STRATEGIC CITATION: A REASSESSMENT

Jeffrey Kuhn, Kenneth Younge, and Alan Marco*

Abstract—The United States patent system is unique in that it requires applicants to cite documents they know to be relevant to the examination of their patent applications. Lampe (2012) presents evidence that applicants strategically withhold 21%–33% of relevant citations from patent examiners, suggesting that many patents are fraudulently obtained. We challenge this view. We first show that Lampe's empirical design is inconsistent with both legal standards and standard operating procedures, including how courts identify strategic withholding. We then compile comprehensive data to reassess the empirical basis for Lampe's main claim. We find no evidence that applicants withhold citations.

I. Introduction

THE United States is unique in requiring a patent applicant to disclose to the U.S. Patent and Trademark Office (USPTO) any document the applicant believes to be relevant to the examination of a patent application. In theory, this "duty to disclose" improves patent examination quality by identifying relevant prior art. However, Lampe (2012) argues that firms strategically withhold 21%–33% of relevant prior art citations. If accurate, Lampe's findings would be alarming, for they suggest that patent applicants (in effect) commit widespread fraud, and presumably thereby obtain more patent protection than is due.

We present evidence that challenges the view that applicants systematically underdisclose; instead, we find no evidence of strategic withholding. We first discuss institutional reasons for believing that Lampe's (2012) methodology biases the results in favor of finding strategic withholding. We then replicate Lampe's (2012) analysis using a larger sample and more accurate and comprehensive data. Finally, we provide empirical evidence suggesting that the methodology is subject to bias based on selection effects, time trends, and firm size.

This article makes several contributions. First, it presents new evidence of interest to policymakers (Mammen, 2009; Kuhn, 2010; Johnson, 2017), for it directly conflicts with earlier results (Lampe, 2012) and challenges the view that the duty of disclosure is ineffective (Taylor, 2012). Second, it contributes to the literature that investigates patent examination as an important aspect of innovation economics (Cockburn, Kortum, & Stern, 2003; Lemley & Sampat, 2012; Frakes & Wasserman, 2015). Third, it contributes to a literature showing how refinements in patent data can support a more nuanced and accurate empirical assessment of innovation (Hall, Jaffe, & Trajtenberg, 2005; Alcacer & Gittel-

man, 2006; Jaffe & De Rassenfosse, 2017; Kuhn, Younge, & Marco, 2020).

II. Institutional Background

A. The Patent Examination Process

A patent examiner determines whether application's claims constitute a new and nonobvious advance over the prior art. To identify prior art, the examiner is required to perform a search and to review applicant submitted references. An applicant is not obligated to search for prior art but may nevertheless know of relevant documents. Since disclosure may not be in applicants' interests, U.S. law imposes a duty to disclose all information known to be material to patentability (37 C.F.R. §1.56). The duty extends to inventors, attorneys, and any other agents of the firm involved in the patent application. Violation can lead to severe penalties, such as unenforceability for the patent and disbarment for complicit attorneys.

When an examiner identifies prior art that justifies rejecting application's claims, she issues an "Office Action" identifying particular locations in specific prior art references where the features of the claim are described. Rejections cabin the scope of patent claims, since most patents are initially rejected, but applicants typically overcome rejections by narrowing the claims (Marco, Sarnoff, & deGrazia, 2019; Kuhn & Thompson, 2019). Examiners rely upon both examiner and applicant references to justify rejections, but the majority of citations *do not* form the basis of rejections. Such citations nevertheless circumscribe the technology, clarify the inventive step, and document both the applicant's disclosure efforts and the examiner's search and examination. We emphasize that the examiner must confirm that he actively reviewed these nonrejection citations and nevertheless considered the claimed invention to represent a novel and nonobvious improvement over them.

The duty of disclosure is intended to prohibit strategic withholding, which may lead to an applicant receiving patent rights that the examiner would not have granted had the examiner known of the withheld information. Lampe (2012) does not conceptually define "strategic withholding," but at a minimum the term implies an active and knowledgeable choice. Thus the term implies that (1) the information in question *was relevant* to the claims of the focal patent, (2) the applicant *knew that* the information was relevant to the examination of the focal patent, (3) the applicant did not submit the information for consideration, and (4) the withholding was intentional rather than the result of oversight. These criteria not only are consistent with the term's ordinary meaning but also accurately reflect the extensive body of law through which courts evaluate claims of strategic withholding (Cotropia, 2009).

B. *Empirically Identifying Strategic Withholding*

Large sample analysis typically precludes the type of judicious evaluation performed by courts, so an empiricist must identify strategic withholding from observational data. Such an approach will only produce reliable results if the empirical definition of strategic withholding (1) is consistent with the theoretical construct and (2) yields a relatively unbiased measure. We argue that nearly all of the citations identified by Lampe's methodology as strategically withheld probably do not meet one or more of the conceptual criteria for strategic withholding, and that both the sample selected (i.e., "relevant" citations) and the dependent variable (i.e., strategic withholding) are likely to be overinclusive and unrepresentative. We therefore contend that Lampe's methodology is flawed in both overall approach and specific selection criteria, and thus likely to lead to biased and unreliable estimates.

Lampe (2012) identifies strategic withholding based on patterns of citations. Lampe defines a citation by patent A of patent B as "relevant" if (1) patent C also cites patent B, (2) patents A and C were assigned to the same firm, and (3) patent C was granted in a calendar year before patent A was filed. Thus a relevant citation (patent B) is one which the firm was aware of (as evidenced by patent C) when it filed the new application (patent A). Lampe defines a relevant citation as "strategically withheld" if it was submitted by the examiner rather than the applicant. That is, Lampe assumes that any examiner citation is strategically withheld if it was previously cited anywhere in the applicant's patent portfolio.

We now examine Lampe's empirical definition of strategic withholding in light of the construct defined in section II.A. The first element of the construct is that the information in question *was relevant* to the examination of the focal patent. We contend that Lampe's assumption that all citations are relevant is invalid, for both applicant and examiner citations.

Kuhn et al. (2020) show that the technological proximity between citing and cited patents has declined substantially over time for applicant citations. The decline is likely an unintended consequence of the duty of disclosure itself—applicants reduce both the compliance cost and the risk of inadvertent noncompliance by simply citing everything, copying hundreds or thousands of citations from patent to related patent without manual review. The vast majority of these citations are ignored by examiners and are not, in fact, relevant to the claims of the citing patent—only about 5% of all citations form the basis of a claim rejection. Lampe's defining a reference as "relevant" merely by virtue of the applicant having cited it is therefore inconsistent with the practicalities of patent examination.

Examiners often cite references not as evidence that the claimed invention is unpatentable, but rather as evidence that the examiner performed an adequate search, or as background technical material. For precisely this reason, courts determine strategic withholding on the basis of whether the withheld information would have been used to support a rejection, and not merely whether the examiner did cite or would have cited the withheld information. This rule is not legalistic, but instead reflects the practicality that applicants cannot be expected to accurately anticipate which of potentially thousands of related background references an examiner may subjectively deem informative. For the same reason, Lampe's defining a citation as "relevant" on the basis of examiner citation alone is also inconsistent with the realities of patent examination.

The second element of the construct—knowledge of relevance—is equally troubling. Companies and inventors with large patent portfolios will have cited many references in the past, and may simply not make the logical connection from a previously cited reference to a newly filed patent application. As noted above, the majority of citations submitted in recent years are likely generated by attorneys copying them across related applications, and nearly half of citations are submitted long after the citing patent is filed (Kuhn et al., 2020). Firms typically employ attorneys to handle patent examination and rarely involve the inventors in any significant way. Indeed, we know of no reason that an inventor would be aware of citations made by attorneys or examiners in previous patents by the same firm or even the same inventor. We therefore question Lampe's assumption that an examiner citation to a reference previously cited in a different patent by the same firm or inventor typically indicates an intentional decision by the applicant to withhold information.

The third element of the construct seems trivial: the identification of a citation as examiner-submitted would seem to constitute evidence that the applicant did not, in fact, submit the information for consideration. However, the USPTO's attribution of the citation's source can be misleading—the USPTO designation (MPEP 1302.12) only indicates whether the reference is *ever* submitted by the examiner, so an applicant-added citation that is re-added by the examiner is nevertheless designated as "cited by examiner." In practice, examiners often ignore applicants' citations in favor of their own search results (Cotropia, Lemley, & Sampat, 2013) and add citations that are highly similar or even identical to those already submitted by the applicant, a particularly common occurrence for rejection citations (Kuhn et al., 2020). The applicant cannot be said to have withheld information in such situations, despite the presence of an examiner citation. Accordingly, even Lampe's assumption that an examiner-submitted citation identifies information not already submitted by the applicant is demonstrably false for some citations.

All of these concerns have the same practical effect. Namely, the empirical definition of "strategically withheld" citations employed by Lampe (2012) is overly broad, and likely encompasses many citations that were, in fact, not actually withheld. At the same time, most of the citations identified as "relevant" are likely *irrelevant* citations that are mechanically copied from patent-to-patent without manual review. Accordingly, variation in both sample selection (i.e., "relevant") and dependent variable (i.e., "strategically withheld") may largely reflect differences in automated compliance strategies rather than intentional document-level

decisions. These automated compliance strategies are likely to vary with firm size, location (i.e., United States versus foreign), and technology, among other characteristics—precisely the factors identified by Lampe as dimensions along which the rate of strategic withholding varies.

III.   Data and Replication

A.   Sources

We obtain bibliographic information on patents, including patent citations, from the PatentsView dataset, which was unavailable at the time of Lampe's analysis. PatentsView provides several advantages over the NBER Patent Data files employed by Lampe (2012), such as inventor disambiguation and improved firm disambiguation. However, the indicator for whether a citation was examiner-added is unavailable prior to 2002, so our sample includes patents granted 2002–2014, whereas Lampe's sample includes patents granted 2001–2002.

The Google Patents Public Datasets provide patent priority data and a correspondence between granted patent number and the patent's pre-grant publication (PGPub) number. Following Kuhn et al. (2020), citations in the sample include those made to PGPubs that were later granted as patents.

We employ patent-to-patent textual similarity data developed by Younge and Kuhn (2016), who compute the patent-to-patent cosine similarity of the full text of every pair of patents under a vector space, term-frequency inverse document frequency model. Kuhn et al. (2020) use these data to evaluate the technological relatedness of different groups of citations.

The PatentsView data set identifies whether an examiner submitted a citation reference, but as discussed in section II.B, this designation should not be interpreted to indicate that the reference was *first added* by the examiner. Kuhn et al. (2020) provide a correction to citation source attribution based on new data from internal USPTO citation submissions forms, which allows for more accurate identification of a citation's original submitter for the period from 2005 through 2014.

We identify patent citations used to support claim rejections from the Office Action Research Dataset for Patents described by Lu, Myers, and Beliveau (2017) for the period 2008–2014 and from the bulk data files published by the USPTO for the period 2005–2008.

All datasets are publicly available or available upon request.

B.   Sample

To reassess Lampe's conclusions, we construct a larger sample to replicate the analysis over a longer time period. We select all utility patents issued from 2002 to 2014, inclusive. Following Lampe (2012), we exclude continuing patents (i.e., continuation, continuation-in-part, divisional, and reissue patents), patents not assigned to firms, and patents as-

signed to more than one firm. Our final sample of patents includes data for 1,746,730 patents.

Next, we select all citations made by these patents that meet Lampe's criteria for relevant citations. The cited patent must have been cited by a different patent that is: (1) issued to the same firm as the citing patent, and (2) issued in a year prior to the year in which the focal citing patent was filed. We exclude citations made to the applicant's own patents (i.e., a self-citation). The final sample includes data for 2,480,248 patent citations.

Because larger firms file many patents and cite many references, some of these previously- cited references may be unfamiliar to inventors and attorneys involved with a later patent application. Accordingly, we follow Lampe (2012) by constructing a subsample of citations that were previously cited in a patent having at least one inventor in common with the focal patent. The common inventor subsample includes data for 784,355 patent citations.

C.   Variable Definitions and Summary Statistics

Table 1 provides summary statistics for the variables used in this study for the full sample and the common inventor subsample. Columns 1, 2, 7, and 8 copy the corresponding values from table 1 of Lampe (2012). Columns 3–6 and 9–12 include statistics for our replication. We would not expect our summary statistics to be identical to Lampe, because we employ different data sources and because our sample includes an overlapping but not identical time span. Nonetheless, the values in the 2002 Replication columns are broadly consistent with the values reported by Lampe (2012).

Following Lampe (2012), *Common inventors* counts the number of times that any inventor of the focal patent was also an inventor on a prior patent that cited the same reference. We calculate this variable using equation (2) in Lampe (2012). Inventors are identified in raw patent data by name and not by a unique identifier. We therefore employ the disambiguated inventor identifier provided by PatentsView to identify previous patents by the same identifier. The common inventor subsample restricts the citations to those for which *Common inventors* ≥ 1.

The variable *Applicant-added* identifies whether the citation was added by the applicant. We find that 71% (81%) of citations in the 2002 replication of the full sample (common inventor subsample) are applicant-added, an increase of 4 (2) percentage points over the value reported by Lampe (2012). This modest difference is likely due to differences between the data sources (e.g., firm disambiguation) leading to differences in sample selection.

To better understand the number of patents "at hazard" of being strategically withheld, we construct counts of the number of unique patents previously cited by the firm and inventor. We construct *previously cited patents (by firm)* by first identifying all patents granted to the same firm as the citing patent in the years before the citing patent was filed (i.e. the firm's prior patents). We then count all unique patents cited

TABLE 1.—DESCRIPTIVE STATISTICS

| | Full sample | | | | | | Common inventor subsample | | | | | |
| | Lampe 2001–2002 | | Replication 2002 | | Replication 2002–2014 | | Lampe 2001–2002 | | Replication 2002 | | Replication 2002–2014 | |
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Citing application year	1,999.18	1.16	1,999.64	1.08	2,006.44	3.61	1,999.33	1.13	1,999.83	0.98	2,006.77	3.61
Cited application year	1,987.40	5.85	1,987.91	7.57	1,993.29	7.26	1,987.03	6.03	1,987.68	5.83	1,992.95	7.40
Citing grant year	2,001.52	0.50	2,002.00	0.00	2,009.98	3.54	2,001.52	0.50	2,002.00	0.00	2,010.12	3.52
Cited grant year	1,989.29	5.75	1,989.81	5.61	1,995.58	7.42	1,989.89	5.92	1,989.54	5.70	1,995.16	7.61
Replication variables												
Applicant-added	0.67	0.47	0.71	0.46	0.81	0.39	0.79	0.41	0.81	0.39	0.90	0.30
Common inventors	1.12	3.89	0.97	2.98	1.78	9.73	3.52	6.26	3.12	4.68	5.62	16.66
Control variables												
Attorney or agent	0.94	0.24	0.94	0.24	0.89	0.32	0.95	0.22	0.95	0.21	0.88	0.33
Non-U.S. firm	0.28	0.45	0.24	0.43	0.19	0.39	0.23	0.42	0.19	0.39	0.15	0.35
Examination time	2.34	1.06	2.42	1.01	3.57	1.60	2.19	1.01	2.27	0.90	3.39	1.54
Citing claims	21.84	18.70	23.32	23.25	22.10	15.13	24.47	24.97	26.19	33.10	23.52	17.08
Cited claims	15.73	14.09	15.76	14.18	19.58	17.27	15.89	13.86	15.81	13.18	19.34	17.03
Previously cited patents												
By firm			10,359	16,283	15,866	33,707			6,273	12,023	8,353	21,933
By common inventor			99	154	265	726			158	193	423	907
Additional variables												
Rejection (102 or 103)					0.04	0.20					0.02	0.15
Applicant-added (cor.)					0.83	0.37					0.91	0.28
Observations	126,340		75,371		2,480,248		40,085		23,382		784,355	

in any of the firm's prior patents. *Previously cited patents (by common inventor)* repeats this analysis for each of the inventors of the citing patent, and represents a count of the union of all citations previously made by those inventors.

For the period from 2008 to 2014, we identify citations used to support claim rejections directly from the Office Action Dataset. For the period from 2005 to 2008, we follow Cotropia et al. (2013) and analyze the raw text of communications (known as "office actions") sent from USPTO patent examiners to applicants to identify rejection citations. We use optical character recognition to convert more than 50 million pages of documents from images to text, and then used natural language processing techniques and regular expressions to identify patent numbers used to support rejections. We find that about 4.2% of citations were used to support a rejection.

Finally, we identify a citation as being applicant-added (corrected) when it meets any of three criteria: (1) the citation was not identified in the USPTO data as examiner-added, (2) the citation was first submitted by the applicant according to internal USPTO records, or (3) the applicant submitted a different reference that is more than 80% textually similar to the focal reference. Although the difference between *applicant-added* and *applicant-added (corrected)* in table 1 may seem small (0.81 versus 0.83), it represents a decrease in examiner citations by about 10% across the sample, and examiner *rejection* citations are even more likely to be corrected (Kuhn et al., 2020).

D. Replication

Lampe's (2012) main result is that applicants withhold between 33% (full) and 21% (common inventor) of relevant citations for a sample of patents granted in 2001 and 2002. For clarity, we note that this result follows directly from table 1 of Lampe (2012). As discussed in section II.B. Lampe (2012) assumes that any citation to a reference that was previously cited by the same firm is strategically withheld if it is submitted by the examiner. Because both the full sample and the common inventor subsample described in Lampe (2012) include only relevant citations, the percentage reported as withheld is simply the percentage submitted by the applicant, subtracted from one hundred. The remainder of the results in Lampe (2012), such as tables explaining variation in which citations are withheld, rely on the validity of this main result. Applying the same criteria and assumptions as Lampe (2012), our 2002 replication sample shows that between 29% (full) and 19% (common inventor) of relevant citations meet Lampe's criteria for strategic withholding.

IV. Sample Evaluation

In section IV.A, we present evidence that the full sample as defined by Lampe (2012) does not lead to reliable estimates of strategic withholding because the methodology results in estimates with an upward bias that increases with firm size. Section IV.B shows that moving to the common inventor subsample does not entirely correct the problem, and does nothing to address several other problems that we identified in section II.B. We investigate and reject various possible corrections in section IV.C, and conclude that Lampe's general methodology is unlikely to lead to reliable estimates of strategic withholding regardless of the sample selection criteria and variable definitions.

FIGURE 1.—DENSITY OF PREVIOUS CITATIONS
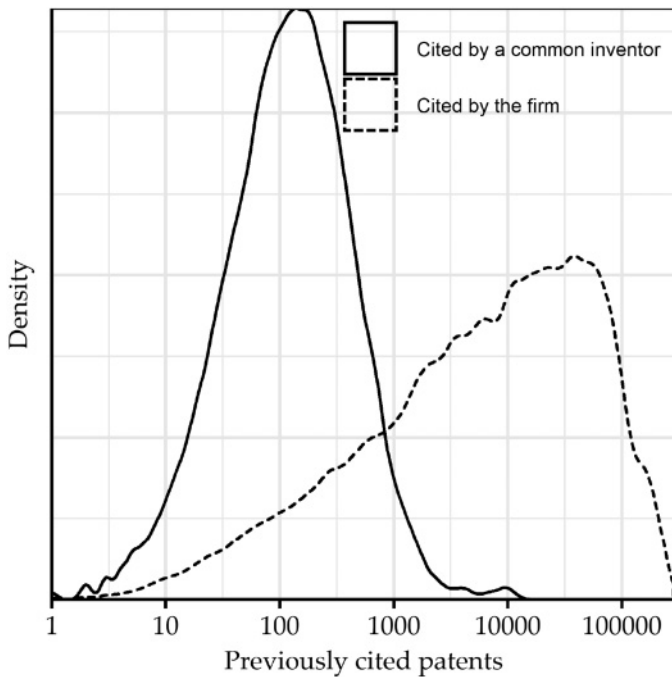
A. Full Sample

In this section, we argue that the full sample as defined by Lampe does not provide a credible basis for investigating when and under what conditions the applicant has strategically withheld citations from the patent office. We begin by noting that strategic withholding implies that a reference cited by a patent examiner was not only known to the patent applicant, but also that the applicant knew that the reference was relevant.

A firm with a large patent portfolio will have previously cited many references. When an examiner cites one of those references in a later-filed application, it is possible that no one at the firm drew a logical connection between the previous citation and the subsequently filed patent. Assuming that *any* examiner citation to a reference previously cited in a patent by the same firm is evidence of strategic withholding biases the results in favor of finding strategic withholding.

Figure 1 illustrates this problem graphically by plotting the number of relevant references for patents in both the full sample and the common inventors subsample. On average, a patent in the full sample is assigned to a firm that previously cited more than 10,000 references. Indeed, some patents are assigned to firms that previously cited over 300,000 references.

If the mere presence of so many previous citations leads to citations being erroneously identified as strategically withheld, then we should expect that for purely mechanical reasons the incidence of strategic withholding increases with the number of previous citations. To test this argument, table 2 includes results from linear probability models es-

timating the probability that a citation in the sample is added by the examiner. In column 1, a 100% increase in the number of previously-cited references corresponds to a 0.039 increase ($p < 0.001$) in the probability that a focal citation is examiner-added and hence appears as strategically withheld. This result is robust to the inclusion of a variety of control variables in column 2, and is economically significant given a mean probability of withholding of 0.17.

One interpretation of these results is that larger firms, such as General Electric, IBM, and Microsoft, are more likely to commit fraud before the patent office than smaller firms, which we do not believe to be a credible conclusion. For instance, we find that under Lampe's definition IBM strategically withholds up to 40% of relevant citations from the patent office. A more reasonable interpretation is that when examiners identify references for patents filed by such firms, those examiner-added references are simply more likely to have been previously cited by the same firm, as a matter of chance.

Measurement error in the dependent variable that is uncorrelated with predictors does not bias estimates. In this context, however, the number of previous citations made by a firm is of course highly correlated with firm size, age, the presence of an attorney, and a variety of other predictors. Moreover, the measurement error is not only located in the dependent variable (i.e., which citations are "withheld"), but also in the sample selection itself (i.e., which citations are "relevant"). For these reasons, we conclude that any estimates based on the full sample approach described in Lampe are likely to be biased, and in particular are likely to substantially overstate the incidence of strategic withholding.

B. Common Inventor Subsample

In this section, we argue that the common inventor subsample is also unlikely to produce unbiased estimates of strategic withholding. Lampe's common inventor subsample restricts the full sample to those citations previously made in a patent by a common inventor within the firm. In theory, this new restriction ensures that at least one of the inventors of the subsequent patent knew of the citation. In practice, as we discussed in section II.B, the inventor in the subsequent patent likely did not know of the earlier citation, because it was likely submitted by an attorney or the examiner, and was unlikely to logically connect the cited reference to the newly filed patent application.

We note that a patent in the common inventors subsample is filed by inventors whose previous patents jointly cite over 420 references, on average. Indeed, as shown in figure 1, some patents are filed by inventors who jointly cite over 10,000 previous references. As the number of previously cited references increases, the likelihood of unintentionally overlooking a technological relationship between a newly filed patent application and a reference previously cited by the firm increases.

TABLE 2.—LINEAR PROBABILITY MODEL REGRESSIONS OF EXAMINER CITATION

	Full Sample		Common Inventor Subsample		Restricted Subsample	
	(1)	(2)	(3)	(4)	(5)	(6)
Previously-cited patents	0.039***	0.035***	0.007***	0.007***	0.032***	0.031***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.001)	(0.001)
Attorney or agent		0.062***		0.041***		0.026*
		(0.001)		(0.001)		(0.013)
Examination time		−0.004***		−0.007***		−0.004
		(0.0001)		(0.0002)		(0.002)
Citing claims		−0.002***		−0.001***		−0.002***
		(0.00002)		(0.00002)		(0.0003)
Cited claims		−0.0003***		−0.0003***		−0.001***
		(0.00001)		(0.00002)		(0.0002)
Non-U.S. firm		0.270***		0.206***		0.175***
		(0.001)		(0.001)		(0.007)
Constant	0.186***	0.149***	0.108***	0.103***	0.603***	0.585***
	(0.0002)	(0.001)	(0.0004)	(0.001)	(0.004)	(0.016)
Observations	2,480,246	2,480,246	784,355	784,355	18,812	18,812
R^2	0.049	0.142	0.003	0.077	0.025	0.062

Previously-cited patents is a logged count of the number of patents cited by the firm in previous calendar years. *Examination time* is measured in years. Standard errors in parentheses. Two-tailed tests: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Columns 3 and 4 of table 2 repeat the analyses in columns 1 and 2, but for the common inventors subsample. In column 3, a 100% increase in the number of previously-cited references corresponds to a 0.007 increase ($p < 0.001$) in the probability that a focal citation is examiner-added, relative to a mean probability of 0.10. This result is also robust to the inclusion of a variety of control variables in column 4. While the coefficients in columns 3 and 4 are lower than in columns 1 and 2, they remain positive and statistically significant. This suggests either that firms with larger patent portfolios are more likely to commit fraud, or that Lampe's methodology overestimates the incidence of strategic withholding for larger firms, even for the common inventor subsample. The common inventor subsample therefore suffers from precisely the same problem as the full sample. Both the sample selection criteria and the dependent variable therefore seem likely to be biased in a way that is correlated with the predictors, leading to biased estimates.

Finally, as further evidence of unreliability, we observe that the rate of strategic withholding in both the full sample and the common inventor subsample declined by more than 50% from 2002 to 2014 (see figure 3a). Because we can identify no reason to believe that the rate of fraud has declined so substantially and smoothly over that 13-year period, we conclude that neither sample provides a reliable estimate of the rate of strategic withholding.

C. Possible Corrections

One problem with even the common inventor subsample is that previous patents by prolific inventors may have cited thousands of references. We could therefore impose an additional selection criteria restricting both relevant and withheld citations to those situations in which the inventors had previously cited fewer than some threshold number of references (e.g., 100 references). Such a restriction, however,

FIGURE 2.—VENN DIAGRAMS FOR ALTERNATIVE SUBSAMPLES SHOWING CITATIONS SELECTED AS RELEVANT, AND PERCENTAGE SUBMITTED BY EXAMINER
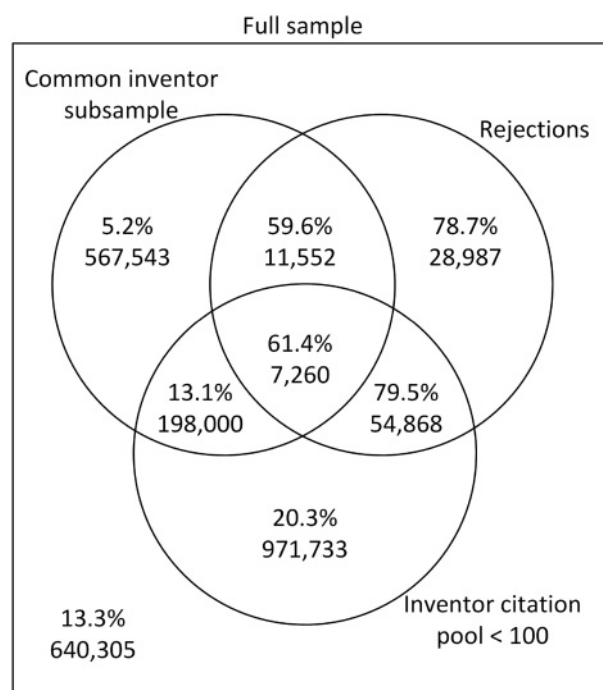
would still fail to address the fact that many citations describe background information that is not particularly relevant to the examination of the citing patent. Accordingly, we could alternatively restrict both relevant and withheld citations to those used to support a rejection of the claims, which are certainly relevant. In this section, we discuss why imposing additional selection criteria such as these is unlikely to lead to remedy the problems with Lampe's methodology and yield credible estimates of strategic withholding.

First, different combinations of selection criteria lead to very different samples and results. Figure 2 shows a Venn

FIGURE 3.—INCIDENCE OF CITATIONS AND PATENTS THAT MEET DEFINITION OF WITHHOLDING
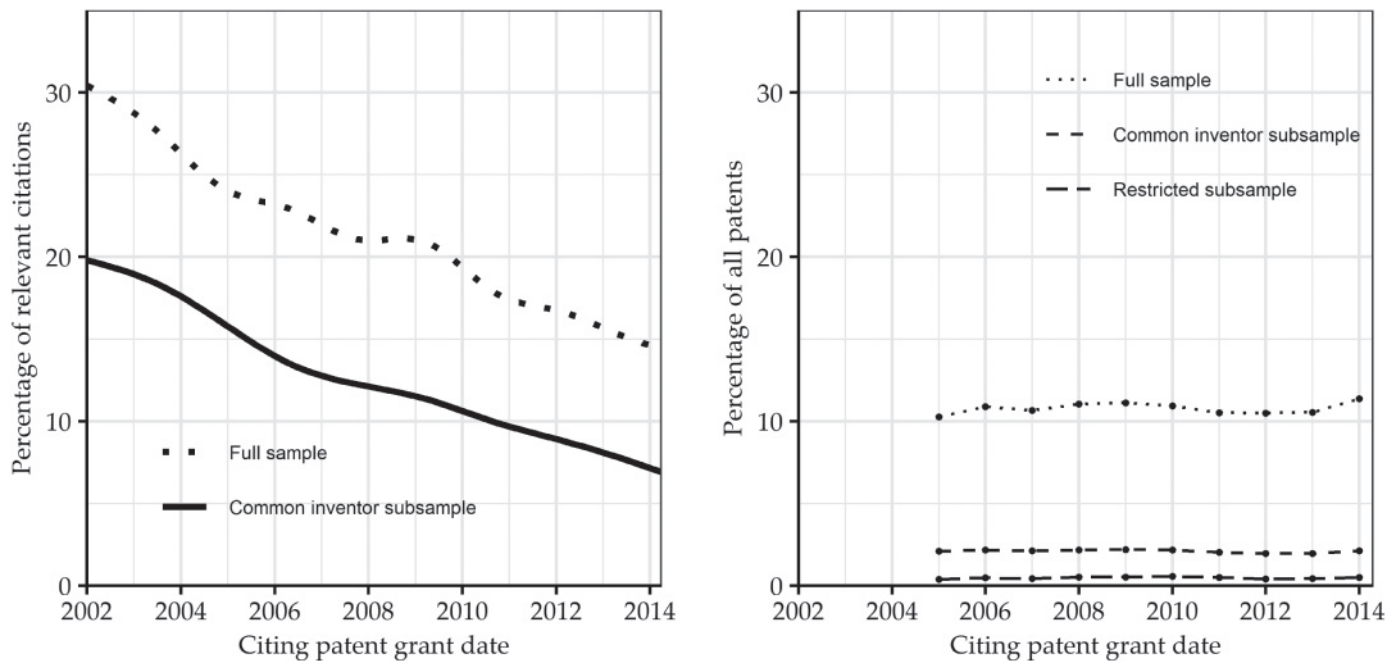
(a) Percent of relevant citations that is examiner-submitted

(b) Percent of all patents with at least once citation that meets the definition of withholding

diagram of the number of citations included in the full sample, the common inventor subsample, and two other subsamples. The rejection subsample restricts the analysis to citations used to support rejections. The inventor citation pool <100 subsample excludes citations, whether or not in the common inventor subsample, when the inventors have previously cited 100 or more references. For all samples, we employ the *applicant-added (corrected)* variable to identify strategic withholding. Estimates of strategic withholding vary from 5.2% to 79.5%, depending on the combination of selection criteria employed. Indeed, these are not the only selection criteria one might use; alternatively one might restrict to the set of citations that share a common attorney, or restrict to citations made by firms below a certain size, or restrict to citations that are textually similar to the citing patent. Because one could reasonably argue for or against each of these selection criteria, the reported outcome is essentially a matter of choice.

Second, even employing all four selection criteria (the restricted sample), we still observe a firm-size effect. As shown in models 5 and 6 of table 2, a 100% increase in the number of previously-cited patents corresponds to a 0.031 increase ($p < 0.001$) in the probability of strategic withholding, relative to a mean probability of 0.061. Accordingly, even the most restrictive selection criteria employed in figure 2 fail to address the problems evident in the full sample and common inventor subsample.

Third, the tradeoff between accuracy and external validity in this context is likely severe. Figure 3b plots the percentage

of all patents that have at least one citation that meets Lampe's definition of strategic withholding, under different sampling approaches. In the most Restricted Subsample, only 7,260 citations over 13 years meet the definition of "relevant," and only 362 patents per year (0.18% of patents in the sample) have even a single withheld citation. At the extreme, we could select a very small sample of litigated patents and determine with some accuracy the rate of strategic withholding for those patents. However, the citations we identified as strategically withheld in that highly selected subsample will not be representative of withheld citations more generally.

Fourth, all subsamples are both overinclusive and underinclusive. They are overinclusive for the reasons discussed in section II.B. However, they are also underinclusive in the sense that many instances of strategic withholding involve the withholding of information such as patents that a firm has not previously cited, nonpatent literature, or foreign patents, none of which are included in either sample. Further, both errors are correlated with predictor variables such as firm size, and additional sample restrictions do not resolve these problems.

Fifth, if interpreted as evidence of strategic withholding, the results presented in table 2 are inconsistent with the institutions related to patent examination. For example, a citation made in a patent by a non-U.S. firm is between 0.175 and 0.270 more probable to be examiner-added. The location in which a firm is incorporated seems unlikely to have such a substantial effect on whether the firm strategically withholds information from the USPTO. A more reasonable interpretation is that non-U.S. firms file more of their patents in

non-U.S. jurisdictions, which would mean that the count variable *previously-cited patents* is a less reliable control for the size of such firms. Further, a citation made in a patent in which the firm is represented by an attorney or agent is between 0.026 ($p < 0.05$) and 0.062 ($p < 0.001$) more probable to be examiner-added. Attorneys and agents owe an independent duty of disclosure and candor to the patent office, and would be risking their disbarment by intentionally withholding information. We therefore expect that the positive coefficient indicates that the presence of an attorney or agent is another indication of firm size, rather than evidence that attorneys and agents are more likely to withhold information from the patent office.

In sum, the methodology employed in Lampe (2012) forces a severe trade-off. With few selection criteria, the samples are overinclusive in ways that may yield severely and unpredictably biased estimates. With more selection criteria, the sample size diminishes substantially without convincingly addressing several of the problems underlying the more general approach. We are skeptical that *any* selection criteria based on publicly available data is likely to lead to a sample suitable for generating relatively unbiased estimates of strategic withholding.

V.   Conclusion

An accurate assessment of applicant citation behavior is necessary for evaluating the costs and benefits of the duty of disclosure, particularly since the U.S. is the only major jurisdiction that imposes this obligation. Lampe's claim that applicants withhold between 21% and 33% of relevant citations provides a powerful argument against the efficacy of disclosure. However, that research relies on two key assumptions.

First, Lampe (2012) assumes that all cited references were indeed relevant to the examination of a patent simply by virtue of having been cited. However, the large majority of cited references do not affect the patent examination process, and indeed most citations are copied from patent to patent with little-to-no manual review. For good reason, courts do not expect applicants to anticipate which of the many different background references an examiner may choose to cite. This first assumption is therefore contrary to the practical realities of patent examination, and suggests that fewer than 5% of the citations identified by Lampe's methodology as "relevant" are entitled to that description.

Second, Lampe (2012) assumes that *any* examiner citation to a reference previously cited in a different patent granted to the same firm or inventor is evidence of strategic withholding. We show that the average firm has cited many references in the past and that the rate of examiner citation increases with the number of previously-cited references, an effect which persists through various sample selection criteria. While a small firm may easily review the citations made in its previously-granted patents, IBM (or even a prolific inventor) cannot be expected to accurately anticipate which

of its more than 300,000 previous citations an examiner may choose to cite in a subsequent patent application. Merely controlling for the number of previous citations is insufficient to address the problem because the bias is embedded in both the definition of the dependent variable (i.e., strategic withholding) and the sample selection criteria (i.e., large firms cite more references) in ways that are correlated with predictors such as the presence of an attorney, a firm's status as U.S. or foreign, and a patent's examination time.

Based on this evidence, we conclude that the large majority of citations identified by Lampe's methodology as "relevant" were in fact not relevant, and that the large majority of citations identified by Lampe's methodology as "strategically withheld" were in fact not strategically withheld, as those terms are typically construed. Moreover, various alternative but reasonable selection criteria lead to very different results, suggesting that under Lampe's methodology the main results are largely driven by the researcher's choices and assumptions rather than the phenomenon of interest. Given that our analysis calls into question assumptions integral to Lampe's results, we are forced to conclude that Lampe's claim that that applicants withhold between 21% and 33% of relevant citations is simply not supported by the evidence. The remainder of Lampe's results rely on the same samples and dependent variable to investigate the determinants of strategic withholding and therefore lack reliability and validity for the same reasons.

REFERENCES

Alcacer, Juan, and Michelle Gittelman, "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations," this REVIEW 88:4 (2006), 774–779.

Cockburn, Iain M., Samuel Kortum, and Scott Stern, "Are All Patent Examiners Equal? The Impact of Examiner Characteristics," in *Patents in the Knowledge-Based Economy* (Washington, DC: National Academies Press, 2003).

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

Cotropia, Christopher A., Mark A. Lemley, and Bhaven Sampat, "Do Applicant Patent Citations Matter?" *Research Policy* 42:4 (2013), 844–854. 10.1016/j.respol.2013.01.003

Frakes, Michael D., and Melissa F. Wasserman, "Does the U.S. Patent and Trademark Office Grant too Many Bad Patents: Evidence from a Quasi-Experiment," *Stan. L. Rev.* 67 (2015), 613.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg, "Market Value and Patent Citations," *The RAND Journal of Economics* 36:1 (2005), 16–38.

Jaffe, Adam B., and Gaétan De Rassenfosse, "Patent Citation Data in Social Science Research: Overview and Best Practices," *Journal of the Association for Information Science and Technology* 68:6 (2017), 1360–1374. 10.1002/asi.23731

Johnson, Eric E., "The Case for Eliminating Patent Law's Inequitable Conduct Defense," *Colum. L. Rev. Online* 117 (2017), 1.

Kuhn, Jeffrey M., "Information Overload at the U.S. Patent and Trademark Office: Reframing the Duty of Disclosure in Patent Law as a Search and Filter Problem," *Yale JL & Tech.* 13 (2010), 89.

Kuhn, Jeffrey M., and Neil C. Thompson, "How to Measure and Draw Causal Inferences with Patent Scope," *International Journal of the Economics of Business* 26:1 (2019), 5–38. 10.1080/13571516.2018.1553284

Kuhn, Jeffrey M., Kenneth A. Younge, and Alan C. Marco, "Patent Citations Reexamined," *RAND Journal of Economics* 51 (2020), 109–132. 10.1111/1756-2171.12307

Lampe, Ryan, "Strategic Citation," this REVIEW 94:1 (2012), 320–333.

Lemley, Mark A., and Bhaven Sampat, "Examiner Characteristics and Patent Office Outcomes," this REVIEW 94:3 (2012), 817–827.

Lu, Qiang, Amanda Myers, and Scott Beliveau, "USPTO Patent Prosecution Research Data: Unlocking Office Action Traits," USPTO Economic working paper (2017).

Mammen, Christian E., "Controlling the Plague: Reforming the Doctrine of Inequitable Conduct," *Berkeley Tech. LJ* 24 (2009), 1329.

Marco, Alan C., Joshua D. Sarnoff, and Charles A. W. deGrazia, "Patent Claims and Patent Scope," *Research Policy* 48 (2019), 103790. 10.1016/j.respol.2019.04.014

Taylor, Priscilla G., "Bringing Equity Back to the Inequitable Conduct Doctrine?" *Berkeley Technology Law Journal* 27 (2012), 349–379.

Younge, Kenneth A., and Jeffrey M. Kuhn, "Patent-to-Patent Similarity: A Vector Space Model," Available at SSRN 2709238 (2016).