Finance continues to have a negative and significant impact in most cases, although the magnitude is smaller than when using the raw citation counts. The assignee type variables in Part II (individual and corporation) generally lack significance, as do their interaction terms, suggesting that citation ratios do not vary across them.

8. Discussion of Findings

In this paper, we used a variety of analyses to examine the quality of financial patents. Overall, our analysis suggests that finance patents are problematic with respect to the citation of prior academic research. Moreover, the study indicates that increased academic citations appear to be linked to a significant reduction in litigation for finance patents. Below we synthesize our results from each of the three sections.

In the first section, we examined the prior art citations of finance and non-finance patents. We discovered that finance patents cite less non-patent prior art and had far fewer citations to non-patent academic prior art relative to each of the comparison groups. Examiners add more non-patent citations to the average finance patent relative to each comparison group, but even after these additions, the total number of citations used in finance patents lags. Even after controlling for a number of variables and separating patents into assignee groups, finance patents still appear weaker across the board with respect to non-patent prior art (and especially academic citations) relative to the comparison groups. We argue that a substantial body of relevant literature exists in the financial arena and that major discrepancies between finance patents and non-finance patents are revealing.

Part II explores whether the lack of academic prior art citations is evenly distributed across awardees and determines that it is not. Patents awarded to non-NPE corporations generally cite more academic prior art than the comparable awards to individuals and NPEs. In addition, the difference in academic prior art citations between the individual/NPE group and corporations holds for finance patents as we focus on the leading journals. While not statistically significant due to the small number of academic citations in finance patents, these patterns of under-citation are qualitatively similar when we restrict our analysis to finance patents held by individuals and NPEs, as opposed to non-NPE corporations.

A dearth of academic citations may suggest that there are few financial advances at the frontier of innovation, or that deficiencies in the examination process discourage citations to highly relevant, potentially patent-defeating, prior art. The latter interpretation is consistent with key findings from Lerner (2002, Table VII): while between 12-26% of the patents in the non-finance samples were reviewed by an examiner with a Ph.D. in a related field, the equivalent figure for finance patents was just 1%.

Finally, Part III examines the impact of these differences in academic citations by exploring the relationship among finance and non-finance patents, academic citations, and litigation. As shown in earlier studies, the finance patents are litigated more often, a fact that has drawn concern from many observers. Furthermore, we find that for finance patents—but not for those in other fields—academic citations are critically related to litigation. Patents with more academic citations had less litigation. The same effect was not seen for non-finance patents.

Together, these findings raise important questions regarding the *type* of innovations being patented in the finance domain and the strength of their academic foundation. Are obvious prior art citations being overlooked by examiners? Do finance patents lack the sophistication seen in other technological areas? Are patented financial innovations somehow separate from the academic domain? Like Lerner (2002), the analyses in this paper also continue to raise concerns about the state of academic research among financial patent applicants and examiners. Our results suggest that the overall absence of links to academic knowledge—a problem that is particularly dramatic for finance patents, especially those awarded to individuals and associated with NPEs—is directly associated with the proliferation of litigation in this area.

While our findings raise questions about the strength of finance patents, more research should be done to conclusively determine the extent of weaknesses in this area. Although beyond the scope of this study, topics of interest might include an examination of the proportion of "wacky" patents in finance relative to other fields[30]; an in-depth assessment of invalidations due to

[30] Czarnitzki et al. 2011 review "wacky" patents, or those that "do not involve a high inventive step or only marginally satisfy the 'non-obviousness' criterion." This study does not, however, examine the frequency or distribution of "wacky" patents across various classes/subclasses.

missing prior art references (e.g., Allison and Lemley 1998); and a critical review of finance patent examination procedures and personnel at the USPTO. Consistent findings across these analyses would serve to highlight the extent of deficiencies in financial patenting.

9. Appendices

9.1 Appendix A: Technical Details

A. Equations and Definitions

We calculate patent originality using a formula drawn from prior literature (Hall et al. 2001). Originality for patent *i* is calculated as:

$$R_i = 1 - \sum_{j}^{n_i} s_{ij}^2$$

Where s_{ij} denotes the proportion of citations made by patent *i* in class j. n_i refers to the total number of patent classes.

B. Constructing the Dataset

The base of our dataset comes from the Harvard Dataverse patent data, specifically the datasets included in Lai et al.'s "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)." This data augments the NBER's patent dataset with the years through 2010 using weekly USPTO patent data. The data were in several pieces, so we downloaded the assignee, citation (one for 1975-1999, another for 2000-2010), class, description, and patent files separately. We also used the researchers' file of full disambiguation results, which includes inventor and assignee numbers. Details regarding the aforementioned datasets can be found in the Methodology section of this paper.[31] All were in delimited spreadsheets, so we began by importing those into Stata and saving them in Stata's native format. We also cleaned the data when necessary and compressed the variable entries whenever possible. Below, we summarize the steps taken to clean and format the data for this analysis.

Determining Finance Patents and Comparison Groups

Once we had tabulated both types of citations for all patents, we identified the two categories of interest: finance classes and the top five non-finance classes (hereafter "Top 5").

<u>Finance Patents</u>

Finance patents were defined as those with primary class 705 and subclasses 35, 36R, 37, 38, or in some cases, 4. Class 705/4 refers to insurance, so to identify relevant patents, we generating a list of keywords that could be used to identify other patents dealing with these topics. To match keywords with patent abstracts in this category we employed Lai et al.'s description data (patdesc.zip).[32] We found a sizable number of observations that had no abstract. In these cases, we substituted the title whenever possible, and dropped the observation when it was not. Finally,

[31] For a full list of datasets, see http://hdl.handle.net/1902.1/15705.

[32] Some observations in the file had spilled the abstract into the cell holding the patent number; others had somehow contaminated the patent number's cell with HTML code. We removed all observations with either of those issues.

we added a leading 0 to the ID numbers—to align with Lai et al.'s other datasets—and removed anything that was not in the correct form for a utility patent.

We ultimately searched the abstracts for the following keywords and roots, only keeping patents containing at least one of them:

- Annuity
- Financ*
- Invest
- Tax
- Interest
- Rate
- Value
- Asset
- Equity
- Fund
- Mutual
- Index
- Cash
- Stock
- Hedge
- Trad*
- Yield
- Profit
- Instrument
- Account
- Portfolio
- Mone*
- Commodit*
- Transaction

Note that Stata searched the abstract for each of these strings, so for example "invest" also found "investment" and "investor." The asterisk (*) is a "wildcard" character and took the place of any number and combination of non-space ASCII characters. For example, trad* would accept trade, trading, and trader. In the code, the wildcard is used after each keyword to make sure we catch plurals and other alternate forms. To avoid issues of case sensitivity, we converted the keywords and the abstracts to lowercase via the *lower* function before performing the searches.

We similarly culled patents that might use a keyword unrelated to finance, and those involving bank back room operations, an exclusion that is consistent with prior literature (Lerner 2002). This was done through the same process, but with what we refer to as negative keywords. These are:

- EFT
- Electronic Funds Transfer
- ATM
- Remote

Comparison Group 1

The top five classes were identified using Lai et al.'s "class" file by tabulating primary classes in order of frequency: 257, 428, 435, 438, and 514 in the Lai et al. dataset (1975-2010).

Comparison Group 2

An alternative top five ("Top 5E"), was chosen by first identifying the top 20 universities in terms of R&D expenditures. The data came from the National Science Foundation, using an excel file that contained university expenditures from FY 2004-10.[33] We dropped 2010 data due to changes in methodology in that year.[34]

To align with assignee data, we aggregated the University of California campuses in the National Science Foundation data to count as a single entity. The top 20 universities are found in Table 12, along with our search terms and the various assignee matches we picked up from these searches. Because multiple assignee names are often associated with a given university (from specific schools within the university, technology transfer operations, affiliated research foundations, etc.), we used search terms to capture these various organizations.[35]

Table 12: Search Terms for Comparison Group 2

University	Search Term(s)	Top Assignee Matches	Total R&D Expenditure 2004-09 ($000)
University of California *(Los Angeles, San Francisco, San Diego, Davis, Berkeley)*	*Must contain* "University of California" & "Regents" or "Technology Transfer"	"(The) University of California The Regents Of"; "(The) Regents of the University of California"; "(The) Regents of the University of California Office of Technology Transfer"	21,846,571
Johns Hopkins University	"Johns Hopkins University"	"(The) Johns Hopkins University"; "(The) Johns	9,469,610

[33] See http://ncsesdata.nsf.gov/profiles/site;jsessionid=AC5C469BCBB35ABB98A288239BB77CC9.prodas2?method=rankingBySource&ds=herd.

[34]The most pertinent change is that prior to the FY 2010 survey, multi-campus institutions were not consistent in reporting measures. While some responded at the aggregate level, others responded at the campus level. In the 2010 survey, all institutions reported at the campus level. For more information, see http://www.nsf.gov/statistics/srvyherd/overview.htm and http://www.ir.vt.edu/work_we_do/PeerInstitutions/HERD_Reporting_Unit_Changes.pdf.

[35] The unique assignee numbers in Lai et al.'s disambiguated dataset are different for these sub-units, so we opted to search by name rather than the assignee number of the main university. We identified the various assignee names from manual searches within the Lai et al. dataset, as well as from the USPTO's "organizational identifiers" for each university at http://www.uspto.gov/web/offices/ac/ido/oeip/taf/univ/org_gr/universities_g.htm.

		Hopkins University School of Medicine”	
University of Wisconsin-Madison	“Wisconsin Alumni Research Foundation”; “University of Wisconsin Madison”	“Wisconsin Alumni Research Foundation” ; “University of Wisconsin Madison”	5,481,125
University of Michigan	“University of Michigan”	“(The) Regents of the University of Michigan”; “University of Michigan”	5,319,029
University of Washington	“University of Washington”	“University of Washington”; “(The) Board of Regents of The University of Washington”	4,618,964
Stanford University	*Must contain* “Stanford” & “University”	“Stanford University Leland Junior The Board of Trustees Of”; “(The) Board of Trustees of the Leland Stanford Junior University”; “Stanford University”	4,223,123
Duke University	“Duke University”	“Duke University (Inc)”; “Duke University Medical Center”	4,174,464
University of Pennsylvania	“University of Pennsylvania”	“University of Pennsylvania”; “(The) Trustees of the University of Pennsylvania”	4,159,418
Ohio State University	“Ohio State University”	“(The) Ohio State University”; “(The) Ohio State University Research Foundation”	4,064,780
Pennsylvania State University	“Pennsylvania State University”; “Penn State Research Foundation”	“(The) Pennsylvania State University”; “(The) Penn State Research Foundation (Inc)”	4,054,311
Massachusetts Institute of Technology	“Massachusetts Institute of Technology”	“(The) Massachusetts Institute of Technology”	3,861,096
Cornell University	“Cornell University”; “Cornell Research Foundation”	“Cornell Research Foundation (Inc)”; “Cornell University”	3,810,620
University of Minnesota	“University of Minnesota”	“University of Minnesota The Regents Of”; “Regents of the University of Minnesota”; “University of Minnesota”	3,808,736
University of Florida	“University of Florida”	“University of Florida Board of Regents”; “University of Florida Board of Regents”; “University of Florida Research Foundation (Inc)/(Incorporation); “University of Florida”	3,535,596
Washington University in St. Louis	“Washington University” [exc. strings that begin with “George”]	“(The) Washington University”; Washington University School of Medicine”; “Washington University In St Louis”	3,352,338
University of Pittsburgh	“University of Pittsburgh”	“University of Pittsburgh”; “University of Pittsburgh of	3,322,083

		the Commonwealth System of Higher Education”	
Texas A&M University	*Must contain* “Texas A” & “M University”	“(The) Texas A&M University System”; “(The) Texas A M University System”; “(The) Texas A & M University System”	3,295,453
Columbia University	“Columbia University”	“Columbia University”; “(The) Trustees of Columbia University in the City of New York”	3,233,214
University of Arizona	“University of Arizona”	“University of Arizona Foundation”; “University of Arizona”; “(The) Arizona Board of Regents on Behalf of the University of Arizona”	3,199,467
University of Illinois at Urbana-Champaign	“University of Illinois” [exc. strings that include Chicago or Springfield]	“(The) Board of Trustees of the University of Illinois”; “University of Illinois”	3,145,162

Using this list, we then identified the five most frequent patent classes, which became the “Top 5E.” These classes are 600, 424, 435, 514, and 530. We refer to this as “Comparison Group 2” when reporting results.

Assignee Data

The next file we brought in was the assignee data. This came from two sources: the actual assignee “fragment” (assignee.zip) and Lai et al.’s final dataset (invpat_final.zip). The fragment contained multiple assignees for most patents, but Lai’s consolidated dataset had only one observation listed per patent. Comparing observations for the same patent numbers from both datasets revealed that the single assignee to be listed with the patent was the one with the greatest value of the variable *asgseq*. We adopted this convention and removed the other assignees, then merged the file into Lai’s dataset.

An issue with the assignee “fragment” data was the variable *asgtype*, which tracked the type of entity that had been assigned the patent—foreign corporation, U.S. individual, etc. Some patents were listed as unassigned, type 1; others were listed as type 0, which is not defined anywhere in the documentation; or were missing a type altogether. Searching Google Patents revealed that the majority of these patents were assigned to their original inventors. We used this fact and the data we had on the inventor’s country of origin to classify these ambiguous assignees as either type 4 or 5, respectively American and foreign individuals.

We also used Google Patents to get information on reassignment events, i.e. when patents changed ownership. This was done using the Legal Events table on each patent’s page, which contains (among other things) the date of each event and a code that categorizes it. Events coded as “AS” are assignment events, so by determining the publication date from the top of the page and comparing it to the event date corresponding to each assignment event, we were able to pull only the assignments that occurred after the patent was published. We identified these as reassignments and integrated them into our data.

Class Data

We continued the cleaning process with the class data (class.zip). There were multiple entries for most patents, but one was designated as the "primary" category. We dropped the others, since this data would only be used for identifying whether the patents were finance-related or in one of the five most common non-finance categories. Unfortunately, almost 10,000 duplicates remained after dropping these entries.

The duplicates fell into two categories—most had the same class and near-identical subclasses, except that one had a period (e.g. 605 and 60.5). This was almost certainly a data entry discrepancy. We kept the entry with the period in the subclass and discarded the other one. A small remainder, about 900 entries (or ~450 patents) had distinct classes or subclasses. Looking up all 450 patents in the Google Patents database was prohibitively time-intensive, so we dropped both entries. Because the dropped patents were an extremely small portion of the data and an even smaller portion of the analyzed subset the effect should be negligible.

Backward and Forward Patent Citations

We then merged the remaining observations by patent number with the citations (citations75_99 and citations00_10), assignee, and description files. Before narrowing the dataset further, we identified the forward and backward citations. Forward citations occur when a later patent refers to the observed patent as prior art. Backward citations refer to the number of patents to which the observed patent refers as prior art.

Counting backwards citations was relatively straightforward—we used the *bysort* command in Stata to consider each patent number as a separate category, then generated a new variable that was equal to "_N," the number of observations in that category. Forward citations were very similar, except that we sorted by "citation," the variable that denoted which patent was *being* cited. In this case, "_N" measured how many times each patent number showed up in that column.

Non-Patent Literature

Once we had the cleaned, final dataset from the Dataverse sources, we gathered additional data from Google Patents. We wrote a type of Python script known as a web scraper and/or web crawler. Essentially, this script takes in a list of patent numbers (a csv file exported from our dataset in Stata) and accesses the source code for Google Patents, then combs through them for the information we want and saves it to a new csv file.

We then repurposed the web scraper to greatly increase the robustness of the list of journals we were using to identify non-patent citations as academic. Almost half of these had some abbreviation of the journal name. Our list of journals comes from SCImago Journal & Country Rank, and is enhanced by including all of the official abbreviations as identified from the Web of Science's list of journal abbreviations (http://images.webofknowledge.com/WOK46/help/WOS/A_abrvjt.html), as well as additional sources such as the National Library of Medicine (NLM) Catalog. We note that while all journals identified within the top 25% were manually searched for at least one common form of abbreviation, the remainder of the journals are primarily limited to those identified by the Web of Science. As a result, the proportion of prior art references to top 25% journals relative to all academic references is inflated.

We considered academic citations in aggregate as well as "leading journals" within their subject areas.[36] We employed three definitions of "leading": journals with SJR scores in the top 1% of their field, top 10%, and top 25%. We then ran a script that identified all matches between journal titles/official abbreviations and the non-patent literature cited in Google Patents.[37]

Importantly, prior to 2001 Google Patents *duplicates* the high majority of non-patent references in each patent. In other words, each non-patent citation appears in the "non-patent citations" section twice, typically with slightly different punctuation. Post-2001, however, we do not encounter this problem.

To limit the number of false positives we employ the following strategy. We shift both the citation and the journal name to lowercase and remove all non-comma punctuation from the journal titles and citations, so there are no issues with case sensitivity or abbreviations inconsistently having/not having periods, colons, and so forth. We also attach common "endings" that following journal publications, found in the following list: '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', 'v', 'summer', 'spring', 'winter', 'fall', 'wntr', 'pp', 'LII', 'LV', 'LX', 'LM', 'XI', 'Elsevi', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'.

In other words, a journal in our list only appears in our results if it precedes an approved ending. This methodology gives us the highest "true match" rate while avoiding any systematic bias in our results across the various samples.

Examiner Added Citations

Another variable that we gathered from Google Patents is examiner-added citations. This detail can be identified from the patents themselves; on January 1, 2001 the USPTO changed reporting procedures to include an asterisk next to citations that were listed by the examiner.[38] To determine examiner-added citations for (a) patent and (b) non-patent citations we therefore counted the number of asterisks on each patent in our dataset.

Importantly, prior to 2001, Google Patents appears to report every patent and non-patent citation as an examiner-added reference. Post-2001, however, we do not encounter this problem. This provided further reason to limit our sample to patents in or after 2001.

---

[36] For a list of the 27 subject areas, see http://www.scimagojr.com/journalrank.php.

[37] We took two steps to deal with atypical forms of journal names. First, we converted the journal names and citations to lowercase to avoid issues with case sensitivity. Second, we removed periods, hyphens, and semicolons from both to avoid the case where, for example, our abbreviation is Journ. of Biochemistry and the citation reads Journ of Biochemistry, leading to an incorrect negative result.

[38] See http://www.uspto.gov/web/patents/patog/week52/OG/TOCCN/item-133.htm.

Part II

Since Part II of the analysis deals with assignee types, we ran into the same problem identified above with respect to assignee types in the Lai et al. data. Patents with assignee class 1, 0, or missing, were treated as individuals.

We determined it was reasonable to re-code patents that were type 0 or missing and simply lacked a value for the variable 'assignee'. There are two categories for individuals: U.S. and foreign. To determine which group the patent belonged in, we used the variable 'country' to create nationality indicators (one of which was 'American') and then called the assignee American or foreign based on the value of the American indicator. We created two matched samples using the comparison groups discussed above, both of which had N = 5,598. This included the 2,799 finance patents in the sample and a non-finance match for each one. Matches were determined by assignee type and grant year; after reducing the sample to our "pool" of finance and comparison group patents, we determined matches using the following process. For each finance patent, all of its potential assignee-grant year matches were labeled and assigned a random number from a uniform distribution over [0, 1]. This excluded any patent that had already been assigned as the chosen match for another finance patent. The non-finance patent with the greatest random number would become a match, and the others were put back into the pool. We use this process for both comparison groups and pool them for Part 3.

Non-Practicing Entities

Our list of over 3,000 NPEs was compiled from several sources. The majority came from PlainSite.org, a website intended for lawyers and finance professionals that we also used for our examiner experience variable. PlainSite had a list of parent companies and additional lists of shell companies held by each. We aggregated these lists and added it to the names we found elsewhere. The second major source was the partial list of NPEs from IPCheckups NPE Tracker List, as discussed in the report.

We identified many subsidiaries even without knowing their exact name due to the nature of our search. We used the *strpos* command (whose name comes from 'string position') in Stata, which returns the index position of one string (in our case, an NPE) in another (an assignee name). If the search term is not found, the value returned is 0. Since our condition for being found was that *strpos* returns a non-zero value, the search term had only to be *in* the string, not an exact match. As a result, a search for "Acacia" found Acacia Media, Acacia Intellectual Property, Acacia Ventures, etc. We also avoided problems with case sensitivity by converting both strings to lowercase before comparing them.

Part III

Our ability to scrape Google Patents and the USPTO website was facilitated by their standard URL templates for patents. Google Patents pages, for example, will always have a URL of the form https://www.google.com/patents/US0000000 (where the seven zeros are a patent number without its leading zero[39]). This meant we could read in a csv of patent numbers, plug each one into the URL stem, and then scrape those pages for the data we wanted. The litigation data, however, came from a website called Westlaw.com that did not use standardized URLs. This prevented us from using the methodology employed on the other sites.

The filings on Westlaw are sourced from the LitAlert database, a popular choice in the patent litigation literature (Lerner 2010; Lanjouw and Schankerman 2002). It includes all filings in each lawsuit, so this variable is a proxy for the intensity of the litigation, not the number of lawsuits filed.

9.2. Appendix B: Brief Background of Additional Variables

To measure quality, we examine a number of patent characteristics that have been used in prior academic literature. Brief descriptions are as follows:

1. **Claims**

   The number of patent claims has been positively associated with the private value of patents, in particular the value of the patent to its owner (Allison et al. 2004; Bessen 2008; Moore 2005). Authors have varying interpretations of this variable. Some argue that having more claims is a proxy for the "scope" or "breadth" of the patent, while others see the number of claims instead as a signal of effort on behalf of the applicant since claim drafting can be very expensive (Moore 2003; Moore 2005).[40] Still others have found that different types of assignees generally include more or fewer claims. In particular, Novelli (2015) found that the number of claims contained in a patent is partially a function of the scientific and related inventive experience of the firm that owns the patent. Finally, claim data may actually be confounded by the nature of the technology itself. As also explained by Lanjouw and Schankerman (1999), more claims will be needed to precisely define property rights in "crowded" technological areas where the potential for infringement increases if patents are overly broad. Despite this lack of consensus regarding the factors underlying the correlation, claims data are often used as a measure of patent value.

2. **Backward patent references**

   Empirical studies have found a positive association between the number of patent prior art references in a patent and its economic value (e.g., Moore 2005; Harhoff et al. 2002). Other studies, however, suggests that patent applicants may load references—disclosing information that is just "remotely relevant to the claimed subject matter" (Cotropia 2009,

[39] The leading zero is used to designate utility patents. We limited our sample exclusively to these patents, so we do lose information or misidentify anything by omitting that first character.

[40] Moore (2003, 1544) stated: "A patentee could file a patent with a single very broad claim or 50 narrow claims. Drafters of patents sometimes file many narrow claims because they cannot succeed with a single broad claim. It is thus impossible *a priori* to determine whether the difference between patents with 15 claims on average and those with 12 claims on average indicates that one set or the other is narrower."

770-772), which therefore says little about the quality of the citations. Lanjouw and Schankerman (1999, 2002) further found that a high number of backward patent citations per claim is likely indicative of technologies in well-developed areas. As a result, patent prior art references are generally considered noisy indicators of a patent's value.

3. Citation originality

Patent originality tracks the breadth of cited technologies. If backward patent citations span a wide/narrow array of patent classes, the originality index will be higher/lower. Originality has been used in a variety of economic references as a positive indicator, with the argument that more original patents build on a broader base of technological fields (Gompers et al. 2005; Lerner et al. 2011; Valentini 2012). The empirical evidence on the positive correlation of originality with patent novelty nonetheless remains murky (Allison et al. 2004; Bessen 2008; Czarnitzki et al. 2011), making this variable an uncertain indicator of patent quality.

4. Forward patent citations count

Forward citations refer to the number of times a patent is cited by later patents. Forward citations indicate that a given patent diffused knowledge that was used to develop further innovations. In other words, forward citations effectively represent a "paper trail" of knowledge flows (Jaffe et al. 1993, 578). Academic literature has consistently found that forward citations are associated, at least to some degree, with private/economic value (Bessen 2008; Harhoff et al. 1999; Hall et al. 2005) and the level of "basicness" of the underlying technology (as suggested in Trajtenberg 1990). There is thus a consensus that forward citations are another generally positive measure of patent quality.

Forward citation counts must be employed and interpreted carefully because these counts are highly time sensitive. As described in Hall et al. (2001, 25-26), patents from different cohorts vary in the intensity of so-called "truncation" issues due to (a) how close award dates are to the end of the sample period and (b) variation in Patent Office practices across time. Because each finance patent in our dataset is matched with a non-finance patent of the same grant year, we avoid differences solely derived from temporal variations.[41]

[41] Hall et al. (2001, 36) discuss the possibility that "field practices" may make comparisons of forward citations difficult and thus propose a potential adjustment technique to normalize citation totals. They explain, however, that "field effects are likely to contain a significant real component."

9.3. Appendix C: Data Limitations

Matching Error

In a number of sections of the study, we match patents, assignee names, and so forth. Automated matching is inherently imperfect when working with data, as human error can slightly alter entries in a way that prevents a good match from being recognized. We take several measures to alleviate this—removing unneeded punctuation, adding abbreviations, adding multiple name variants, and converting to titles to the same case—but in the process some false positives were identified and/or applicable data was missed. To limit these results, we excluded matches for the following journal names contained in the SCImago database: Communications; Age; Review; Management; Society; Futures; Oncology; Materials; Review; Immunology; City; System; Syntax; Surgery; RN. Journals were removed if the number of false positives generated was greater than or equal to the number of true positives (i.e. it was doing more harm than good).[42] While most of these titles were almost exclusively false positives in our matching algorithm, Surgery was a much closer call with a nearly even true-false score. Given that Class 600 (Surgery) is included in Comparison Group 2, our results for Comparison Group 2 are likely to be especially conservative.

Missing Data

Lai et al. provide their dataset in fragments, as discussed in Appendix A. We matched them based on patent number, but not every variable was available for every patent. We might find citation data without class data, for example, or vice versa. Since we needed both of those variables, we had to drop any observation missing either of them. This was true for several other variables, but given that we only used patents from the time period after the USPTO began using computers, it is unlikely that the patents we dropped were systematically related. This means our sample is unlikely to be biased by some nonrandom selection.

Finance Identification

When generating the finance indicator variable, we take the class/subclass combination 705/4 and then search for some keywords based on prior literature. This is an imperfect identification process, so a few of the patents we consider to be in the finance category could be debated.

Different Sources

Our data are based on Lai et al.'s dataset, but we generate several variables ourselves from information available online. Because the Lai dataset only goes to 2010, the scraped variables are available for more years. This means we work from different base data when generating them. Additionally, since we used multiple websites, it is possible that there are even differences in available years between, for example, Google Patents and Lai et al.

Journal Abbreviations

[42] This removal may have meant that other "problem journals" remained in the dataset. We nonetheless find that such journals are unlikely to have a systematic bias and this would be unlikely to impact the direction of the results.

As previously mentioned, our list of full journal titles comes from SCImago Journal & Country Rank. Many patent applicants, however, do not cite full journal names, but rather their abbreviations. To identify common journal abbreviations for the publications listed in the SCImago database, we first include all of the official abbreviations as identified from the Web of Science's list of journal abbreviations (http://images.webofknowledge.com/WOK46/help/WOS/A_abrvjt.html). Next, for the full list of top 25% journals, we used additional sources such as the National Library of Medicine (NLM) Catalog to identify, when available, at least one common form of each journal's respective abbreviation (roughly 96% of journals are searched with at least one abbreviation). The remainder of the journals are primarily limited to those abbreviations contained in the Web of Science list. Roughly 56% of journals outside of the top 25% are searched without any abbreviation. As a result, the proportion of prior art references to top 25% journals relative to all academic references is over-estimated.

NPE Names

A final limitation is the existence of slightly varied NPE names that are not substrings of the listed assignee name. If we had "Company XYZ" in our list, "Company X" or "Company XY" would not be flagged. Assignee names are not standardized, so as with the journal titles, abbreviations and typos will not be found. As NPEs are not generally well-known, they may be more likely to have variations on their names and therefore escape our searches. Caution should be exercised when drawing conclusions from the presence or absence of NPEs in a population.

Bibliography

Alcacer, Juan, Michelle Gittelman, and Bhaven Sampat. "Applicant and Examiner Citations in US Patents: An Overview and Analysis." *Research Policy* 38, no. 2 (2009): 415-427.

Allen, Lynne B. "The Patentability of Computer Programs: Merrill Lynch's Patent for a Financial Services System." *Ind. LJ* 59, no. 4 (1984): 633-657.

Allison, John R. and Starling D. Hunter. "On the Feasibility of Improving Patent Quality One Technology at a Time: The Case of Business Methods." *Berkeley Tech. LJ* 21, (2006): 729.

Allison, John R. and Mark A. Lemley. "Empirical Evidence on the Validity of Litigated Patents." *American Intellectual Property Law Association (AIPLA) Quarterly Journal* 26, no. 3 (1998): 185-275.

Allison, John R., Mark A. Lemley, and Joshua H. Walker. "Extreme Value Or Trolls on Top? The Characteristics of the most Litigated Patents." *University of Pennsylvania Law Review* 158, no. 1 (2009).

Allison, John R. and Emerson H. Tiller. "The Business Method Patent Myth." *Berkeley Tech. LJ* 18, no. 4 (2003): 987-1084.

Barnett, Philip and Claudia Lascar. "Comparing Unique Title Coverage of Web of Science and Scopus in Earth and Atmospheric Sciences." *Issues in Science and Technology Librarianship* 70, (Summer 2012).

Bessen, James. "The Value of US Patents by Owner and Patent Characteristics." *Research Policy* 37, no. 5 (2008): 932-945.

Callaert, Julie, Bart Van Looy, Arnold Verbeek, Koenraad Debackere, and Bart Thijs. "Traces of Prior Art: An Analysis of Non-Patent References found in Patent Documents." *Scientometrics* 69, no. 1 (2006): 3-20.

Chinchilla-Rodríguez, Zaida, Grisel Zacca-González, Benjamín Vargas-Quesada, and Félix Moya-Anegón. "Latin American Scientific Output in Public Health: Combined Analysis using Bibliometric, Socioeconomic and Health Indicators." *Scientometrics* 102, no. 1 (2015): 609-628.

Cimini, Giulio, Andrea Gabrielli, and Francesco Sylos Labini. "The Scientific Competitiveness of Nations." *PloS One* 9, no. 12 (December 10, 2014).

Cockburn, Iain, Bronwyn H. Hall, Woody Powell, and Manuel Trajtenberg. "Patent Data Project-NSF Proposal." (2005).

Cohen, Julie E. "Reverse Engineering and the Rise of Electronic Vigilantism: Intellectual Property Implications of Lock-Out Programs." *S. Cal. L. Rev.* 68 (1995): 1091-1202.

Cotropia, Christopher A. "Modernizing Patent Law's Inequitable Conduct Doctrine." *Berkeley Tech. LJ* 24 (2009): 723-783.

Cotropia, Christopher A., Mark A. Lemley, and Bhaven Sampat. "Do Applicant Patent Citations Matter?" *Research Policy* 42, no. 4 (2013): 844-854.

Czarnitzki, Dirk, Katrin Hussinger, and Cédric Schneider. "'Wacky' Patents Meet Economic Indicators." *Economics Letters* 113, no. 2 (2011): 131-134.

Falagas, M. E., V. D. Kouranos, R. Arencibia-Jorge, and D. E. Karageorgopoulos. "Comparison of SCImago Journal Rank Indicator with Journal Impact Factor." *FASEB Journal* 22, no. 8 (Aug 2008): 2623-2628.

Gompers, Paul, Josh Lerner, and David Scharfstein. "Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999."*The Journal of Finance* 60, no. 2 (2005): 577-614.

González-Pereira, Borja, Vicente P. Guerrero-Bote, and Félix Moya-Anegón. "A New Approach to the Metric of Journals' Scientific Prestige: The SJR Indicator." *Journal of Informetrics* 4, no. 3 (2010): 379-391.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." (2001).

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg. "Market Value and Patent Citations." *RAND Journal of Economics* (2005): 16-38.

Harhoff, Dietmar, Francis Narin, Frederic M. Scherer, and Katrin Vopel. "Citation Frequency and the Value of Patented Inventions." *Review of Economics and Statistics* 81, no. 3 (1999): 511-515.

Harhoff, Dietmar, Frederic M. Scherer, and Katrin Vopel. "Citations, Family Size, Opposition and the Value of Patent Rights." *Research Policy* 32, no. 8 (2003): 1343-1363.

Hunter III, Starling David. "Have Business Method Patents Gotten a Bump Rap: Some Empirical Evidence." *Journal of Information Technology Theory and Applications* 6, no. 1 (2004).

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty. "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." *AEA Papers and Proceedings* (May 2000): 215-218.

Kim, Kathleen. "Legal Research Site Names 2,000 Patent Trolls." *Inc.com* (December 2012).

Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, Lee Fleming, "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)", Harvard Dataverse, V5, 2011, at http://hdl.handle.net/1902.1/15705.

Lanjouw, Jean O. and Mark Schankerman. "Enforcing Intellectual Property Rights." *Working Paper 8656* (December 2001).

———. "The Quality of Ideas: Measuring Innovation with Multiple Indicators." *NBER Working Paper no. 7345* (September 1999).

Lanjouw, Jean O. and Mark Schankerman. "Enforcing Patent Rights: An Empirical Study." (2002).

Lerner, Josh. "Where does State Street Lead? A First Look at Finance Patents, 1971 to 2000." *Journal of Finance* 57, no. 2 (2002): 901-930.

Lerner, Josh, Morten Sorensen, and Per Strömberg. "Private Equity and Long-Run Investment: The Case of Innovation." *Journal of Finance* 66, no. 2 (2011): 445-477.

Lerner, Josh. "The Litigation of Financial Innovations." *Journal of Law and Economics* 53, no. 4 (2010): 807-831.

Li, Guan-Cheng, Ronald Lai, Alexander D'Amour, David M. Doolin, Ye Sun, Vetle I. Torvik, Z. Yu Amy, and Lee Fleming. "Disambiguation and Co-Authorship Networks of the US Patent Inventor Database (1975–2010)." *Research Policy* 43, no. 6 (2014): 941-955.

Mañana-Rodríguez, Jorge. "A Critical Review of SCImago Journal & Country Rank." *Research Evaluation* (2015): 1-12.

Moore, Kimberly A. "Xenophobia in American Courts." *Northwestern University Law Review* 97, no. 4 (2003): 1497-1550.

Moore, Kimberly A. "Worthless Patents." *Berkeley Technology Law Journal* 20, no. 4 (September 2005): 1521-1552.

Narin, F. and D. Olivastro. "Technology Indicators Based on Patents and Patent Citations." Chap. 15, In *Handbook of Quantitative Studies of Science and Technology*, edited by van Raan, A. F. J., 465-508. Amsterdam: Elsevier Science Publishers B.V., 1988.

Novelli, Elena. "An Examination of the Antecedents and Implications of Patent Scope." *Research Policy* 44, no. 2 (2015): 493-507.

Organisation for Economic Co-operation and Development. *OECD Patent Statistics Manual*. Paris: OECD, 2009.

Rantanen, Jason. "Patents, Litigation and Reexaminations." *Patentlyo.com* (December 29, 2011).

Raskind, Leo J. "State Street Bank Decision: The Bad Business of Unlimited Patent Protection for Methods of Doing Business, the." *Fordham Intell. Prop. Media & Ent. LJ* 10, no. 1 (1999): 61-104.

Roach, Michael and Wesley M. Cohen. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." *Management Science* 59, no. 2 (2013): 504-525.

Spulber, Daniel F. "Should Business Method Inventions be Patentable." *J. Legal Analysis* 3, no. 1 (Spring 2011): 265-340.

Thomas, John R. "Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties." *U. Ill. L. Rev.* (2001): 305-353.

Trajtenberg, Manuel. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *RAND Journal of Economics* (1990): 172-187.

Valentini, Giovanni. "Measuring the Effect of M&A on Patenting Quantity and Quality." *Strategic Management Journal* 33, no. 3 (2012): 336-346.