# Measuring the Economic Impact of Science with In-Text Patent Citations*

Kevin A. Bryan[†] Yasin Ozcan[‡] Bhaven Sampat[§]

September 15, 2018

**Abstract**

We introduce, validate, and provide a public database of a new measure of the knowledge inventors draw on: scientific references in patent specifications. These references are common and algorithmically extractable. Critically, they are very different from the "front page" prior art commonly used to proxy for inventor knowledge. Only 24% of front page citations to academic articles are in the patent text, and 31% of in-text citations are on the front page. We explain these differences by describing the legal rules and practice governing citation. Empirical validations suggest that in-text citations appear to more accurately measure real knowledge flows, consistent with their legal role.

*Parts of this project were supported by the Office of the Director, National Institutes of Health (OD), Office of Behavioral and Social Sciences Research (OBSSR) and the National Institute on Aging (AG) under Award Number R24 AG048059 to the National Bureau of Economic Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH or NBER.

[†]University of Toronto, Rotman School of Management.

[‡]NBER

[§]Columbia University.

# 1 Introduction

What prior knowledge do inventors and firms use as inputs to their research? These spillovers are at the core of urban economics, growth, and the economics of innovation. The empirical challenge is that "knowledge flows...are invisible; they leave no paper trail by which they may be measured and tracked" (Krugman [1991]). Trying to make these flows visible, researchers have surveyed firms (Levin et al. [1987], Cohen et al. [2000]), written qualitative industry histories (von Hippel [1988]), and investigated inventor biographies (Khan et al. [2014], Moser [2005]). However, by far the most commonly-used measure has been prior art cited on the "front page" of granted patents (Jaffe et al. [1993a], Narin [1994], Trajtenberg [1990]).

Several pioneering studies, beginning in the 1990s, investigated the impact of science on industrial innovation by looking at front page citations from private sector patents to academic patents (Jaffe et al. [1993b], Henderson et al. [1998]). But patents also cite non-patent references, including scientific literature which for two reasons may represent an even more promising way to evaluate the effects of public science. First, non-patent citations are less likely to come from examiners than patent-patent citations [Lemley and Sampat, 2012]. Moreover, since most academic research is disseminated through publications (rather than patents) patent citations to publications capture a broader range of potential impact than patent-patent citations [Agrawal and Henderson, 2002].

Recently, several research teams have taken advantage of computational advances and made progress in extracting front page citations to academic research and linking them to scientific literature databases. These data have been used for analyzing the impact of NIH research on private sector patenting (Azoulay et al. [2015], Li et al. [2017]), the impact of publicly funded energy research on applied technology (Popp [2017]), the economic impact of universities (Jefferson et al. [2018]), and the distance between science and technology in different fields (Ahmadpoor and Jones [2017]). Funding agencies such as the NSF and NIH are also increasingly using these indicators for evaluation purposes.

We argue that there is a fundamental problem in interpreting front page citations as knowledge flows. These citations are not simply a list of earlier patents, academic articles, and other documents which were relevant to the invention. Rather, they derive from a legal "duty of disclosure" requiring inventors to list documents material to the patentability of their claims. Material to patentability means there is no need to cite tools that help build the invention, information used to avoid unpromising paths, basic facts that focus effort, or research suggesting a technological hole or a market need. That is, front page citations in patents are not analogous to references in academic papers.[1]

However, patents contain an alternative measure: citations in the specification text itself. Patent specifications by law must include enough detail that someone "skilled in the art" could replicate the invention. As Figure 1 shows, inventors fulfill this requirement in part by citing literature while describing the background of their invention, why it is novel, and how it is built. These descriptive parts of patents are often largely written by the inventors themselves, rather than by patent attorneys who focus on the more legally consequential claims and prior art disclosure. Both their legal purpose and their practical construction suggest that in-text citations should better measure the real knowledge inventors use to motivate and construct their inventions.

Despite this theoretical advantage in tracking knowledge flows, innovation researchers have not used in-text citations for two reasons. First, there is a folk belief, which we will show is untrue, that citations in the text almost always appear on the front page. Second, since patents do not have bibliographies and in-text citations have no standardized format, they are difficult to extract.[2] We develop a practical algorithm for extracting

[1]See Meyer [2000] on the differences between academic and front page patent citation patterns.

[2]The only previous paper using individual in-text references, by two of the present authors, investigated citations to articles in 44 medical journals while studying the effect of the NIH open access mandate (Bryan and Ozcan [2018]). The only large-scale institutional attempt we know of to extract in-text references is an EPO trial beginning in 2006 to include a "summary of references for the reader's convenience" with in-text patent and academic references attached to patent pdfs (EPO [2007]). We are unaware of any research in economics or innovation that has used this measure. At an aggregate level, Tamada et al. [2006] attempts to find "regular expressions" in the text that look like references to academic science, then counts these across various classes in Japanese patents.

these citations, and use it to develop a public database of all 2,779,258 front page and in-text citations to every article since 1984 in 260 journals across 20 fields. The distinction between the two types of citations is dramatic. Only 31% of in-text citations appear on the front page, and only 24% of front page citations appear in the patent text.[3]

After describing the formal and practical differences in the legal origins of in-text versus front page citations, we perform three empirical investigations. First, we examine "patent-paper pairs" where a biotechnology patent derives from the same research as an academic article (Murray and Stern [2007]). References in the underlying academic article bibliography are cited 68 to 138% more often in the text of corresponding patents than on the front page. Second, we show that the number of in-text citations to academic research in a firm's patents is correlated with the firm R&D manager's stated reliance on public sector research and open science, using data from the survey in Roach and Cohen [2013]. . Third, in-text citations actually perform *worse* than front page citations as a proxy for value. Patents with more front page scientific citations are more valuable according to four separate measures, and the link between in-text citations and value is generally either nonexistent or less strong. This ought not be surprising: even if in-text references better proxy for knowledge flows, this does not imply that they better proxy for everything front page references have been used to study. Indeed, we might expect inventors and their attorneys to perform more comprehensive prior art searches for patents they expect will be more valuable (Sampat [2010]).

This paper has the primary purpose of introducing researchers to in-text citations and their properties. We see this tool as complementing the recent broad expansion in techniques exploiting patent text. The first generation of patent-based studies largely used count measures like patent classes, the number of forward and backward front page citations, and so on. A series of recent papers uses modern computing power and natural language processing to develop statistically-usable metrics from the raw text of patents.

[3] Our publication data contains all articles in selected journals from 1984 to 2016, and our patent dataset includes all public US patents as of May 2018. ADD GITHUB ADDRESS HERE.

For example, Kelly et al. [2017] identifies impactful patents by examining text that is unlike existing patents but similar to future patents, Kuhn and Thompson [2017] uses the length of patent claims to measure the scope of patents, and Kaplan and Vakili [2015] use topic modeling to identify breakthrough patents.

## 2 The Empirics of In-Text Citations

The major practical difficulty with using in-text references is that they have no standard format. Patents do not have bibliographies, and references to publications are made in the flow of text. For instance, US6551784 writes that "methods for aligning sequences using the CLUSTAL program are well described by Higgins and Sharp in *Gene*, 73: 237-244 (1988) and in CABIOS 5: 151-153 (1989))." Extracting that reference requires knowing both that there are two references to academic publications in this sentence, and that the CABIOS article refers to one written by Higgins and Sharp. Some in-text references are incredibly vague, such as "Genomic DNA was obtained from leaf tissue according to Doyle and Doyle (1987)" in US6483012.

The most natural approach to follow is a coarsened match between article metadata and text in a patent. Coarse matches permit mis-spellings, various combinations of metadata, and so on.[4] The problem in our setting is that identifying a chunk of text as containing a scientific reference is itself a challenging problem in the absence of a bibliography or the one-line-represents-one-citation format of front page citations. We would therefore need to apply the coarsened algorithm directly to over a terabyte of text. In practice, our algorithm, described in detail in the Online Appendix, is much faster and much less likely to generate false positives. Further, our algorithm captures many citations where metadata are not located near each other in the patent text. For example, in US6605754, "Comai et al

[4]E.g., US6190856 has an article by Erkki Koivunen cited as "Korvunen"; these mis-spellings are not uncommon. US6130090 cites a Bradley and Liu paper giving the year as 1996 instead of the correct 1997. US630824 cites a 1989 paper as being published in "Genetics" when it actually appeared in the journal "Genomics"; ironically, this article was written by one of the inventors!

have previously described a chimeric plant promoter combining elements of the CaMV35S and the mannopine synthase (mas) promoters (1990, Plant Mol Biol, 15:373-381)" is correctly identified despite the author name being nowhere near the other metadata, and the Higgins and Sharp paper in CABIOS discussed in the previous paragraph is also found.

We use this procedure to extract all in-text and front page references to every research article published between 1984 and 2016 in 260 prominent journals drawn from fields as diverse as biology, medicine, engineering, computer science, and social science.[5] These 3,389,853 articles have been cited collectively 2,779,258 times in patents granted since 1984, with 1,568,516 citations of the front page and 1,210,742 in-text. 10.1% of the articles are cited at least once, with the probability of being cited highest for biomedical articles. 6.3% are cited at least once in-text, and 8.1% are cited at least once on the front page. Very recent articles are, of course, unlikely to be have been cited yet, so these figures understate the general citation propensity. For articles published in 2000, nearly 16 percent have been cited at least once in the full-text and/or on the front-page of issued patents.

The critical fact about in-text citations is their lack of overlap with front page citations: only 24% of the front page citations are cited in-text in the same patent, and only 31% of the in-text citations are cited on the front page.[6] That is, patentees use the two types of citations in very different ways.

Table 1 shows summary statistics on front-page and full-text citations. The lack of overlap between in-text and front page citations occurs across patentee types, geographies, time, and industry. 24% of front page citations by American inventors appear in-text, and 31% of in-text citations appear on the front page; for foreign inventors, the percentages are 24% and 32%. Patents assigned to academic research institutions have overlap of 31% and 36%, while those assigned to organizations outside academia have overlap of 21% and 29%. So-called "triadic" patents, which are filed in the US, Europe, and Japan- (this is

[5]A complete list of the journals, and the criteria for selection, is in the Online Appendix.

[6]This difference is not a result of misclassifications by the matching algorithm. In Section 4, we perform two robustness checks where in-text and front page citations are classified by hand; within these sets the overlap is similarly small.

often used as in indicator for more important patents) have overlap of 23% and 32%, while nontriadic patents have overlap of 26% and 30%.[7] Patents in medical or biotechnology classes have overlap of 27% and 31%, while those in other fields have overlap of 19% and 32%.[8] To ensure our results are not being driven by patentees who "flood" examiners with hundreds of references, we can restrict to patents with fewer than 20 front-page and 20 in-text references; the overlaps in this restricted set are 26% and 30%.

Not only is there little overlap between in-text and front page citations whether or not we restrict to academic, domestic, biomedical, or "important" triadic inventions, but it is also the case that these in-text citations are nearly all divulged by the inventor at the time of their initial patent application. Examining public initial applications which are available for post-2001 patents, *every* in-text cite in an application remains in the eventually granted patent, and only 7% of in-text citations in the grant were not in the initial application.

In the Online Appendix, we provide complete details on our algorithm and the sample of journals we attempt to match to patents, show that the relative distributions of in-text and front-page citations are similar and highly skewed, give examples of how our algorithm treats various types of citations, and describe more precisely how we classify "university", "biomedical" and other groups mentioned above.

# 3 Why Front Page and In-Text Citations Differ

It has long been known that front page citations may miss references to prior scientific work used by inventors. Indeed, one of the first papers to empirically examine front page citations notes that in-text references "may be more related to the history, usefulness, and development of the invention." Nonetheless, they used front page citations since they are

[7] Triadic patents are often considered a proxy for high value patents. It is therefore interesting to note that while triadic and non-triadic patents have an almost identical number of in-text citations conditional on having at least one (5.79 and 5.59, respectively), triadics have far more front page citations (6.65 and 4.68). We return to this point in the third empirical exercise in Section 4.

[8] Medical patents make up 59% of patents citing at least one article in-text, and 51% of those citing at least one on the front page.

"far easier to extract" (Narin and Noma [1985]). Summary statistics on these front page citations, especially for citations to academic research, have been reported in OECD documents and in the NSF *Science and Engineering Indicators* since the 1990s. The number of front page citations is one measure used by funders, including the NIH, to investigate the impact of their grants. Front page citations to patents are widely used as measures of knowledge flows between inventors, and as proxies for the importance, value, or quality of the cited inventions (Jaffe and de Rassenfosse [2017]). Recent research suggests front-page citations to academic patents are not strongly correlated with survey-based indicators of the extent to which firms rely on public research; front-page citation to non-patent literature is more strongly correlated [Roach and Cohen, 2013]. However qualitative and historical studies suggest that even front page citations to science appear to map poorly into the directly measured prior knowledge of inventorsTijssen [2002], Meyer [2000]).[9]

To understand more clearly why patents cite such a radically different set of prior knowledge in their specification text than on the front page, and what this difference means for scholars of innovation, let us examine how patentees legally ought to use these citations, and how they do so in practice.

Consider first front page citations, whose legal origin lies in patentees' "duty of disclosure". Everyone involved in filing a patent "has a duty...to disclose to the Office all information known to that individual to be material to patentability" (37 CFR 1.97). That is, *any* individual involved in the filing of the patent, whether the initial inventors, a drafter, a patent agent, or a patent attorney, must disclose on an Information Disclosure Statement (IDS) any prior publications they know which are relevant to the novelty and/or non-obviousness of the invention.[10] These individuals must cite prior art even if they learn of

[9]Nearly half of front page citations to patents are added by examiners (Alcácer et al. [2009]). Only around 5% of front page citations to academic articles are added by the examiner (Lemley and Sampat [2012]), so the examiner problem is less severe for these types of citations, though of course it still unclear whether lawyers, inventors, or someone else inside the inventing firm added a given reference.

[10]Brasseler, U.S.A. I, L.P. v. Stryker Sales Corp., Fed. Cir. 2001: "Once an attorney, or an applicant has notice that information exists that appears material and questionable, that person cannot ignore that notice in an effort to avoid his or her duty to disclose."

it years long after the invention is complete.[11] If known prior art is not disclosed, there is a risk the patent office will find "inequitable conduct," which is grounds for unenforceability of a patent (Cotropia et al. [2013]). The documents on the IDS, alongside relevant prior art found by a patent examiner, make up the front page citations in a granted patent. Front page citations therefore represent a list of prior documents, known by any person involved in preparing the patent, which might be relevant to the novelty or non-obviousness of the patent's claims.

In-text citations have a completely different legal origin. The specification must "enable" the invention by describing its background, showing how it solves a useful problem, and showing how a person "skilled in the art" can make and use it without excessive experimentation. The applicant is not required to cite anything formally in the patent specification text, since she can simply describe the invention's background and method of construction using text and graphics. Often, it is easier to "incorporate by reference" aspects of the background and method.[12] For instance, a patent application for a new cancer drug could describe the method of discovery by writing "we inject our mice with cancer using the technique developed by Smith (2017)." The reference to Smith replaces lengthy details on how exactly that injection method works.[13] These in-text citations, therefore, serve a role more like academic citations than front page citations. Knowledge flows like basic motivating facts, open scientific puzzles, and tools used to construct the invention, are often part of an invention's method and background but not material to patentability.

[11] Another complicating factor is that there is a strategic aspect to whether applicants search for or cite prior art (Lampe [2012], Sampat [2010]) and heterogeneity among patent examiners in the extent to which they do so (Lemley and Sampat [2012]).

[12] See 37 CFR 1.71 and 37 CFR 1.57.

[13] The situation is slightly different outside the United States. European patents, for instance, do not have a duty of disclosure, and therefore tend to have fewer front page citations; they also operate under the requirement that the patent specification is interpretable by a more specialized reader than a U.S. patent, and hence may also see differences with in-specification references. That said, the nature of in-text citations in European patents may not be much different than in the United States. European specifications must "indicate the background art which, as far as is known to the applicant, can be regarded as useful to understand the invention" (EPC Rule 42.1(b)). We have not validated the empirical properties of in-text citations for non-U.S. patents, but at least by the letter of the law, the qualitative distinction between front page and in-text citations is the U.S. and Europe is not large.

## 3.1 Patent strategy and practice

In addition to laws on the books discussed above, patent practice and legal strategy will shape what we see cited on the front-page vs. in the full-text practice.

Based on our understanding from legal scholars and practitioners, inventors are typically more involved in drafting patent specifications than they are in prior art searches. And once drafted, the patent specification typically doesn't change much (technically it cannot without filing a continuation-in-part, which may or may not benefit from the priority date of the original application; MPEP 201.07). So in-text citations generally are generated around the time the original application is drafted, sometimes based on citations in provisional applications or scientific articles accompanying the patent.

As noted above, front page citations come from applicant information disclosure statements (IDS, PTO form 1449) and the examiner search that typically follows (reported on the PTO Form 892). In contrast to full-text citations, these front page citations come in over the course of prosecution process, and include inputs from attorneys who prepare the IDS based on information from inventors and their own prior art searches, and examiners' own searches at the USPTO.[14]

Importantly while the "duty of candor" applies to the entire patent process, failure to enable an invention would not consititute a violation of it (though may result in an enablement rejection) whereas failure to disclose known prior art on an IDS would. (MPEP 37 CFR 1.56). Violation of the duty of candor involves severe penalties for applicants and their attorneys, and could render the subject patent unenforceable. In this context, applicants (and attorneys) many err on the side of caution in their front page citations, since there is no penalty for citing too much. There is also some argument that attorneys may "flood the patent office" to hide actual relevant references [Taylor, 2012], though there is no strong evidence on this. On the other hand, it is also known that in some fields a substantial share

[14] As an empirical matter front-page citations to non-patent literature are less likely to be listed as "added by examiner" than front-page citations to patents [Lemley and Sampat, 2012]. Based on published PTO data about 4 percent of non-patent references are from examiners, compared to 40 percent of patent references.

of patents include little or no no applicant provided art, possibly because inventors don't read competitors' patents (for fear of willful infringement liability) or don't care much about the validity of any given patent, but instead about accumulating large portfolios.[15]

These strategic aspects of front-page citation practice, and the legal meaning of front-page citations above would seem to call into question some of the early assumptions in citation analysis, such as Trajtenberg's assumption that the front-page citation process "apparently does generate the right incentives to have all relevant patents cited, and only those" [Trajtenberg, 1990].

But there are potential strategic aspects to in-text citation as well. Inventors and attorneys may want to provide enough information to satisfy the disclosure requirement, but not enough to "actually" enable competitors to practice the invention.[16] Both legal scholarship and the economics literature provide mixed evidence on how much patents actually disclose, and how seriously applicants take the disclosure requirement [Ouellette, 2011, Devlin, 2009, Fromer, 2008, Cohen et al., 2000] though none of the literature focuses on in-text citation practices per se.

Finally, though some attorneys and law firms consider it best legal practice to reference everything in the specification on the front-page (because there is little cost to doing so) there is no legal requirement to do so, since (as emphasized in the previous section) prior art citations and citations in the specification have different purposes. Only the in-specification citations that also bear on novelty and/or non-obviousness need to be actually need to be cited on the front-page, though actual citing practices may vary by firms, attorneys, and even the importance of specific inventions.

[15]The duty of candor does not require affirmative search, only disclosure of known prior art. See Sampat [2010] for more discussion.

[16]As a practical matter, only 35 percent of applications get a rejection on Section 112 grounds - failure to meet disclosure - while 72 percent do for non-obviousness reasons, i.e. in light of prior art on the IDS or found by examiners. See Frakes and Wasserman [2017].

# 4 Validating In-Text Citations

We have shown that in-text citations are algorithmically extractable, have little overlap with front page citations, and legally ought better track real knowledge flows. To confirm that final hypothesis empirically, we perform two validations linking both types of citations to non-patent measures of the invention's underlying science. We also examine how both types of citations relate to commonly used measures of patent value.

## 4.1 In Patent-Paper Pairs, In-Text References Better Match the Base Article

Our first validation asks, for patents where there is an academic article describing the same invention, do in-text or front page citations more closely match the knowledge flows cited in the article? Murray and Stern [2007] collects 171 articles in the journal *Nature Biotechnology* with a related patent filed at least partially on the basis of that article, as judged by a reader with subject-matter expertise. If in-text citations are a superior measure of the type of knowledge flows represented by academic citations, the research cited in the academic article should appear more frequently in the patent text than the front page.

As our algorithmic method for identifying in-text citations requires starting with a fixed list of academic articles, for this comparison we instead read each patent manually, counting the total number of in-text and front page references to academic articles in *any* journal, and the number of references in the original article's bibliography that are also cited in-text or on the front page.[17]

The biotechnology patents in this sample cite academic work much more heavily than the modal patent. The 171 patents have a mean of 26.9 (median: 15) front page references to academic articles, and a mean of 41.5 (median: 29) in-text references. The

[17] Note that extracting these references by hand means we are able to handle the non-trivial number of patents with typos on references, such as misspelled author names, misstated years, or obscure journal abbreviations. Reassuringly, the overwhelming majority of references we find by hand are ones that our algorithm would match if given the proposed academic article.

majority of references of each type are unrelated to the research cited in the corresponding *Nature Biotechnology* article; this is both because the patent text is generally very long and detailed compared to the academic writeup, and because the patent often covers an invention broader than the particular result in the underlying article.

On average, each *Nature Biotechnology* article in the Stern-Murray sample has about 30 referenced articles. Of these, an average of 6.9 articles are also mentioned in the text of the corresponding patent, while only 4.1 are cited on the patent's front page. That is, a citation in the underlying *Nature Biotechnology* article is 68% more likely to be found in the corresponding patent's specification text than in the front page citations. For the median patent-paper pair, the difference is even more stark. The front page of the median patent contains only 6.3% of the corresponding article's academic references, while the text contains 13.0% of these references, a 108% increase.

More than 25% of patents in the Murray-Stern sample have *zero* front page references which match a reference in the article bibliography, and 43.3% have no more than a single such reference. Patents with only zero or one in-text citation matching the article bibliography are far less common, at 10.5% and 25.7% respectively. That is, a researcher who relied on front page citations rather than in-text citations to investigate knowledge flows would be almost 2.5 times more likely to incorrectly conclude that patent did not rely on *any* of the knowledge contained in the corresponding academic paper's references. The correlation of the total number of in-text and front page references matching the article bibliography for a given patent-paper pair is only 0.48, though this correlation overstates the overlap; even when there are, for instance, 3 in-text and 2 front page citations that match the article's bibliography, those 5 citations are often entirely distinct. Indeed, the in-text academic references and front page academic citations are identical in only 3 of the 171 patents.

Do in-text citations contain more of the corresponding article's academic citations simply because the inventor has lazily copy-and-pasted parts of the background from the article into the patent? In our experience manually reading both the article and the patent, it was very rare to find identical language. Only 5 of the 171 patents contained every academic reference from the corresponding article in the patent text, and in only 12.9% of the patents were even half of the underlying article's citations included in the patent text. A reader may ask why patents that make up a patent-paper pair do not cite *all* of the references in the original article. There are two reasons. First, the patent in general does not describe precisely the same invention and claims as the result described in the original article; rather, the original article often describes a single claim of the invention. Second, when manually examining these patent-paper pairs, similar basic science is often cited with different yet scientifically-equivalent references in the article and the patent.

## 4.2 References Compared to Survey Evidence of Knowledge Transfers

Our second validation asks, are front page or in-text citations of academic research better correlates of firms' stated reliance on public sector research in a large survey? The Carnegie Mellon Survey (Cohen et al. [2002]) of industrial R&D managers asked how much their firm relies on public sector spillovers for their inventions, as well as a series of questions about their reliance on "open science" like conferences, books and articles, versus "closed science" like contract work with academics. A follow-up study counted front page citations to public sector patents and non-patent literature in surveyed firm's patents (Roach and Cohen [2013]). The former doesn't correlate at all with the R&D manager's stated response on the percent of a firm's research using public sector knwledge, and the latter correlates relatively weakly.

To check whether in-text citations to academic research may better predict firm's actual stated use of public sector knowledge, we manually count all in-text and front page references to journal publications in all 6,148 patents filed by 614 surveyed firms between 1991 and 1993. There are 8,307 total front page citation of academic journal articles, and 9,296 in-text cites. The raw correlation between the two count measures, at the individual patent level, is .52.

Figure 2 plots the correlation between the number of in-text citations or front page citations to academic research and the R&D manager's estimate of whether less than 10%, 10-40%, 40-60%, 60-90%, or greater than 90% of their research projects rely on public sector knowledge. This plot is monotonically and strongly increasing for the in-text measure, and increasing though imprecise for the front page measure.

In Table 2, we show that, in line with Figure 2, the in-text measure explains more of the survey response variance across six separate regressions. Simple correlations or comparisons of $R^2$ do not formally permit us to choose among different models. The literature on Bayesian model selection provides a formal test (Raftery [1995], Kass and Raftery [1995]). The idea is the following. Take two potentially non-nested models, such as two regressions of different measures of citations on a survey measure of actual public sector knowledge transfer. Consider the relative likelihood of the data given the model in question, and adjust for the sample size and number of explanatory variables. A good model makes the data more likely while being parsimonious. A particular measure, the Bayesian Information Criterion (BIC), is identical to considering the posterior likelihood of two models with a "unit information" prior (Raftery [1998]).

A difference in the BIC of 6, by a standard rule of thumb (Raftery [1995]), is "strong" evidence for one model over another. In particular, when the difference is BIC is more than 6, one model is at least 20 times more likely to explain the data observed than another. Note the final row in Table 1: in-text citations are more strongly predictive for *every* model, whether without controls, or after controlling for covariates like industry and the number of scientists at each firm.[18]

[18] Online Appendix Table 1 shows that in-text citations are also a better proxy for the "open science" factor in Roach and Cohen [2013], measuring the reliance of a firm's R&D on publicly available science.

## 4.3 In-text References and the Value of Patents

The previous two validations suggest that in-text citations more accurately represent knowledge flows, in line with the formal legal purpose of front page versus in-text citations. This does not mean that front page citations have no use for scholars of innovation. Prior research suggests that front page scientific references can serve as a proxy for high-value patents, as measured by forward citations or other metrics (Fleming and Sorenson [2004], Sorenson and Fleming [2004].[19] While the interpretation of this result is unclear (for example, Sampat [2010] suggests that firms have incentives to search for prior art more diligently for more important inventions), comparing how front page references and in-text references to science respectively correlate with patent value can help us better understand the information contained in each measure.

To do so, we collect data on all front page and in-text citations from 489,346 patents issued between 2006 and 2008 to articles published in the 260 journals described in Section 2. Of these patents, most (93 percent) cite no scientific articles from our set. Of patents with front page references to a scientific article, 57 percent also cite at least one scientific article in text. And of the patents with a full text reference, 69 percent cite at least one article on the front page.

We also collected data on four different measures of invention value: (1) forward citations in later patents; (2) the stock market reaction to patent issuance (Kogan et al. [2017]); (3) whether the patents were renewed to at least year 8; and (4) whether the patents are part of triadic patent families. Prior research has used each of these measures as an indicator of patent value.

Table 3 shows summary statistics on each of the variables in this model. Tables 4, 5, and Online Appendix Tables 2 and 3 show results from OLS regressions relating the

[19]Other research shows a weak or even negative relationship between extent of science citation in firms patents and forward citations (Trajtenberg et al. [1992], Gittelman and Kogut [2003], Cassiman et al. [2008]). Patent-to-patent citations by inventors/attorneys tend to be focused on canonical inventions in an area and high-quality prior patents, while examiner-added citations focus on similarity (Moser et al. [2018]).

value measures to the number of front page and in-text references (Models 1-3) and to indicators for whether there were any front page or in-text references (Models 4-6). We find front page backward citations to science are positively correlated with forward citations (Table 3) and with whether the patent is part of a triadic patent family (Table 4), consistent with prior research (Sorenson and Fleming [2004], Fleming and Sorenson [2004]). To our knowledge the relationship between front page science references and stock market reaction to patent issue nor maintenance decisions has been examined before, and here the relationships are less robust across specifications (Online Appendix Tables 2 and 3). That said, for all four measures of value, in most specifications, front page citations are more strongly related to value than in-text citations.[20]

The precise mechanisms for this are unclear. It may be that a patent's similarity or proximity to science, captured by front page science citations cited as prior art material to patentability, is more predictive of the private value of a patent to a firms than is whether the patent is based on science. This could be true, for example, if scientific inputs cited in text were in the public domain and available to competitors as well. Alternatively, it may be that applicants submit more front-page prior art or search more intensively for their more important inventions, to "bulletproof" these patents against validity challenges or guard against duty of candor violations, whereas for reasons discussed above this is not necessary to do for in-text citations. Whatever the reasons, this final validation emphasizes that front-page and full-text citations are fundamentally different, and each potentially useful for measuring different concepts.

# 5 Concluding Remarks

Our results should not be interpreted as saying that the oft-used front page citation measure has no valid use. It is, after all, true that front page citations help measure patent

[20]Further, a formal Bayesian model selection procedure as in the previous subsection strongly prefers models using front page citations to proxy for forward citations and triadic patents.

value, and that these citations are useful in investigating the *similarity* of patents (e.g., Ahmadpoor and Jones [2017]). That said, front page citations in the strict legal sense neither measure underlying knowledge used in making an invention, nor delineate knowledge known by the inventor themselves as opposed to that known by their lawyer. The situation is very different with in-text citations, which are used explicitly to point to prior literature relevant to a patent's method and background. Assuming that front page citations are simply a noisy measure of in-text citations is incorrect: the overlap between the two measures is only 24% to 31%, and the magnitude and significance of each type of citation as proxies for various real outcomes in Section 4 vary enormously even with large N sample sizes.

In addition to having completely different legal uses, in-text citations possess two practical benefits compared to front page citations. First, non-granted patent applications do not have any front page citations listed.[21] For studies that require the use of contemporaneous data, it is often infeasible to wait five or more years for patents to be granted. Second, pre-1947 U.S. patents do not have front page citations, while in-text citations can, in theory, be extracted for patents going back to the 1800s. For example, U.S. patent 2,295,481 A, applied for in 1939 by a scientist at Merck, contains no front page at all, but cites in the specification text just like modern patents: "Thus, Domagk (Deutsche Med. Wochsch., 61, 250, 1935) claimed that Prontosil, a derivative of diazotized sulphanilamide, was moderately effective against pneumococci, especially of Type III."

We have shown that in-text citations can be accurately and comprehensively extracted from patents, and that these citations closely correlate with actual knowledge flows in multiple empirical validations. Therefore, we suggest that future work relying on patents as a "paper trail of knowledge" should use in-text rather than front page citations. A public database covering 260 journals for over three decades is available alongside this

[21] The information disclosure statements with applicant prior art citations can be filed throughout the application process, and examiner searches are conducted after the application is filed. More practically, neither of these is readily available in machine readable form.

paper. That said, an institutional project establishing a more complete and open-access database along the lines of the existing NBER front page citation database would be particularly useful.

As for future research, the actual text of patents remains an incredibly underutilized resource. Rather than relying on count measures or features like a patent's class, machine learning methods (e.g., Mullainathan and Speiss [2017], Gentzkow et al. [2017]) can "read" the text of the patent and hence uncover information on precisely what knowledge a patent recombines, the exact way certain types of knowledge were used in the invention, and so on. In-text citations should prove value not just in better capturing actual knowledge flows, but in the ability to use the words around those citations to understand exactly how, when, and why inventors build on the past.

# References


Ajay Agrawal and Rebecca Henderson. Putting patents in context: Exploring knowledge transfer from mit. *Management science*, 48(1):44–60, 2002.

Mohammad Ahmadpoor and Benjamin F. Jones. The dual frontier: Patented inventions and prior scientific advance. *Science*, 2017.

Juan Alcácer, Michelle Gittelman, and Bhaven Sampat. Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, 38(2):415–427, 2009. ISSN 00487333. doi: 10.1016/j.respol.2008.12.001.

Pierre Azoulay, Joshua S Graff Zivin, Danielle Li, and Bhaven N Sampat. Public r&d investments and private-sector patenting: evidence from nih funding rules. *The Review of Economic Studies*, 2015.

Kevin A. Bryan and Yasin Ozcan. The Impact of Open Access Mandates on Invention. 2018.

Bruno Cassiman, Reinhilde Veugelers, and Pluvia Zuniga. In search of performance effects of (in) direct industry science links. *Industrial and Corporate Change*, 17(4):611–646, 2008.

Wesley Cohen, Richard Nelson, and John Walsh. Links and impacts: The influence of public research on industrial r&d. *Management Science*, 2002.

Wesley M Cohen, RR Nelson, and John P Walsh. Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not). *NBER Working Paper*, (7552):50, 2000. ISSN 17561663. doi: 10.1093/dnares/dsr014. URL `www.nber.org/papers/w7552`.

Christopher A. Cotropia, Mark A. Lemley, and Bhaven Sampat. Do applicant patent citations matter? *Research Policy*, 42(4):844–854, 2013. ISSN 00487333. doi: 10.1016/j.respol.2013.01.003.

Alan Devlin. The misunderstood function of disclosure in patent law. *Harv. JL & Tech.*, 23: 401, 2009.

EPO. Cited References in European Patent Documents, 2007. URL `http://www.wipo.int/export/sites/www/cws/en/taskforce/citation{_}practices/docs/epo{_}citation{_}practice{_}summary.pdf`.

Lee Fleming and Olav Sorenson. Science as a map in technological search. *Strategic Management Journal*, 25(8-9):909–928, 2004.

Michael D. Frakes and Melissa F. Wasserman. Is the time allocated to review patent applications inducing examiners to grant invalid patents? evidence from microlevel application data. *Review of Economics and Statistics*, 99(3):550–563, 2017.

Jeanne C Fromer. Patent disclosure. *Iowa L. Rev.*, 94:539, 2008.

Matt Gentzkow, Bryan Kelly, and Matt Taddy. Text as data. *NBER Working Paper*, 2017.

Michelle Gittelman and Bruce Kogut. Does good science lead to valuable knowledge? biotechnology firms and the evolutionary logic of citation patterns. *Management Science*, 49(4):366–382, 2003.

Rebecca Henderson, Adam B Jaffe, and Manuel Trajtenberg. Universities as a source of commercial technology: a detailed analysis of university patenting, 1965–1988. *Review of Economics and statistics*, 80(1):119–127, 1998.

Adam Jaffe, Manuel Trajtenberg, and Rebecca Handerson. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations Author ( s ): Adam B . Jaffe , Manuel Trajtenberg and Rebecca Henderson. *The Quarterly Journal of Economics*, 108(3): 577–598, 1993a. ISSN 0033-5533. doi: 10.2307/2118401.

Adam B. Jaffe and Gaetan de Rassenfosse. Patent citation data in social science research:

Overview and best practices. *Journal of the Association for Information Science and Technology*, 2017.

Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598, 1993b.

Osmat A Jefferson, Adam Jaffe, Doug Ashton, Ben Warren, Deniz Koellhofer, Uwe Dulleck, Aaron Ballagh, John Moe, Michael DiCuccio, Karl Ward, et al. Mapping the global influence of published research on industry and innovation. *Nature biotechnology*, 36(1): 31, 2018.

Sarah Kaplan and Keyvan Vakili. The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36(10):1435–1457, 2015. ISSN 10970266. doi: 10.1002/smj.2294.

Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 1995.

Bryan Kelly, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. Measuring Technological Innovation over the Long Run. *Working Paper*, pages 1–57, 2017.

B Zorina Khan, Lee Branstetter, Colleen Chien, Claude Diebolt, Rochelle Dreyfuss, Naomi Lamoreaux, Petra Moser, Adam Mossoff, Tom Nicholas, Alessandro Nuvolari, Patricio Saíz, and Ted Sichelman. Inventing in the Shadow of the Patent System: Evidence from 19th-centrury Patents and Prizes for Technological Innovations. *NBER Working Paper #20731*, 2014. doi: 10.3386/w20731.

Leonid Kogan, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2): 665–712, 2017.

Paul R. Krugman. *Geography and trade*. MIT Press, Cambridge, MA, 1991. ISBN 9780262111591.

Jeffrey Kuhn and Neil Thompson. The Ways We've been Measuring Patent Scope are Wrong: How to Measure and Draw Causal Inferences with Patent Scope. 2017.

Ryan Lampe. Strategic citation. *Review of Economics and Statistics*, 2012.

Mark A. Lemley and Bhaven Sampat. Examiner Characteristics and Patent Office Outcomes. *Review of Economics and Statistics*, 94(3):817–827, 2012. ISSN 0034-6535.

Richard C Levin, Alvin K Klevorick, Richard R Nelson, and Sidney G Winter. Appropriating the Returns from Industrial Research and Development; Comments and Discussion. *Brookings Papers on Economic Activity*, (3):783, 1987. ISSN 0007-2303. doi: 10.2307/2534454.

Danielle Li, Pierre Azoulay, and Bhaven N Sampat. The applied value of public investments in biomedical research. *Science*, 356(6333):78–81, 2017.

Martin Meyer. What is special about patent citations? Differences between patent and scientific citations. *Scientometrics*, 49(1):93–123, 2000. ISSN 01389130 (ISSN). doi: 10.1023/A:1005613325648. URL `http://dx.doi.org/10.1023/A:1005613325648`.

Petra Moser. How do patent laws influence innovation? Evidence from nineteenth-century world's fairs, 2005. ISSN 00028282.

Petra Moser, Joerg Ohmstedt, and Paul Rhode. Patent citations - an analysis of quality differences and citing practices in hybrid corn. *Management Science*, 2018.

Sendhil Mullainathan and Jann Speiss. Machine learning: An applied econometric approach. *Journal of Economic Persepectives*, 2017.

Fiona Murray and Scott Stern. Do formal intellectual property rights hinder the free flow of scientific knowledge?. An empirical test of the anti-commons hypothesis. *Journal of*

*Economic Behavior and Organization*, 63(4):648–687, 2007. ISSN 01672681. doi: 10.1016/j.jebo.2006.05.017.

F. Narin. Patent bibliometrics. *Scientometrics*, 30(1):147–155, 1994. ISSN 01389130. doi: 10.1007/BF02017219.

F. Narin and E. Noma. Is technology becoming science? *Scientometrics*, 7(3-6):369–381, 1985. ISSN 01389130. doi: 10.1007/BF02017155.

Lisa Larrimore Ouellette. Do patents disclose useful information. *Harv. JL & Tech.*, 25:545, 2011.

David Popp. From science to technology: the value of knowledge from different energy research institutions. *Research Policy*, 46(9):1580–1594, 2017.

Adrian E. Raftery. Bayesian model selection in social resarch. *Sociological Methodology*, 1995.

Adrian E. Raftery. Bayes factors and bic: Commets on weakliem. *University of Washington Technical Report*, 1998.

Michael Roach and Wesley M. Cohen. Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research. *Management Science*, 59(2):504–525, 2013. ISSN 0025-1909. doi: 10.1287/mnsc.1120.1644. URL `http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1120.1644`.

Bhaven Sampat. When do patent applicants search for prior art? *Journal of Law and Economics*, 2010.

Olav Sorenson and Lee Fleming. Science and the diffusion of knowledge. *Research policy*, 33(10):1615–1634, 2004.

Schumpeter Tamada, Yusuke Naito, Fumio Kodama, Kiminori Gemba, and Jun Suzuki. Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics*, 68(2):289–302, 2006. ISSN 01389130. doi: 10.1007/s11192-006-0112-2.

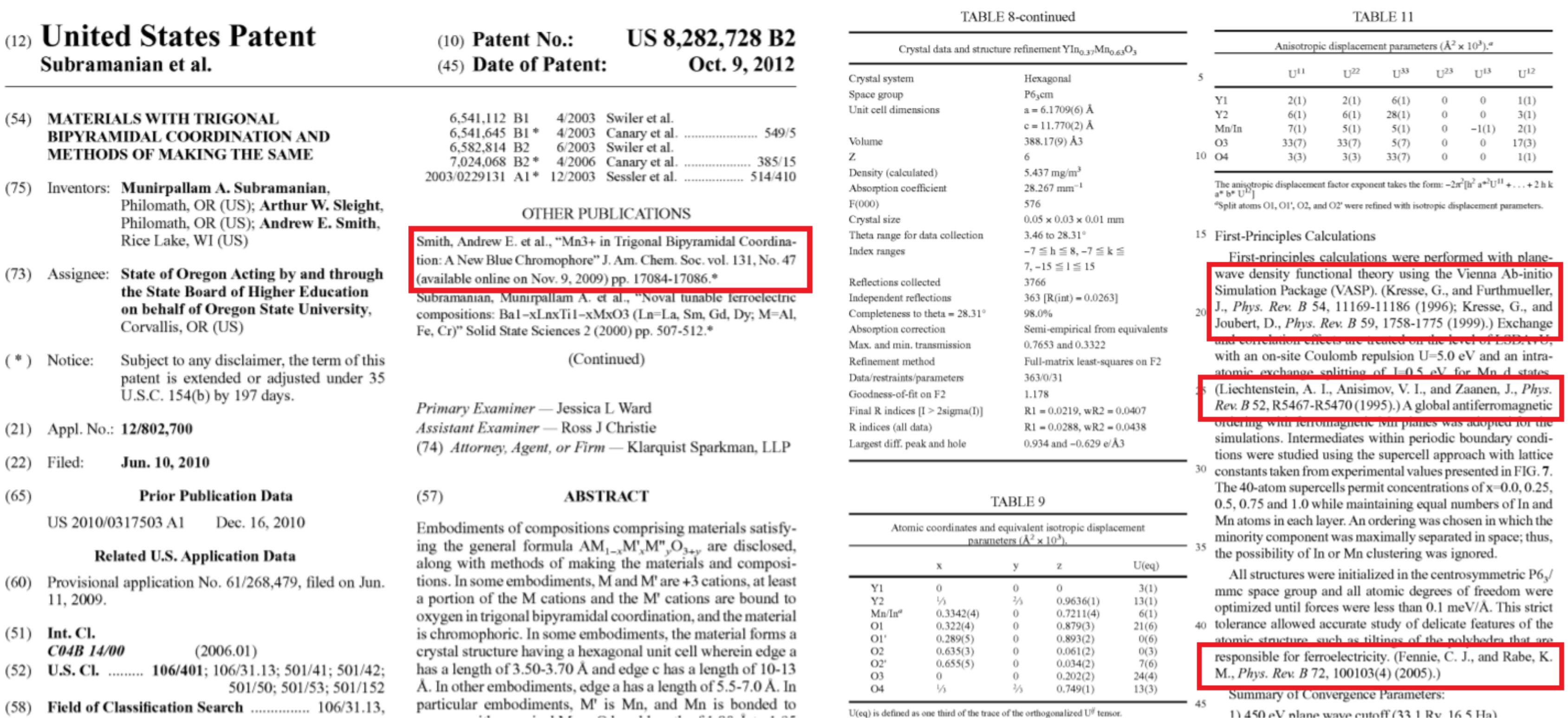
Robert Brendan Taylor. Burying. *Mich. Telecomm. & Tech. L. Rev.*, 19:99, 2012.

Robert J.W Tijssen. Science dependence of technologies: evidence from inventions and their inventors. *Research Policy*, 31(4):509–526, 2002. ISSN 00487333. doi: 10.1016/S0048-7333(01)00124-X. URL `http://linkinghub.elsevier.com/retrieve/pii/S004873330100124X`.

Manuel Trajtenberg. A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The RAND Journal of Economics*, 21(1):172, 1990. ISSN 07416261. doi: 10.2307/2555502. URL `http://doi.wiley.com/10.2307/2555502`.

Manuel Trajtenberg, Rebecca Henderson, and Adam Jaffe. Ivory tower versus corporate lab: An empirical study of basic research and appropriability. Technical report, National Bureau of Economic Research, 1992.

Eric von Hippel. *The Sources of Innovation*. Oxford University Press, 1988.



(12) **United States Patent**
**Subramanian et al.**

(10) **Patent No.: US 8,282,728 B2**
(45) **Date of Patent: Oct. 9, 2012**

(54) **MATERIALS WITH TRIGONAL BIPYRAMIDAL COORDINATION AND METHODS OF MAKING THE SAME**

(75) Inventors: **Munirpallam A. Subramanian**, Philomath, OR (US); **Arthur W. Sleight**, Philomath, OR (US); **Andrew E. Smith**, Rice Lake, WI (US)

(73) Assignee: **State of Oregon Acting by and through the State Board of Higher Education on behalf of Oregon State University**, Corvallis, OR (US)

( * ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 197 days.

(21) Appl. No.: **12/802,700**

(22) Filed: **Jun. 10, 2010**

(65) **Prior Publication Data**

US 2010/0317503 A1 Dec. 16, 2010

**Related U.S. Application Data**

(60) Provisional application No. 61/268,479, filed on Jun. 11, 2009.

(51) **Int. Cl.**
***C04B 14/00*** (2006.01)

(52) **U.S. Cl.** ......... **106/401**; 106/31.13; 501/41; 501/42; 501/50; 501/53; 501/152

(58) **Field of Classification Search** .............. 106/31.13,

| | | | | |
|---|---|---|---|---|
| 6,541,112 | B1 | 4/2003 | Swiler et al. | |
| 6,541,645 | B1 * | 4/2003 | Canary et al. | 549/5 |
| 6,582,814 | B2 | 6/2003 | Swiler et al. | |
| 7,024,068 | B2 * | 4/2006 | Canary et al. | 385/15 |
| 2003/0229131 | A1 * | 12/2003 | Sessler et al. | 514/410 |

OTHER PUBLICATIONS

Smith, Andrew E. et al., "Mn3+ in Trigonal Bipyramidal Coordination: A New Blue Chromophore" J. Am. Chem. Soc. vol. 131, No. 47 (available online on Nov. 9, 2009) pp. 17084-17086.*

Subramanian, Munirpallam A. et al., "Noval tunable ferroelectric compositions: Ba1−xLnxTi1−xMxO3 (Ln=La, Sm, Gd, Dy; M=Al, Fe, Cr)" Solid State Sciences 2 (2000) pp. 507-512.*

(Continued)

*Primary Examiner* — Jessica L Ward
*Assistant Examiner* — Ross J Christie
(74) *Attorney, Agent, or Firm* — Klarquist Sparkman, LLP

(57) **ABSTRACT**

Embodiments of compositions comprising materials satisfying the general formula $AM_{1-x}M'_xM''_yO_{3+y}$ are disclosed, along with methods of making the materials and compositions. In some embodiments, M and M' are +3 cations, at least a portion of the M cations and the M' cations are bound to oxygen in trigonal bipyramidal coordination, and the material is chromophoric. In some embodiments, the material forms a crystal structure having a hexagonal unit cell wherein edge a has a length of 3.50-3.70 Å and edge c has a length of 10-13 Å. In other embodiments, edge a has a length of 5.5-7.0 Å. In particular embodiments, M' is Mn, and Mn is bonded to

TABLE 8-continued

| Crystal data and structure refinement $YIn_{0.37}Mn_{0.63}O_3$ | |
|---|---|
| Crystal system | Hexagonal |
| Space group | $P6_3cm$ |
| Unit cell dimensions | a = 6.1709(6) Å |
| | c = 11.770(2) Å |
| Volume | 388.17(9) Å3 |
| Z | 6 |
| Density (calculated) | 5.437 mg/m$^3$ |
| Absorption coefficient | 28.267 mm$^{-1}$ |
| F(000) | 576 |
| Crystal size | 0.05 × 0.03 × 0.01 mm |
| Theta range for data collection | 3.46 to 28.31° |
| Index ranges | $-7 \leqq h \leqq 8, -7 \leqq k \leqq 7, -15 \leqq l \leqq 15$ |
| Reflections collected | 3766 |
| Independent reflections | 363 [R(int) = 0.0263] |
| Completeness to theta = 28.31° | 98.0% |
| Absorption correction | Semi-empirical from equivalents |
| Max. and min. transmission | 0.7653 and 0.3322 |
| Refinement method | Full-matrix least-squares on F2 |
| Data/restraints/parameters | 363/0/31 |
| Goodness-of-fit on F2 | 1.178 |
| Final R indices [I > 2sigma(I)] | R1 = 0.0219, wR2 = 0.0407 |
| R indices (all data) | R1 = 0.0288, wR2 = 0.0438 |
| Largest diff. peak and hole | 0.934 and −0.629 e/Å3 |

TABLE 9

Atomic coordinates and equivalent isotropic displacement parameters ($Å^2 \times 10^3$).

| | x | y | z | U(eq) |
|---|---|---|---|---|
| Y1 | 0 | 0 | 0 | 3(1) |
| Y2 | ⅓ | ⅔ | 0.9636(1) | 13(1) |
| Mn/In[a] | 0.3342(4) | 0 | 0.7211(4) | 6(1) |
| O1 | 0.322(4) | 0 | 0.879(3) | 21(6) |
| O1' | 0.289(5) | 0 | 0.893(2) | 0(6) |
| O2 | 0.635(3) | 0 | 0.061(2) | 0(3) |
| O2' | 0.655(5) | 0 | 0.034(2) | 7(6) |
| O3 | 0 | 0 | 0.202(2) | 24(4) |
| O4 | ⅓ | ⅔ | 0.749(1) | 13(3) |

U(eq) is defined as one third of the trace of the orthogonalized $U^{ij}$ tensor.

TABLE 11

Anisotropic displacement parameters ($Å^2 \times 10^3$).[a]

| | $U^{11}$ | $U^{22}$ | $U^{33}$ | $U^{23}$ | $U^{13}$ | $U^{12}$ |
|---|---|---|---|---|---|---|
| Y1 | 2(1) | 2(1) | 6(1) | 0 | 0 | 1(1) |
| Y2 | 6(1) | 6(1) | 28(1) | 0 | 0 | 3(1) |
| Mn/In | 7(1) | 5(1) | 5(1) | 0 | −1(1) | 2(1) |
| O3 | 33(7) | 33(7) | 5(7) | 0 | 0 | 17(3) |
| O4 | 3(3) | 3(3) | 33(7) | 0 | 0 | 1(1) |

The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2 a^{*2}U^{11} + \ldots + 2\,h\,k\,a^*\,b^*\,U^{12}]$
[a]Split atoms O1, O1', O2, and O2' were refined with isotropic displacement parameters.

First-Principles Calculations

First-principles calculations were performed with plane-wave density functional theory using the Vienna Ab-initio Simulation Package (VASP). (Kresse, G., and Furthmueller, J., *Phys. Rev. B* 54, 11169-11186 (1996); Kresse, G., and Joubert, D., *Phys. Rev. B* 59, 1758-1775 (1999).) Exchange … with an on-site Coulomb repulsion U=5.0 eV and an intra-… (Liechtenstein, A. I., Anisimov, V. I., and Zaanen, J., *Phys. Rev. B* 52, R5467-R5470 (1995).) A global antiferromagnetic … simulations. Intermediates within periodic boundary conditions were studied using the supercell approach with lattice constants taken from experimental values presented in FIG. **7**. The 40-atom supercells permit concentrations of x=0.0, 0.25, 0.5, 0.75 and 1.0 while maintaining equal numbers of In and Mn atoms in each layer. An ordering was chosen in which the minority component was maximally separated in space; thus, the possibility of In or Mn clustering was ignored.

All structures were initialized in the centrosymmetric $P6_3$/mmc space group and all atomic degrees of freedom were optimized until forces were less than 0.1 meV/Å. This strict tolerance allowed accurate study of delicate features of the … responsible for ferroelectricity. (Fennie, C. J., and Rabe, K. M., *Phys. Rev. B* 72, 100103(4) (2005).)

Summary of Convergence Parameters:

**Figure 1:** Example of front page citations (left) versus in-text citations (right)

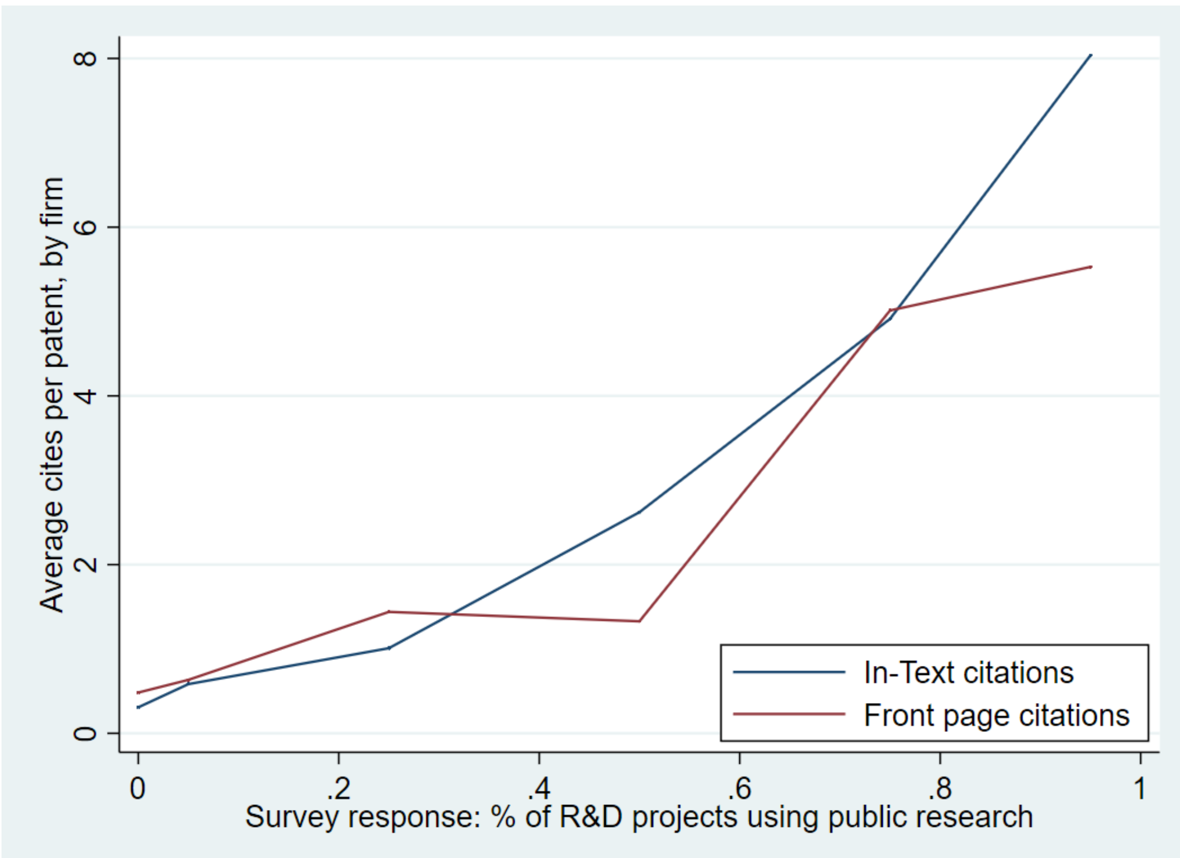



**Figure 2:** Survey response is per firm to "what fraction of your unit's R&D projectors rely on public sector knowledge", on a five point scale: 0-10%, 10-40%, 40-60%, 60-90%, 90-100%.

**Table 1:** Summary statistics on in-text versus front-page citations

| | All | US | Non-US | Univ | Non-Uni | Triadic | Non-Tri | Biomed | Non-BM |
|---|---|---|---|---|---|---|---|---|---|
| N of Patents | 341799 | 226742 | 115057 | 77959 | 263840 | 178191 | 163511 | 168472 | 173230 |
| Avg. # of In-Text | 3.55 | 4.05 | 2.54 | 5.28 | 3.03 | 3.83 | 3.23 | 5.42 | 1.72 |
| Avg. # of Front Page | 4.60 | 5.22 | 3.34 | 6.21 | 4.11 | 5.41 | 3.70 | 6.24 | 2.98 |
| Share of In-Text on FP | 31.1 | 30.9 | 32.0 | 35.6 | 28.7 | 32.0 | 30.0 | 30.7 | 32.4 |
| Share of FP In-Text | 24.0 | 24.0 | 24.3 | 30.1 | 21.1 | 22.7 | 26.2 | 26.7 | 18.6 |

**Table 2:** Ordered logit models relating percent of firms' R&D projects using public research to science references

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| In-Text Cites per Patent | 0.0882*** | | 0.0619*** | | 0.0582*** | |
| | (0.0122) | | (0.0171) | | (0.0161) | |
| Front-Page Cites per Patent | | 0.0892*** | | 0.0470* | | 0.0404 |
| | | (0.0140) | | (0.0223) | | (0.0228) |
| Total Firm Patents | | | | | 0.133 | 0.143 |
| | | | | | (0.0878) | (0.0887) |
| Fraction Scientists | | | | | 1.589*** | 1.555*** |
| | | | | | (0.326) | (0.334) |
| Observations | 615 | 615 | 615 | 615 | 614 | 614 |
| Pseudo $R^2$ | 0.020 | 0.015 | 0.046 | 0.043 | 0.053 | 0.050 |
| *BIC* | 1680.7 | 1687.8 | 1629.6 | 1635.7 | 1628.4 | 1634.4 |
| Industry Controls | No | No | Yes | Yes | Yes | Yes |

Standard errors in parentheses

Firm-level ordered logit with s.e. clustered by industry; depvar is % of firms' R&D projects using public research

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table 3:** Summary statistics for value vs. science reference analyses

| VARIABLES | (1)<br>N | (2)<br>mean | (3)<br>sd | (4)<br>min | (5)<br>max |
|---|---|---|---|---|---|
| Number of front page science refs | 489,346 | 0.332 | 2.533 | 0 | 195 |
| Number of full text science refs | 489,346 | 0.331 | 2.568 | 0 | 224 |
| Number of overlapping science refs | 489,346 | 0.178 | 1.772 | 0 | 170 |
| Any full text science refs? | 489,346 | 0.0513 | 0.221 | 0 | 1 |
| Any front page science refs? | 489,346 | 0.0596 | 0.237 | 0 | 1 |
| Forward citations | 489,346 | 9.577 | 24.36 | 0 | 2,120 |
| Maintained at 8? | 489,346 | 0.572 | 0.495 | 0 | 1 |
| Stock Reaction | 199,986 | 10.94 | 30.40 | 0.000237 | 1,457 |
| Triadic Patent? | 489,346 | 0.285 | 0.451 | 0 | 1 |

**Table 4:** OLS models relating forward citations to science references

| VARIABLES | (1)<br>Forward | (2)<br>Forward | (3)<br>Forward | (4)<br>Forward | (5)<br>Forward | (6)<br>Forward |
|---|---|---|---|---|---|---|
| Number of front page science refs | 0.3*** | | 0.3*** | | | |
| | (0.02) | | (0.02) | | | |
| Number of full text science refs | | 0.1*** | 0.007 | | | |
| | | (0.01) | (0.01) | | | |
| Any front page science refs? | | | | 6.4*** | | 6.1*** |
| | | | | (0.3) | | (0.3) |
| Any full text science refs? | | | | | 3.9*** | 0.7* |
| | | | | | (0.4) | (0.4) |
| issyear = 2007 | -1.9*** | -1.9*** | -1.9*** | -1.9*** | -1.9*** | -1.9*** |
| | (0.08) | (0.08) | (0.08) | (0.08) | (0.08) | (0.08) |
| issyear = 2008 | -3.9*** | -3.9*** | -3.9*** | -3.9*** | -3.9*** | -3.9*** |
| | (0.08) | (0.08) | (0.08) | (0.08) | (0.08) | (0.08) |
| Constant | 11*** | 11*** | 11*** | 11*** | 11*** | 11*** |
| | (0.06) | (0.06) | (0.06) | (0.06) | (0.06) | (0.06) |
| Observations | 489,346 | 489,346 | 489,346 | 489,346 | 489,346 | 489,346 |
| R-squared | 0.091 | 0.090 | 0.091 | 0.092 | 0.090 | 0.092 |
| Patent class FE | Yes | Yes | Yes | Yes | Yes | Yes |
| BIC | 4467047.51 | 4467491.42 | 4467060.39 | 4466099.53 | 4467184.33 | 4466101.66 |

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

**Table 5:** OLS models relating whether a patent is in a triadic family to science references

| VARIABLES | (1) Triad? | (2) Triad? | (3) Triad? | (4) Triad? | (5) Triad? | (6) Triad? |
|---|---|---|---|---|---|---|
| Number of front page science refs | 0.006*** | | 0.007*** | | | |
| | (0.0003) | | (0.0003) | | | |
| Number of full text science refs | | -0.0002 | -0.003*** | | | |
| | | (0.0003) | (0.0003) | | | |
| Any front page science refs? | | | | 0.1*** | | 0.1*** |
| | | | | (0.003) | | (0.004) |
| Any full text science refs? | | | | | 0.04*** | -0.01*** |
| | | | | | (0.004) | (0.005) |
| issyear = 2007 | -0.007*** | -0.006*** | -0.007*** | -0.007*** | -0.006*** | -0.007*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| issyear = 2008 | -0.01*** | -0.01*** | -0.01*** | -0.01*** | -0.01*** | -0.01*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant | 0.3*** | 0.3*** | 0.3*** | 0.3*** | 0.3*** | 0.3*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Observations | 489,346 | 489,346 | 489,346 | 489,346 | 489,346 | 489,346 |
| R-squared | 0.114 | 0.113 | 0.114 | 0.115 | 0.113 | 0.115 |
| Patent class FE | Yes | Yes | Yes | Yes | Yes | Yes |
| BIC | 550821.01 | 551418.98 | 550739.87 | 550319.26 | 551301.72 | 550318.84 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

# Appendix: Additional Results

**Table A.1:** Bayesian Model of Patent Citations Predicting Firms' Use of Open Science in Research

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| In-Text Cites per Patent | 0.0338*** | | 0.0255** | | 0.0216*** | |
| | (0.00583) | | (0.00721) | | (0.00535) | |
| Front-Page Cites per Patent | | 0.0304*** | | 0.0163** | | 0.0117* |
| | | (0.00504) | | (0.00518) | | (0.00538) |
| Total Firm Patents | | | | | 0.102** | 0.107** |
| | | | | | (0.0298) | (0.0294) |
| Fraction Scientists | | | | | 0.625** | 0.643** |
| | | | | | (0.188) | (0.198) |
| Constant | -0.0295 | -0.0287 | -0.0454*** | -0.0433*** | -0.167*** | -0.170*** |
| | (0.0505) | (0.0530) | (0.00245) | (0.00211) | (0.0245) | (0.0256) |
| Observations | 615 | 615 | 615 | 615 | 614 | 614 |
| $R^2$ | 0.037 | 0.025 | 0.122 | 0.113 | 0.145 | 0.138 |
| *BIC* | 1534.0 | 1541.6 | 1470.4 | 1476.8 | 1465.6 | 1470.9 |
| Industry Controls | No | No | Yes | Yes | Yes | Yes |

Standard errors in parentheses

All regressions are firm-level OLS with s.e. clustered by industry; depvar is continuous factor in Roach and Cohen [2013]

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table A.2:** OLS models relating stock market reaction to science references

| VARIABLES | (1) Stock | (2) Stock | (3) Stock | (4) Stock | (5) Stock | (6) Stock |
|---|---|---|---|---|---|---|
| Number of front page science refs | 0.10** | | 0.10** | | | |
| | (0.04) | | (0.04) | | | |
| Number of full text science refs | | 0.02 | -0.004 | | | |
| | | (0.04) | (0.04) | | | |
| Any front page science refs? | | | | 0.3 | | -0.7 |
| | | | | (0.5) | | (0.5) |
| Any full text science refs? | | | | | 2.8*** | 3.2*** |
| | | | | | (0.7) | (0.8) |
| issyear = 2007 | 1.5*** | 1.5*** | 1.5*** | 1.5*** | 1.5*** | 1.5*** |
| | (0.1) | (0.1) | (0.1) | (0.1) | (0.1) | (0.1) |
| issyear = 2008 | 6.7*** | 6.7*** | 6.7*** | 6.7*** | 6.7*** | 6.7*** |
| | (0.2) | (0.2) | (0.2) | (0.2) | (0.2) | (0.2) |
| Constant | 8.4*** | 8.4*** | 8.4*** | 8.4*** | 8.3*** | 8.3*** |
| | (0.08) | (0.08) | (0.08) | (0.08) | (0.08) | (0.08) |
| Observations | 199,986 | 199,986 | 199,986 | 199,986 | 199,986 | 199,986 |
| R-squared | 0.126 | 0.126 | 0.126 | 0.126 | 0.126 | 0.126 |
| Patent class FE | Yes | Yes | Yes | Yes | Yes | Yes |
| BIC | 1906330.95 | 1906339.77 | 1906343.14 | 1906339.67 | 1906298.57 | 1906307.13 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table A.3:** OLS models relating whether maintained to year 8 to science references

| VARIABLES | (1) Maintained | (2) Maintained | (3) Maintained | (4) Maintained | (5) Maintained | (6) Maintained |
|---|---|---|---|---|---|---|
| Number of front page science refs | 0.0006* | | 0.001*** | | | |
| | (0.0003) | | (0.0003) | | | |
| Number of full text science refs | | -0.002*** | -0.002*** | | | |
| | | (0.0003) | (0.0003) | | | |
| Any front page science refs? | | | | -0.04*** | | -0.03*** |
| | | | | (0.004) | | (0.004) |
| Any full text science refs? | | | | | -0.06*** | -0.04*** |
| | | | | | (0.004) | (0.005) |
| issyear = 2007 | 0.0003 | 0.0003 | 0.0003 | 0.0005 | 0.0004 | 0.0004 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| issyear = 2008 | -0.007*** | -0.007*** | -0.007*** | -0.007*** | -0.007*** | -0.007*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Constant | 0.6*** | 0.6*** | 0.6*** | 0.6*** | 0.6*** | 0.6*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Observations | 489,346 | 489,346 | 489,346 | 489,346 | 489,346 | 489,346 |
| R-squared | 0.076 | 0.076 | 0.076 | 0.076 | 0.076 | 0.076 |
| Patent class FE | Yes | Yes | Yes | Yes | Yes | Yes |
| BIC | 661556.99 | 661525.07 | 661520.23 | 661408.7 | 661365.09 | 661328.82 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

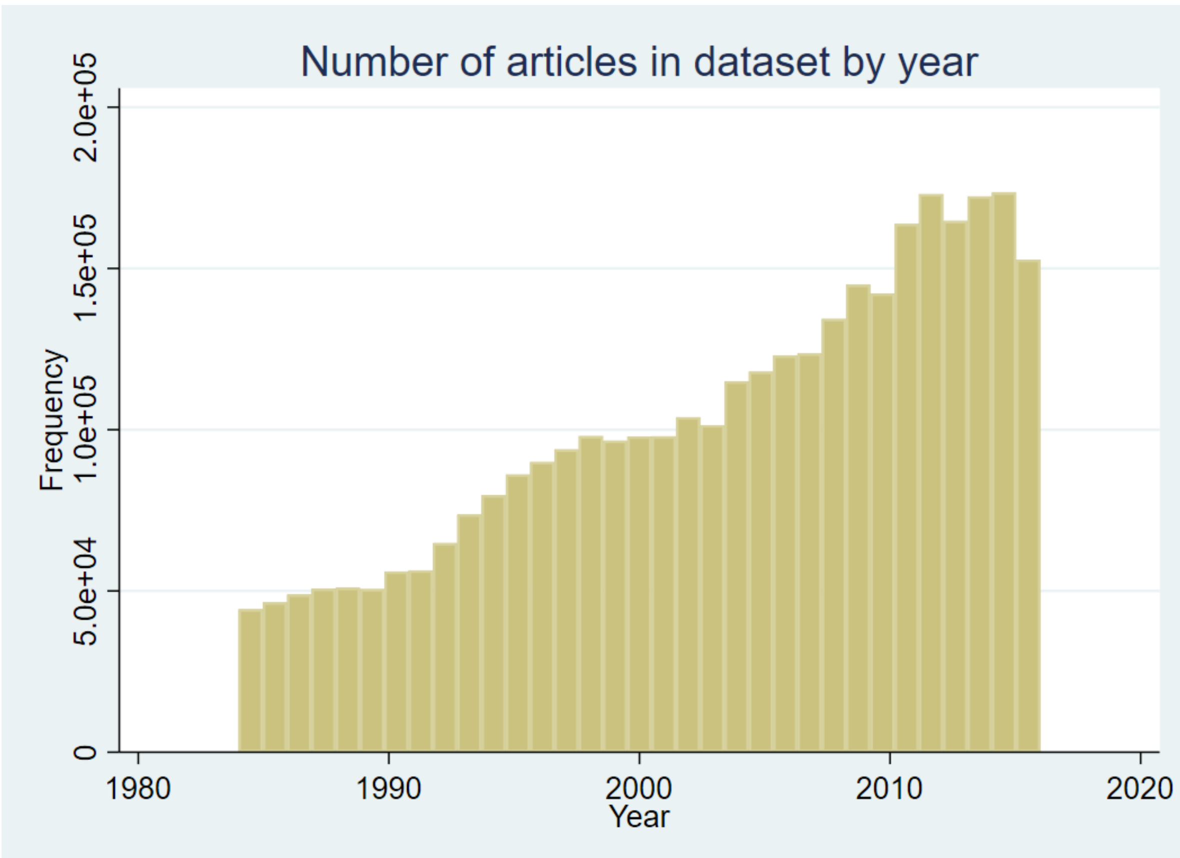



**Figure A.1:** Articles in Web of Science dataset across 260 journals by year.

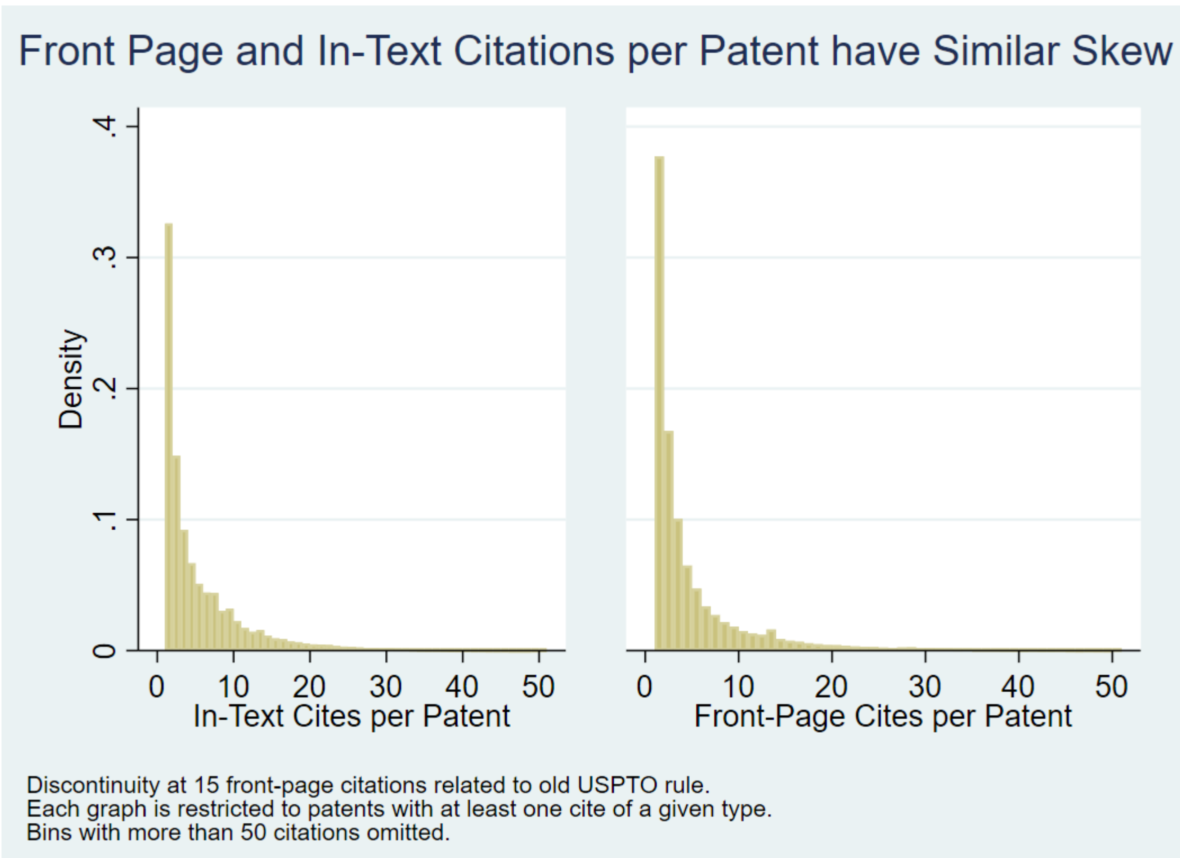



**Figure A.2:** The distribution of front page and in-text citations per patent are both highly skewed.