NBER WORKING PAPER SERIES

THE USE AND MISUSE OF PATENT DATA:
ISSUES FOR CORPORATE FINANCE AND BEYOND

Josh Lerner
Amit Seru

Working Paper 24053
http://www.nber.org/papers/w24053

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2017

The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond
Josh Lerner and Amit Seru
NBER Working Paper No. 24053
November 2017
JEL No. G30,O34

<u>ABSTRACT</u>

Patents and citations are powerful tools for understanding innovative activity inside the firm, and are increasingly use in corporate finance research. But due to the complexities of patent data collection and the changing spatial and industry composition of innovative firms, biases may be introduced. We highlight several patent-level biases induced by truncation of reported patent awards and citations, affecting estimates of time trends and patterns across technology classes and regions. We then introduce measures of patent and citation biases. When aggregated at the firm level, these survive popular methods of adjustment and are correlated with firm-level characteristics. We show that these issues can lead to problematic – and ex ante predictable – inferences, using several examples from prominent streams of finance literature that use patent data. We suggest a number of concrete steps that researchers can employ to avoid biased inferences.

Josh Lerner
Harvard Business School
Rock Center 214
Soldiers Field
Boston, MA 02163
and NBER
jlerner@hbs.edu

Amit Seru
Stanford Graduate School of Business
Stanford University
655 Knight Way
Stanford, CA 94305
and NBER
aseru@stanford.edu

Introduction

In the past several years, an increasing number of papers in the finance, accounting, and related literatures have made use of patent data. This growth has reflected the broadening of the topics seen as relevant to corporate finance researchers. As Zingales (2000) argued,

> the wave of initial public offerings of purely human capital firms, such as consultant firms, and even technology firms whose main assets are the key employees, is changing the very nature of the firm…. The changing nature of the firm forces us to reexamine much of what we take for granted in corporate finance.

Not only is innovation critical in many cases to firm survival—witness the fates of firms that failed to innovate successfully, such as Kodak, Motorola, and Xerox—but it illustrates the critical issues that motivate corporate finance theory more generally. Topics such as uncertainty, information asymmetries, and the intangibility of assets are central when it comes to financing innovative firms and projects.

The growth of interest in this topic among finance researchers can be seen from Figure 1, which depicts the number of papers identified in Google Scholar that both cite at least one article in the *Journal of Finance* and contain the phrase "patent citations," the references in patents to earlier work added by patent examiners and inventors that serve as "property markers" delineating the scope of the granted claims. The steady growth in interest is apparent, with more than a hundred-fold increase in the number of papers since 1996. Even when these papers are presented as a fraction of all papers citing a *Journal of Finance* paper, the share has increased substantially, from less than one-tenth of one percent to approaching one percent. In Appendix 1, we list almost 70 papers using patent data, which have appeared in one of the "top three" finance journals—the *Journal of Finance*, the *Journal of Financial Economics*, and the *Review of Financial Studies*—between 2005 and 2017.

In many cases, the papers have used these data to shed fresh insights onto important problems. But in other instances, the interpretation of the results has been marred by a failure to understand some of the peculiarities of patents and patent data. These misunderstandings have led to conclusions that are not robust to the use of alternative methodologies. Moreover, the biases that appear are frequently highly predictable.

The presence of such mistakes is understandable. The patent application and review process is extremely complex. The construction and features of the key database used for patent research—which originated at the National Bureau of Economic Research in 1999 and has been updated to 2006—have not been as fully documented as would be desirable. Rather, much of the knowledge about the use of patent data has been an oral tradition, shared in workshops of the NBER's Productivity, Innovation and Entrepreneurship group.

This paper is an attempt to rectify this omission. This paper consists of three primary parts. After an introductory first section for researchers less familiar with patent data, we begin in the second section by discussing the key patent-level features that can lead to problematic inferences from these data. These have to do with the truncation of patent data in ways that vary with the technology class being pursued and the region and industry of the inventor. We document the reasons for these distortions and how they can affect researchers. We also highlight the two broad classes of corrections used to address these issues, the "fixed-effects" and "quasi-structural" approaches.

In the third section, we explore the consequences of these biases in patent and citation data for firm-level analyses. We use information on patent grants in the 2006 NBER patent database and compare it with the newer data on patents granted to the same firms applied for during the same time period. The newer data is collected through the end of 2012 using the method employed

3

in Kogan, et al. (2017). It therefore gives us a time window post-2006 to assess if patents that were applied for in earlier years did eventually get granted. The difference in the actual patents granted relative to what was recorded in the NBER data is what we call "patent bias". Similarly, we compute "citation bias," where we compare citations to patents in the NBER data that ends in 2006 with the citations garnered by the same patents over a longer period (until 2012). We first demonstrate that these biases are large and systematic: they are present more dramatically in recent years, in some technology classes, in some industries, and in some regions. We show that the popular methods in the literature to account for these biases only partially adjust for them.

After characterizing these biases in detail, we explore how they impact inferences when analyzing patenting activity at the firm level. One solution for accounting for these biases at the firm level is to ignore them. The rationale could be that, when these biases are aggregated at the firm level, they end up being classical measurement error: i.e., they do not impact coefficients of the explanatory variables when patents or citations are used as dependent variables. We show that this is not the case. In particular, these biases at the firm level are strongly correlated with firm characteristics that are of key interest to researchers, both when we look at unadjusted and adjusted biases. In particular, market capitalization, the R&D-to-sales ratio, and the ratio of cash to total assets are positively associated with patent and citation bias, while bid-ask spread is negatively associated with such biases. These biases are also related to technological and regional characteristics associated with the firm. Thus, in many empirical settings where firm-level innovation is explored, several inferences about the phenomenon under study might be driven by non-classical measurement error.

To illustrate how these issues can lead to problematic – and ex ante predictable – inferences, we use several examples from prominent published papers that use patent data. In

particular, we highlight the implications for analyses of the impact on innovation of (a) banking deregulation, (b) firm cash holdings, and (c) the adoption of state antitakeover protection. In each case, we show that a failure to properly control for the issues highlighted earlier can lead to incorrect – and ex ante predictable – inferences.

In the penultimate section, we discuss four additional issues that may affect corporate finance researchers who wish to use patent data. These issues are perhaps more subtle, but may also have a substantial impact. We end with a checklist that finance and other management researchers may wish to use as they formulate a research project using patent data.

It may be argued that the issues confronting users of patent data are similar to ones that those analyzing almost any contemporaneous database face. For instance, issues of truncation are commonplace: for a discussion of these issues in research into financial misconduct, see Dyck, Morse, and Zingales (2010) and Karpoff, et al. (2014). But the dramatic changes in the direction and location of technological innovation (and patenting practice) over recent decades have led to a situation where these data limitations lead to highly predictable biases in the results of patent-based analyses. Given the frequency with which these issues have surfaced in the published articles and working papers using patents in the finance literature, it is the goal of this article to document these biases' characteristics and consequences.

It is also worth highlighting what this paper does not do. It is not a review of the key empirical features of patent grants and their economic applications: Jaffe and Trajtenberg's classic volume (2002) is the "go to" reference for such an analysis.[2] Nor is it a review paper summarizing

---

[2] This paper overlaps to some extent with a subsequent (at least in term of initial presentations and draft) paper by Dass, Nanda, and Xiao (2017), but our paper takes a more general approach to these issues and provides a more detailed description of the biases and why they survive popular adjustments in the literature.

the crucial works using patent data. Far more details about the patent application process can be found in many legal texts. The paper does not attempt to explore the issues associated with non-U.S. patent data, whose use in some cases can alleviate some (though not all) of the issues highlighted here. This decision is rooted in the twin desires to keep this paper manageable in length and the fact that the overwhelming majority of the papers in the finance literature listed in Appendix 1 have analyzed U.S. data. Despite these limitations, it is our hope that this work proves helpful to academics in corporate finance and related fields seeking to exploit the richness available in patent data.

1. The Necessity and Nature of Patent Data

This section provides a general introduction to patent data. The researcher more experienced with patent data may wish to skip directly the second section, where we turn to a discussion of the major potential biases associated with its use. The less experienced reader may wish to refer to the description of the patent award process in Part A of the Online Appendix.

The Necessity of Patent Data

It might be thought that innovation can be studied by examining research and development expenditures. But these measures are highly imperfect for three reasons:

- First, firms need only report R&D if the expenditures are "material." It is thus difficult to interpret non-responders: does this mean that no research was performed, or that it was for some reason interpreted as immaterial? For instance, under U.S. tax law, service firms are generally unable to take advantage of the R&D tax credit. As a result, institutions such as

major investment banks, which may employ dozens of PhDs and have well-defined new product development groups, nonetheless often report no R&D expenditures.

- Second, R&D expenditures are typically not broken down by product line or geography: rather, firms just give an indication of activity at a firm-wide level. Thus, any detailed analysis of divisional or geographic differences within a firm will be stymied.

- Finally, R&D expenditures are an innovative input, rather than an output. The effectiveness of the research may vary tremendously. For instance, between the 2001 and 2011 fiscal years, Nokia spent more than three times the amount on R&D than Apple did, yet languished in its ability to introduce innovative, market relevant products.

All these problems can, at least in theory, be addressed by patent data. Most key discoveries are protected through patent filings. Patent awards provide a wealth of technological, geographical, and industry data. The relationship between an invention's economic importance and patent data, particularly citations, is well documented. All these considerations are leading to a greater interest in patent data by financial economists and related management researchers.

Patent Data for Researchers

The first, most fundamental U.S. database is from the U.S. Patent and Trademark Office (USPTO) itself. This information can be accessed online at http://www.uspto.gov. The database covers patents awarded between 1976 and today. Earlier patents are included as well, but only in PDF format.

These data pose several issues. First, there is no identifier that uniquely flags each applicant. Moreover, a huge number of variants of each name appear. In part, this reflects inconsistency on the part of the applicants, but it also reflects sloppiness on the part of USPTO.

For instance, among the patent assignment data contains several hundred variants of IBM, differing in punctuation, spelling, and the use of corporate and legal suffixes (Thoma, et al., 2010).

An additional problem is that these data are difficult to use. While it is possible to extract the records into a file suitable for regression analysis using PERL or a similar program, these data are not "user friendly."[3]

The NBER Patent Citation Dataset—created under the leadership of Bronwyn Hall, Adam Jaffe, and Manuel Trajtenberg (HJT)—was designed to address these difficulties, as well as the fact that USPTO data then (and still now) is not in the easiest form for undertaking research. The original database sought to capture the key information on each utility patent awarded between 1963 and 1999 in a readily accessible database. (About 99% of all patents issued are utility patents; there are also design, plant, and a few other specialized categories of patents.)

The authors created a number of original measures. They assembled a six-class classification scheme, which consolidated the many hundreds of patent classes employed by the USPTO into broad categories, such as drugs and medical, computers and communications, mechanical, and so forth. The authors recorded the grant date and the application year (though only the final application date). They computed the generality and originality of patent issues, two citation-based proxies for the fundamental nature of awards.

---

[3] In point of fact, the USPTO makes it difficult to do such data extractions through limitations on daily downloads. Some authors have figured out ways to circumvent these restrictions, though consequences of Aaron Schwartz's similar efforts to download the Federal PACER database suggests some of the potential risks of this approach (Singel, 2009). The Google patent data discussed below are much more conducive to such extractions.

The primary contribution of the NBER database, however, related to their research on the patent assignees. In particular, the authors looked carefully at the first assignee of each patent (other assignees were ignored). They first reported the USPTO's broad classification of the patent assignees, indicating whether the assignees were individuals, governments, and other entities, as well as domestic and foreign. They then sought to link the U.S. publicly traded entities to their Compustat CUSIP identifiers. They used the 1989 Compustat file, so the coverage deteriorates over time: for patents granted in the mid-1980s, about 65% of all patents with a U.S. inventor were matched, but among those granted in 1999, the share falls below 50%. This attrition reflects the entry into patenting of numerous firms that were not publicly listed in the 1980s.

The authors also looked at subsidiaries by using the 1989 edition of the *Who Owns Whom* directory (now known as the *D&B WorldBase® - Who Owns Whom*). While this volume identified the major operating subsidiaries of firms, it did not really capture the maze of financial and shell corporations to which firms use to obscure the ownership of patent awards.

This project also examined citations. The main data set also tabulated the numbers of citations made and received (between 1975 and 1999) for each award. The pairs of citing and cited patents were recorded in a separate data file (again, restricted to those made by patents awarded between 1975 and 1999).

There have been a number of updated versions of the NBER data, which seek to address a variety of these limitations. The most recent of these is the file extending through 2006, compiled under the leadership of Bronwyn Hall and Jim Bessen (https://sites.google.com/site/patentdataproject/Home). This effort addressed several of the issues associated with the earlier data base (Bessen, 2009):

- Patents and citations were included through the end of 2006.

- If a patent had multiple assignees, all were all included.

- Patents other than utility awards were included in the sample.

- International Patent Classification classifications were included, in addition to detailed U.S. subclasses.

The most significant progress, however, was made on matching assignees to Compustat identifiers (using GVKEYs, the more "permanent" of the two firm identifiers used in Compustat). In particular, they took the following steps which substantially increased the details about and number of matched patents:

- Creating a separate file that links cases where a firm has multiple GVKEYs (e.g., due to the fact that it has multiple securities trading) to a single firm identifier, which was then associated with each patent.

- Adding an ownership chain. To the extent that firms were acquired and their ownership changed, the successive GVKEYs were identified. It should be noted that these were identified through the Thomson SDC M&A database, which misses many smaller transactions involving private firms, but should have good coverage of publicly traded firms such as the ones tracked here.

- Extending the number of matches between assignee names and Compustat. This was done by:

  o Using the same mapping between Compustat identifiers and assignee names as employed in the 1989 data-set, and applying it to the more recent awards.

  o Using a computerized algorithm which stripped suffixes (e.g., Inc. or LLC) and standardized abbreviations, and identifying exact matches between Compustat and patent assignee names.

o Identifying inexact matches that were nonetheless assigned a high probability of being matches by the program, and then manually examining the Compustat entries and patent records.

The 2006 data update did not, however, revisit the mapping between parent and subsidiary firms. To the extent that the mapping in the 1989 firm no longer was accurate due to subsequent acquisitions and divestitures, corporations might be assigned fewer or more patents than they actually received. Of course, as long as firms assign patents to the ultimate corporate parent rather than subsidiaries, this issue should not surface.

Since the completion of the 2006 NBER database, there have been a number of efforts to update and enhance these data. One notable effort is the U.S. Patent Inventor Database (also known as the HBS Patent Database), which sought to rationalize the (frequently inconsistently reported) names of individual inventors. The database's features are described in Li, et al. (2014).

Another issue, not addressed by either of the NBER patent datasets, relates to earlier patents. Pre-1963 patents are not available from either version of the database (the 2006 version only extends back to 1976), and are included on the USPTO web site only in scanned (PDF) form. However, these patents have been digitized by Google, albeit imperfectly due to the limitations of its text recognition software. These have been used in a variety of recent papers, including Moser and Voena (2012) and Kogan, et al. (2017). Kogan, et al. make this dataset available at https://iu.app.box.com/patents. In these cases, the authors have done manual matches to the publicly listed firms, as no concordance exits.

The state of development of data from other patent offices is much less mature. The area with the greatest activity has been in Europe. The European Patent Office makes its data available on-line and through CDs, and certain aspects of the European patent system (e.g., re-examinations)

have been the subject of academic scrutiny (Graham, et al., 2004). But the development of an EPO research database remains a work in progress: an initial mapping of UK firms' filings has been undertaken by Grid Thoma and co-authors, as well as a mapping between the names in Bureau van Dyck's Amadeus dataset and European patent assignees (http://www.epip.eu/datacentre.php). While there have been recent efforts to make Chinese and Japanese data available online as well, this information remains much less well scrutinized.

2. The Central Challenges

We will now highlight several issues with patent data that researchers in corporate finance and related disciplines face. The first set of pitfalls derive from the impact of changes in patenting over time and truncation; the second, from differences across technology classes; and the third, from differences across regions. As we will discuss, while these problems are known to some researchers – and there are some popular ways to deal with these – it is difficult to account for these problems entirely when conducting "firm-level" analysis in corporate finance and related research. Knowing the nature of these biases, however, does allow one to ex ante predict how inferences in various empirical settings might be impacted. As we will discuss in Section 6, one can conduct a battery of tests to ensure that the inferences established in a given empirical setting are not significantly impacted by these biases.

2.A The Impact of Time

A failure to properly correct to the time period of the award can lead to two difficulties. The first of these is engendered by the changing pattern of patenting over time; the second, by the truncation of entries in the standard databases. We will discuss each of these two issues in turn.

12

The past three decades have seen a dramatic acceleration in patenting activity in the United States and elsewhere in the world. Figure 2(a) depicts the number of patent awards in the U.S. in the NBER 2006 database, and highlights the three-fold increase between 1975 and 2006. If we look at patent applications (whether ultimately successful or not) during the same period, there is a four-fold increase, as Figure 2(c) reveals. (This data series, unlike the others in the paper, is drawn from the annual reports of the USPTO, not the NBER database.)

Practitioner accounts suggest that this increase in patent filings was a response to the increase in patent rights, rather than a reflection of an endogenous shift in the amount of innovation. This shift towards a more "pro-patent" policy has been effected partially through legislation – e.g., the Computer Software Protection of Act of 1980 and the Semiconductor Chip Protection Act of 1984 – but even more so through the decisions of the Court of Appeals for the Federal Circuit (CAFC). When the CAFC was created in 1982, its stated purpose was to be a streamlined venue for treating patent cases in a systemized manner. But as Merges (1992) notes,

> While the CAFC was ostensibly formed strictly to unify patent doctrine, it was no doubt hoped by some (and expected by others) that the new court would make subtle alterations in the doctrinal fabric, with an eye to enhancing the patent system. To judge by results, that is exactly what happened.

This claim is supported through a comparison of CAFC's rulings with previous appellate decisions in patent infringement cases. Between 1953 and 1978, circuit courts affirmed 62% of district court decisions holding patents to be valid and infringed, and reversed 12% of the decisions holding patents to be invalid or not infringed (Koenig, 1980). In the years 1982-90, the CAFC affirmed 90% of district court decisions holding patents to be valid and infringed, and reversed 28% of the judgments of invalidity or non-infringement (Harmon, 1991).

This secular trend in patenting has a profound impact on patent analyses. To the extent that a comparison without proper controls is made in which one group of firms is primarily engaged in patenting from the 1970s, while the other set of firms generates patents in the 1980s and 1990s, it will lead almost inevitably to the conclusion that the latter group of firms has experienced a greater increase in innovation. As we will argue, this secular trend in patenting is not uniform across technology classes or regions. Thus, simple adjustments such as time fixed effects will fail to fully account for such interactions, which can then bias inferences in predictable ways.

The second critical time effect has to do with truncation of patent data. The patent literature has generally focused on analyzing patent filings by the application year, rather than the award year. The motivation is that firms, eager to protect their intellectual property, will tend to file for patents soon after they discoveries are made. The gap between the date at which the patent is applied for and issued, however, is a product of many other considerations, such as the area of technology covered by the patent and the contemporaneous state of the patent office. To eliminate this noise, looking at patent by application date seems a more reasonable approach.

This "adjustment" is, however, not sufficient to account for the truncation problem. In particular, any analysis of patent filings near the end of the database needs to control for truncation. This is illustrated in Figure 2(b), which depicts the number of patents in the NBER 2006 dataset by application year. Because this database only reports the number of patents that issued by the end of 2006, there is a dramatic tail-off: the number of applications peaks in 2001. This has nothing to do with the actual number of filings (which, as Figure 2(c) reveals, actually continued to rise steadily), but instead with the delays in issuing patents.[4]

---

[4] It might be argued that it would be even more defensible to look at the original patent filing date: that is, the date the original patent filing was made, before taking into account divisions,

This truncation issue is even more severe with it comes to computing citations. It is rare for a patent to be cited by another patent filing before the cited patent has issued. Even after issue, the reaction is not instantaneous, as the citing patents themselves have to work their way through the application process. As Figure 2(d) reveals (again drawn from the 2006 patent data), the rate of citations per patent peaked for patents filed in 1986 and began a rapid slide by the mid-1990s. As a result, more recent cohorts of patents will be mechanically less cited, even if their degree of innovativeness does not decline.

This truncation is also clearly illustrated in Table 1, which compares the citations per patent in the 2006 NBER patent database for the cohorts of patents applied for between 1975 and 1991 and between 1992 and 2006. The mean patent in the more recent cohort has only one-third the number of citations than its older peers.

The truncation is not uniformly distributed across technology classes in which patents are generated. Figures 3(a) and 3(b) provide an illustration of the truncation issue, by reporting citations per patent for older and recent cohorts of electronics and chemical patents. In each case, there is a dramatic tail-off in citations in later years for the younger cohorts: for many of the patents in the younger cohorts, there has not been sufficient time for these patents to garner citations in the later years. We can anticipate that if we were revisit this distribution for the younger cohort (patents awarded between 1992 and 2006) in a later year, the distribution of citations by year will more closely resemble that of the older cohort. The tail-off is more dramatic for patents in the electronics subcategory.

---

continuations-in-part, and the like. But given that the various versions of NBER patent databases do not readily allow such a determination, and that such a step would doubtless intensify the truncation problems that we discuss in this essay, such an alternative approach appears impractical.

There are two primary responses to the truncation issue seen in the literature. The first of these we term the "fixed-effect" approach. We term the simplest approach, pioneered by Jaffe and Trajtenberg (1996) and HJT (2001, 2005), time adjusted: one estimates a distribution function of the citations over time using non-truncated data, then infers what the truncated data should look like. This approach was originally developed for the use with aggregate patent data, but has since been applied to individual patent data as well (e.g., Chemmanuer and Xuan's 2014 analysis of hostile takeovers). It might be thought that such inferences would be extremely noisy at the individual patent level, as very small differences in early citation rates – especially if they vary across technology class and spatially – could be amplified through such an imputation approach. As we will discuss below, differences across technologies may also pose problems. Finally, work by Nicholas (2008) and Kolev (2013) suggests that more fundamental patents have much longer "half-life" of citations than more routine extensions, which might lead us to worry that such inferences could introduce systematic biases.[5]

A variant, also pioneered by HJT (2001), which we term time and tech class adjusted, is to look at patents and citations relative to those awarded in the same technology class and year. Detailed descriptions of these adjustments are discussed in Online Appendix B. For instance, Seru (2014), in his analysis of the impact of conglomerates on innovation, uses as the ratio of the number of citations per patent for each firm to the mean citations per patent in the same cohorts as the firm's patents. Ideally, this approach will control not just for truncation problems, but also adjust

---

[5] A related approach is to only use citations in a short window after a patent award. For instance, Lerner, Sorensen and Stromberg (2011) only look at citations in the three years after awards. This avoids some of the issues delineated above, but as we see from Figure 3, early citations only capture a very small number of total citations. Thus, the information that is discarded through such an approach is potentially quite significant.

for the shifts engendered by changes in patent office policy and technological fluctuations. However, it is still subject to distortions: a single early citation may lead to a large ratio. The issue of important patents having differing citation profiles over time (as discussed above) is also a problem here. Moreover, as Hall and co-authors (2001) suggest, by undertaking such a normalization, one may be sweeping away information: for instance, if a key innovation leads to a substantial burst in innovation in a given industry.

A second class of approach is what we term the "quasi-structural" one. In summary, this method "models" the distribution of citations based on citing year effects, cited year effects and propensity to cite fixed effects for different technology classes. For brevity, we produce analyses with adjustments using citing year and cited year effects. The analysis based on propensity to cite adjustment is relegated to Online Appendix B. This method also has its share of potential problems, since it is hard to model distributions over time, which may be altered as new patents and technologies are added in the models. Details about this adjustment are discussed in Online Appendix B.

In conclusion, the time lag between the filing of a patent application and its subsequent grant results in a mechanical tail-off in patent grants towards the end of the sample. Moreover, it may be a decade or longer after an original patent is filed before when can get a good sense of how influential it is. While it is possible to adjust the number of patent grants and number of patent citations received in early years based on historical patterns – and thus project the total number of patents or amount of citations likely to be ultimately received – these estimates are quite imprecise and potentially biased. These measurement problems are compounded by the fact that these patterns may not be uniformly distributed across patents in different technology classes and generated in different geographic regions. As we will demonstrate, the measurement problems due

to truncation can have very substantial implications for empirical analyses that look at the consequences of relatively recent events on innovation.

2.B The Impact of Technology Class

A second concern has to do with a failure to adjust for the technological class of the discovery. The propensity to patent across technologies and industries can vary dramatically, and as a result the "density" of patents in given areas may be very different. Thus, some patents will be heavily cited due to their technological location, rather than their fundamental innovativeness.

These issues are illustrated in Figures 4(a) and 4(b), which compare awards in several fields. All figures in this section will use a similar structure. The first panel compares patent assigned to the HJT "computers and communications" classification – referred to henceforth as computers – with their chemicals category. The second panel contrasts electronics and mechanical patents. The figures make clear that computers and electronics patents experienced a much more dramatic run-up in patenting activity in the 1980s, 1990s, and early 2000s.

Table 2 illustrates this pattern by comparing the citations for the patents from 1976 to 2006 in the chemicals, computers, mechanical, and electronics cohorts. The substantially greater number of citations for computer and electronics is apparent, an effect that is driven by patents at the top end of the citation distribution.

We have already seen that the distribution of patents in technology classes differs across time. There are also different application-grant lags across technology classes. These could reflect differences in the amount of technical complexity, which may lead to the diffusion of technologies occurring at different rates.

Figures 5(a) and 5(b) provide an initial way of illustrating the differing truncation of patent citations across industries. It shows that the distribution of patent citations to computer and electronics firms is skewed left-ward: more citations happen sooner after the award.[6]

We illustrate these patterns in Table 3, which shows the relative rate at which citations drop off for patents in the chemicals, computers, mechanical, and electronics cohorts. To undertake this calculation, we use all patents applied for between 1975 and 2006 that were included in the NBER 2006 database: thus, there are many more patents included in the calculation of the number of citations one year after the patent grant than thirty years after. We then compute the share of citations to these patents by year since the award. Thus, the more rapid tail-off in the citations to computer and electronics patents may be due to two factors: (a) the more rapid obsolescence of these technologies, leading to a reduced propensity to cite older awards, and (b) the more recent vintage of the typical patent in this area. As we will demonstrate, the measurement problems due to not accounting for differential patterns of patenting and citing across technology classes can have very substantial implications for empirical analyses, especially ones that explore innovation across firms in similar industries using different technologies.

2.C The Impact of Region

The third major problem is related to the second: the differences in innovative activities can vary dramatically across regions. Many patent-based analyses exploit regional differences in

---

[6] These results can also be illustrated in a different way. Online Appendix F (Panel A) shows the distribution of the lag (in years) between application and grant date for patent applications across technology classes. There is some heterogeneity across classes, with computer patents having the longest lag. Panel B shows the lag (in years) between citing and cited patents across technology classes. Again, there is considerable heterogeneity across technology classes.

order to identify effects: e.g., by looking at when different states adopt certain policies. However, the distribution of innovative activities across states is quite different, reflecting more general agglomeration effects. Moreover, as will become clear, regional differences could interact with type of technology class in which patenting is conducted, as well as the time period in which innovation was performed. Thus, any analysis that hopes to explain differences in innovativeness across firms but does not control differential patenting across regions is likely to result in problematic inferences.

To illustrate the severity of the problem, we proceed here by comparing patents by assignees in a state that has frequently been on the cusp of business-friendly policy reforms, Delaware, with California and Massachusetts, which have been accused of being at the other extreme. As Figure 6 reveals, patenting only increased between 1997 and 2000 by 3% for inventors in Delaware, while for inventors in the other two states, the increase was by 15%.

Figure 7 shows that patents with assignees in California and Massachusetts were more likely to be cited. The figure also highlights the drop-off in citations is more concentrated for the California and Massachusetts patents in the very last years of the sample.

Tables 4 and 5 illustrate the same patterns. Table 4 shows that the total number of citations across these states differs: as discussed above, patents with assignees living in California and Massachusetts are more heavily cited, particularly at the upper end of the distribution. In Table 5, we show that citations take place earlier in the patents' lives in California and Massachusetts relative to Delaware awards. Again, these comparisons suggest that any naïve correction for the truncation of citations may lead to incorrect inferences.

Of course, behind these differences are considerable disparities in the industry composition of the firms active in these states. The computer and electronics firms, whose patenting we

discussed above, are far more likely to be located in California and Massachusetts. Moreover, the mixture of industries across states changes over time. As we will demonstrate, not accounting for industry composition across regions, and simply looking at the impact of a "regional" policy change, might lead us to spuriously conclude that these policies affected innovative activity.

3. Are Popular Patent-Level Adjustments for Biases Sufficient for Firm-Level Analyses?

While several methods are available to account for biases due to time, technology class, and region, these methods are primarily for adjusting for biases at the patent level. Most research in corporate finance is at the firm level. We now demonstrate that the popular methods available to account for these biases at the patent-level may not be sufficient when one aggregates patents at the firm level. In particular, we will demonstrate that the residual measurement problems that emerge are not pure noise, but rather are related systematically to firm characteristics. This makes it difficult to disentangle firm-level factors that truly impact innovative activity from spurious measurement error induced due to the problems discussed above.

To illustrate the nature of the problem, we estimate the firm-level bias that is created due to the truncation issues. We do so by computing the difference – both unadjusted and adjusted for truncation using popular methods – between patenting and citation activity of a firm in a given year as recorded by the end of the 2006 NBER data relative to the patenting and citation activity of the same firm in the same year as recorded in "our data" that tracks patents granted through the end of 2012. Our dataset is constructed by scraping the patent records directly from 1976 through 2012, using a similar procedure as in Kogan, et al. (2017).

More specifically, we construct the unadjusted "patent bias" for each firm-year by comparing the number of patents for each firm filed in each application year in our data (thus, which have been granted by 2012) and in the NBER 2006 dataset (i.e., granted by 2006). It should be noted that this measure will understate the true extent of the truncation problem since there will be patents granted subsequent to 2012 based on applications from 2006 and before. We repeat a similar exercise to compute "citation bias" for each firm-year: we compare the number of citations to all the patents for each firm filed in each application year in our data (i.e., citations in patents granted by 2012 to applications filed by a firm in a given year and granted by 2006) and in the NBER 2006 dataset (i.e., citations in patents granted by 2006 to applications filed by a firm in a given year and granted by 2006).

Because we will relate these firm-level biases to firm characteristics, we confine our analysis to publicly listed firms. As will become clear, we will explore how these biases relate to the following characteristics: Firm Size (Log_Size), Market Value to Book Value (Log_M2B), R&D Investment to Sales ratio (Log_RD2Sale), Cash to Assets ratio (Log_Cash2Asset), Investment to Assets ratio (IA), Return on Assets (ROA), Return on Equity (ROE), Sales Growth (SG), Market Leverage (Log_LEV), Net Stock Issues (NSI), and Bid-Ask Spread (Log_Spread). Online Appendix C details the exact definitions of these variables and how they were constructed. There are a total of 1807 publicly listed patenting firms in our sample, with 1443 firms having no missing information.

We then explore the patent and citation biases at the firm level. Here, we provide analysis both for unadjusted biases as well as adjusted ones, where the adjustment includes the popular methods used in the literature discussed in Section 2.A – i.e., the fixed-effect methods and the

quasi-structural approach. We start by discussing how these unadjusted and adjusted biases differ across time, technology class, region, and industry.

3.A Unadjusted and Adjusted Biases in Publicly Traded Firms across Time

Figure 8, Panel A illustrates the firm patent bias over time. We sum the patent bias across publicly traded firms by application year. The resulting patent bias is more severe for more recent patents than older ones. This is because many patents that are applied close to 2006 end up getting granted between 2007 and 2012. While the adjustments – the time fixed effects and the time and tech class fixed effects – help alleviate some of the truncation problem, a significant portion of the bias remains. Several thousand missing patents remain unaccounted for, even after the adjustments.

Figure 8, Panel B shows the citation bias at the firm level over time. We again sum citation bias across publicly traded firms by year. Here, the trend is a bit different. The bias is most severe during the year 1998, which is eight years before the end of the sample period in 2006. Using information on citations granted to these patents for another six years past 2006 – which implies that we track citations for these patents 14 years after issuance – yields a large number of subsequent citations to these patents that were not captured until 2006. In contrast, the bias is not as large for patents granted as of 2006. This is likely because tracking citations for six years after issuance – that is between 2006 and 2012 – is not long enough of a time period to capture the bulk of subsequent citations, as citations tend to peak with some lag (see HJT, 2001). Thus, we are likely severely underestimating the true extent of citation bias for the patents that are granted towards the end of the sample. While the adjustments – the time fixed effects, the time and class fixed effects, and the quasi-structural method – are useful in alleviating some of the bias, a

significant bias remains. In essence, adjustments using historical data do not fully account for the time-varying dynamics at the firm level in both patents and citations.

3.B Unadjusted and Adjusted Biases in Publicly Traded Firms across Technology Class

Figure 9 (Panels A to C) shows bias at the firm level for different technology classes, again unadjusted and time-adjusted and time- and tech class-adjusted. Firms are assigned to a particular technology class in a given year based on modal primary patent class of patents produced by the firm in that year (using the U.S. patent classification system). We then sum the bias across publicly traded firms in each technology class. It is clear that the most severe patent biases are concentrated in the computer and the electronics classes. This reflects the explosion of patents in these sectors relative to other classes, especially towards the end of the sample.

Similar patterns emerge when we adjust the bias in Panel B and Panel C. Interestingly, when we adjust by the fixed-effect method, some classes display a "negative bias": the adjusted number of patents exceeds the actual number issued through 2012. This pattern may reflect the failure of the fixed effects to fully capture the rapid acceleration of computer-based patenting and the declining share of other classes.

Figure 10 shows the citation bias at the firm level for different technology classes. As before, we assign publicly traded firms into technology classes in each year and sum the citation biases. Compared with the large unadjusted citation biases in Panel A, the adjustments in Panels B through E do help. However, as was the case before, a significant part of the bias remains, particularly when it comes to computer patents. Finally, the pattern of citation bias peaking earlier in time than patent bias, discussed in Section 3.A, emerges across technology classes.

To the extent that citation and patent bias across technology classes illustrated in this section varies within and across firms, granted patents and their citations within and across firms will be less comparable as we get closer to the end of the sample period in the NBER dataset.[7]

3.C Unadjusted and Adjusted Biases in Publicly Traded Firms across Regions

Figure 11 illustrates the distribution of patent bias at the firm level across different states. We assign firms to different states based on the modal US state or territory of the assignee recorded by USPTO at the time of the application.[8] We then sum the bias by state across publicly traded firms. As can be observed, patent bias is mainly concentrated in states like California, New York, Texas, and Washington. This reflects not only the size of the states (there are many more patent applications), but also the concentration of computer firms in these states. The patterns in the unadjusted data (Panel A) remain even after adjustment in Panels B and C. Interestingly, again, we see "negative bias" in some states: we end up with fewer patents relative to what is estimated using historical data in certain states.

Figure 12 shows the distribution of citation bias at the firm level across different states. We again sum the citation bias in a state across publicly traded firms, with firms assigned to states each year as discussed above. Similar to patent bias, states like California, New York, and Texas

---

[7] Another way to illustrate this issue is based on Online Appendix G. The figure shows the distribution of granted patent applications (Panel A) and the mean number of citations per patent (Panel B) for the six HJT technology classes. This analysis is undertaken at the firm level, with firms assigned to the modal class of patents granted to a given firm in a given year. The figure highlights, when aggregated at firm level, the substantial heterogeneity in the volume and time trends of patents and citations.

[8] In a few rare instances, firms report multiple assignee states for a patent. We randomly picked a state in such situations. Doing this procedure several times assured us that the inferences made in this section are not sensitive to this choice.

suffer most from citation bias. Adjustments help to some degree, but significant bias remains. As we noted before, patent and citation bias by state is particularly evident among recent patents.

3.D Unadjusted and Adjusted Biases in Publicly Traded Firms across Industries.

It is useful at this stage to ask if firms in some industries might suffer more from these biases. As we have seen from our analysis so far, these biases occur more in some technology classes and regions than others. To the extent that firms in some industries are more active in certain technology classes and regions, we might expect these biases at the industry level as well.

Figure 13 shows the patent bias at the firm level across different industries, where the industries are defined using NAICS code. (In Online Appendix D, we plot the corresponding figures using Standard Industrial Classification (SIC) codes.) We assign firms into different industries at time of the patent application using Compustat. We then sum the bias by industry across publicly traded firms. Patent bias is mainly present in industries like manufacturing (especially class 33, which includes computer and communications equipment manufacturing), information technology, and services (which includes computer systems design and R&D services). The adjustments reduce the biases, but considerable distortions still exist. A comparison of Panels B and C shows that adding controls for technology class has very little impact on the biases across these industries: either these classes are too crude, or the differences across industries in patenting growth are largely orthogonal to the controls.

Figure 14 illustrates the citation bias at the firm level across different industries (again, see Online Appendix D for the analysis using SIC codes). The Y axis depicts the citation bias for each industry, again summed across publicly traded firms. As illustrated in Panel A, unadjusted citation

bias is mainly concentrated in manufacturing, information technology, and services. It proves difficult to eradicate even with fixed-effect and quasi-structural adjustments (Panels B through E).

3.E Bias and Firm Characteristics

Our analysis so far has illustrated the bias in patent and citation counts at the firm level that varies across time, technology class, region, and industry. An adjustment for such "measurement error" in patenting at the firm level could be to estimate regressions that account for time- and firm-invariant characteristics using time and firm fixed effects. This approach will be sufficient if such measurement error is time-invariant within a given firm or constant across firms in a given year. But our discussion so far, illustrating the complex time-varying nature of these firm-level biases across time, technology class, regions, and industries, suggests that this may not be true.

To illustrate this more formally, we estimate fixed-effect OLS regressions in Tables 6 and 7. The unit of observation in each case is firm-year observations of patent and citation bias between 1976 and 2006, with the results reported in first three columns. In each table, we also report results in the next six columns for two sub-periods between 1976 and 2006. There are three panels in Table 6, where we employ as our dependent variables unadjusted patent bias (Panel A), patent bias adjusted for time fixed effects (Panel B), and patent bias adjusted for time and technology class fixed effects (Panel C). The dependent variables are computed as the adjusted or unadjusted difference in log of one plus number of successful patents filed by a firm in a given year as of 2012 ("our data") and log of one plus number of successful patents filed by that firm in the same year as of the end of sample in the NBER 2006 dataset. Logarithms are taken to account for skewness

in patenting activity.[9] The sub-periods chosen are 1976-1996 and 1997-2006, since Figure 8A shows that patent bias becomes more severe from 1997 onwards.

Similarly, there are four panels in Table 7, where we iteratively employ as dependent variables unadjusted citation bias (Panel A), citation bias adjusted for time fixed effects (Panel B), citation bias adjusted for time and technology class fixed effects (Panel C), and citation bias adjusted by the quasi-structural method (Panel D). The dependent variables are computed as the adjusted or unadjusted difference in the log of one plus the number of citations to all patents of a firm applied for in a given year and granted by 2006 in our data and the log of one plus the number of citations to the same set of successful patents of that firm in the same application year in the NBER 2006 dataset. Restricting the successful patents from our data to only those that are granted by 2006 allows for comparison with successful patents in the NBER 2006 data. Logs are taken to account for skewness in citation activity. The sub-periods chosen are 1976-1990 and 1991-2006, since Figure 8B shows that citation bias becomes more severe from 1991 onwards.

In these specifications, we also employ time, technology class, industry, and firm fixed effects to account for characteristics that might be driving the biases in patent and citation activity. We include a host of firm-level variables (described in detail in Online Appendix C) to assess how the patent and citation bias might be related to these characteristics. This allows us to explore if – conditional on adjustments for time and technology class and accounting for time, technology, industry, and firm fixed effects – these biases are orthogonal to firm characteristics. Put another way, can the measurement error in patent and citation counts at the firm level due to truncation

---

[9] We do not use for the analysis in this section firms without any patenting activity between 1976 and 2006.

issues be called "classical" measurement error? If the error is classical, we may not have to worry about such biases confounding inferences about our explanatory variables in many cases.

Several facts emerge from the analysis. First, we see that various firm-level measures—size, R&D to sales, market-to-book ratio, cash to total assets, and spread—are economically and statistically related to these biases in most specifications. For instance, size, R&D-to-sales, leverage, and cash to total assets are by-and-large strongly positively related to both patent and citation bias. Spread is also by-and-large strongly negatively related to these biases. This relationship holds across specifications that account for time, technology, and industry fixed effects in full sample, as well as for the changing mixture of firms. This is especially true when we estimate the regressions for the entire sample or for the sub-sample where patent and citation bias are severe.

The economic magnitudes are important as well. For instance, a one standard deviation change in size is associated with a one-fifth standard deviation change in patent bias (column 7 in Table 6, Panel A). Similarly, a one standard deviation change in R&D-to-sales ratio is related to about one-tenth standard deviation change in unadjusted patent bias (column 7 in Table 6, Panel A). Along the same lines, a one standard deviation change in size and R&D-to-sales ratio is associated with a one-fourth and one-eleventh standard deviation change in the unadjusted citation bias respectively (column 7 in Table 7, Panel A). We find similar economic magnitudes when we use adjusted patent and citation biases as dependent variables in other panels.

One can rationalize the relationships in these specifications. Larger firms, as well as those that spend heavily on R&D, might produce more complex patents that require a longer time to be approved. Consequently, these patents might take longer to accrue citations, leading to a positive citation bias. One can make analogous arguments for why higher cash-to-assets (greater financial

29

strength) and lower spreads (higher liquidity) might be related to these biases. Importantly, these relationships hold after we adjust for patent and citation truncation. Even after accounting for truncation using different techniques, the variation in patent and citation bias is systematically related to firm characteristics. It is important to reiterate that the relationships between the firm-level biases in patents and citations and firm characteristics also hold after conditioning on firm and time fixed effects.

Second, as mentioned earlier, in columns 4 through 9 of Table 6 and Table 7, we analyze separately patents that are likely to suffer the most from the biases (i.e., those that are granted towards the end of the sample) and those that are likely to suffer the least from the biases (i.e., those that are granted towards the beginning of the sample). We find that the patterns on the overall sample are largely driven by the younger patents, which might be expected to have more of the biases. This reiterates the importance of time in impacting inferences as outlined in Section 2.A.

Finally, the log of the total number of granted patents in the same class as the modal class of the firm's patents (log (class patents) and the log of the total number of granted patents in the same state as the modal state of the assignee on firm's patents (log (state patents)) are also related to unadjusted and adjusted patent bias in several of the specifications in Table 6. Similarly, these two overall activity measures, log (class cites) and log (state cites), are also related to unadjusted and adjusted citation bias in several of the specifications in Table 7. These findings underscore the impact of technology class and region on inferences, as discussed in Section 2.B and Section 2.C.

Taken together, this analysis shows that the truncation problems are quite complex when patents are aggregated at the firm level. These biases are not accounted for by usual adjustment methods in the literature. Adding fixed effects at the firm, industry, and year level also are not

sufficient. Substantial patent and citation bias remains, which is systematically related to firm characteristics such as size, R&D intensity, leverage, cash-to-assets ratio, and spread. It is also related to technological and regional characteristics of firm's patenting. As a result, several inferences that researchers might attribute to a phenomenon that they are studying may instead be driven by non-classical measurement errors. In Section 4, we will argue that, in many empirical settings where firm-level innovation is explored, the impact of the correlation between the innovation measurement error and firm characteristics is real and ex ante predictable.

3.F Robustness: Newer Ways of Adjustment

In our final analysis of Section 3, we assess how robust our inferences are to the new method for adjustment of truncation proposed by Jaffe and Rassenfosse (2017). The authors suggest comparing patenting or citation activity with the "group of patents" to which the patent of interest belongs. We follow this approach and assess how it does when we aggregate patents at the firm level for publicly traded firms. The group of patents considered for this adjustment, following this new approach, is all those granted to all the publicly traded firms. This differs from the earlier adjustments, where the comparison set is the entire population of patents. In interest of brevity, we present this analysis in Online Appendix E.

In Figure E1, Panel A and Panel B show the distribution of the firm patent and citation bias over time. One can immediately observe that compared with Panel A and B in Figure 8, the patent and citation biases are alleviated somewhat. However, significant biases still remain. To further explore this issue, we examine these biases across classes (Figure E2), regions (Figure E3), and industries (Figure E4). While the biases are alleviated somewhat relative to those using all classes,

significant biases still remain. Even after applying adjustments, the biases are accentuated for some technology classes (e.g., computers) and regions (e.g., California).

Finally, we also explore whether these biases, when aggregated at the firm level, are related to firm characteristics, after applying various adjustments and using year, technology, industry, and firm fixed effects. As reported in Tables E1 and E2, we continue to find that various biases are systematically related to firm level characteristics. As before, variables such as size of the firm and R&D expenditures strongly predict variation in patent and citation biases at the firm level.

4. The Impact on Inference

When we first posed the censorship problems in Section 2, these problems may have seem like abstract ones, more of theoretical interest to economists of innovation than anything else. In Section 3, we showed how these biases are systematically correlated with firm characteristics of interest to financial economists and others. In this section, we illustrate the impact of these correlations by examining three typical applications of patent data in corporate finance research. We highlight the extent to which the results are sensitive to the potential biases highlighted above.

The first example we present illustrates the impact of neglecting changes in patenting patterns over time, as described in Section 2.A. We look at what has become a popular topic for scrutiny in recent years: the impact of bank deregulation in the United States on innovation. The existing literature (e.g., Amore, et al., 2013; Cornaggia, et al., 2015; Chava et al., 2013; Hombert and Matray, 2017), has assessed the consequences for innovation of various bank reforms.

To examine the effect of bank deregulation, we create a firm-year level panel data set with about 10,000 observations for the period between 1978 and 1995. The dependent variable, similar to the earlier studies, is citations per patent across all the patents produced by a given firm in that

year; the key independent variable of interest is a dummy variable denoting if the state had undergone interstate banking deregulation.[10] We identify deregulation events following the earlier literature. Standard errors are clustered at the state level. Controls include state and year fixed effects, as well as firm-level measures including firm size, asset tangibility, market to book, and leverage. The results are presented in Table 8.

When we run a regression over the entire time period (1975 to 1995) in column (1), the coefficient on bank deregulation is positive. (Note that in this and subsequent analyses, we truncate the observations more than a decade prior to the end of the patent dataset, in order to minimize the effects of the patent citation truncation.) The changes appear to stimulate more innovation, as proxied for by the greater number of citations per patent.

When we split the sample in columns (2) and (3) into firms incorporated in states that underwent banking deregulation prior to and after January 1, 1985, however, a very different picture emerges. The deregulation dummy takes a strongly positive sign in the pre-1985 sample, but a negative one of almost equal absolute magnitude in the years afterwards. The unadjusted count of citations per patent by year for the two sub-samples is plotted in Figure 15. Note that while in the earlier specification we had included year dummies, they did not capture the effect which is so clearly demonstrated in this figure.

What explains these dramatic differences across time in the consequences of bank deregulation on innovation? In part, this pattern likely reflects the fact that while patenting was relatively static during the initial period, during the second half it underwent a dramatic acceleration (with a corresponding increase in patent citations). Moreover, the composition of the

---

[10] The results are similar if we use log of one plus citations per patent as the dependent variable instead.

populations in the two samples differs. This analysis suggests that critical importance of assessing heterogeneity of the samples across time periods, rather than simply assuming a uniform effect.

The second illustrative analysis emphasizes the importance of properly controlling for technology class in these analyses, as discussed in Section 2.B. To explore this, we look at the relationship between firms' cash holdings and their innovative output, as measured by the number of citations per patent. This topic is one which has attracted increasing interest in the finance literature in recent years. (See, for instance, Almeidia, Hsu, and Li, 2013; Lyandres and Palazzo, 2016; and Schroth and Szalay, 2010.) These relevant studies reach a variety of conclusions.

Again, we create a firm-year panel data set with around 25,000 observations between 1975 and 1995. The dependent variable is, as is often used in this literature, is the log of one plus citations per patent across all the patents produced by a given firm in that year; the key independent variable of interest is the firm's ratio of cash to assets in a given year, which is taken from Compustat. Standard errors are clustered at the firm level. Controls include industry (at the three-digit SIC level) and year fixed effects, as well as firm-level controls for firm size, tangibility, market-to-book, and leverage. The results are presented in Table 9.

The basic pattern, shown in column (1), is that firms with more cash have more highly cited patents. But while the industry fixed effects are important, they again do not fully capture the heterogeneity in the sample. In particular, when we add (in column (2)) as an independent variable the average number of citations received by the universe of patents in the modal technology class of the patents filed by the firm in a given year, the R-squared of the regression increases from 0.22 to 0.28.

Moreover, there are enormous variations across industries. When separate regressions are run for patents across various broad modal technology classes at the firm level, the magnitude of

the coefficient is almost four times greater for computers (column (4)) than it is for drugs (column (5)). This exercise suggests the importance of including both industry and technology class effects, It also seems critical to assess the heterogeneity of firms in different technology classes across the various sub-samples being compared.

The third analysis looks at the impact of region, as discussed in Section 2.C. As we mentioned above, many analyses have been run at the state level, without adequate controls. To illustrate this point, we examine the impact of the adaptation of state anti-takeover laws on innovation, again a topic that has attracted extensive exploration (for instance, Atanassov, 2013; Becker-Blease, 2011; and Chakraborty, Rzakhanova, and Sheikhb, 2014).

Again, we create a firm-year panel data set with over 37,000 observations between 1982 and 1999. The dependent variable, as is often used in this literature, is log of one plus citations per patent across all the patents awarded to a given firm in that year; the key independent variable of interest is a dummy denoting whether the firm was incorporated in a state that had adopted state antitakeover legislation as of a given year, which is taken from the sources employed in these papers. Standard errors are clustered at the firm level. Controls include state and year dummy variables. The results are presented in Table 10.

The basic analysis, reported in the first column, suggests that firms produce less-cited patents after the adoption of anti-takeover laws. Once again, the state fixed effects are important and often statistically significant, but fail to account for significant heterogeneity across the states. We illustrate this by presenting two additional regressions, one excluding firms incorporated in California and Massachusetts (the second column) and one with only firms in those two states (the third column). When we exclude these two states, the magnitude of the coefficient on anti-takeover protection drops by six-fold and is only marginally statistically significant. Meanwhile, the

35

regressions using firms from these two states have a coefficient that is of the opposite sign and an order of magnitude higher. Figure 16 presents the citations per patent in California and Delaware by year graphically, and highlights the same differences.

Interestingly, California never adopted state anti-takeover legislation. It is essentially a control state in this analysis, yet drives the key findings: the reason why anti-takeover protection matters is because the other states that adopted these provisions did not experience the explosive growth in citations per patent that California did. As discussed above, much of the growth in patenting and citations in California was due to the heavy representation of the computer and electronics industries in that state, which experienced a boom in patenting and citations during this period. Once again, it is absolutely critical to assess heterogeneity across firms when exploring regional effects.

These three examples are meant to be merely illustrative. There are many other topics we could explore, in order to highlight the challenges that these problems can pose to corporate finance researchers when using patent data. Examples include:

- Broad policy shifts on a national level, which had particular effect some classes of firms. An example would be the consequences for innovation of a shift such as NASDAQ decimalization.

- Other policy changes where states may differ systematically in implementing, and where these differences may be associated with the mixture of innovative activities in that state. An example is reforms of labor rules.

- Corporate finance decisions that may be associated with firm and industry characteristics, such as the decision to issue equity in the public markets.

5. Additional Issues

The analysis above have highlighted how the truncation of patent data can trip up researchers. These issues are not exhaustive; there are also a number of other problems that can bedevil corporate finance researchers using patent data. In this section, we will briefly and qualitatively review four such additional challenges.

5.A The Impact of Firm Exits

The first of these challenges stems from the truncation of patenting in the years prior to a firm's acquisition or liquidation. This problem derives from the timing of patent assignments. As noted above, the assignment of the patent is not typically done until about the time a patent issues. This makes determining which firm is responsible for an invention difficult in the case where a firm has been acquired, liquidated, or there has been a corporate reorganization. In particular, the patent reveals who the owner is at the time of issuance, rather than at the time of invention. In some cases, the acquiring firm may assign patents to its subsidiary, but often all assignments are made to the new parent. It may be possible to determine which entity was responsible with coming up with the idea by examining the location of the inventor or to whom the earlier awards by that inventor were assigned. But in many cases, it is impossible to tell.

As a result, there is a significant reporting bias in the patent data. Beginning several years before a merger or acquisition, the applications of the ultimately subsumed entity are likely to begin disappearing from the patent database. Because these patents will often be assigned to the successor entity once they issue, our picture of the acquired and acquiring firms' patenting may be distorted. Moreover, since acquired firms are not representative along numerous dimensions, this reporting bias may distort our inferences.

The issues also appear when a unit is bought out from or spun out of a corporate parent. In these instances, the spun-out entities' patents may not be assigned to itself, but rather to the old parent. For instance, consider a divisional buyout. The patents applied for more than five years before the buyout are likely to be issued before the private equity investment. In most instances, these will be assigned to the corporate parent. But after that, assignments are unclear: the patents may be assigned to either the spun-out entity or the corporate parent, depending on the agreement in place between the seller and buyer (which is typically not available to the researcher). Again, the reported number of patent applications is likely to be distorted for several years around the transaction. Lerner, Sorensen and Stromberg (2011) explore the impact of several corrections: e.g., looking at patent filings by individuals who subsequently patent for the spun-off entities that were assigned to the corporate parent, and adding these back into the total for the spin-offs. Their analysis suggests that the results are very sensitive to the corrections used.

It should be noted that the USPTO operates a reassignments database, which is supposed to capture transfers of ownership across firms. This has been explored in a variety of papers, most notably by Carlo Serrano (Serrano, 2010; Galasso, Schankerman, and Serrano, 2013; etc.). But the coverage of these transactions appears problematic. This can be illustrated by an analysis by Filippo Mezzanotti of a random sample of 90 firms (with 475 patents) that were acquired in 2004 (identified via Capital IQ) and were granted with at least one patent in previous 15 years:

- 52% of patents were re-assigned to the buyer or reassigned as part of a noted "merger."
- 3% patents were sold before the merger.
- 17% were not assigned to the merger partner, but were assigned to financial institutions, apparently as a security interest in a pre-merger financing.
- 24% did not get re-assigned at all.

- 4% cannot be categorized under any of these categories.

5.B Misleading Assignments

A second problem relates to deliberately deceptive assignment practices. In other cases, corporations will assign patents to subsidiaries and shell corporations that they control. While the motivations for this are multiple, a common motivation is tax stratagems that can shift earnings to low-cost nations through patent licensing agreements. (For an example of a firm that allegedly engaged in this practice on a large scale, see "Glaxo to Settle Tax Dispute with IRS over U.S. Unit for $3.4 Billion," http://online.wsj.com/article/SB115798715531459461.html (accessed December 21, 2016)). It might be anticipated that firms that have major offshore operations – which are likely to be more readily able to utilize tax-avoidance strategies – will disproportionately employ these strategies. Another potential driver may be the desire to obscure new strategic innovation initiatives: by patenting under the name of a little-known subsidiary, a firm may throw off its competitors.

For example, 3M Company, formerly known as the Minnesota Mining and Manufacturing Company, spun off a new subsidiary called "3M Innovative Properties Company" in 1998 to help to obtain, protect and manage intellectual property rights. This has been a prolific patenting subsidiary, with over 7000 patents granted in the USPTO database. Similarly, Microsoft Technology Licensing (MTL) was founded in 2014 to offer patent, licensing, and legal support services for Microsoft Corporation. More than 1000 patents are under the name of the MTL.

While the above examples can be captured by careful name matching (Autor, et al., 2016, is a state-of-the-art example), many others cannot. Firms that specialize in aggressively litigating awards, such as Intellectual Ventures and Acacia, have begun to assign their patents to thousands

of shell companies and subsidiaries. The apparent motivations for this practice are two-fold. First, by assigning the patent to an obscure entity, they may minimize the probability that a rival will identify it and request an Inter Partes or Post Grant Review (challenges within the USPTO that take shortly after the time of issue). Another possibility is that, were they to assign the patent to themselves, other firms might anticipate a lawsuit and cease innovating in that area, or otherwise invent around their awards (Lerner, 1995).

5.C Concordance Limitations

These problems with patent assignment are compounded by the process by which firms are mapped to their identifiers. The first NBER concordance matched up CUSIP identifiers and patent numbers using the status of the firms and their affiliates as of 1989. Patents assigned to entities they were affiliated with in in 1989; if they were not public then, no identifier was assigned. Moreover, the identifications of subsidiaries were based on the *Who Owns Whom* volume: while this includes many operating subsidiaries, it is not exhaustive, particularly when it comes to legal entities which do not have operating components.

The problem for corporate finance researchers is that the matching procedure that the N/BER used is likely to produce biases associated with the market for corporate control. For instance, patents are likely to be undercounted when firms employ complex organizational structures. Similarly, firms whose identity changes due to a reorganization (for instance, around the time of an acquisition, financing, or bankruptcy) are likely to have misattributed patents. Finally, firms that do not go public until after 1989 may not be identified. These biases will become progressively more severe, the further one moves from the 1989 match date.

The second match in the 2006 database, based on the names of the assignees, also poses some potential issues. In particular, the developers simply used the existing sets of matches, rather than updating them with more recent editions of *Whom Owns Whom*. (Of course, the process of doing such a match over multiple years would have been a very major undertaking.) As a result of the changing structure of firms over time, it will be likely that there will be an increasing disparity between where patents were actually assigned and what the databases suggest happened.

5.D Strategic Citations

Patent lawyers and examiners argue that patent citations have a complex set of drivers, many of which may not reflect either the presence of knowledge flows or economic importance. For instance, patent examiners end up adding the same patent as a citation to a number of different patents. One scenario when this may happen is if the examiner is aware of a patent (an "exemplary patent") that does a particularly good job of describing an aspect of the field of invention or has a particularly exhaustive list of references (articles, other patents, *etc.*). Cockburn, Kortum, and Stern (2002) show that examiners tend to cite patents that they themselves examined, as well as those that are particularly well written.

In addition, firms may cite patents which are only tangentially relevant to the claims held by their competitors. Such a move, were they to end up in litigation with a rival, would be extremely unlikely to influence the judge or the appellate court. But lawyers suggest that firms often believe that these citations will profoundly influence the jury that will hear the infringement case. Juries may see the patent office as omniscient, and believe that examiners carefully review all cited patents. (In point of fact, practitioners believe that examiners typically pay very little attention to the citations.) As a result, firms are likely to cite the patents belonging to their closest

competitors whom they anticipate meeting in court. Other accounts suggest patent lawyers sometimes urge weak applicants to employ the "kitchen sink" approach to citations: to cite a wide variety of prior art, burying the relevant citations under a mountain of irrelevant prior art in the hopes that the time-pressed examiner will not discover it.

6. A Checklist for Researchers

Patent data is a powerful tool for glimpsing innovative activity inside the black box of the firm. As such, it should be a valuable tool for those undertaking corporate finance and related research. At the same time, due to the complexities of the patent filing process, the changing spatial and industry composition of firms conducting innovation, and the databases that track these activities, the translation of this information into research conclusions is not simple.

This essay highlighted the features, strengths, and limitations of the key sources of patent data available to researchers. We began by emphasizing the biases induced by truncation of patent awards, affecting estimates of patterns across time, technology classes, regions, and industries. We then introduced the concepts of patent and citation biases, and showed how these patent-level distortions can be correlated with key firm-level characteristics that are of interest to researchers.

It is worth reiterating that the issues confronting users of patent data are similar to ones that those analyzing almost any contemporaneous database face. However, as we have highlighted, the dramatic changes in the direction and location of technological innovation (and patenting practice) over recent decades have led to a situation where these data limitations have led to highly predictable biases in the results of patent-based analyses. We used several examples based on analyses in prominent published papers that use patent data to highlight that a failure to properly control for these issues can lead to incorrect – and ex ante predictable – inferences.

There is no easy answer to adjusting for these problems: the dynamic nature of technological change and patent policy means that any formulaic set of adjustments would soon be out-of-date. Rather, to avoid these pitfalls, we would suggest that researchers employ the following steps to avoid the types of issues we have highlighted here:

- Focus on biases due to differences/changes in composition of the sample. As we have highlighted here, spurious results regarding the impact of policy changes and firm financing decisions can result when changes in the composition of sub-samples are ignored.

- Look across sub-samples for variations in patenting and citation practices due to:

    - Differences/changes in composition over time.

    - Differences/changes in composition across technology classes and industries.

    - Differences/changes in composition across states.

- Compute patent and citation biases using patents granted to the same firms and applied for during the same time period, employing different versions of patent data. Assess how these biases are related to policy or firm choices being studied.

- Evaluate the robustness of estimates with respect to differences and changes in composition and when interpreting effects. It is particularly important to check for the impact of selection biases, especially when evaluating policies or choices where much of the activity of interest lies within a decade of the end of the database.

- Think more broadly about potential biases that the broader issues with patent data highlighted in Section 4 can introduce in these analyses.

To help researchers, we have captures these considerations in the form of ten key questions, which we list in Table 11.

References

Almeida, Heitor, Po-Hsuan Hsu, and Dongmei Li, 2013, "Less is More: Financial Constraints and Innovative Efficiency," Unpublished working paper, University of Illinois at Urbana-Champaign, University of Hong Kong, and University of South Carolina.

Amore, Mario, Cedric Schneider, and Alminas Zaldokas, 2013, "Credit Supply and Corporate Innovation," *Journal of Financial Economics* 109, 835-855.

Atanassov, Julian, 2013, "Do Hostile Takeovers Stifle Innovation? Evidence from Antitakeover Legislation and Corporate Patenting," *Journal of Finance* 68, 1097–1131.

Autor, David, David Dorn, Gordon Hanson, Gary Pisano, and Pian Shu, 2016, "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents," Working Paper no. 22879, National Bureau of Economic Research.

Becker-Blease, John R., 2011, "Governance and Innovation," *Journal of Corporate Finance* 17, 947–958.

Bessen, James, 2009, "NBER PDP Project User Documentation: Matching Patent Data to Compustat Firms," Unpublished working paper, Boston University.

Cornaggia, Jess, Yifei Mao, Xuan Tian, Xuan and Brian Wolfe, 2015, "Does Banking Competition Affect Innovation?," *Journal of Financial Economics* 115, 189–209.

Chakraborty, Atreya, Zaur Rzakhanova, Shahbaz Sheikhb, 2014, "Antitakeover Provisions, Managerial Entrenchment and Firm Innovation," *Journal of Economics and Business* 72, 30–43.

Chava, Sudheer, Alexander Oettl, Ajay Subramanian, and Krishnamurthy Subramanian, 2013, "Banking Deregulation and Innovation." *Journal of Financial Economics* 109, 759-774.

Chemmanuer, Thomas, and Xuan Tian, 2014, "Anti-Takeover Provisions, Innovation, and Firm Value: A Regression Discontinuity Analysis?," Unpublished working paper, Boston University and Indiana University.

Cockburn Iain M., Samuel Kortum, and Scott Stern, 2002, "Are All Patent Examiners Equal? The Impact of Examiner Characteristics," Working Paper no. 8980, National Bureau of Economic Research.

Dass, Nishant, Vikram Nanda, and Steven C. Xiao, 2017, "Truncation Bias Corrections in Patent Data: Implications for Recent Research on Innovation," *Journal of Corporate Finance* 44, 353-374

Dyck, Alexander, Adair Morse, and Luigi Zingales, 2010, "Who Blows the Whistle on Corporate Fraud?," *Journal of Finance* 65, 2213-53.

Galasso, Alberto, Mark Schankerman, and Carlos Serrano, 2013, "Trading and Enforcing Patent Rights," *RAND Journal of Economics* 44: 275–312.

Graham, Stuart J. H., Bronwyn H. Hall, Dietmar Harhoff, and David C. Mowery, 2004, "Prospects for Improving U.S. Patent Quality via Post-Grant Opposition," *Innovation Policy and the Economy* 4, 115-144.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, 2001, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," Working Paper 8498, National Bureau of Economic Research.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, 2005, "Market Value and Patent Citations," *Rand Journal of Economics* 36, 16-38.

Harmon, Robert L., 1991, *Patents and the Federal Circuit,* Washington, Bureau of National Affairs.

Hombert, Johan, and Adrien Matray, 2017, "The Real Effects of Lending Relationships: Evidence from Banking Deregulation and Innovation," *Review of Financial Studies* 30, 2413-45.

Jaffe, Adam B., and Gaétan de Rassenfosse, 2017, "Patent Citation Data in Social Science Research: Overview and Best Practices," *Journal of the Association for Information Science and Technology* 68, 1360-74.

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

Jaffe, Adam B., Manuel Trajtenberg, and Michael Fogarty, 2000, "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *American Economic Review Papers and Proceedings* 90, 215-218.

Karpoff, Jonathan M., Allison Koester, D. Scott Lee, and, Gerald S. Martin, 2014, "Database Challenges in Financial Misconduct Research," Research Paper No. 2012-15, Georgetown McDonough School of Business.

Koenig, Gloria K., 1980, *Patent Invalidity: A Statistical and Substantive Analysis*, New York, Clark Boardman.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, "Technological Innovation, Resource Allocation, and Growth," *Quarterly Journal of Economics* 132 665–712.

Kolev, Julian, 2013, "Credit Constraints and their Impact on Innovation: Evidence from Venture Capital Exits," Unpublished working paper, Harvard University.

Lerner, Josh, 1995, "Patenting in the Shadow of Competitors," *Journal of Law and Economics* 38, 563–595.

Lerner, Josh, Morten Sorensen, and Per Strömberg, 2011, "Private Equity and Long-Run Investment: The Case of Innovation," *Journal of Finance* 66, 445-477.

Li, Guan-Cheng, Ronald Lai, Alexander D'Amour, David M. Doolin, Ye Sun, Vetle I. Torvik, Amy Z. Yu, and Lee Fleming, 2014, "Disambiguation and Co-Authorship Networks of the U.S. Patent Inventor Database (1975–2010)," *Research Policy* 43, 941-955.

Lyandres, Evgeny, and Dino Palazzo, 2016, "Cash Holdings, Competition, and Innovation," *Journal of Financial and Quantitative Analysis* 51, 1823-1861.

Merges, Robert P., 1992, *Patent Law and Policy: Cases and Materials*, Charlottesville, Virginia, Michie Company.

Moser, Petra, and Alessandra Voena, 2012, "Compulsory Licensing: Evidence from the Trading with the Enemy Act," *American Economic Review* 102, 396-427.

Nicholas, Tom, 2008, "Does Innovation Cause Stock Market Runups? Evidence from the Great Crash," *American Economic Review* 98, 1370-96.

Schroth, Enrique, and Dezsö Szalay, 2010, "Cash Breeds Success: The Role of Financing Constraints in Patent Races," *Review of Finance* 14, 73-118.

Serrano, Carlos, 2010, "The Dynamics of the Transfer and Renewal of Patents," *Rand Journal of Economics* 41, 686–708.

Seru, Amit, 2014, "Firm Boundaries Matter: Evidence from Conglomerates and R&D Activity," *Journal of Financial Economics* 111, 381-405.

Singel, Ryan, 2009, "FBI Investigated Coder for Liberating Paywalled Court Records," *Wired* October 5, http://www.wired.com/threatlevel/2009/10/swartz-fbi/, accessed January 4, 2015.

Thoma, Grid, Salvatore Torrisi, Alfonso Gambardella, Dominique Guellec, Bronwyn Hall, and Dietmar Harhoff, 2010, "Methods and Software for the Harmonization and Integration of Datasets: A Test Based on IP-Related Data and Accounting Databases with a Large Panel of Companies at the Worldwide Level," Unpublished Working Paper.

Zingales, Luigi, 2000, "In Search of New Foundations," *Journal of Finance* 55, 1623-1653.

Appendix 1: Papers using Patent Citation Data in Major Finance Journals, 2005-2017

Journal of Finance

Atanassov, Julian, 2013, "Do Hostile Takeovers Stifle Innovation? Evidence from Antitakeover Legislation and Corporate Patenting," *Journal of Finance* 68, 1097-1131.

Bartram, Söhnke M., Gregory Brown, and René M. Stulz, 2012, "Why Are U.S. Stocks More Volatile?," *Journal of Finance* 67, 1329-1370.

Bena, Jan, and Kai Li, 2014, "Corporate Innovations and Mergers and Acquisitions," *Journal of Finance* 69, 1923-1960.

Bernstein, Shai, 2015, "Does Going Public Affect Innovation?," *Journal of Finance* 70, 1365-1403.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, 2016, "The Impact of Venture Capital Monitoring," *Journal of Finance* 71, 1591–1622.

Gompers, Paul, Josh Lerner, and David Scharfstein, 2005, "Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999," *Journal of Finance* 60, 577-614.

Haslem, Bruce, 2005, "Managerial Opportunism during Corporate Litigation," *Journal of Finance* 60, 2013-2041.

Hirshleifer, David, Angie Low, and Siew Hong Teoh, 2012, "Are Overconfident CEOs Better Innovators?," *Journal of Finance* 67, 1457-1498.

Hoberg, Gerard, and Gordon Phillips, 2010, "Real and Financial Industry Booms and Busts," *Journal of Finance* 65, 45-86.

Hoberg, Gerard, Gordon Phillips, and Nagpurnanand Prabhala, 2014, "Product Market Threats, Payouts, and Financial Flexibility," *Journal of Finance* 69, 293-324.

Hsu, David H., 2004, "What Do Entrepreneurs Pay for Venture Capital Affiliation?," *Journal of Finance* 59, 1805-1844.

Kaplan, Steven N., Berk A. Sensoy, and Per Strömberg, 2009, "Should Investors Bet on the Jockey or the Horse? Evidence from the Evolution of Firms from Early Business Plans to Public Companies," *Journal of Finance* 64, 75-115.

Kumar, Praveen, and Dongmei Li, 2016, "Capital Investment, Innovative Capacity, and Stock Returns," *Journal of Finance* 71, 2059–2094.

Kung, Howard, and Lukas Schmid, 2015, "Innovation, Growth, and Asset Prices," *Journal of Finance* 70, 1001-1037.

Lerner, Josh, Morten Sorensen, and Per Strömberg, 2011, "Private Equity and Long-Run Investment: The Case of Innovation," *Journal of Finance* 66, 445-477.

Journal of Financial Economics

Acharya, Viral, and Zhaoxia Xu, 2017, "Financial Dependence and Innovation: The Case of Public versus Private Firms," *Journal of Financial Economics* 124, 223-243.

Amore, Mario D., Cédric Schneider, and Alminas Žaldokas, 2013, "Credit Supply and Corporate Innovation," *Journal of Financial Economics* 109, 835-855.

Balsmeier, Benjamin, Lee Fleming, and Gustavo Manso, 2017, "Independent Boards and Innovation," *Journal of Financial Economics* 123, 536-557.

Baranchuk, Nina, Robert Kieschnick, and Rabih Moussawi, 2014, "Motivating Innovation in Newly Public Firms," *Journal of Financial Economics* 111, 578-588.

Bena, Jan, Miguel A. Ferreira, Pedro Matos, and Pedro Pires, 2017, "Are Foreign Investors Locusts? The Long-Term Effects of Foreign Institutional Ownership," *Journal of Financial Economics* 126, 122-146.

Benson, David, and Rosemarie H. Ziedonis, 2010, "Corporate venture capital and the returns to acquiring portfolio companies," *Journal of Financial Economics* 98, 478-499.

Blanco, Iván, and David Wehrheim, 2017, "The Bright Side of Financial Derivatives: Options Trading and Firm Innovation," *Journal of Financial Economics* 125, 99-119.

Campello, Murillo, and Janet Gao, 2017, "Customer Concentration and Loan Contract Terms," *Journal of Financial Economics* 123, 108-136.

Chang, Xin, Kangkang Fu, Angie Low, and Wenrui Zhang, 2015, "Non-Executive Employee Stock Options and Corporate Innovation," *Journal of Financial Economics* 115, 168-188.

Chava, Sudheer, Alexander Oettl, Ajay Subramanian, and Krishnamurthy V. Subramanian, 2013, "Banking Deregulation and Innovation," *Journal of Financial Economics* 109, 759-774.

Cornaggia, Jess, Yifei Mao, Xuan Tian, and Brian Wolfe, 2015, "Does Banking Competition Affect Innovation?," *Journal of Financial Economics* 115, 189-209.

Cremers, K.J. Martijn, Lubomir P. Litov, and Simone M. Sepe, 2017 "Staggered Boards and Long-Term Firm Value, Revisited," *Journal of Financial Economics* forthcoming.

Custódio, Cláudia, and Daniel Metzger, 2014, "Financial Expert CEOs: CEO's Work Experience and Firm's Financial Policies," *Journal of Financial Economics* 114, 125-154.

Dyreng, Scott D., Bradley P. Lindsey, and Jacob R. Thornock, 2013, "Exploring The Role Delaware Plays as a Domestic Tax Haven," *Journal of Financial Economics* 108, 751-772.

Faleye, Olubunmi, Rani Hoitash, and Udi Hoitash, 2011, "The Costs of Intense Board Monitoring," *Journal of Financial Economics* 101, 160-181.

Gomes-Casseres, Benjamin, John Hagedoorn, and Adam B. Jaffe, 2006, "Do Alliances Promote Knowledge Flows?," *Journal of Financial Economics* 80, 5-33.

Gu, Tiantian, 2017, "U.S. Multinationals and Cash Holdings," *Journal of Financial Economics* 125, 344-368.

He, Jie, and Xuan Tian, 2013, "The Dark Side of Analyst Coverage: The Case of Innovation," *Journal of Financial Economics* 109, 856-878.

Higgins, Matthew J., and Daniel Rodriguez, 2006, "The Outsourcing of R&D through Acquisitions in the Pharmaceutical Industry," *Journal of Financial Economics* 80, 351-383.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, 2013, "Innovative Efficiency and Stock Returns," *Journal of Financial Economics* 107, 632-654.

Hsu, Po-Hsuan, 2009, "Technological Innovations and Aggregate Risk Premiums," *Journal of Financial Economics* 94, 264-279.

Hsu, Po-Hsuan, Xuan Tian, and Yan Xu, 2014, "Financial Development and Innovation: Cross-Country Evidence," *Journal of Financial Economics* 112, 116-135.

Humphery-Jenner, Mark, Ling L. Lisic, Vikram Nanda, and Sabatino D. Silveri, 2016, "Executive Overconfidence and Compensation Structure," *Journal of Financial Economics* 119, 533-558.

Jones, Christopher S., and Selale Tuzel, 2013, "Inventory Investment and the Cost Of Capital," *Journal of Financial Economics* 107, 557-579.

Lerner, Josh, 2006, "The New New Financial Thing: The Origins of Financial Innovations," *Journal of Financial Economics* 79, 223-255.

Lerner, Josh, Antoinette Schoar, Stanislav Sokolinski, and Karen Wilson, 2017, "The Globalization of Angel Investments: Evidence across Countries," *Journal of Financial Economics* forthcoming.

Mukherjee, Abhiroop, Manpreet Singh, and Alminas Žaldokas, 2017, "Do Corporate Taxes Hinder Innovation?," *Journal of Financial Economics* 124, 195-221.

Nanda, Ramana, and Matthew Rhodes-Kropf, 2013, "Investment Cycles and Startup Innovation," *Journal of Financial Economics* 110, 403-418.

Nanda, Ramana, and Tom Nicholas, 2014, "Did Bank Distress Stifle Innovation during the Great Depression?," *Journal of Financial Economics* 114, 273-292.

Ozmel, Umit, David T. Robinson, and Toby E. Stuart, 2013, "Strategic Alliances, Venture Capital, and Exit Decisions in Early Stage High-Tech Firms," *Journal of Financial Economics* 107, 655-670.

Qiu, Jiaping, and Chi Wan, 2015, "Technology Spillovers and Corporate Cash Holdings," *Journal of Financial Economics* 115, 558-573.

Segal, Gill, Ivan Shaliastovich, and Amir Yaron, 2015, "Good and Bad Uncertainty: Macroeconomic and Financial Market Implications," *Journal of Financial Economics* 117, 369-397.

Seru, Amit, 2014, "Firm Boundaries Matter: Evidence from Conglomerates and R&D Activity," *Journal of Financial Economics* 111, 381-405.

Sunder, Jayanthi, Shyam V. Sunder, and Jingjing Zhang, 2017, "Pilot CEOs and Corporate Innovation, *Journal of Financial Economics* 123, 209-224.

Review of Financial Studies

Acharya, Viral V., and Krishnamurthy V. Subramanian, 2009, "Bankruptcy Codes and Innovation," *Review of Financial Studies* 22, 4949-4988.

Acharya, Viral V., Ramin P. Baghai, and Krishnamurthy V. Subramanian, 2014, "Wrongful Discharge Laws and Innovation," *Review of Financial Studies* 27, 301-346.

Ball, Eric, Hsin Hui Chiu, and Richard Smith, 2011, "Can VCs Time the Market? An Analysis of Exit Choice for Venture-backed Firms," *Review of Financial Studies* 24, 3015-3138.

Celikyurt, Ugur, Merih Sevilir, and Anil Shivdasani, 2014, "Venture Capitalists on Boards of Mature Public Firms," *Review of Financial Studies* 27, 56-101.

Chari, Anusha, Paige P. Ouimet, and Linda L. Tesar, 2010, "The Value of Control in Emerging Markets," *Review of Financial Studies* 23, 1741-1770.

Chemmanur, Thomas J., Elena Loutskina, and Xuan Tian, 2014, "Corporate Venture Capital, Value Creation, and Innovation," *Review of Financial Studies* 27, 2434-73.

Cohen, Lauren, Karl Diether, and Christopher Malloy, 2013, "Misvaluing Innovation," *Review of Financial Studies* 26, 635-666.

Dass, Nishant, Omesh Kini, Vikram Nanda, Bunyamin Onal, and Jun Wang, 2014, "Board Expertise: Do Directors from Related Industries Help Bridge the Information Gap?," *Review of Financial Studies* 27, 1533-92.

Fang, Lily H., Josh Lerner, and Chaopeng Wu, 2017, "Intellectual Property Rights Protection, Ownership, and Innovation: Evidence from China," *Review of Financial Studies* 30, 2446–77.

Frésard, Laurent, Ulrich Hege, and Gordon Phillips. 2017, "Extending Industry Specialization through Cross-Border Acquisitions," *Review of Financial Studies* 30, 1539–82.

Gan, Jie, Yan Guo, and Chenggang Xu, 2017, "Decentralized Privatization and Change of Control Rights in China," *Review of Financial Studies* forthcoming.

He, Jie, and Jiekun Huang, 2017, "Product Market Competition in a World of Cross-Ownership: Evidence from Institutional Blockholdings," *Review of Financial Studies* 30, 2674–2718.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, 2017, "Innovative Originality, Profitability, and Stock Returns," *Review of Financial Studies* forthcoming.

Hombert, Johan, and Adrien Matray, 2017, "The Real Effects of Lending Relationships: Evidence from Banking Deregulation and Innovation," *Review of Financial Studies* 30, 2413-45.

Irvine, Paul J., and Jeffrey Pontiff, 2009, "Idiosyncratic Return Volatility, Cash Flows, and Product Market Competition," *Review of Financial Studies* 22, 1149-1177.

Kerr, William R., Josh Lerner, and Antoinette Schoar, 2014, "The Consequences of Entrepreneurial Finance: Evidence from Angel Financings," *Review of Financial Studies* 27, 20-55.

Ouimet, Paige P., 2013, "What Motivates Minority Acquisitions? The Trade-Offs between a Partial Equity Stake and Complete Integration," *Review of Financial Studies* 26, 1021-1047.

Phillips, Gordon M., and Giorgo Sertsios, 2017, "Financing and New Product Decisions of Private and Publicly Traded Firms," *Review of Financial Studies* 30, 1744–1789.

Tian, Xuan, and Tracy Yue Wang, 2011, "Tolerance for Failure and Corporate Innovation," *Review of Financial Studies* 27, 211-255.

Figure 1: Evidence from Google Scholar

This figure presents the number of entries in Google Scholar with the phrases "patent citations" and "*Journal of Finance*" (left) along with the percent share of the phrase "*Journal of Finance*" (right) over time (in years). The sample covers the years 1996 through 2016. The graphic shows there is a growing interest in papers on patents, as there is a steady growth even when these papers are presented as a fraction of all papers citing a *Journal of Finance* paper. Source: Google Scholar database.
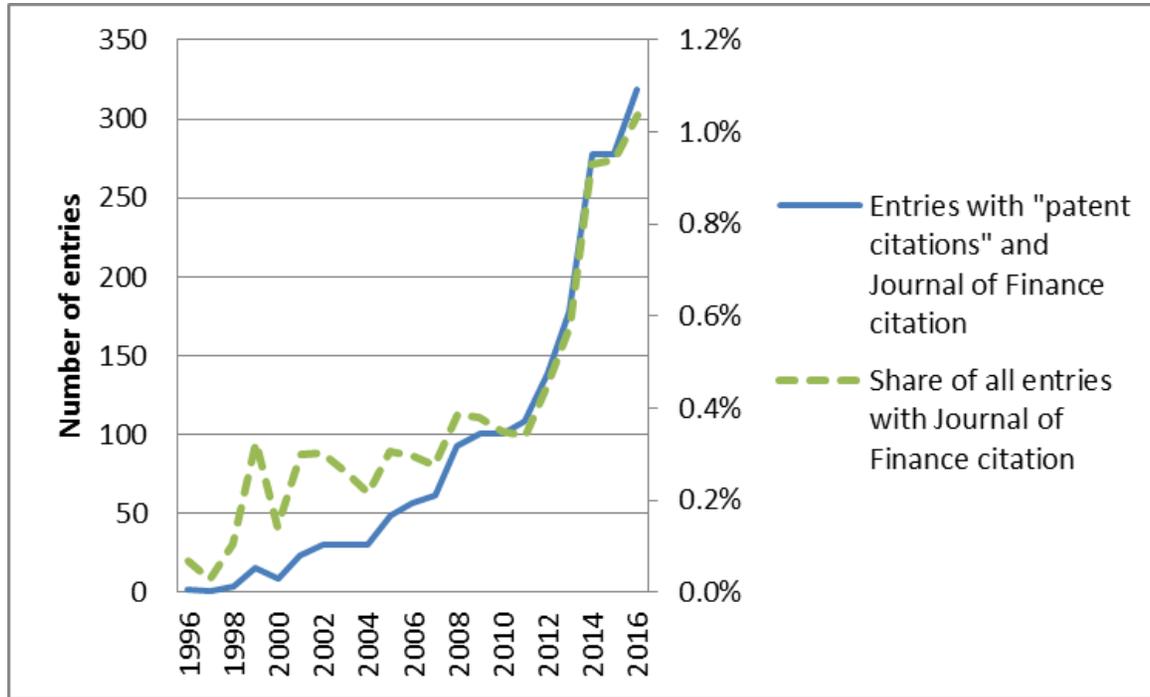
Figure 2: U.S. Grants, Successful Patent Applications, Total Patent Applications, and Citations per Patent

The figure shows the number of grants (Panel A), the number of successful applications (Panel B), and the citations per patent (Panel D) dataset from 1975 through 2006 in the NBER 2006 patent database. The total number of applications in Panel C is defined by the number of successful and unsuccessful patent applications and is from the USPTO. Panel A highlights the three-fold increase in grants, while Panel C shows how total applications increased four-fold. Panels B and D highlight the truncation of the NBER database, which only reports the patents issued by the end of 2006, and saw citations peak at 1985 before falling rapidly. Source: NBER 2006 patent dataset and the USPTO website.
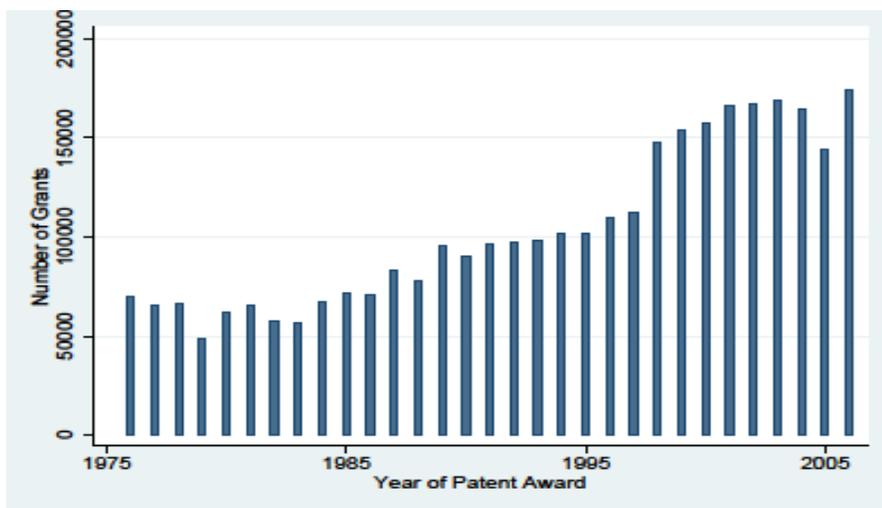
(a) U.S. patent awards over time, by award year
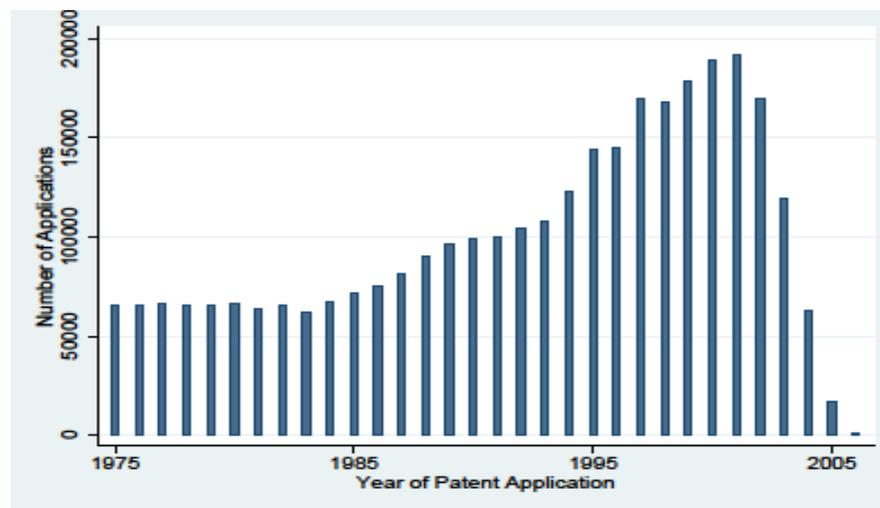
(b) U.S. successful patent applications over time, by application year
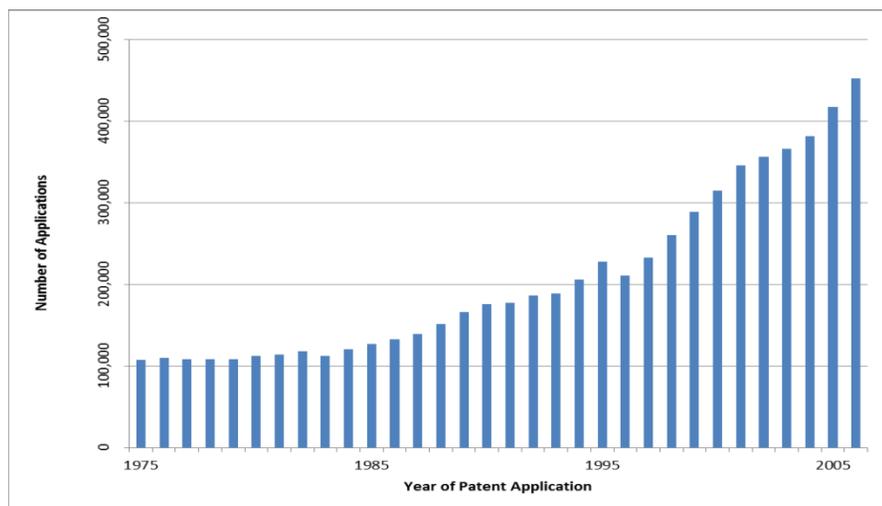
(c) Actual patent applications (successful and unsuccessful), by application year
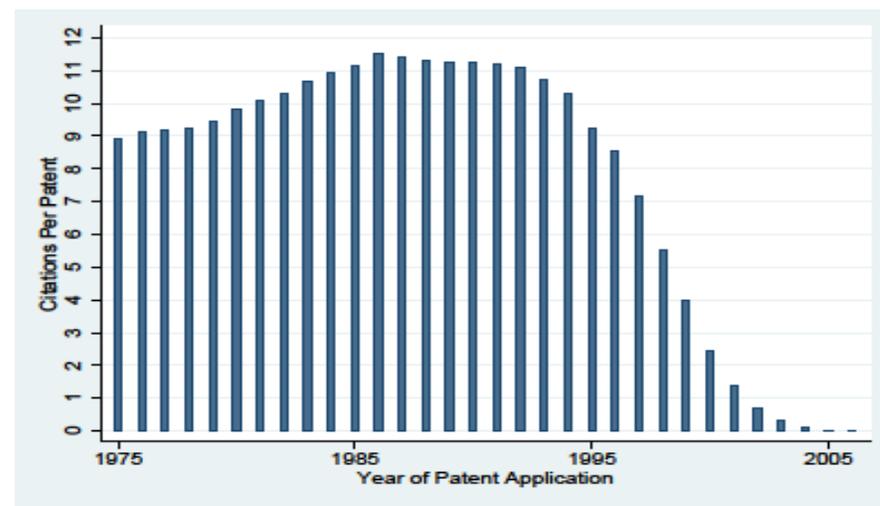
(d) Citations per patent over time, by application year

Figure 3: Citation Distribution

Panel A of this figure graphs the density of citations over the years after a patent in electronics HJT subcategory was issued pre- and post-1991. The figure in Panel B plots a similar density for patents in chemicals HJT subcategory. These highlight the truncation problem in the dataset. Source: NBER 2006 database.
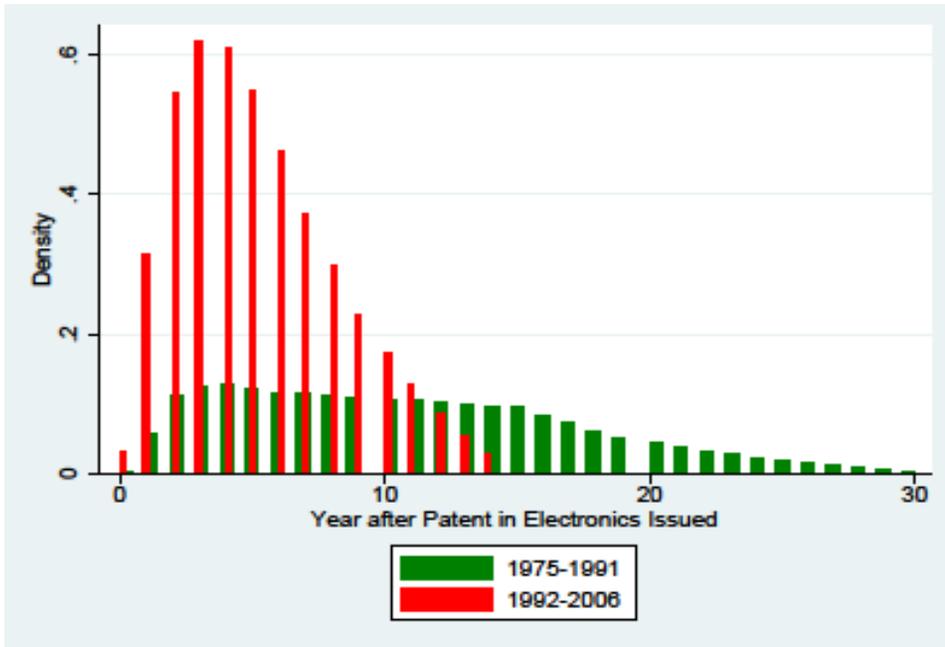
(a) Citation distribution for electronics patent cohorts
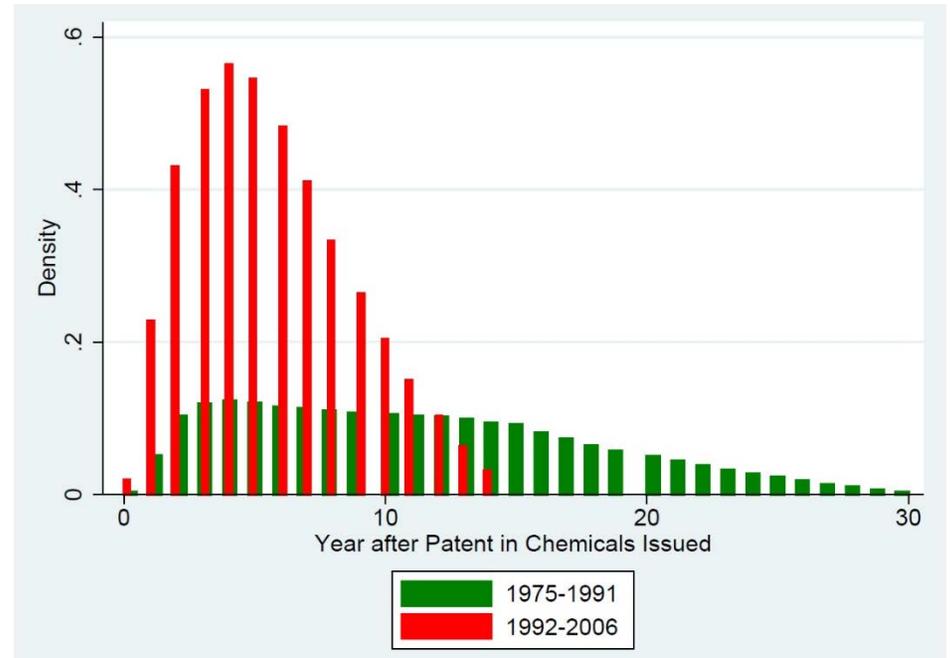
(b) Citation distribution for chemicals patent cohorts

Figure 4: U.S. Successful Applications Comparisons

Panel A of this figure graphs the number of successful applications in the "computers and communications" and "chemicals" HJT subcategories and the year of the patent application. Panel B graphs the number of successful applications in the "electronics" and "mechanical" subcategories and the year of the patent application. The two graphs make clear that computers and electronics experienced a dramatic run-up in patenting activity in the 1980s to early 2000s. Conversely, the chemicals and mechanical cohorts have had less dramatic increases in the number of applications, and were both ultimately surpassed by computers and electronics, respectively. These figures highlight the heterogeneity in truncation bias in successful applications across various industry subcategories. Source: NBER 2006 patent dataset.
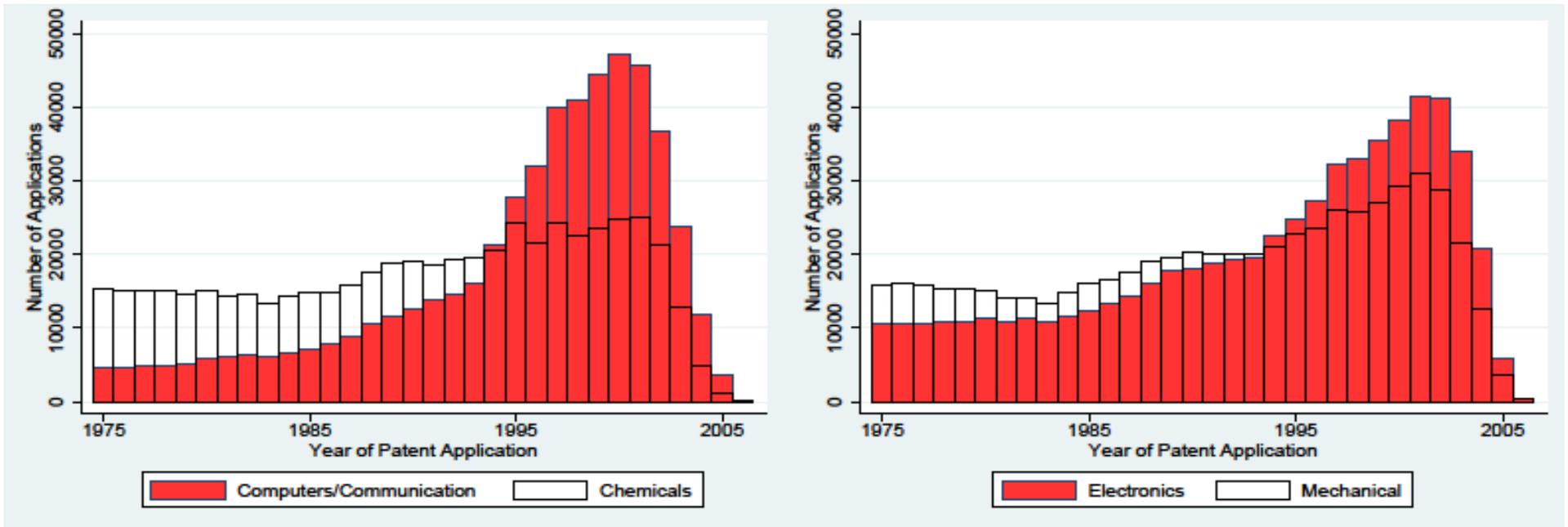
(a) U.S. successful patent applications, chemicals versus computers

(b) U.S. successful patent applications, electronics versus mechanical

Figure 5: Drop-off in Citations in Recent Patents

Panel A plots the citations per patent for patents in the "computers/communication" and "chemicals" HJT subcategories and the year of patent application. Panel B shows a similar plot for patents in the "electronics" and "mechanical" subcategories. Both figures show there are dramatic differences in citation rates across technologies, as computer and (more weakly) electronics patents are more heavily cited than their counterparts in other industries. These figures highlight the heterogeneity in truncation bias in citations received by patents across various industry subcategories. Source: NBER 2006 patent dataset.
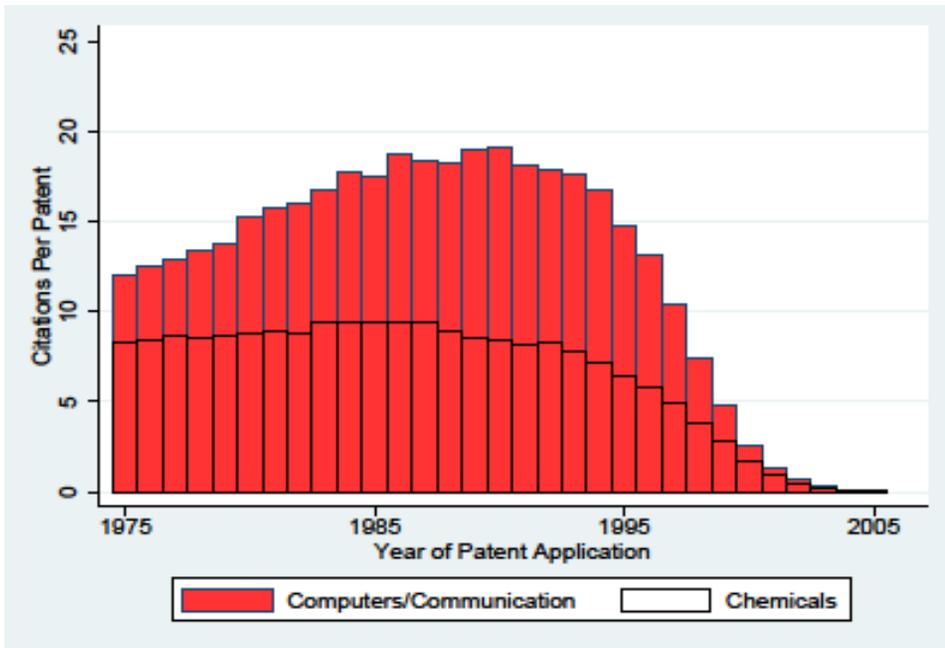
(a) Chemicals versus Computers
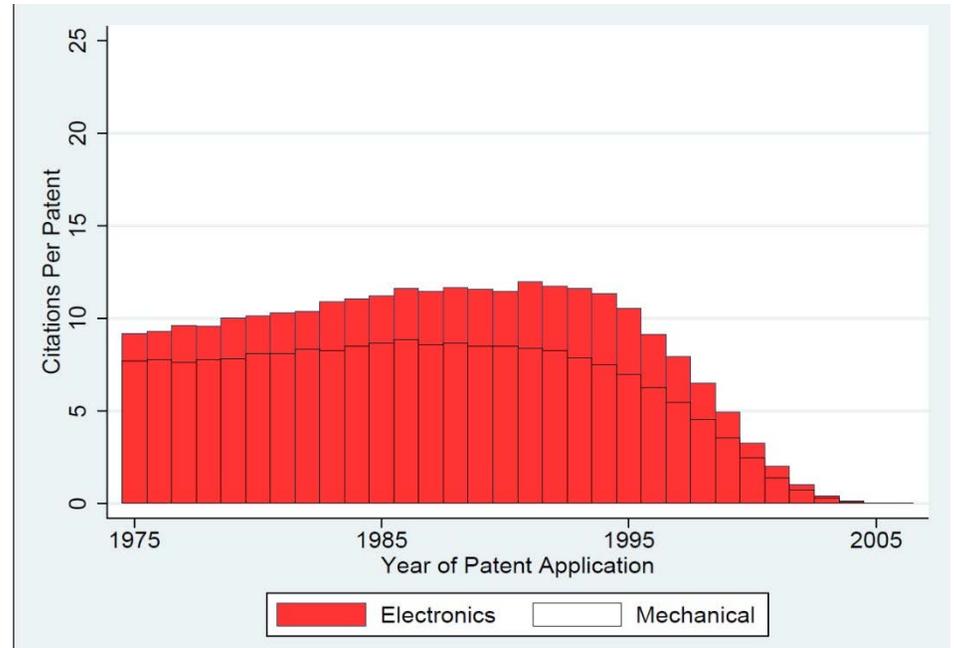
(b) Electronics versus Mechanical

Figure 6: U.S. Successful Patent Applications, by State of Assignee

This figure presents the number of successful applications over time by the state of assignee, with Delaware (DE) in red and California/Massachusetts (CA/MA) in white. The graph shows that growth in patenting is far from uniform geographically, as there was a dramatic increase in successful patents from assignees in CA and MA over time, whereas patents stayed more or less the same in DE. These comparisons suggest that any naïve correction for the truncation of patents which does not account for such dramatic regional differences may lead to incorrect inferences. Source: NBER 2006 patent dataset.
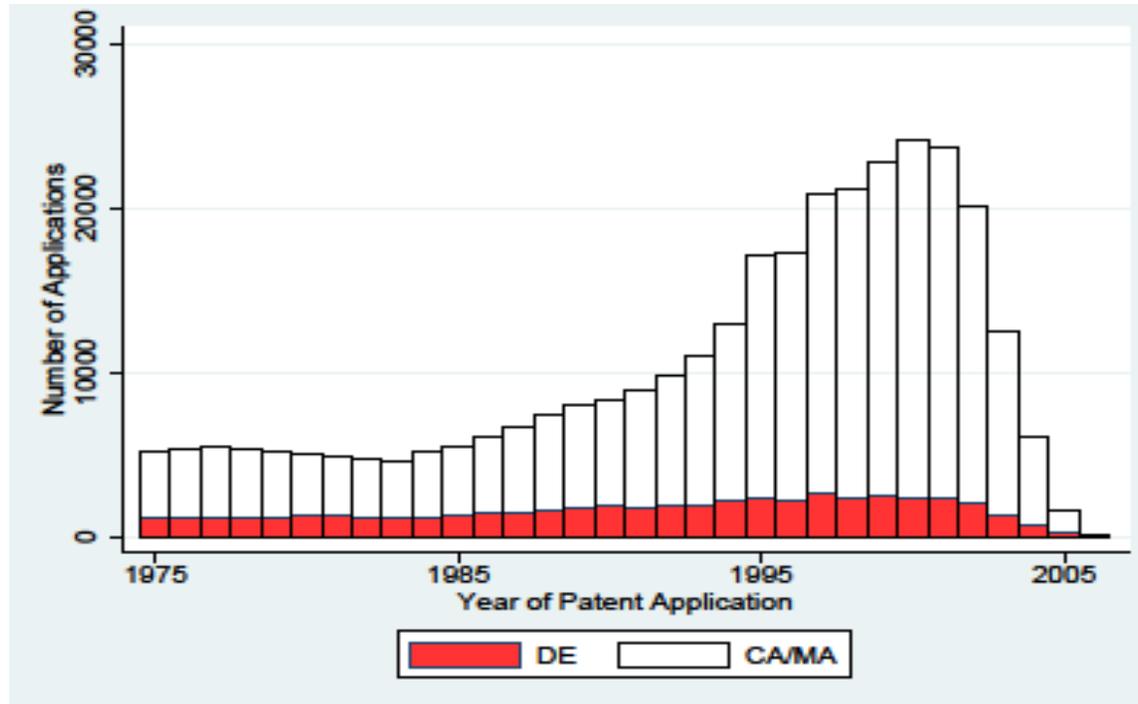
Figure 7: Drop-off in Citations in Recent Patents, by State of Assignee

This figure presents the number of citations per patent by state of assignee, with Delaware (DE) shown in red and California/Massachusetts (CA/MA) shown in white. Citations per patent are also geographically different: patents by assignees from CA and MA were far more likely to get cited than those from DE. These comparisons suggest that any naïve correction for the truncation of citations which does not account for such dramatic regional differences may lead to incorrect inferences. Source: NBER 2006 patent dataset.
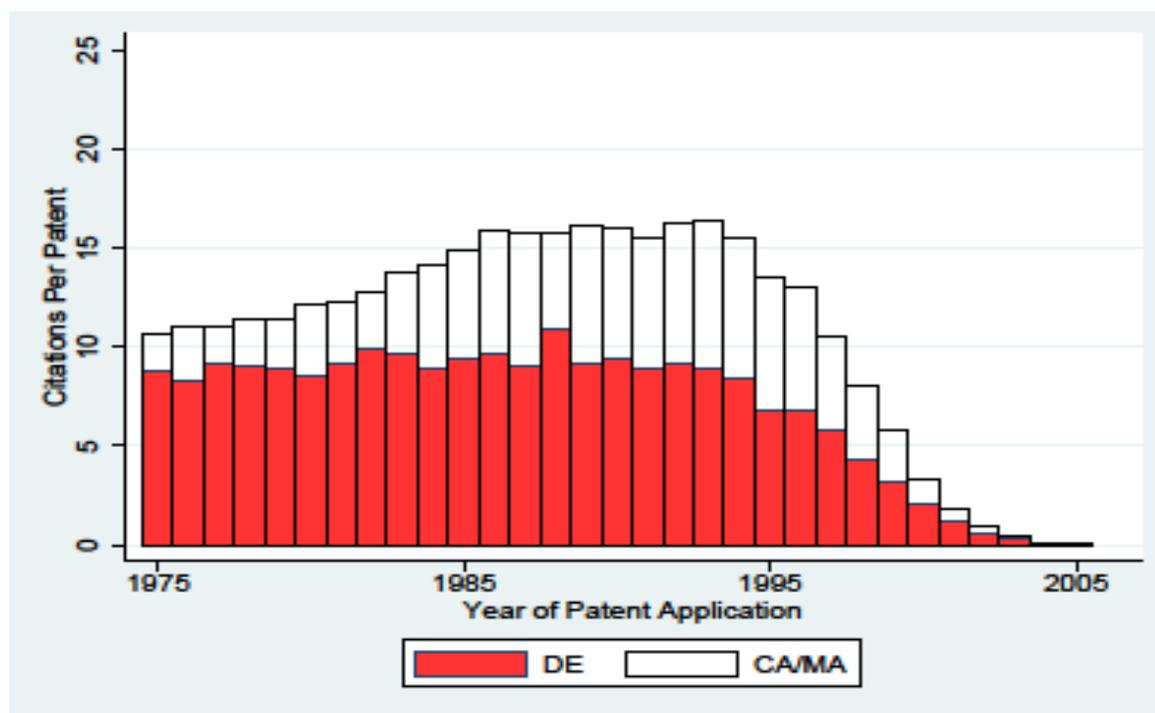
Figure 8A: Distribution of Firm Patent Bias (Unadjusted and Adjusted) over Time

This figure presents the distribution for patent bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012. To compute the unadjusted patent bias for each firm-year, we compare the number of patents for each firm filed in each application year in our data (thus, which have been granted by 2012) and in the NBER 2006 dataset (i.e., granted by 2006). We sum patent bias by year across publicly traded firms. The adjustments use the time fixed effect and the time and technology class fixed effect methodologies, with details discussed in Appendix B. Sources: NBER 2006 patent and our datasets.
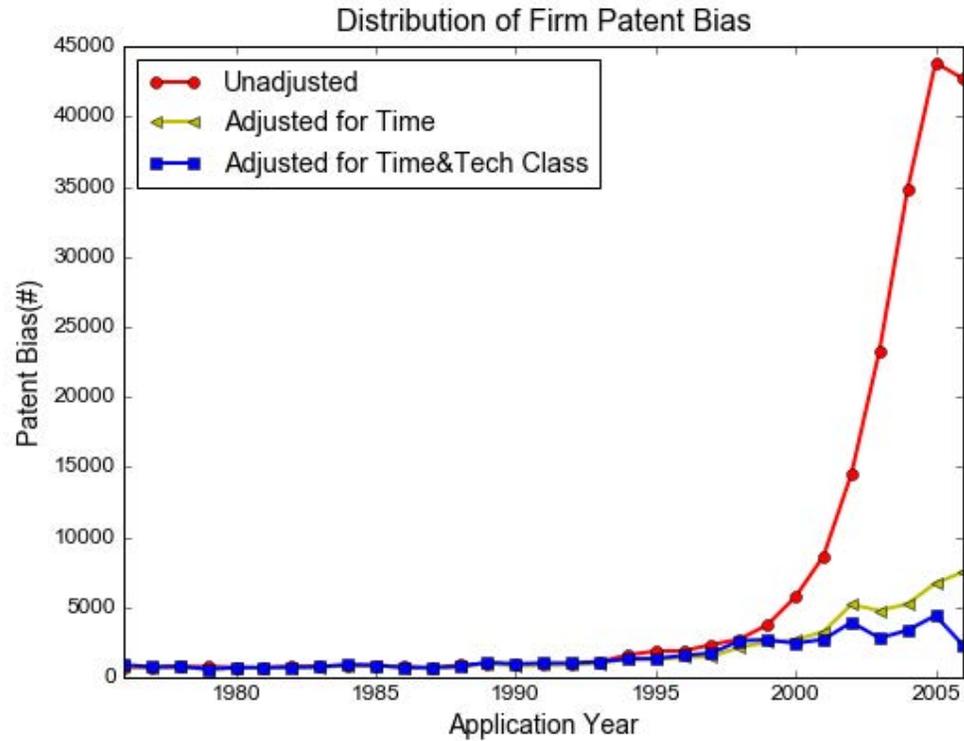
Figure 8B: Distribution of Firm Citation Bias (Unadjusted and Adjusted) over Time

This figure presents the distribution for citation bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012. To compute the unadjusted citation bias for each firm-year, we compare the number of citations to all the patents for each firm filed in each application year in our data (i.e., citations in patents granted by 2012 to applications filed by a firm in a given year and granted by 2006) and in the NBER 2006 dataset (i.e., citations in patents granted by 2006 to applications filed by a firm in a given year and granted by 2006). We sum citation bias by year across publicly traded firms. The adjustments use the time fixed effect methodology, the time and technology class fixed effect methodology, and the quasi-structural approach (citing year), with details discussed in Appendix B. The lines for the time fixed effect methodology and the time and technology class fixed effect methodology are almost superimposed due to the scale. Sources: NBER 2006 patent and our datasets.
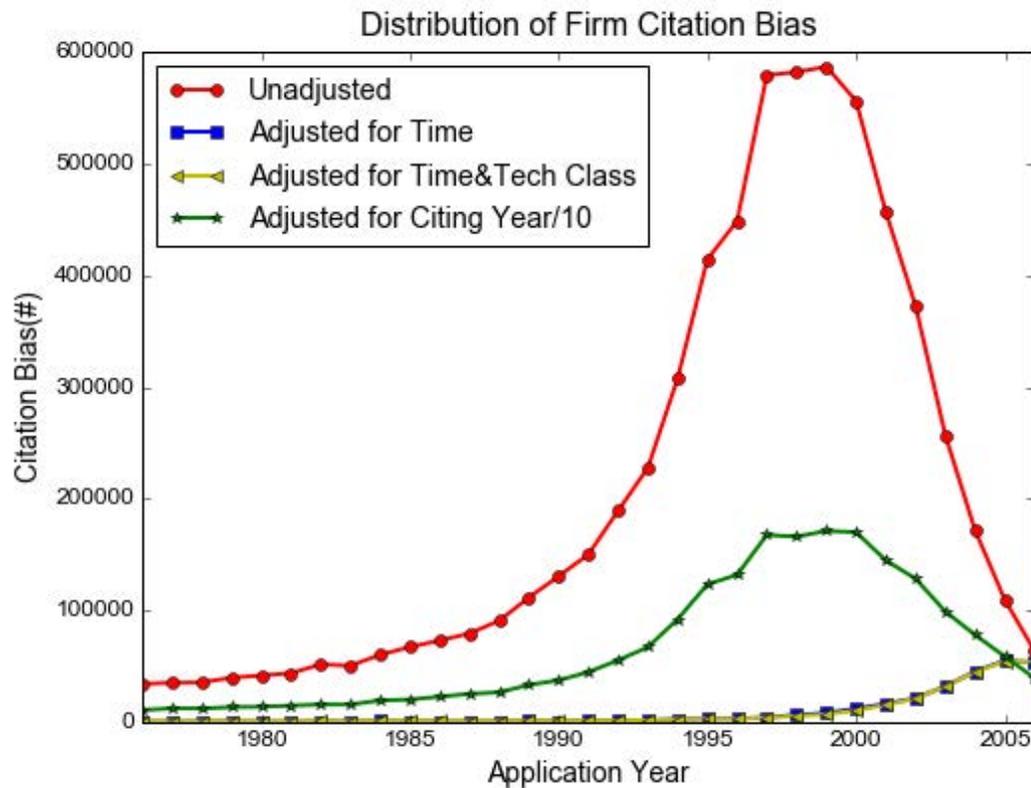
Figure 9: Firm Patent Bias (Unadjusted and Adjusted) across HJT Technology Classes

This figure presents the distribution for patent bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012 in different HJT technology classes. To compute the unadjusted patent bias for each firm-year, we compare the number of patents for each firm filed in each application year in our data (thus, which have been granted by 2012) and in the NBER 2006 dataset (i.e., granted by 2006). A firm is assigned to a particular technology class in a given year based on the modal primary patent class of patents produced by that firm in that year, based on the U.S. patent classification system. We sum patent bias in a technology class across publicly traded firms. The adjustments use the time fixed effect and the time and technology class fixed effect methodologies, with details discussed in Appendix B. Sources: NBER 2006 patent and our datasets.

(a)

Time Series of Adjusted Firm Patent Bias(Time)

Time Series of Adjusted Firm Patent Bias(Time & Tech Class)
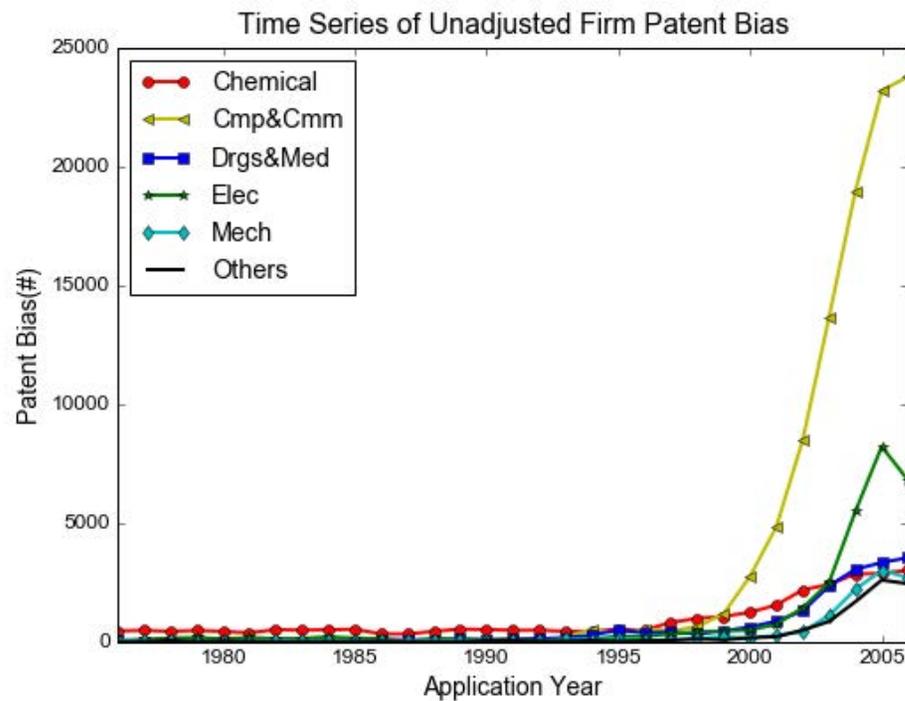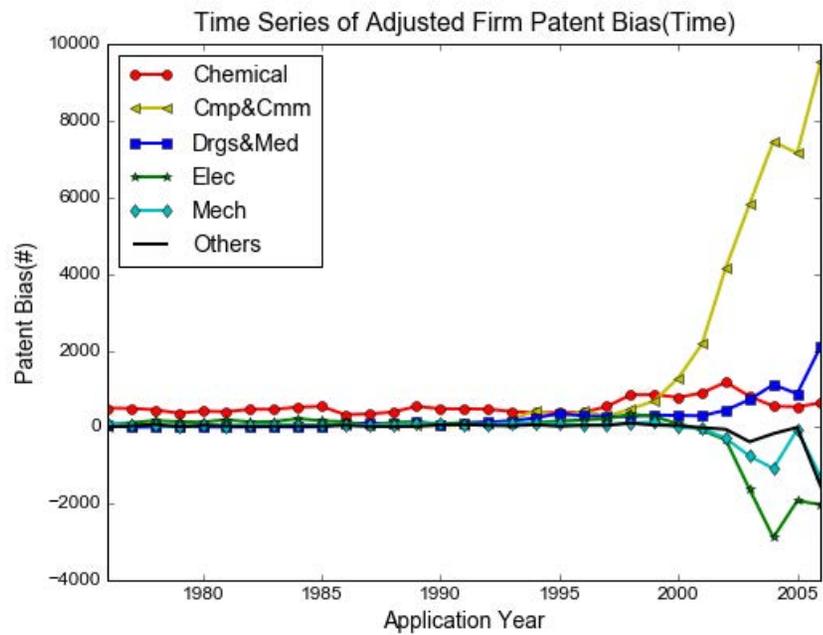
(b)
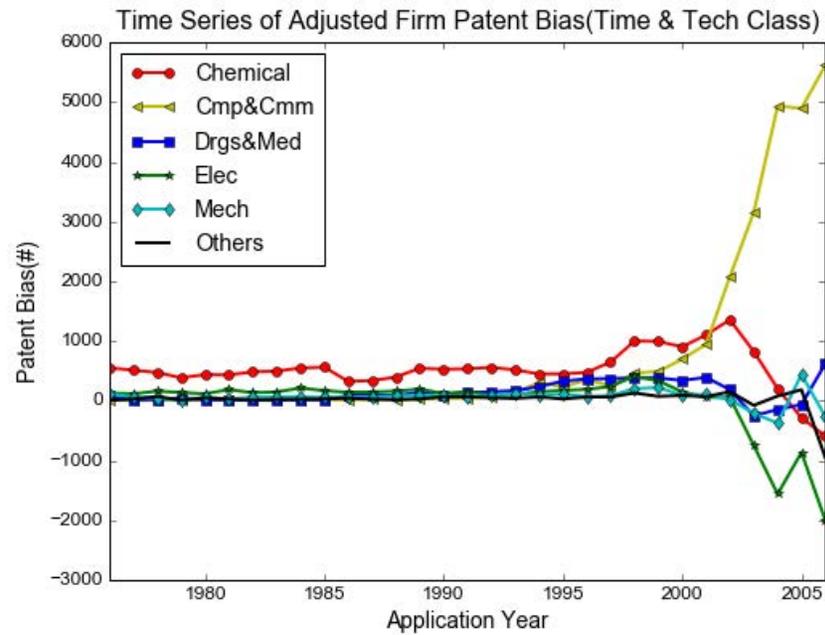
(c)

Figure 10: Firm Citation Bias (Unadjusted and Adjusted) across HJT Technology Classes

This figure presents the distribution for citation bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012 in different HJT technology classes. To compute the unadjusted citation bias for each firm-year, we compare the number of citations to all the patents for each firm filed in each application year in our data (i.e., citations in patents granted by 2012 to applications filed by a firm in a given year and granted by 2006) and in the NBER 2006 dataset (i.e., citations in patents granted by 2006 to applications filed by a firm in a given year and granted by 2006). A firm is assigned to a particular technology class in a given year based on the modal primary patent class of patents produced by that firm in that year, based on the U.S. patent classification system. We sum citation bias in a technology class across publicly traded firms. The adjustments use the time fixed effect methodology, the time and technology class fixed effect methodology, and the citing year effect adjustment using the quasi structural method, with details discussed in Appendix B. Sources: NBER 2006 patent and our datasets.

(a)

Time Series of Adjusted Firm Citation Bias(Time)

(b)

Time Series of Adjusted Firm Citation Bias(Time & Tech Class)

(c)

Time Series of Adjusted Firm Citation Bias(Citing Year)
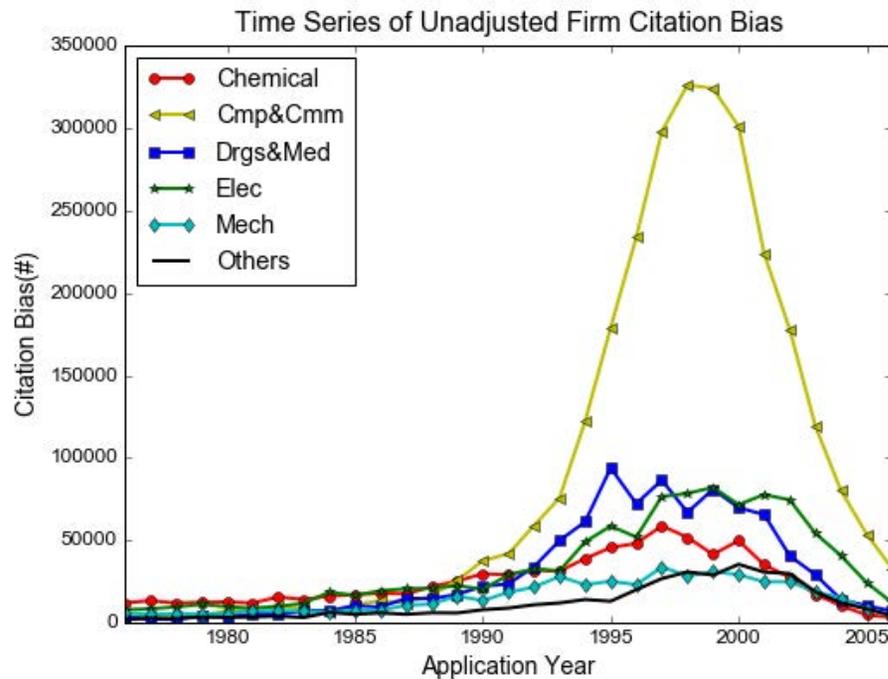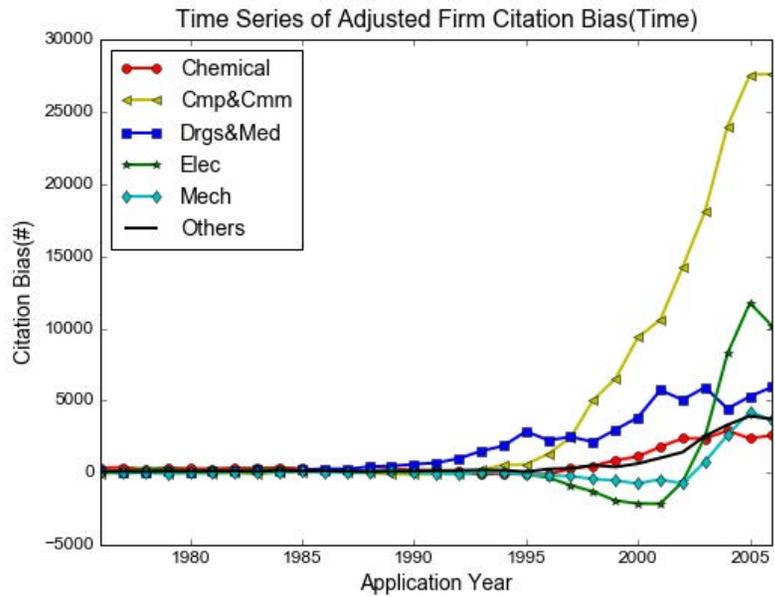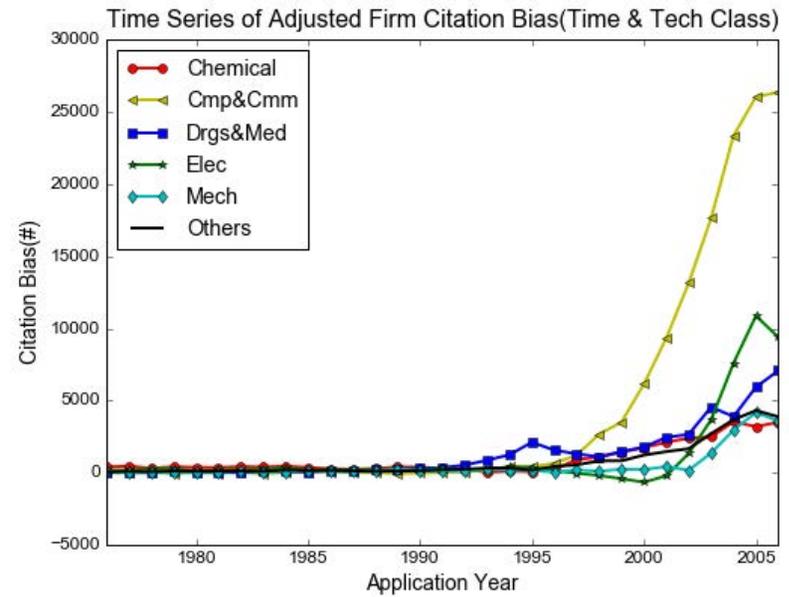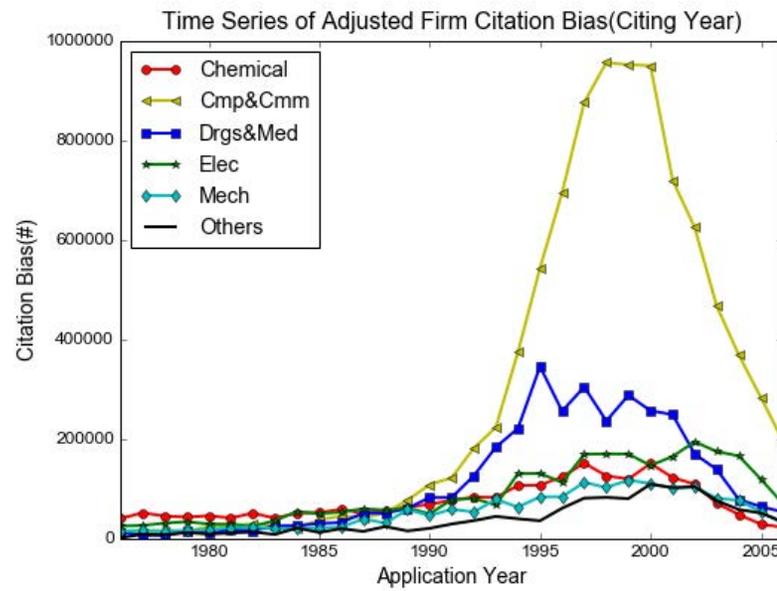
(d)

Figure 11: Firm Patent Bias (Unadjusted and Adjusted) across States

This figure presents the distribution for patent bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012 in different states. To compute the unadjusted patent bias for each firm-year, we compare the number of patents for each firm filed in each application year in our data (thus, which have been granted by 2012) and in the NBER 2006 dataset (i.e., granted by 2006). A firm is assigned to a particular state in a given year based on modal state of the assignee across patents granted to the firm at the time of the patent filing. We sum patent bias in each state across publicly traded firms. The adjustments use the time fixed effect and the time and technology class fixed effect methodologies, with details discussed in Appendix B. Sources: NBER 2006 patent and our datasets.
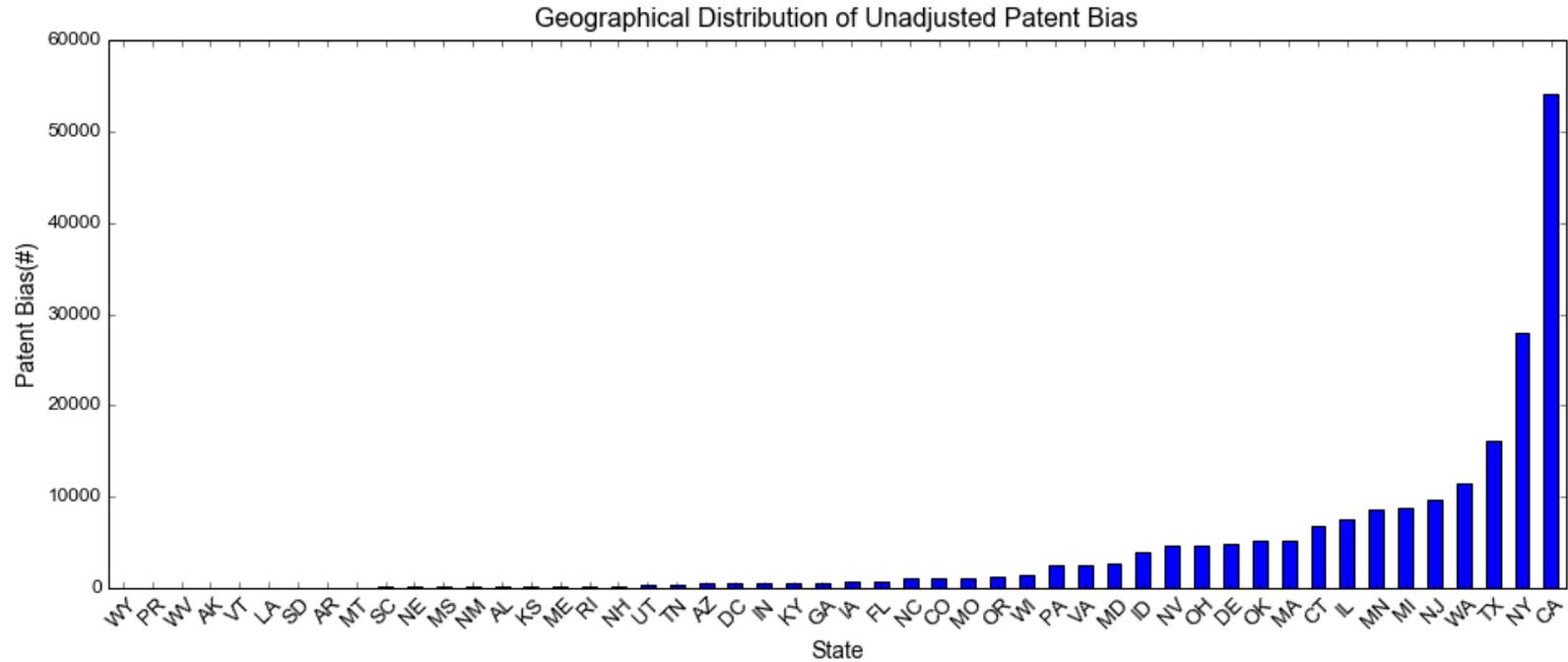

(a)
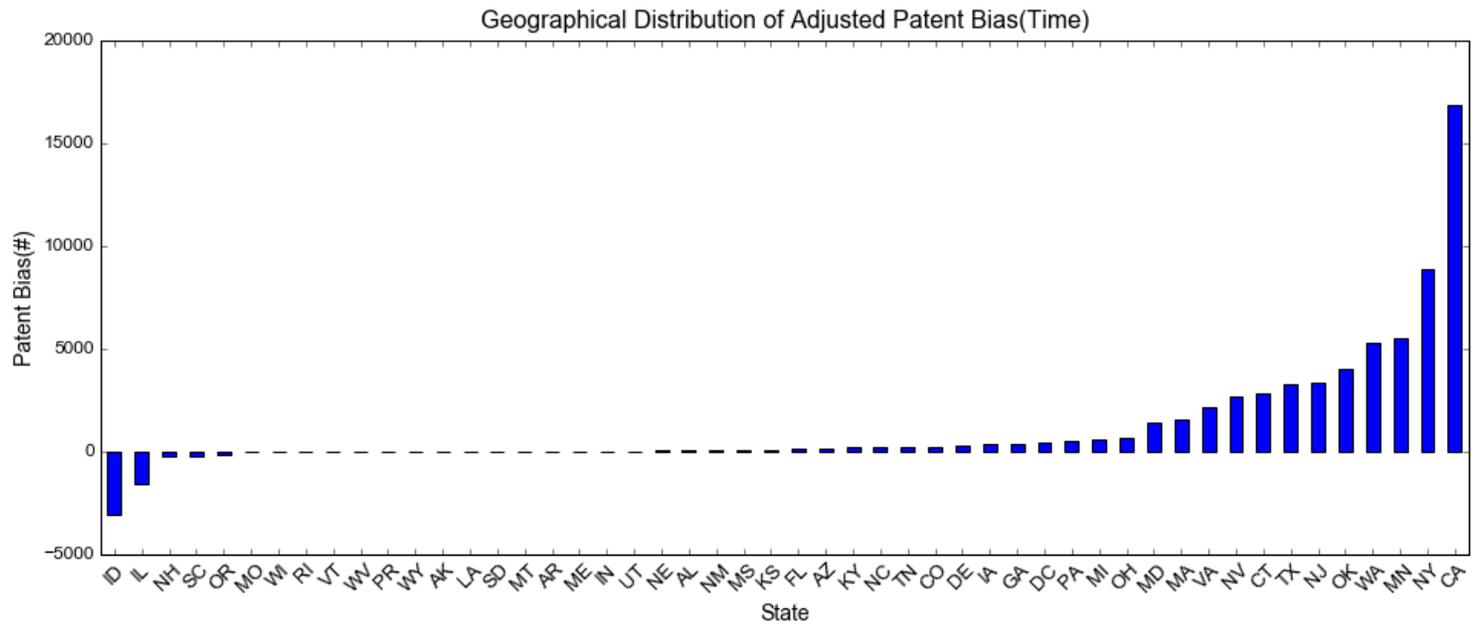
(b)
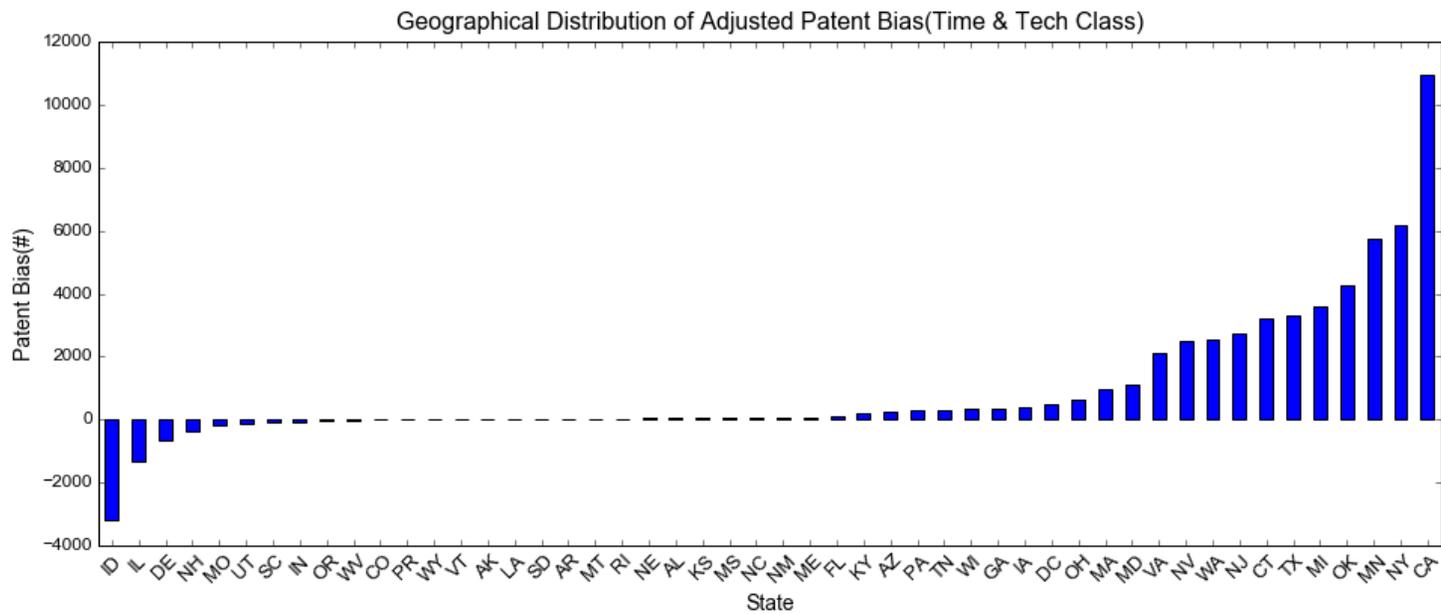
(c)

Figure 12: Firm Citation Bias (Unadjusted and Adjusted) across Statess

This figure presents the distribution for citation bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012 in different states. To compute the unadjusted citation bias for each firm-year, we compare the number of citations to all the patents for each firm filed in each application year in our data (i.e., citations in patents granted by 2012 to applications filed by a firm in a given year and granted by 2006) and in the NBER 2006 dataset (i.e., citations in patents granted by 2006 to applications filed by a firm in a given year and granted by 2006). A firm is assigned to a particular state in a given year based on modal state of the assignee across patents granted to the firm at the time of the patent filing. We sum citation bias in each state across publicly traded firms. The adjustments use the time fixed effect, the time and technology class fixed effect methodologies, and the citing year effect adjustment using the quasi structural method, with details discussed in Appendix B. Sources: NBER 2006 patent and our datasets.
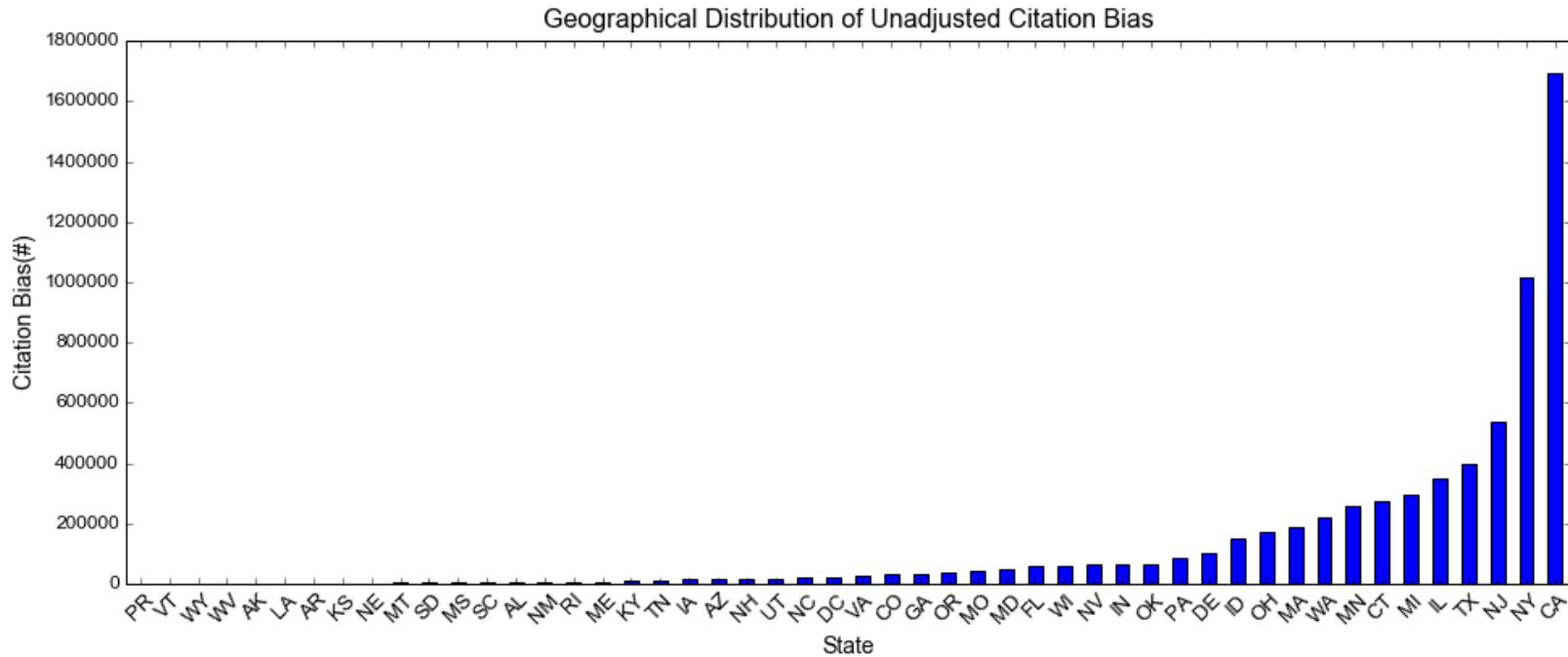
(a)

Geographical Distribution of Adjusted Citation Bias(Time)
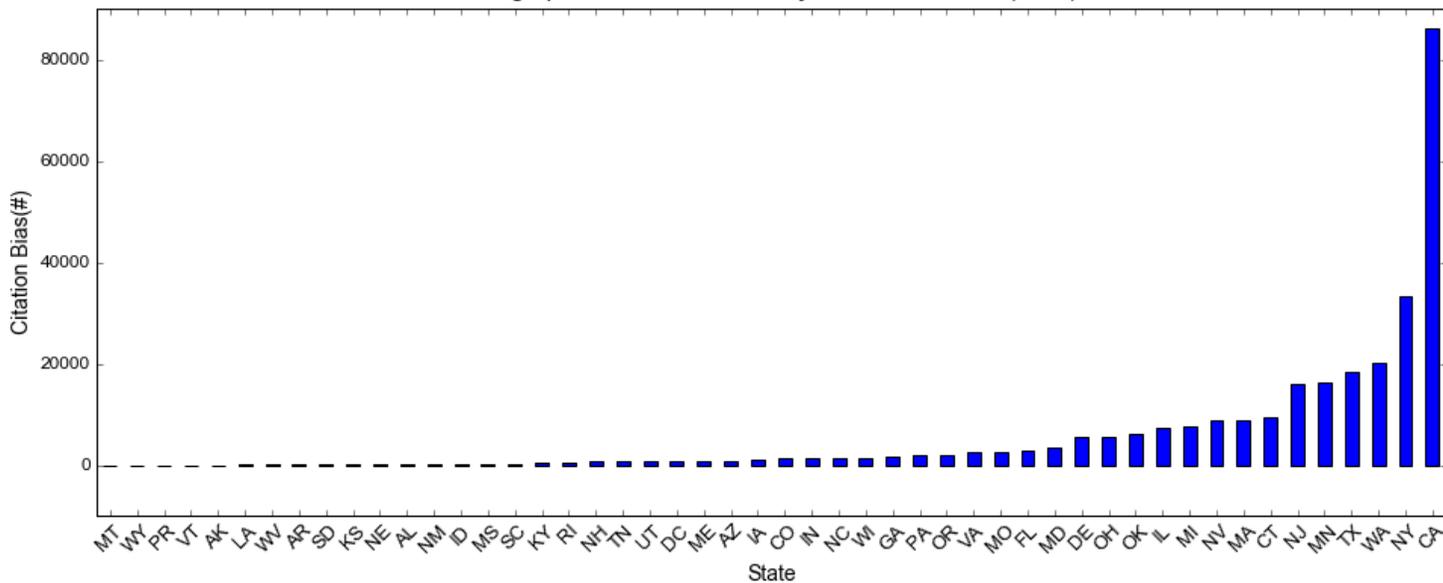
(b)

Geographical Distribution of Adjusted Citation Bias(Time & Tech Class)

(c)

Geographical Distribution of Adjusted Citation Bias(Citing Year)
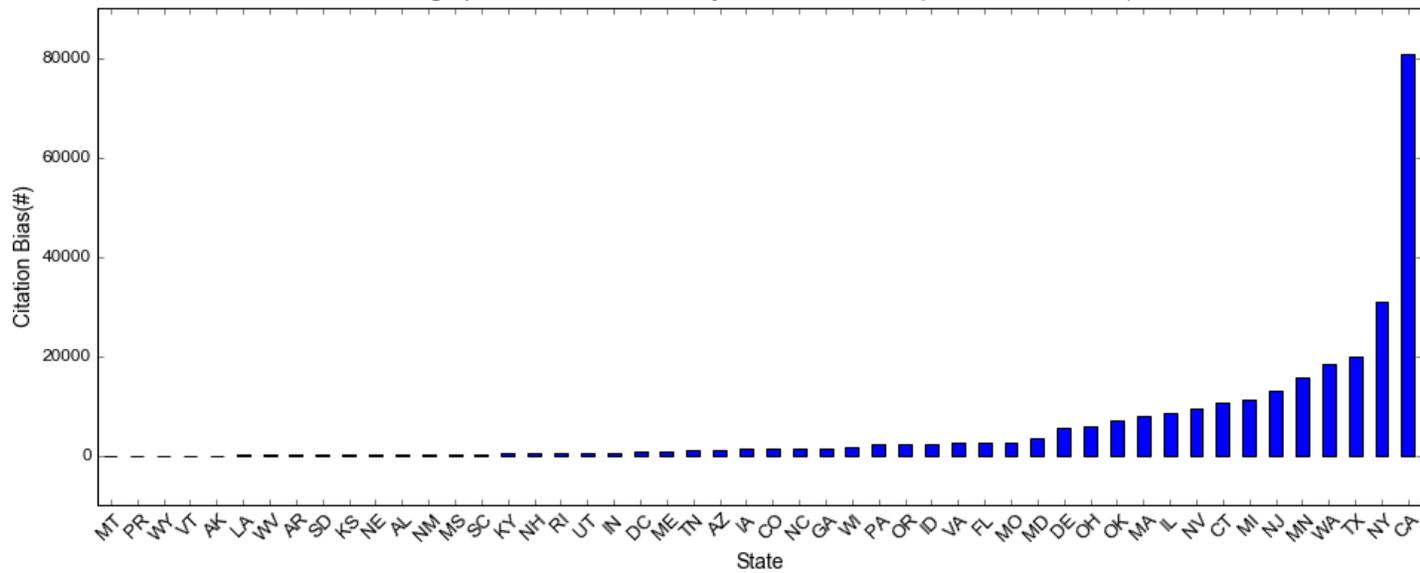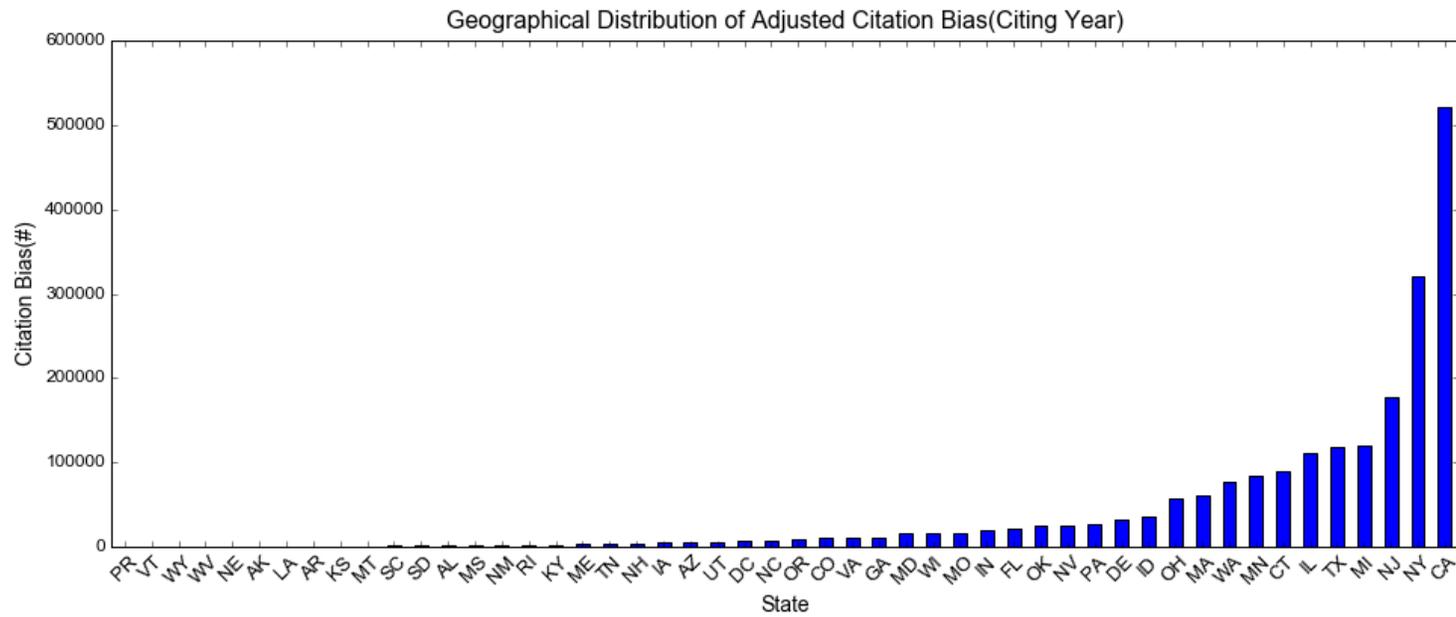
(d)

Figure 13: Firm Patent Bias (Unadjusted and Adjusted) across Industries (NAICS)

This figure presents the distribution for patent bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012 in different industries (classified using NAICS code). The unadjusted patent bias of a given firm is computed as the difference in number of patents ultimately granted to a firm in a given application year as of 2012 ("our data") and the number of patents ultimately granted to that firm in the same application year as of the end of sample in the NBER 2006 dataset. We assign firms into different industries at time of the patent application using Compustat. We sum patent bias in each industry across publicly traded firms. The adjustments use the time fixed effect and the time and technology class fixed effect methodologies, with details discussed in Appendix B. Sources: NBER 2006 patent and our datasets.
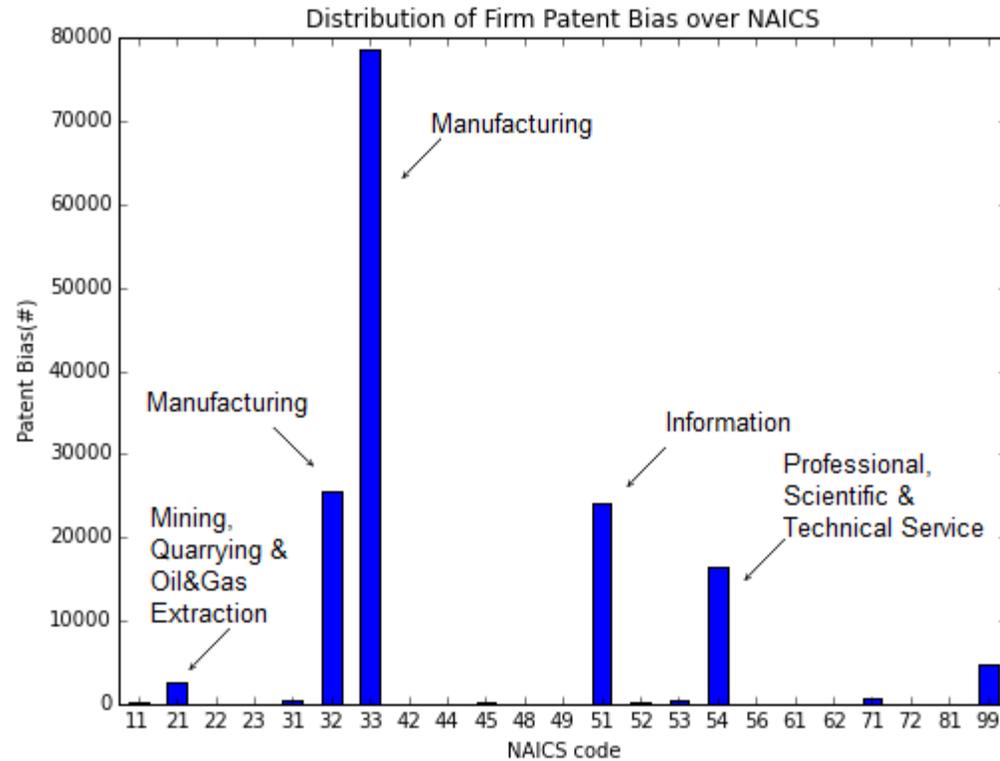

(a)
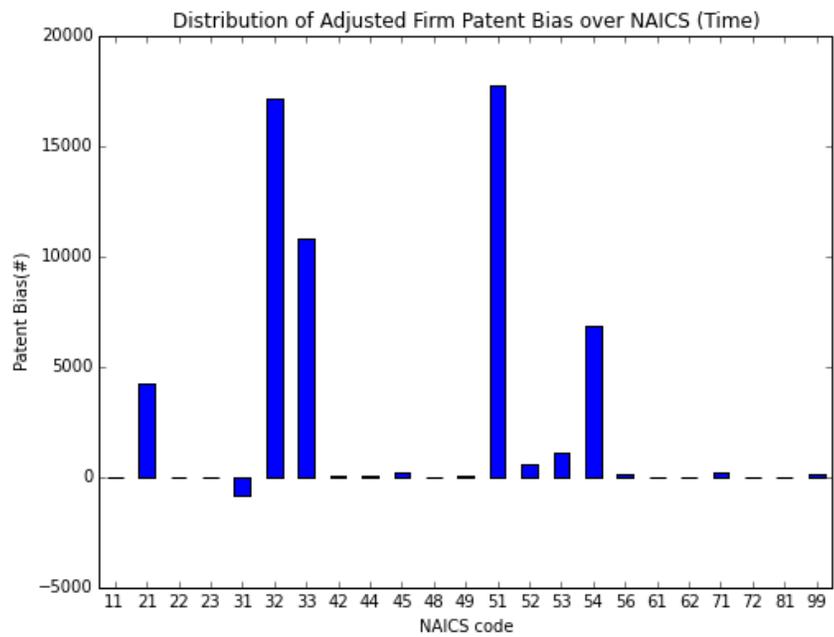
(b)
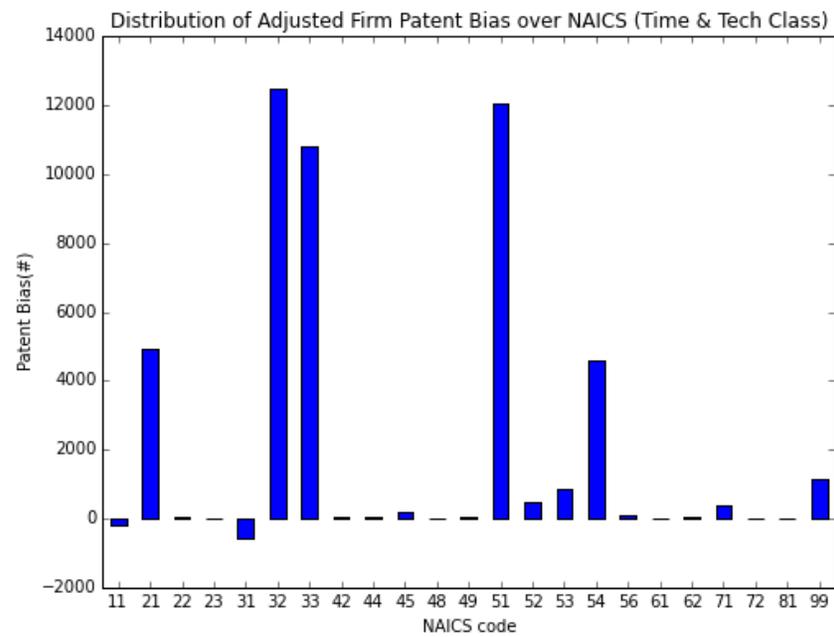
(c)

Figure 14: Firm Citation Bias (Unadjusted and Adjusted) across Industries (NAICS)

This figure presents the distribution for citation bias aggregated at the firm-year level from patents granted to public firms from 1976 through 2012 in different industries (classified using NAICS code). To compute the unadjusted citation bias for each firm-year, we compare the number of citations to all the patents for each firm filed in each application year in our data (i.e., citations in patents granted by 2012 to applications filed by a firm in a given year and granted by 2006) and in the NBER 2006 dataset (i.e., citations in patents granted by 2006 to applications filed by a firm in a given year and granted by 2006). We assign firms into different industries at time of the patent application using Compustat. We sum citation bias in each industry across publicly traded firms. The adjustments use the time fixed effect methodology, the time and technology class fixed effect methodology, and the citing year effect adjustment using the quasi structural method, with details discussed in Appendix B. Sources: NBER 2006 patent and our datasets.


(a)

Distribution of Adjusted Firm Citation Bias (Time) over NAICS

(b)

Distribution of Adjusted Firm Citation Bias (Time & Tech Class) over NAICS

(c)

Distribution of Adjusted Firm Citation Bias (Citing Year) over NAICS
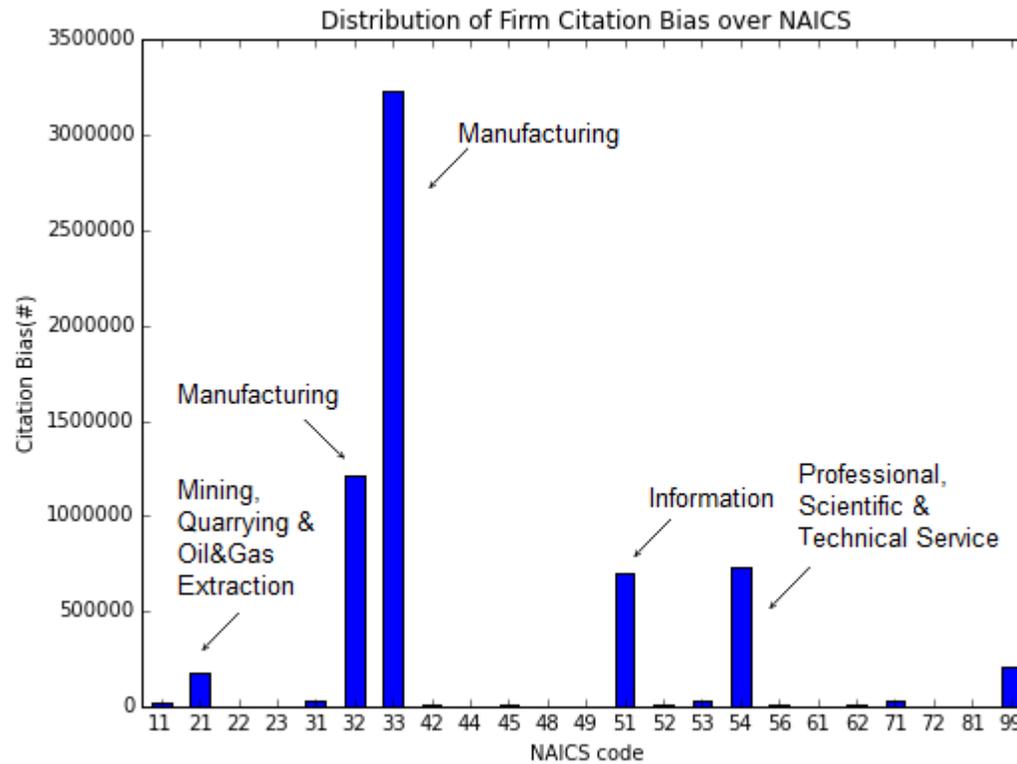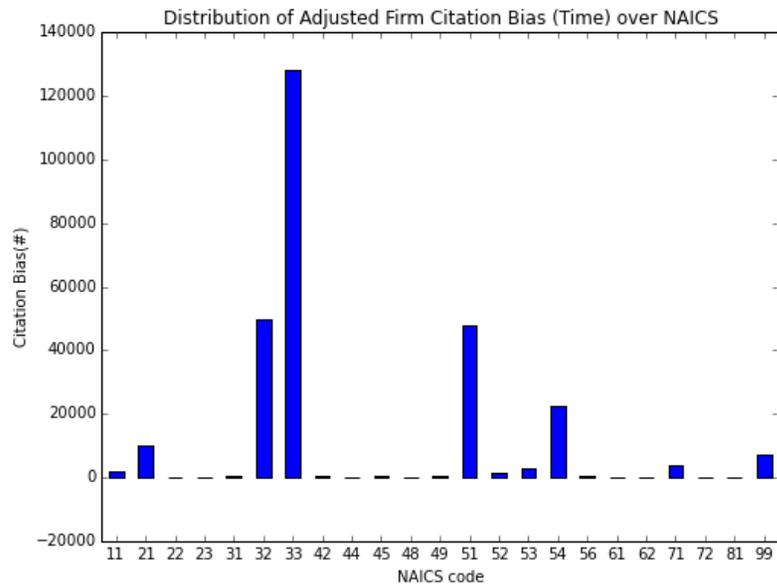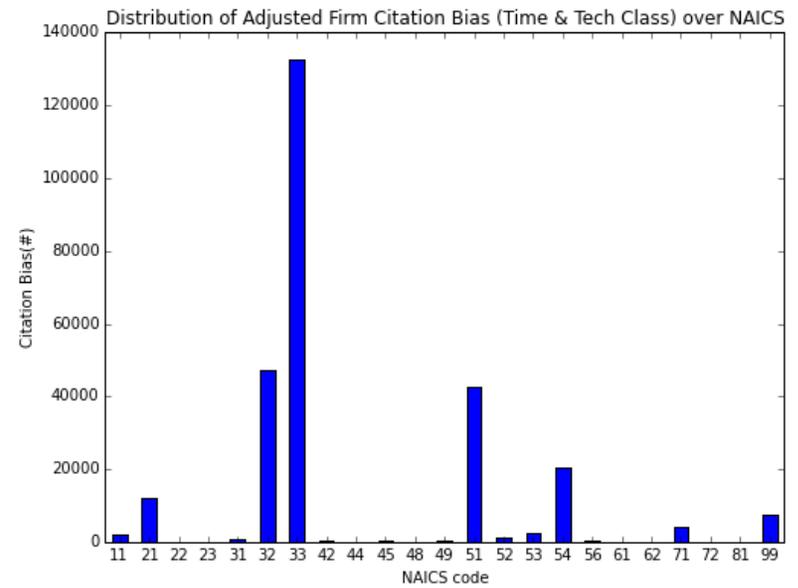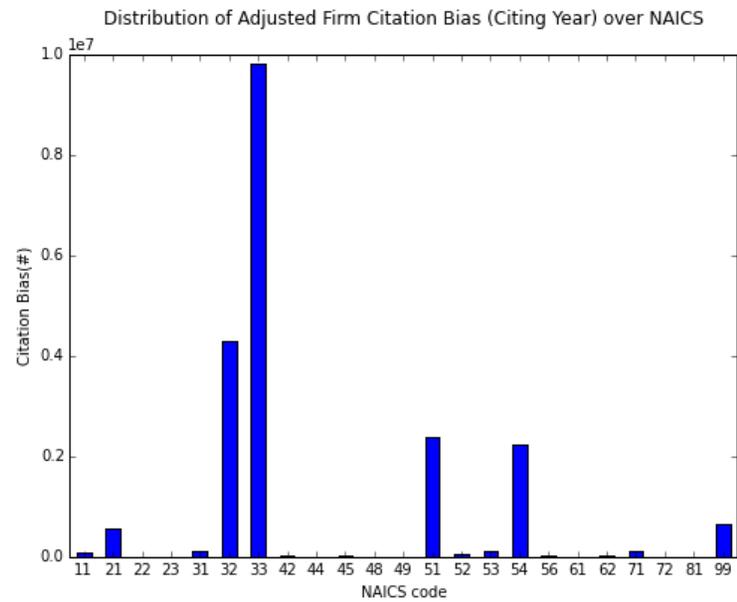
(d)

Figure 15: Bank Deregulation and Innovation

This figure plots the average citations per patent by year for patent applications before and after deregulation in 1985 for two groups: patents by firms incorporated in states that experienced deregulation pre-1985 and patents by firms incorporated in states that experienced deregulation post-1985. The results suggest that though there is a secular increase in cites per patents in both the groups, firms incorporated in states experiencing laws earlier see a larger effect when comparing pre regulation to post regulation. More importantly, while there is the familiar tapering in citations per patent towards the end of the sample, firms incorporated in states experiencing deregulation later see a smaller drop. This implies that simple time fixed effects may not be able to capture the substantial heterogeneity across group of firms in different states over time. Source: NBER 2006 patent dataset.
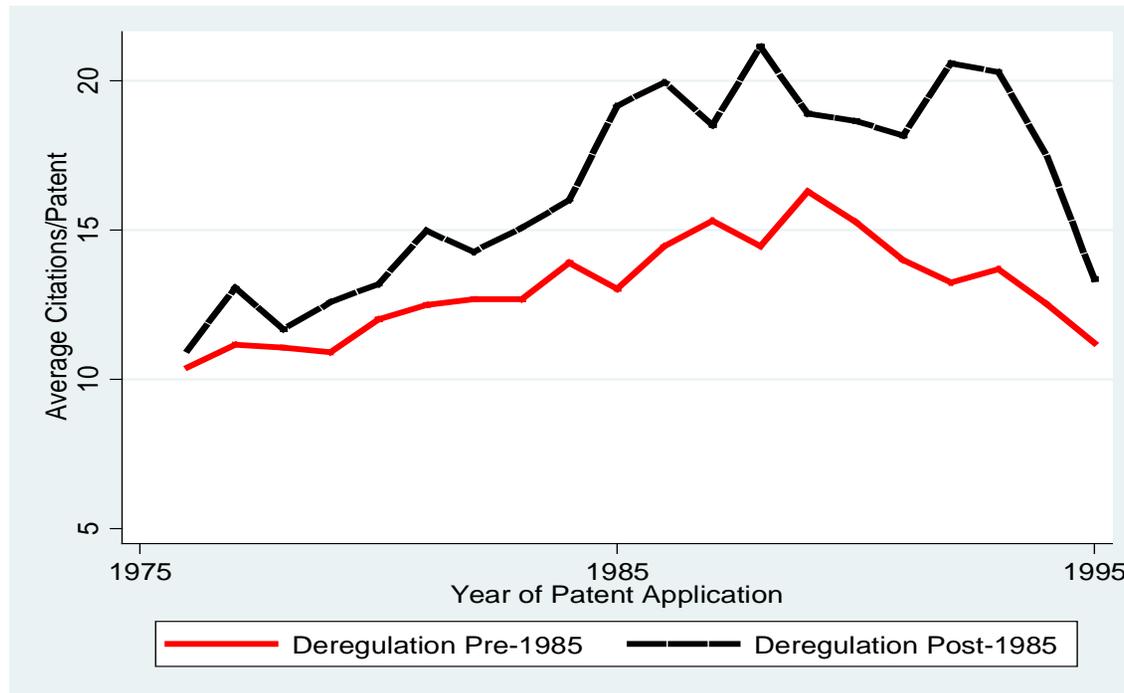
Figure 16: Anti-takeover Laws and Innovation

This figure presents the average citations per patent by year of patent application for firms incorporated in California (CA) and Delaware (DE). Together these figures suggest that anti-takeover laws, passed in DE in the middle of the time period, tend to decrease citations per patent for patents created by firms incorporated in DE. However, the magnitudes need to be interpreted more closely, since the substantial portion of drop in citations per patent after the anti-takeover laws is driven by patterns in patenting by firms incorporated in CA (which did not enact such a law). In other words, anti-takeover laws seem to matter to a large degree because the firms incorporated in other states which adopted these provisions did not experience the explosive growth in citations per patent that CA did. These patterns imply that simple state fixed effects will not account for such region-specific time trends. Source: NBER 2006 patent dataset.
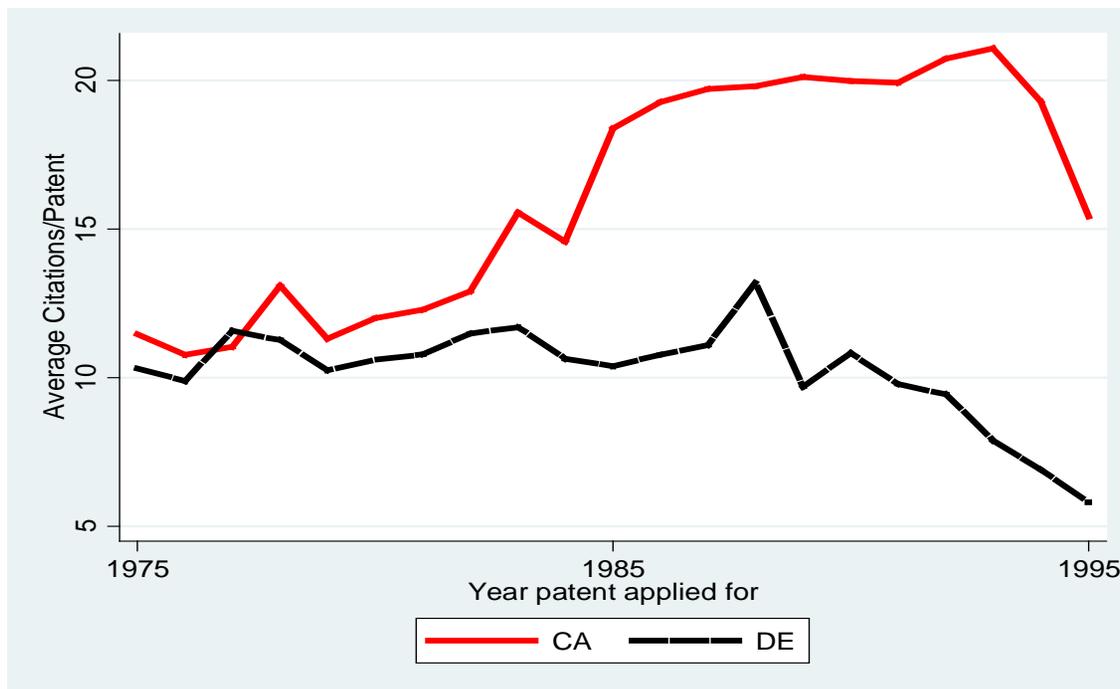
Table 1: Citation Distribution for Patent Cohorts

This table presents the distribution for citations to granted patents that were applied for between 1975 to 1991 and 1992 to 2006. The table shows that the distribution of citations is significantly different across the two patent cohorts. The mean patent in the 1975-1991 cohort gets twice as many citations as the mean in the 1992-2006 cohort. Similarly, the heavily cited patents (95th percentile and above) in the 1975-1991 cohort get at least 10 more citations than their equivalent counterparts in the 1992-2006 cohort. Source: NBER 2006 database.

Percentiles	Citations	
	Successful Patents applied for between 1975-1991	Successful Patents applied for between 1992-2006
1	0	0
5	0	0
10	1	0
25	3	0
50	6	2
75	13	6
90	23	13
95	33	21
99	69	50
Mean	10.5	5.3

Table 2: Citation Distribution by Technology Class

This table presents the distribution for citations to granted patents in different technology classes that were applied for between 1975 and 2006. The table shows that the distribution of citations is significantly different across patents granted in the technology class areas of "Mechanical" (HJT category 5), "Chemical" (HJT category 1), "Electronics" (HJT category 4) and "Computers/Communications" (HJT category 2). The citation distribution varies significantly across areas, especially at the top end of the distribution. Patents in the top end of the distribution of computers/communications and electronics areas tend to be cited more heavily relative to equivalent patents in the other categories depicted in the table. The lower end of the distribution looks similar across categories. Source: NBER 2006 database.

Percentiles	Citations			
	Successful Patents applied for "Mechanical" Technology Class	Successful Patents applied for "Chemicals" Technology Class	Successful Patents applied for "Electronics" Technology Class	Successful Patents applied for "Computers/Communication" Technology Class
	(1)	(2)	(3)	(4)
1	0	0	0	0
5	0	0	0	0
10	0	0	0	0
25	1	1	1	1
50	3	3	4	4
75	8	8	9	11
90	15	16	18	24
95	21	24	27	37
99	41	49	55	79
Mean	6.1	6.5	7.3	9.5

Table 3: Temporal Distribution of Citations

This table presents the temporal distribution for citations to granted patents in different technology classes that were applied for between 1975 and 2006. The table shows that the temporal distribution of citations is significantly different across patents granted in the technology class areas of "Mechanical" (HJT category 5), "Chemical" (HJT category 1), "Electronics" (HJT category 4) and "Computers/Communications" (HJT category 2). Patents in the computers/communications and electronics areas tend to be cited faster relative to equivalent patents in the other categories. A majority of citations (90th percentile) occur by around 13 years in the computers/communications area but an equivalent number of citations take five more years for patents in mechanical and chemical areas. Source: NBER 2006 database.

| Year after Patent Grant | **Distributions of Citations** | | | |
	Successful Patents applied for "Mechanical" Technology Class	Successful Patents applied for "Chemicals" Technology Class	Successful Patents applied for "Electronics" Technology Class	Successful Patents applied for "Computers/Communication" Technology Class
	(1)	(2)	(3)	(4)
0	0.35	0.27	0.46	0.44
1	4.63	3.79	5.51	5.71
2	12.45	10.60	14.54	15.65
3	21.22	18.76	24.77	26.95
4	30.01	27.34	34.93	38.35
5	38.21	35.67	44.21	49.05
6	45.59	43.30	52.31	58.48
7	52.15	50.23	59.30	66.37
8	57.97	56.40	65.31	72.80
9	63.15	61.87	70.40	77.76
10	67.83	66.76	74.81	81.86
11	72.01	71.12	78.63	85.14
12	75.77	75.00	81.90	87.86
13	79.14	78.46	84.73	90.09
14	82.13	81.48	87.14	91.95
15	84.77	84.16	89.23	93.48
16	87.08	86.54	91.02	94.76
17	89.13	88.69	92.58	95.84
18	90.90	90.55	93.92	96.70
19	92.47	92.22	95.05	97.40
20	93.84	93.69	96.01	97.96
21	95.04	94.99	96.83	98.42
22	96.08	96.12	97.53	98.81
23	96.98	97.06	98.12	99.11
24	97.75	97.85	98.61	99.36
25	98.41	98.52	99.02	99.55
26	98.95	99.04	99.35	99.71
27	99.36	99.43	99.60	99.82
28	99.67	99.72	99.80	99.91
29	99.88	99.90	99.93	99.97
30	100	100	100	100
	3,676,031	3,562,889	4,582,956	5,082,461

Table 4: Citation Distribution by State of Assignees

This table presents the distribution for citations to granted patents with assignees in California/Massachusetts (CA/MA) and Delaware (DE) that were applied for between 1975 and 2006. The table shows that the distribution of citations is significantly different across the patents produced by assignees in CA/MA relative to those produced by assignees in DE. The average (median) patent produced in CA/MA garners about three citations (one citation) more than that produced in DE. The differences become starker at the higher end of the distribution with heavily cited patents (95th percentile and above) produced in the CA/MA getting at least 13 more citations than their counterparts in DE. Source: NBER 2006 database.

Percentiles	Citations	
	Successful Patents applied for with assignees in CA/MA	Successful Patents applied for with assignees in DE
1	0	0
5	0	0
10	0	0
25	1	1
50	4	3
75	11	8
90	24	17
95	37	24
99	81	51
Mean	9.69	6.83

Table 5: Temporal Distribution of Citations by State of Assignees

This table presents the temporal distribution for citations to granted patents produced by assignees in CA/MA and DE that were applied for between 1975 and 2006. The table shows that the temporal distribution of citations is significantly different across patents produced in the two groups. Patents produced by assignees in CA/MA tend to be cited faster relative to equivalent patents produced in DE. A majority of citations (90th percentile) occur by around 15 years for patents produced in CA/MA but equivalent citations take three more years for patents produced in DE. Source: NBER 2006 database.

Year after Patent Grant	Distributions of Citations	
	Successful Patents applied for with inventors in "CA/MA"	Successful Patents applied for with inventors in "DE"
	(1)	(2)
0	0.34	0.26
1	4.51	3.77
2	12.72	10.76
3	22.62	19.12
4	32.91	27.89
5	42.79	36.55
6	51.65	44.42
7	59.27	51.53
8	65.73	57.77
9	70.96	63.29
10	75.46	68.14
11	79.24	72.55
12	82.48	76.30
13	85.27	79.71
14	87.63	82.69
15	89.63	85.24
16	91.36	87.47
17	92.83	89.48
18	94.08	91.17
19	95.14	92.72
20	96.05	94.07
21	96.82	95.28
22	97.51	96.31
23	98.08	97.20
24	98.58	97.95
25	98.99	98.60
26	99.32	99.07
27	99.58	99.43
28	99.78	99.73
29	99.92	99.91
30	100	100
	3,821,650	343,085

Table 6: Patent Bias and Firm Characteristics

This table presents OLS regressions relating unadjusted patent bias at the firm level with different firm characteristics. The dependent variable is the unadjusted patent bias of a given firm in that year for years 1976-2006 (columns 1-3) and for subsamples, 1976-1996 (columns 4-6) and 1997-2006 (columns 7-9). The dependent variable is computed as the difference in log of one plus number of successful patents filed by a firm in a given year as of 2012 ("our data") and the log of one plus number of successful patents filed by that firm in the same year as of the end of sample in the NBER 2006 dataset. Logs are taken to account for skewness in patenting activity. Control variables and their construction are described in Appendix C. Robust t-tests are reported in the parenthesis. Sources: NBER 2006 patent and our datasets.

Panel A: Unadjusted Patent Bias

	Unadjusted Patent Bias (1976-2006)			Unadjusted Patent Bias (1976-1996)			Unadjusted Patent Bias (1997-2006)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log_Size	0.0543***	0.0486***	0.0160*	0.0543***	0.0486***	0.0160*	0.0543***	0.0486***	0.0160*
	(12.44)	(10.83)	(1.76)	(12.44)	(10.83)	(1.76)	(12.44)	(10.83)	(1.76)
Log_M2B	0.0613***	0.0680***	0.0508***	0.0613***	0.0680***	0.0508***	0.0613***	0.0680***	0.0508***
	(3.97)	(4.39)	(3.37)	(3.97)	(4.39)	(3.37)	(3.97)	(4.39)	(3.37)
Log_RD2Sale	0.0241***	0.0358***	0.0441***	0.0241***	0.0358***	0.0441***	0.0241***	0.0358***	0.0441***
	(3.68)	(5.25)	(3.91)	(3.68)	(5.25)	(3.91)	(3.68)	(5.25)	(3.91)
Log_Cash2Asset	0.0196***	0.0189***	0.0147**	0.0196***	0.0189***	0.0147**	0.0196***	0.0189***	0.0147**
	(3.20)	(3.09)	(2.40)	(3.20)	(3.09)	(2.40)	(3.20)	(3.09)	(2.40)
Log_LEV	0.0769***	0.0769***	0.0502***	0.0769***	0.0769***	0.0502***	0.0769***	0.0769***	0.0502***
	(5.00)	(4.98)	(2.79)	(5.00)	(4.98)	(2.79)	(5.00)	(4.98)	(2.79)
IA	-0.0289	-0.0201	0.00398	-0.0289	-0.0201	0.00398	-0.0289	-0.0201	0.00398
	(-0.38)	(-0.26)	(0.06)	(-0.38)	(-0.26)	(0.06)	(-0.38)	(-0.26)	(0.06)
ROA	-0.124*	-0.0676	-0.0635	-0.124*	-0.0676	-0.0635	-0.124*	-0.0676	-0.0635
	(-1.80)	(-0.98)	(-0.90)	(-1.80)	(-0.98)	(-0.90)	(-1.80)	(-0.98)	(-0.90)
ROE	0.0942**	0.102***	0.130***	0.0942**	0.102***	0.130***	0.0942**	0.102***	0.130***
	(2.57)	(2.78)	(4.03)	(2.57)	(2.78)	(4.03)	(2.57)	(2.78)	(4.03)
SG	-0.0123	-0.0176	-0.0455**	-0.0123	-0.0176	-0.0455**	-0.0123	-0.0176	-0.0455**
	(-0.51)	(-0.73)	(-2.16)	(-0.51)	(-0.73)	(-2.16)	(-0.51)	(-0.73)	(-2.16)
NSI	-0.191**	-0.223**	-0.327***	-0.191**	-0.223**	-0.327***	-0.191**	-0.223**	-0.327***
	(-2.15)	(-2.53)	(-4.24)	(-2.15)	(-2.53)	(-4.24)	(-2.15)	(-2.53)	(-4.24)
Log_Spread	-0.0361*	-0.0310	-0.0783***	-0.0361*	-0.0310	-0.0783***	-0.0361*	-0.0310	-0.0783***
	(-1.80)	(-1.54)	(-4.07)	(-1.80)	(-1.54)	(-4.07)	(-1.80)	(-1.54)	(-4.07)
Log(state patents)	0.000932	0.00112	-0.0109***	0.000932	0.00112	-0.0109***	0.000932	0.00112	-0.0109***
	(0.58)	(0.70)	(-2.89)	(0.58)	(0.70)	(-2.89)	(0.58)	(0.70)	(-2.89)
Log(class patents)	0.0783***	0.0828***	0.00786***	0.0783***	0.0828***	0.00786***	0.0783***	0.0828***	0.00786***
	(4.68)	(4.98)	(3.10)	(4.68)	(4.98)	(3.10)	(4.68)	(4.98)	(3.10)
Observation	14503	14503	14503	14503	14503	14503	14503	14503	14503
R^2	0.434	0.441	0.682	0.434	0.441	0.682	0.434	0.441	0.682
Firm Fixed Effect			Yes			Yes			Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Fixed Effect	Yes	Yes		Yes	Yes		Yes	Yes	
NAICS Fixed Effect	Yes			Yes			Yes		
SIC Fixed Effect		Yes			Yes			Yes	

Table 6: Patent Bias and Firm Characteristics (contd.)

This table presents OLS regressions relating time fixed effect method adjusted patent bias at the firm level with different firm characteristics. The dependent variable is the time fixed effect adjusted patent bias of a given firm in that year for years 1976-2006 (columns 1-3) and for subsamples, 1976-1996 (columns 4-6) and 1997-2006 (columns 7-9). The dependent variable is computed as the difference in the log of one plus number of successful patents filed by a firm in a given year as of 2012 ("our data") and the log of one plus number of successful patents filed by that firm in the same year as of the end of sample in the NBER 2006 dataset. Logs are taken to account for skewness in patenting activity. Control variables and their construction are described in Appendix C. Details about the time fixed effect adjustment can be found in Appendix B. Robust t-tests are reported in the parenthesis. Sources: NBER 2006 patent and our datasets.

Panel B: Adjusted Patent Bias (Time)

	Adjusted Patent Bias (Time) (1976-2006)			Adjusted Patent Bias (Time) (1976-1996)			Adjusted Patent Bias (Time) (1997-2006)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log_Size	0.0138***	0.00740	0.0177*	-0.0152***	-0.0187***	0.0151**	0.0380***	0.0304***	0.0719***
	(2.89)	(1.50)	(1.71)	(-2.58)	(-3.10)	(2.00)	(5.43)	(4.23)	(3.48)
Log_M2B	0.0356**	0.0424**	0.0308*	0.0960***	0.0931***	0.0269**	-0.00371	0.00311	-0.00594
	(2.11)	(2.49)	(1.80)	(3.81)	(3.67)	(2.01)	(-0.16)	(0.14)	(-0.24)
Log_RD2Sale	0.00970	0.0252***	0.0299**	-0.00916	-0.0115	-0.00321	0.0220**	0.0508***	0.0582***
	(1.35)	(3.37)	(2.33)	(-0.97)	(-1.18)	(-0.37)	(2.20)	(4.84)	(2.65)
Log_Cash2Asset	0.0109	0.0102	0.00386	0.0295***	0.0319***	-0.00275	0.00296	-0.000769	0.00385
	(1.62)	(1.52)	(0.55)	(3.78)	(4.09)	(-0.68)	(0.29)	(-0.08)	(0.31)
Log_LEV	0.0267	0.0259	0.0248	0.0695***	0.0641***	0.0361**	0.00750	0.00240	0.0820**
	(1.58)	(1.53)	(1.21)	(2.93)	(2.69)	(2.31)	(0.32)	(0.10)	(2.55)
IA	-0.0500	-0.0416	-0.0188	0.0477	0.0757	0.0227	-0.0381	-0.0352	-0.0382
	(-0.60)	(-0.50)	(-0.25)	(0.49)	(0.77)	(0.51)	(-0.30)	(-0.28)	(-0.31)
ROA	-0.0251	0.0449	0.0606	-0.442***	-0.437***	0.109	0.127	0.249**	0.0753
	(-0.33)	(0.59)	(0.76)	(-3.58)	(-3.54)	(1.61)	(1.30)	(2.55)	(0.67)
ROE	0.0573	0.0634	0.0941**	0.147**	0.143**	-0.00586	0.00595	0.00686	0.0719
	(1.43)	(1.58)	(2.57)	(2.28)	(2.24)	(-0.21)	(0.11)	(0.13)	(1.40)
SG	-0.0198	-0.0235	-0.0735***	-0.00619	-0.0110	-0.0276	-0.0306	-0.0357	-0.0740**
	(-0.75)	(-0.89)	(-3.07)	(-0.16)	(-0.29)	(-1.61)	(-0.87)	(-1.02)	(-2.18)
NSI	0.0688	0.0298	-0.168*	-0.0212	-0.0602	-0.0926	0.114	0.0650	-0.0982
	(0.71)	(0.31)	(-1.91)	(-0.16)	(-0.47)	(-1.63)	(0.84)	(0.49)	(-0.75)
Log_Spread	-0.0387*	-0.0307	-0.0822***	-0.0208	-0.0140	-0.00810	-0.0746**	-0.0586*	-0.146***
	(-1.76)	(-1.39)	(-3.75)	(-0.85)	(-0.57)	(-0.71)	(-2.17)	(-1.69)	(-3.64)
Log(state patents)	0.00245	0.00257	-0.00832*	-0.00397	-0.00365	-0.00113	0.00436*	0.00402*	0.0101
	(1.39)	(1.46)	(-1.94)	(-1.53)	(-1.40)	(-0.37)	(1.86)	(1.73)	(1.38)
Log(class patents)	0.0595***	0.0654***	0.00402	0.0238	0.0281	0.000274	0.432***	0.435***	-0.00165
	(3.25)	(3.59)	(1.39)	(0.78)	(0.93)	(0.15)	(2.80)	(2.84)	(-0.36)
Observation	14503	14503	14503	5908	5908	5908	8595	8595	8595
R²	0.208	0.217	0.520	0.056	0.076	0.878	0.241	0.256	0.551
Firm Fixed Effect			Yes			Yes			Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Fixed Effect	Yes	Yes		Yes	Yes		Yes	Yes	
NAICS Fixed Effect	Yes			Yes			Yes		
sic Fixed Effect		Yes			Yes			Yes	

Table 6: Patent Bias and Firm Characteristics (contd.)

This table presents OLS regressions relating time and tech class fixed effect method adjusted patent bias at the firm level with different firm characteristics. The dependent variable is the time and tech class fixed effect adjusted patent bias of a given firm in that year for years 1976-2006 (columns 1-3) and for subsamples, 1976-1996 (columns 4-6) and 1997-2006 (columns 7-9). The dependent variable is computed as the difference in the log of one plus number of successful patents filed by a firm in a given year as of 2012 ("our data") and the log of one plus number of successful patents filed by that firm in the same year as of the end of sample in the NBER 2006 dataset. Logs are taken to account for skewness in patenting activity. Control variables and their construction are described in Appendix C. Details about the time and tech class fixed effect adjustment can be found in Appendix B. Robust t-tests are reported in the parenthesis. Sources: NBER 2006 patent and our datasets.

Panel C: Adjusted Patent Bias (Time and Tech Class)

	Adjusted Patent Bias (Time & Tech Class) (1976-2006)			Adjusted Patent Bias (Time & Tech Class) (1976-1996)			Adjusted Patent Bias (Time & Tech Class) (1997-2006)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log_Size	0.0138***	0.00740	0.0177*	-0.0152***	-0.0187***	0.0151**	0.0380***	0.0304***	0.0719***
	(2.89)	(1.50)	(1.71)	(-2.58)	(-3.10)	(2.00)	(5.43)	(4.23)	(3.48)
Log_M2B	0.0356**	0.0424**	0.0308*	0.0960***	0.0931***	0.0269**	-0.00371	0.00311	-0.00594
	(2.11)	(2.49)	(1.80)	(3.81)	(3.67)	(2.01)	(-0.16)	(0.14)	(-0.24)
Log_RD2Sale	0.00970	0.0252***	0.0299**	-0.00916	-0.0115	-0.00321	0.0220**	0.0508***	0.0582***
	(1.35)	(3.37)	(2.33)	(-0.97)	(-1.18)	(-0.37)	(2.20)	(4.84)	(2.65)
Log_Cash2Asset	0.0109	0.0102	0.00386	0.0295***	0.0319***	-0.00275	0.00296	-0.000769	0.00385
	(1.62)	(1.52)	(0.55)	(3.78)	(4.09)	(-0.68)	(0.29)	(-0.08)	(0.31)
Log_LEV	0.0267	0.0259	0.0248	0.0695***	0.0641***	0.0361**	0.00750	0.00240	0.0820**
	(1.58)	(1.53)	(1.21)	(2.93)	(2.69)	(2.31)	(0.32)	(0.10)	(2.55)
IA	-0.0500	-0.0416	-0.0188	0.0477	0.0757	0.0227	-0.0381	-0.0352	-0.0382
	(-0.60)	(-0.50)	(-0.25)	(0.49)	(0.77)	(0.51)	(-0.30)	(-0.28)	(-0.31)
ROA	-0.0251	0.0449	0.0606	-0.442***	-0.437***	0.109	0.127	0.249**	0.0753
	(-0.33)	(0.59)	(0.76)	(-3.58)	(-3.54)	(1.61)	(1.30)	(2.55)	(0.67)
ROE	0.0573	0.0634	0.0941**	0.147**	0.143**	-0.00586	0.00595	0.00686	0.0719
	(1.43)	(1.58)	(2.57)	(2.28)	(2.24)	(-0.21)	(0.11)	(0.13)	(1.40)
SG	-0.0198	-0.0235	-0.0735***	-0.00619	-0.0110	-0.0276	-0.0306	-0.0357	-0.0740**
	(-0.75)	(-0.89)	(-3.07)	(-0.16)	(-0.29)	(-1.61)	(-0.87)	(-1.02)	(-2.18)
NSI	0.0688	0.0298	-0.168*	-0.0212	-0.0602	-0.0926	0.114	0.0650	-0.0982
	(0.71)	(0.31)	(-1.91)	(-0.16)	(-0.47)	(-1.63)	(0.84)	(0.49)	(-0.75)
Log_Spread	-0.0387*	-0.0307	-0.0822***	-0.0208	-0.0140	-0.00810	-0.0746**	-0.0586*	-0.146***
	(-1.76)	(-1.39)	(-3.75)	(-0.85)	(-0.57)	(-0.71)	(-2.17)	(-1.69)	(-3.64)
Log(state patents)	0.00245	0.00257	-0.00832*	-0.00397	-0.00365	-0.00113	0.00436*	0.00402*	0.0101
	(1.39)	(1.46)	(-1.94)	(-1.53)	(-1.40)	(-0.37)	(1.86)	(1.73)	(1.38)
Log(class patents)	0.0595***	0.0654***	0.00402	0.0238	0.0281	0.000274	0.432***	0.435***	-0.00165
	(3.25)	(3.59)	(1.39)	(0.78)	(0.93)	(0.15)	(2.80)	(2.84)	(-0.36)
Observation	14503	14503	14503	5908	5908	5908	8595	8595	8595
R²	0.208	0.217	0.520	0.056	0.076	0.878	0.241	0.256	0.551
Firm Fixed Effect			Yes			Yes			Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Fixed Effect	Yes	Yes		Yes	Yes		Yes	Yes	
NAICS Fixed Effect	Yes			Yes			Yes		
sic Fixed Effect		Yes			Yes			Yes	

Table 7: Citation Bias and Firm Characteristics

This table presents OLS regressions relating unadjusted citation bias at the firm level with different firm characteristics. The dependent variable is the unadjusted citation bias of a given firm in that year for years 1976-2006 (columns 1-3) and for subsamples, 1976-1990 (columns 4-6) and 1991-2006 (columns 7-9). The dependent variable is computed as the difference in the log of one plus number of citations to all patents of a firm applied for in a given year and granted by 2006 in our data and the log of one plus number of citations to the same set of successful patents of that firm in the same application year in the NBER 2006 dataset. Restricting the successful patents from our data to only those that are granted by 2006 allows for comparison with successful patents in the NBER 2006 data. Logs are taken to account for skewness in citations activity. Control variables and their construction are described in Appendix C. Robust t-tests are reported in the parenthesis. Sources: NBER 2006 patent and our datasets.

Panel A: Unadjusted Citation Bias

	Unadjusted Citation Bias (1976-2006)			Unadjusted Citation Bias (1976-1990)			Unadjusted Citation Bias (1991-2006)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log_Size	0.124***	0.121***	0.0634***	-0.0487***	-0.0504***	0.0280	0.177***	0.175***	0.0763***
	(15.69)	(14.89)	(3.81)	(-3.41)	(-3.40)	(1.29)	(19.11)	(18.38)	(3.41)
Log_M2B	0.0996***	0.104***	0.0875***	0.214***	0.215***	0.00198	0.0283	0.0339	0.0265
	(3.58)	(3.73)	(3.17)	(2.94)	(2.92)	(0.05)	(0.92)	(1.10)	(0.81)
Log_RD2Sale	0.0428***	0.0516***	0.0934***	0.0263	0.0204	-0.0117	0.0544***	0.0664***	0.116***
	(3.64)	(4.19)	(4.53)	(1.06)	(0.79)	(-0.51)	(4.10)	(4.77)	(4.42)
Log_Cash2Asset	0.0512***	0.0516***	0.0559***	0.0696***	0.0726***	-0.0127	0.0431***	0.0410***	0.0578***
	(4.64)	(4.66)	(4.97)	(3.53)	(3.67)	(-1.24)	(3.33)	(3.17)	(3.98)
Log_LEV	0.0984***	0.0900***	0.115***	0.125*	0.103	-0.0196	0.0769**	0.0682**	0.109***
	(3.55)	(3.23)	(3.48)	(1.90)	(1.55)	(-0.46)	(2.48)	(2.19)	(2.75)
IA	-0.0571	-0.0102	-0.0986	0.186	0.220	0.200*	-0.0757	-0.0128	-0.0167
	(-0.42)	(-0.07)	(-0.81)	(0.74)	(0.87)	(1.84)	(-0.47)	(-0.08)	(-0.11)
ROA	-0.204	-0.182	-0.325**	-1.297***	-1.326***	-0.0134	-0.126	-0.0933	-0.154
	(-1.64)	(-1.46)	(-2.52)	(-3.66)	(-3.74)	(-0.06)	(-0.93)	(-0.69)	(-1.05)
ROE	0.0437	0.0452	0.155***	0.353**	0.306*	-0.0358	-0.0418	-0.0417	0.112*
	(0.66)	(0.69)	(2.63)	(2.01)	(1.75)	(-0.47)	(-0.58)	(-0.58)	(1.65)
SG	0.0744*	0.0620	0.0306	-0.0757	-0.0742	-0.114**	0.0818*	0.0644	0.0358
	(1.71)	(1.43)	(0.79)	(-0.68)	(-0.67)	(-2.36)	(1.73)	(1.37)	(0.82)
NSI	-0.590***	-0.628***	-0.614***	0.246	0.139	-0.0630	-0.683***	-0.719***	-0.595***
	(-3.69)	(-3.93)	(-4.34)	(0.68)	(0.39)	(-0.41)	(-3.86)	(-4.06)	(-3.61)
Log_Spread	-0.0554	-0.0538	-0.132***	-0.0665	-0.0553	-0.0162	-0.0736*	-0.0676	-0.203***
	(-1.53)	(-1.48)	(-3.73)	(-1.12)	(-0.93)	(-0.62)	(-1.70)	(-1.55)	(-4.32)
Log(state cites)	0.00473	0.00430	-0.0196***	-0.0138*	-0.0150**	-0.00106	0.00654**	0.00576*	-0.00661
	(1.60)	(1.45)	(-2.79)	(-1.86)	(-2.00)	(-0.12)	(2.02)	(1.78)	(-0.76)
Log(class cites)	0.173***	0.177***	0.0221***	-0.0437	-0.0300	-0.00132	0.0685	0.0740	0.0257***
	(5.14)	(5.30)	(4.66)	(-0.25)	(-0.18)	(-0.29)	(1.05)	(1.13)	(4.42)
Observation	14503	14503	14503	3090	3090	3090	11413	11413	11413
R²	0.322	0.328	0.606	0.093	0.115	0.892	0.271	0.278	0.561
Firm Fixed Effect			Yes			Yes			Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Fixed Effect	Yes	Yes		Yes	Yes		Yes	Yes	
NAICS Fixed Effect	Yes			Yes			Yes		
SIC Fixed Effect		Yes			Yes			Yes	

Table 7: Citation Bias and Firm Characteristics (contd.)

This table presents OLS regressions relating time fixed effect adjusted citation bias at the firm level with different firm characteristics. The dependent variable is the time fixed effect adjusted citation bias of a given firm in that year for years 1976-2006 (columns 1-3) and for subsamples, 1976-1990 (columns 4-6) and 1991-2006 (columns 7-9). The dependent variable is computed as the adjusted difference in the log of one plus number of citations to all patents of a firm applied for in a given year and granted by 2006 in our data and the log of one plus number of citations to the same set of successful patents of that firm in the same application year in the NBER 2006 dataset. Restricting the successful patents from our data to only those that are granted by 2006 allows for comparison with successful patents in the NBER 2006 data. Logs are taken to account for skewness in citations activity. Control variables and their construction are described in Appendix C. Details about the time fixed effect adjustment can be found in Appendix B. Robust t-tests are reported in the parenthesis. Sources: NBER 2006 patent and our datasets.

Panel B: Adjusted Citation Bias (Time)

	Adjusted Citation Bias (Time) (1976-2006)			Adjusted Citation Bias (Time) (1976-1990)			Adjusted Citation Bias (Time) (1991-2006)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log_Size	0.0825***	0.0794***	0.0256*	-0.0160*	-0.0139	0.0236*	0.113***	0.110***	0.0372**
	(13.55)	(12.68)	(1.88)	(-1.91)	(-1.60)	(1.85)	(15.23)	(14.40)	(1.98)
Log_M2B	0.0506**	0.0561***	0.0549**	0.141***	0.137***	0.00510	0.00799	0.0153	0.0214
	(2.35)	(2.60)	(2.43)	(3.30)	(3.18)	(0.23)	(0.32)	(0.62)	(0.78)
Log_RD2Sale	0.0281***	0.0381***	0.0697***	0.00553	-0.000308	-0.00844	0.0367***	0.0505***	0.0868***
	(3.09)	(4.00)	(4.13)	(0.38)	(-0.02)	(-0.62)	(3.44)	(4.51)	(3.96)
Log_Cash2Asset	0.0234***	0.0234***	0.0299***	0.0356***	0.0409***	-0.00424	0.0172*	0.0152	0.0303**
	(2.75)	(2.74)	(3.26)	(3.08)	(3.52)	(-0.70)	(1.65)	(1.46)	(2.48)
Log_LEV	0.0391*	0.0334	0.0726***	0.0781**	0.0643*	-0.00770	0.0225	0.0168	0.0774**
	(1.83)	(1.55)	(2.69)	(2.02)	(1.65)	(-0.31)	(0.90)	(0.67)	(2.33)
IA	-0.0350	0.00593	-0.00705	0.00185	0.0543	0.101	-0.00989	0.0370	0.0735
	(-0.33)	(0.06)	(-0.07)	(0.01)	(0.37)	(1.58)	(-0.08)	(0.29)	(0.59)
ROA	-0.0325	-0.00497	-0.0719	-0.604***	-0.636***	-0.0673	0.0102	0.0498	0.0261
	(-0.34)	(-0.05)	(-0.68)	(-2.90)	(-3.06)	(-0.55)	(0.09)	(0.46)	(0.21)
ROE	0.0295	0.0334	0.112**	0.257**	0.230**	0.0215	-0.0272	-0.0235	0.0837
	(0.58)	(0.65)	(2.33)	(2.49)	(2.25)	(0.49)	(-0.47)	(-0.41)	(1.47)
SG	0.0390	0.0278	-0.0157	-0.0395	-0.0466	-0.0675**	0.0409	0.0262	-0.0129
	(1.16)	(0.83)	(-0.50)	(-0.60)	(-0.72)	(-2.38)	(1.08)	(0.69)	(-0.35)
NSI	-0.368***	-0.403***	-0.372***	0.105	0.0463	-0.0738	-0.427***	-0.465***	-0.359***
	(-2.98)	(-3.27)	(-3.22)	(0.50)	(0.22)	(-0.81)	(-3.00)	(-3.27)	(-2.59)
Log_Spread	-0.0513*	-0.0478*	-0.0959***	-0.0186	-0.0109	0.00458	-0.0769**	-0.0694**	-0.161***
	(-1.84)	(-1.70)	(-3.33)	(-0.53)	(-0.31)	(0.30)	(-2.21)	(-1.97)	(-4.09)
Log(state cites)	0.00382*	0.00331	-0.0151***	-0.0110**	-0.0116***	-0.00390	0.00502*	0.00426	-0.00642
	(1.67)	(1.44)	(-2.62)	(-2.52)	(-2.63)	(-0.75)	(1.93)	(1.64)	(-0.88)
Log(class cites)	0.104***	0.109***	0.0122***	-0.0151	-0.00113	0.00113	-0.0285	-0.0235	0.0131***
	(4.02)	(4.21)	(3.16)	(-0.15)	(-0.01)	(0.42)	(-0.54)	(-0.45)	(2.68)
Observation	14503	14503	14503	3090	3090	3090	11413	11413	11413
R²	0.349	0.355	0.576	0.087	0.111	0.891	0.345	0.352	0.570
Firm Fixed Effect			Yes			Yes			Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Fixed Effect	Yes	Yes		Yes	Yes		Yes	Yes	
NAICS Fixed Effect	Yes			Yes			Yes		
SIC Fixed Effect		Yes			Yes			Yes	

Table 7: Citation Bias and Firm Characteristics (contd.)

This table presents OLS regressions relating time and tech class fixed effect adjusted citation bias at the firm level with different firm characteristics. The dependent variable is the time and tech class fixed effect adjusted citation bias of a given firm in that year for years 1976-2006 (columns 1-3) and for subsamples, 1976-1990 (columns 4-6) and 1991-2006 (columns 7-9). The dependent variable is computed as the adjusted difference in the log of one plus number of citations to all patents of a firm applied for in a given year and granted by 2006 in our data and the log of one plus number of citations to the same set of successful patents of that firm in the same application year in the NBER 2006 dataset. Restricting the successful patents from our data to only those that are granted by 2006 allows for comparison with successful patents in the NBER 2006 data. Logs are taken to account for skewness in citations activity. Control variables and their construction are described in Appendix C. Details about the time and tech class fixed effect adjustment can be found in Appendix B. Robust t-tests are reported in the parenthesis. Sources: NBER 2006 patent and our datasets.

Panel C: Adjusted Citation Bias (Time and Tech Class)

	Adjusted Citation Bias (Time & Tech Class) (1976-2006)			Adjusted Citation Bias (Time & Tech Class) (1976-1990)			Adjusted Citation Bias (Time & Tech Class) (1991-2006)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log_Size	0.0839***	0.0798***	0.0238*	-0.0152*	-0.0129	0.0242*	0.116***	0.111***	0.0371**
	(13.72)	(12.67)	(1.74)	(-1.77)	(-1.46)	(1.89)	(15.45)	(14.44)	(1.97)
Log_M2B	0.0566***	0.0624***	0.0592***	0.153***	0.152***	0.00850	0.00936	0.0168	0.0260
	(2.62)	(2.87)	(2.61)	(3.51)	(3.43)	(0.38)	(0.38)	(0.67)	(0.94)
Log_RD2Sale	0.0302***	0.0396***	0.0662***	0.00552	-0.000997	-0.0132	0.0397***	0.0529***	0.0859***
	(3.30)	(4.15)	(3.91)	(0.37)	(-0.06)	(-0.97)	(3.70)	(4.71)	(3.90)
Log_Cash2Asset	0.0198**	0.0190**	0.0248***	0.0336***	0.0385***	-0.00646	0.0139	0.0112	0.0256**
	(2.31)	(2.21)	(2.68)	(2.84)	(3.25)	(-1.07)	(1.33)	(1.07)	(2.09)
Log_LEV	0.0519**	0.0454**	0.0801***	0.0937**	0.0805**	-0.000708	0.0338	0.0273	0.0873***
	(2.41)	(2.09)	(2.96)	(2.37)	(2.02)	(-0.03)	(1.35)	(1.09)	(2.61)
IA	-0.0266	0.0102	-0.00596	0.00182	0.0509	0.102	-0.00648	0.0370	0.0659
	(-0.25)	(0.10)	(-0.06)	(0.01)	(0.34)	(1.59)	(-0.05)	(0.29)	(0.52)
ROA	-0.0341	-0.00377	-0.0929	-0.574***	-0.610***	-0.103	0.00925	0.0535	0.00653
	(-0.35)	(-0.04)	(-0.88)	(-2.69)	(-2.87)	(-0.84)	(0.08)	(0.49)	(0.05)
ROE	0.0268	0.0303	0.108**	0.270**	0.244**	0.0373	-0.0338	-0.0306	0.0797
	(0.52)	(0.59)	(2.23)	(2.56)	(2.32)	(0.84)	(-0.58)	(-0.53)	(1.40)
SG	0.0288	0.0178	-0.0245	-0.0368	-0.0450	-0.0692**	0.0304	0.0161	-0.0247
	(0.85)	(0.53)	(-0.77)	(-0.55)	(-0.68)	(-2.42)	(0.80)	(0.42)	(-0.67)
NSI	-0.335***	-0.368***	-0.337***	0.0901	0.0301	-0.103	-0.383***	-0.418***	-0.316**
	(-2.70)	(-2.97)	(-2.90)	(0.41)	(0.14)	(-1.13)	(-2.67)	(-2.92)	(-2.28)
Log_Spread	-0.0417	-0.0398	-0.0908***	-0.0153	-0.00706	0.00773	-0.0646*	-0.0587*	-0.152***
	(-1.48)	(-1.41)	(-3.14)	(-0.43)	(-0.20)	(0.50)	(-1.85)	(-1.66)	(-3.83)
Log(state cites)	0.00308	0.00254	-0.0165***	-0.0114**	-0.0120***	-0.00494	0.00444*	0.00367	-0.00728
	(1.34)	(1.10)	(-2.85)	(-2.56)	(-2.66)	(-0.94)	(1.70)	(1.40)	(-1.00)
Log(class cites)	0.0904***	0.0947***	0.00748*	-0.0233	-0.00609	0.000129	0.00605	0.0113	0.00749
	(3.46)	(3.64)	(1.92)	(-0.23)	(-0.06)	(0.05)	(0.11)	(0.21)	(1.53)
Observation	14503	14503	14503	3090	3090	3090	11413	11413	11413
R²	0.346	0.352	0.575	0.078	0.102	0.894	0.343	0.351	0.569
Firm Fixed Effect			Yes			Yes			Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Fixed Effect	Yes	Yes		Yes	Yes		Yes	Yes	
NAICS Fixed Effect	Yes			Yes			Yes		
SIC Fixed Effect		Yes			Yes			Yes	

Table 7: Citation Bias and Firm Characteristics (contd.)

This table presents OLS regressions relating adjusted citation bias at the firm level with different firm characteristics. The adjustment is for citing year effect using the quasi-structural method. The dependent variable is the adjusted citation bias of a given firm in that year for years 1976-2006 (columns 1-3) and for subsamples, 1976-1990 (columns 4-6) and 1991-2006 (columns 7-9). The dependent variable is computed as the adjusted difference in the log of one plus number of citations to all patents of a firm applied for in a given year and granted by 2006 in our data and the log of one plus number of citations to the same set of successful patents of that firm in the same application year in the NBER 2006 dataset. Restricting the successful patents from our data to only those that are granted by 2006 allows for comparison with successful patents in the NBER 2006 data. Logs are taken to account for skewness in citations activity. Control variables and their construction are described in Appendix C. Details about the citing year effect adjustment using the quasi structural method can be found in Appendix B. Robust t-tests are reported in the parenthesis. Sources: NBER 2006 patent and our datasets.

Panel D: Adjusted Citation Bias (Quasi Structural Method)

	Adjusted Citation Bias (Citing Year) (1976-2006)			Adjusted Citation Bias (Citing Year) (1976-1990)			Adjusted Citation Bias (Citing Year) (1991-2006)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log_Size	0.0485***	0.0443***	0.0427*	-0.0769***	-0.0826***	0.0116	0.0889***	0.0873***	0.0718**
	(4.68)	(4.15)	(1.91)	(-4.33)	(-4.49)	(0.39)	(7.19)	(6.85)	(2.35)
Log_M2B	0.110***	0.112***	0.134***	0.252***	0.261***	0.0169	0.0415	0.0435	0.0749*
	(3.00)	(3.05)	(3.60)	(2.78)	(2.85)	(0.33)	(1.01)	(1.05)	(1.67)
Log_RD2Sale	0.0259*	0.0375**	0.0685**	0.0401	0.0400	-0.00283	0.0292	0.0434**	0.0896**
	(1.67)	(2.31)	(2.47)	(1.30)	(1.25)	(-0.09)	(1.64)	(2.33)	(2.50)
Log_Cash2Asset	0.0546***	0.0553***	0.0558***	0.0809***	0.0851***	-0.0186	0.0446***	0.0419**	0.0623***
	(3.76)	(3.80)	(3.69)	(3.30)	(3.46)	(-1.33)	(2.58)	(2.42)	(3.13)
Log_LEV	0.0565	0.0450	0.158***	0.156*	0.129	-0.0112	0.0185	0.00713	0.171***
	(1.55)	(1.22)	(3.57)	(1.91)	(1.56)	(-0.19)	(0.45)	(0.17)	(3.15)
IA	0.0213	0.0932	-0.0426	0.358	0.373	0.324**	-0.00997	0.0787	-0.00226
	(0.12)	(0.51)	(-0.26)	(1.14)	(1.19)	(2.21)	(-0.05)	(0.37)	(-0.01)
ROA	-0.223	-0.188	-0.297*	-1.503***	-1.584***	0.122	-0.109	-0.0662	-0.131
	(-1.36)	(-1.15)	(-1.72)	(-3.40)	(-3.60)	(0.43)	(-0.61)	(-0.36)	(-0.65)
ROE	0.0444	0.0391	0.210***	0.387*	0.322	-0.0766	-0.0446	-0.0536	0.173*
	(0.51)	(0.45)	(2.66)	(1.77)	(1.48)	(-0.75)	(-0.46)	(-0.56)	(1.87)
SG	0.126**	0.113**	0.0169	-0.0998	-0.0712	-0.106	0.141**	0.122*	0.0260
	(2.20)	(1.98)	(0.33)	(-0.72)	(-0.52)	(-1.62)	(2.23)	(1.92)	(0.43)
NSI	-0.531**	-0.577***	-0.586***	0.399	0.226	-0.0122	-0.649***	-0.690***	-0.563**
	(-2.52)	(-2.75)	(-3.08)	(0.88)	(0.50)	(-0.06)	(-2.74)	(-2.91)	(-2.49)
Log_Spread	-0.0908*	-0.0872*	-0.159***	-0.0735	-0.0674	-0.0181	-0.115**	-0.102*	-0.216***
	(-1.91)	(-1.82)	(-3.36)	(-0.99)	(-0.91)	(-0.51)	(-1.98)	(-1.74)	(-3.36)
Log(state cites)	0.00543	0.00462	-0.0230**	-0.0103	-0.0131	0.000280	0.00680	0.00556	-0.00956
	(1.39)	(1.18)	(-2.43)	(-1.11)	(-1.41)	(0.02)	(1.57)	(1.28)	(-0.81)
Log(class cites)	0.107**	0.116***	0.0159**	-0.105	-0.102	-0.00408	-0.0693	-0.0624	0.0203**
	(2.43)	(2.64)	(2.49)	(-0.49)	(-0.48)	(-0.65)	(-0.79)	(-0.71)	(2.56)
Observation	14503	14503	14503	3090	3090	3090	11413	11413	11413
R²	0.324	0.329	0.590	0.091	0.116	0.871	0.278	0.283	0.544
Firm Fixed Effect			Yes			Yes			Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Fixed Effect	Yes	Yes		Yes	Yes		Yes	Yes	
NAICS Fixed Effect	Yes			Yes			Yes		
SIC Fixed Effect		Yes			Yes			Yes	

Table 8: Bank Deregulation and Innovation

This table reports OLS estimates of regressions where the dependent variable is the citations per patent produced by a given firm in that year. The variable *Post Interbank Deregulation* is a dummy variable denoting if the state had yet undergone banking deregulation. Other controls include firm size, leverage, asset tangibility, market-to-book, and the R&D-to-sales ratio. The regressions include year and state fixed effects. Column (1) shows the results with the basic set of controls over the entire sample period. Column (2) and Column (3) restrict the sample period to pre-1985 and post-1985, respectively. As can be observed, while the deregulation dummy is positive in the entire sample and in the pre-1985 period, the post-1985 deregulation dummy is negative. The estimation period is 1975 to 1995. Standard errors are in parentheses and clustered at the state level. Source: NBER 2006 database.

	Citations/Patent		
	(1)	(2)	(3)
Post Interbank Deregulation	0.018**	0.102***	-0.119***
	(2.25)	(3.64)	(4.57)
Clustering Unit	State	State	State
Other Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Deregulation Event	All Years	Pre 1985	Post 1985
Observations	10772	3949	6823

Table 9: Impact of Technology Class

This table reports OLS estimates of regressions where the dependent variable is the log of one plus the citations per patent produced by a given firm in that year. The variable *Cash/Assets* is the firm's ratio of cash to assets in a given year, as reported by Compustat. The variable *Citation/Patent Tech Class* is the number of citations per patent for a given technology class in that year. Other controls include firm size, leverage, asset tangibility, market-to-book, and the R&D-to-sales ratio. The regressions include year and industry (3 digit SIC) fixed effects. Column (1) shows the effect of the cash-to-asset ratio given controls and fixed effects, while excluding the citations per patent. Column (2) runs the same regression as in Column (1), but now includes the citations per patent at the technology class level as an additional explanatory variable. The variation that is explained in the data increases substantially. Columns (3) through (8) restrict the sample to a specific technology class, namely Chemicals, Computers, Drugs, Electronics, Mechanical, and Others. The basic pattern, shown in the first regression, is that firms with more cash have more highly cited patents. However, when looking through the remaining columns, we find an enormous variation across industries, with the coefficient for computers almost four times larger than the coefficient for drugs. The estimation period is 1975 to 1995. Standard errors are in parentheses and clustered at the firm level. Source: NBER 2006 database.

	Log (1+Citations/Patent)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cash/Assets	7.71***	6.00***	3.99***	11.64***	3.18***	5.91***	5.66***	8.36***
	(14.01)	(11.11)	(3.02)	(7.91)	(2.97)	(4.96)	(2.88)	(5.50)
Citation/Patent Tech Class		1.41***	1.24***	1.06***	1.01***	1.21***	1.27**	0.66
(in '00000)		(47.2)	(4.59)	(5.57)	(3.74)	(6.72)	(2.59)	(0.94)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24502	24502	3756	4458	3860	3819	3790	4819
Primary Tech Class	All	All	Chemicals	Computers	Drugs	Electrical	Mechanical	Others
R-squared	0.22	0.28	0.27	0.34	0.36	0.30	0.15	0.20

Table 10: Anti-Takeover Laws and Innovation

This table reports OLS estimates of regressions where the dependent variable is the log of one plus the citations per patent produced by a given firm in that year. The variable *Post Anti-Takeover* is a dummy variable denoting whether the firm was located in a state that had antitakeover legislation in place in a given year. Other controls include firm size, leverage, asset tangibility, market-to-book, and the R&D-to-sales ratio. The regressions include year and state fixed effects. Column (1) shows the results with the basic set of controls over the entire sample period. Column (2) and Column (3) restrict the sample to firms not in California or Massachusetts and those only in California and Massachusetts, respectively. Standard errors are in parentheses and clustered at the state level. Source: NBER 2006 database.

	Log (1+Citations/Patent)		
	(1)	(2)	(3)
Post Anti-Takeover	-0.0872***	-0.013*	0.241***
	(2.79)	(1.86)	(2.69)
Clustering Unit	Firm	Firm	Firm
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Sample	All States	Not CA/MA	CA/MA
Observations	37577	28685	8892
R-squared	0.077	0.028	0.014

Table 11: Checklist for Analyses

This table presents the final checklist for the analyses based on the principles developed and discussed in the paper.

1. To what extent are the key policy changes occurring around times when patenting and citations per patent accelerated?
2. Are firms in industries that experienced a surge of patenting or in citations per patent (e.g., computers and electronics) included in one of the sub-populations being analyzed?
3. To what extent are firms in states that experienced a surge of patenting or citations per patent (e.g., California and Massachusetts) included in one of the sub-populations being analyzed?
4. Are firms with features akin to those that experienced a surge in patenting or citations per patent (e.g., those with a high market-to-book value) included in one of the sub-populations being analyzed?
5. To what extent are the patterns consistent across the sample (e.g., across the entire period under study), or do they vary with in ways that might be associated with unobserved factors that may be driving patenting practice? Are the coefficients of the effect consistent across the sample, or being driven by a sub-group?
6. May the results be driven by selection biases, due to the researchers' inability to observe pending patents or not-yet cited patents? Are the results robust to treat these truncation biases in different ways? Do patent and citation biases explain the results?
7. Could the results be driven by the exit of firms, and the likelihood that some or all of the patents pending at the time of exit will not be assigned to this firm, but rather to a successor entity? To what extent may this exit truncation problem be linked to the phenomenon under study?
8. Is there any way to ascertain the extent to which the firms under study may be engaging in misleading assignment practices, in order to disguise their technological strategy from competitors?
9. To what extent may the limitations of the concordances between patent assignees and firms be systematically affecting the results of the analysis (a consideration particularly relevant when the implications of the market for corporate control on innovation are being studied)?
10. Do the citation practices of the firms under study differ significantly from the norm, which might suggest that firms are engaging in strategic use of citations?