The Ways We've been Measuring Patent Scope are Wrong: How to Measure and Draw Causal Inferences with Patent Scope[1]

Jeffrey M. Kuhn
UNC Chapel Hill Kenan-Flagler Business School

Neil C. Thompson
MIT Sloan School of Management

Abstract

According to surveys, the top 10% of patents are worth more than a thousand times as much as the bottom 10%. This isn't surprising since a patent's value derives from its ability to exclude rival products from the market, and patents vary widely in their ability to do this. At the same time as there is great variation in the value of patents that do issue, there is almost no variation between patents that are 'just' worth filing and those that aren't – both are nearly worthless. This presents a challenge for innovation scholars, because we have much better tools for measuring whether or not something is patented than we do for how much scope a patent protects. Thus our empirical tools are weakest precisely where they are most important.

This paper presents an easy-to-use measure of patent scope that is grounded both in patent law and in the practices of patent attorneys. Our measure counts the number of words in the patents' first claim. The longer the first claim, the *less* scope a patent has. This is because a longer claim has more details – and *all* those details must be met for another invention to be infringing. Hence, the more details there are in the patent, the greater are the opportunities for others to invent around it. We validate our measure by showing both that patent attorneys' subjective assessments of scope agree with our estimates, and that the behaviour of patenters is consistent with it. Our validation exercise also allows us to examine the performance of previous measures of patent scope: the number of patent classes, the number of citations made by future patents, and the number of claims in a patent. We find them all to be uninformative (no useful correlation with scope) or misleading (*negative* correlation with scope).

To facilitate drawing causal inferences with our measure, we show how it can be used to create an instrumental variable, patent examiner *Scope Toughness*, which we also validate. We then demonstrate the power of this instrument by examining standard-essential patents. We show that an (exogenous) diminishment of patent scope leads to patents being much less likely to be declared standard-essential.

[1] The authors would like to thank Tim Simcoe, Iain Cockburn, Scott Stern, Rui de Figueiredo, Noam Yuchtman, Kenneth Younge, and participants at the NBER Productivity Seminar, the MIT TIES Seminar, the Searle Conference on Innovation Economics, the MIT Laboratory for Innovation Seminar, and the EPFL MTEI Seminar for helpful comments. The Hoover Institution Working Group on Intellectual Property, Innovation, and Prosperity provided generous financial support during the preparation of this paper. They would also like to acknowledge Ben Roin's contributions to discussions of this topic.

I. Introduction

> *"Broad patent claims are similar to a property fence; the bigger the scope of the claims, the more intellectual property is covered by the claims" - ARC Intellectual Property Law blog*

This paper provides two tools for patent scholars to help them study patents and their value. The first is a measure of *patent scope*, the extent of the legal coverage that a patent provides. The second is an econometric instrument, a measure of a patent examiner's *scope toughness* on the extent of scope given to a patent, which allows for causal inference on the effect of getting more or less patent scope.

To ensure that our measures are capturing the phenomena that we intend, we ground them in (i) the details of patent law, (ii) the rules of the U.S. Patent and Trademark Office (USPTO), and (iii) the practice of attorneys prosecuting patents. We also do extensive validation work to show that our measures accord with the views of practitioner experts and with the observed behavior of patent filers. Next, we compare our measure with other approaches for measuring patent scope suggested in the literature. Finally, we provide an example of how our instrument can be used, showing that patents with (exogenously) narrower scope are less likely to be declared standards-essential.

Patents are valuable because of what they protect. If they can exclude competitors from a valuable market or from important technologies, they can generate greater profit margins or faster growth for the patent holder, becoming extremely valuable. In contrast, if they cannot restrict competitors, their value can be negligible. The enormous range of patent values can be seen in Figure 1, based on a survey of patent holders by Gambardella, Harhoff, and Verspagen (2008).

--------Insert Figure 1 about here--------

The ability of a patent to exclude others arises from a particular part of a patent: the claims. "Patent claims define the scope of patent protection. Whether a claim covers a product accused of infringement frequently determines whether a patent holder will be rewarded or left with nothing." (DiPietrantonio, 2011). If those claims are to be valuable then they must be sufficiently *broad* – that is, they must not only stop a competitor from directly copying the invention but also prevent that competitor from producing a readily substitutable variation of the invention. For example, a patent on an engine would be worth little if a competitor could escape that protection by making the same engine in a different color. Patent attorney Thomas Kulaga explains: "Broad patents typically offer more protection against infringers than a narrow patent" because they are "oftentimes very hard to design around". (Kulaga).

We show that the breadth of a patent's scope can be measured by counting the number of words in its first claim[2], with *more* words corresponding to *less* scope.[3] This relationship holds because a competitor's offering must meet every condition of a claim in order to infringe it, so a longer claim implies more conditions that must be met for a patent to be violated. For example, a patent on an "engine" would be broader than a patent on a "car engine" and both would be broader than a patent on a "200 horsepower car engine." This correspondence between claim length and patent scope is widely acknowledged by practitioners, who report that "longer claims are necessarily more narrow." (Quinn, 2015)

Ours is not the first attempt to measure patent scope. Common alternatives have included (1) the number of patent classes to which a patent is assigned, (2) the number of citations from future patents to that one, and (3) the number of claims listed on the patent. We test our measure, and these alternatives, against the judgment of patent attorneys as to the scope of patent coverage. We find that our measure is highly predictive ($p < 0.001$) and explains nearly half of all the variation in patent scope, as judged by experts. In contrast, we find that the number of patent classes assigned to a patent is *negatively* correlated with patent scope, the reverse of what is typically assumed. We find that the number of citations to a patent is only incredibly weakly (and statistically insignificantly) related to patent scope, and that in practice it only explains 0.04% of the variance in patent scope, as judged by experts. We do find that the number of claims in a patent is predictive of patent scope, but that this relationship is much weaker than that of our measure and only explains 7% of the variation in patent scope. From these findings we conclude that our measure is highly predictive of patent scope, and that traditional measures are either misleading or lack explanatory power.

Inspired by the work of Sampat and Williams (2015) we also use our measure to construct an econometric instrument, *scope toughness*. This reflects how tough a patent examiner at the USPTO is on giving scope when evaluating their patents. We construct our instrument in such a way that it is exogenous to the patent being evaluated, and therefore can be used for causal inference insomuch as patent examiner assignment is random (see Sampat and Williams (2015) and Righi and Simcoe (2017) for discussions of this).

Importantly, our measure of *scope toughness* is notably different from the measure introduced by Sampat and Williams (2015). Their measure is based on the likelihood of patents being granted at all. In cases where technology is very discrete, such as gene patenting where they apply it, we believe

[2] Technically we are interested in the first *independent* claim, although by law the first claim in a patent must be independent.
[3] Concurrent work done independently by Marco, Sarnoff and deGrazia (Working Paper) has also proposed a related measure.

that theirs is a good measure. However, we do not find it compelling for more general application because, as Figure 1 shows, a large number of patents have very little value. If one examiner is tougher than another, it is presumably because they are disallowing more of these low-value patents. This may cause large differences in the number of patents that are allowed, but it will reflect only a very small share of the value being protected by the patent system, certainly less than 5% and perhaps more like 1-2%.[4,5] More important for the operation of the patent system is how patent examiners treat valuable patents. These are patents with little to no danger of not being awarded a patent of some sort, but where differences in scope can vary their value enormously. Our measure captures this margin – and we show that, as might be expected, patent examiner behavior on patents of little value is notably different than their behavior on valuable ones.

Finally, we provide an example of the usage of our *scope toughness* instrument. We examine how exogenous changes in patent scope that arise from random assignment of patent examiners affect the likelihood of patents being declared standards essential. We find that a one standard deviation increase in examiner toughness is associated with a 7.1%*** lower probability of a patent being declared standard-essential.

The purpose of our paper is provide useful tools for patent scholars that are grounded in the details of the legality and practice of patent prosecution and that are properly validated with experts. The following sections lay out these arguments and provide the empirical support for them.

II. Measuring patent scope

A. Previous work

Scholars have long recognized the importance of patent scope. (Chang, 1995; Cohen & Lemley, 2001; Gilbert & Shapiro, 1990; Green & Scotchmer, 1995; Klemperer, 1990; Lerner, 1994; Merges & Nelson, 1990; Scotchmer, 1996). As noted by Merges and Nelson (1990), "The economic significance of a patent depends on its scope: the broader the scope, the larger the number of competing products and processes that will infringe the patent." However, empirical analysis of the effects of patent scope presents significant measurement challenges due to the difficulties involved in accurately quantifying such an amorphous and idiosyncratic construct.

[4] Based on calculations by the authors on the Gambardella et al data.
[5] Sampat and Williams (2015)'s finding of very little effect on downstream outcomes is consistent with this.

An ideal measure of patent scope would accurately measure the extent of the legal right to exclude afforded by a patent. It would also be both validated and easy to use without being a subject matter expert. Further, it would be computable ex ante at the time a patent is issued. Finally, such a measure would facilitate both performing cross-patent comparisons and drawing causal inferences about the effects of patent scope. Scholars have proposed various measures of patent scope, including counting the number of patent citations received by a patent, the number of claims in a patent, or the number of distinct International Patenting Classification (IPC) subclasses to which a patent is assigned. (Lerner, 1994). Our analysis shows that all of these measures are problematic.

The legal basis of patent citations does not support its use as a direct measure patent scope. Patent citations are selected by patent applicants and examiners ostensibly based on the content of the cited patent's description of the invention (the specification) rather than the cited patent's claims (the protection provided by the patent). Therefore, a well-described patent may garner more citations, irrespective of patent scope. Moreover, patent citations present numerous empirical challenges as a measure of claim scope. First, citations are accrued over time and are thus unavailable for computing a measure of patent scope immediately after a patent issues. Second, citations are subject to numerous selection effects specific to firms, technologies, patent families, and time. (Kuhn, Younge, & Marco, 2016; Marco, 2007). Third, patent citations have also been used to measure constructs as diverse as private patent value, knowledge flows, and technological impact, and using them to measure patent scope risks conflating scope with these other constructs. This conflation is particularly problematic for causal inference because it means that any regression of patent citations on commercial outcomes would need to control for all these other effects to get an unbiased estimate. Since many of these factors are co-determined (e.g. market growth leading to more company investment and more competitors citing the patent), such controls will almost always be insufficient.

The number of claims fares no better as a measure of patent scope. A patent's legal right to exclude is defined by its broadest claims, which are referred to as "independent" claims. If a patent includes more than a handful of independent claims, or if the independent claims differ substantially in their scope, the patent examiner will typically force the patent applicant to split the patent application by filing a "divisional application" on the theory that substantially different claims should be subject to distinct examination processes. Most claims in a typical patent are "dependent" claims. These serve as fallback positions in the event that an independent claim is ruled invalid, but by law each dependent claim must be narrower than the independent claim on which it depends. Because applicants pay extra for each claim beyond 20, patents with many claims may cover more valuable technology but are not necessarily any broader than patents with fewer claims.

A more nuanced measure of patent scope proposed in the literature is the number of IPC classes to which a patent is assigned. (Lerner, 1994). The logic behind this measure is that patents that relate to many different technologies are broader than those that are specific to a more limited range of technology. However, a patent may be assigned to many IPC classes if it lies at a technological border, even if it represents only an incremental improvement on prior approaches and thereby affords only a limited legal right to exclude. Further, patent classes are complex, discrete, and vary in granularity across technology. For these and other reasons, the literature includes substantial debate about their use in empirical applications. (Younge & Kuhn, 2015).

In sum, none of these measures accurately captures the scope of the legal right to exclude afforded by a patent's claims. Each of the proposed measures is also subject to selection effects that render the measures unsuitable to drawing causal inferences. Finally, to the knowledge of the authors, none of the proposed measures have been validated against expert evaluations. Later in the paper, we demonstrate that each of these measures also correlates poorly with our validated measure of patent scope.

B. Perspective of a filer

The legal right to exclude afforded by a patent is defined by its claims. "Each claim is in a very concise single-sentence format." (Yan, 2013). Under the "every element" test, a competing product or process must exhibit every element included in a patent claim in order to infringe that claim. For this reason, "[v]irtually every word in a claim is important." (Yan, 2013).

Patents directed to similar technology can vary considerably in claim scope. According to one practitioner, "Patent ***claims*** can be 'broad'—protecting a whole concept that would be difficult for someone else to duplicate without infringing your patent, or 'narrow'—protecting only a very specific configuration of your invention." (Goldstein). Another practitioner notes that "[a] broader claim is usually more valuable than a narrower claim. In the context of patent claims, 'broader' means 'reciting fewer features'." (Funnell, 2014). This is because infringers must meet every condition of the patent to be infringing—the more conditions, the more room there is to invent around. "Every patent practitioner knows that shorter claims are usually broader than long claims. This can be stated as a 'rule of thumb:' If your claim is longer than your thumb, it is too easy for an infringer to get around it." (Beem, 2015).

Under U.S. patent law, the broadest claim, which is necessarily an independent claim, should be presented first.[6] Subsequent independent claims are thus typically either more narrow or identical in scope. For instance, a first independent claim may cover a method, while a later independent claim may encompass a computer that performs the method. If a subsequent independent claim differs significantly in scope from the first independent claim, the patent examiner typically requires the applicant to pursue those different claims in different patent applications.[7] In addition to these independent claims, a patent may include some number of dependent claims that serve as fallback positions and which, by law, must be narrower than their corresponding independent claims.[8] For these reasons, in this article we focus on the first independent claim as providing the best indication of a patent's scope.[9]

A patent's scope is defined by a back-and-forth process known as patent examination. An applicant files a patent with a full technical description and an initial set of claims. The patent is then sorted into an appropriate examining division at the USPTO based on its technology and assigned to an examiner. Exactly which examiner is assigned to which patent and how this happens is an important question, which we revisit in Section V.C. But once an examiner is assigned, they search the prior art to identify reasons to reject the patent, such as the claims reciting a concept that is either not new or is obvious in view of what is already known.

A patent applicant whose claims are rejected cannot add any new matter to the technical description of the patent application but can amend the claims (typically by narrowing them) to overcome the rejection. A patent examiner can never finally reject an application—only the applicant can decide to stop pursing a patent application. Most patent applications (about 80%) are initially rejected, but most patent applications (again about 80%) ultimately issue as patents. As we will show, the patent examination process causes the claims of most patent applications to be narrowed considerably between filing and issuance. Further, as noted by Cockburn, Kortum, and Stern (2002), Sampat and Williams (2015), and others, examiners vary considerably in their characteristics, which provides variation useful for constructing a causal instrument.

---

[6] 37 C.F.R. 1.75(g).
[7] 35 U.S.C. § 121.
[8] 35 U.S.C. § 112(d).
[9] Alan Marco, Joshua Sarnoff and Charlie deGrazia have pursued an alternate calculation that focuses on the shortest independent claim. We do not adopt this definition in part because we judge that in most situations it will be equivalent but also because it adds considerable complexity to empirical analysis. For example, in this article we compare the scope before and after patent examination. However, because claims can be deleted or re-ordered during examination, the, the shortest claim at filing may be completely different from the shortest claim at issuance, whereas the first claim at issuance is likely to be an amended version of the first claim at filing. Accordingly, focusing on the first claim allows for more straightforward analysis with more interpretable results.

C. Our measure

We measure patent scope using the number of words in the first independent claim of a patent. This measure provides several advantages. From an institutional standpoint, it is grounded in both the legal impact of patent claims and anecdotal evidence from practitioners about how patent claims are drafted. From an empirical standpoint, it is clearly observable both at the time a patent application is filed and when the resulting patent is issued.

Although concurrent, independent work has proposed claim length as a measure of patent scope (see, e.g., (Harhoff, 2016); Marco, Sarnoff, and deGrazia (2016)[10]), to our knowledge it has neither been validated against expert judgments nor employed for causal inference. We will present evidence in Section IV that claim length accurately reflects the opinion of expert reviewers regarding claim scope. Further, as will be discussed in Section V, variation in patent examiner characteristics can be used to support causal inferences regarding the effects of patent scope.

Another advantage of claim length is that it is a characteristic shared by all patents. However, different technologies are described differently. Further, institutional details arising from administrative organizational structure may give rise to further between-technology differences in claim structure. To parcel out these differences, we normalize claim length at the level of the USPTO examining division, known as an "art unit". The resulting measure is standardized in the sense that a patent having a claim scope value of 1 is one standard deviation narrower than other patents in the same area of technology.

III. Data

A. Data Sources

The data for this project comes from four principal sources. First, to observe the changes made to patent claims during their prosecution, we collect and parse the text of both the published U.S. patent application and the issued U.S. patent for each of 947,191 patents. Second, to observe various other characteristics of these patents relevant to our analysis, we link each patent with bibliographic data extracted from metadata XML files published by the USPTO. Third, for our validation exercise, we find seven patent attorneys who generously agreed to volunteer their time, give them 140 randomly selected U.S. patents (in total) to review, and have them complete a questionnaire for each patent they review

[10] The measure noted by Harhoff (2016) provides very similar results as ours, but with a slightly different specification.

reporting on the patent's scope relative to other patents in the same area of technology. Fourth, we rely on firm disambiguation data provided by Hanley (2015) to link each patent with a firm identifier.

We link each patent with claims data from the USPTO patent claims dataset and bibliographic data extracted from bulk data documents published by the USPTO. The bulk data files contain a wealth of information about each patent such as the name of the patent examiner who examined the patent and the dates on which a patent was filed, published, and issued.

We partition these data based on the patent examining division of the USPTO. Patents are divided into eight administrative units, known as Patent Technology Centers, responsible for examining utility patent applications. The technology centers correspond to broad areas of technology, such as "Communications" and "Semiconductors" and help to organize patent examiners into logical units. The technology centers include:

1. Biotechnology and Organic Chemistry
2. Chemical and Materials Engineering
3. Computer Architecture, Software, and Information Security
4. Computer Networks, Multiplex Communication, Video Distribution, and Security
5. Communications
6. Semiconductors, Electrical and Optical Systems and Components
7. Transportation, Construction, Electronic Commerce, Agriculture, National Security and License & Review
8. Mechanical Engineering, Manufacturing, Products

Each technology center includes subdivisions, known as "art units", which focus on a more-granular technology, such as "Internal Combustion Engines" and "Electrical Heating." Together, the eight technology centers include ~450 art units. Each art unit includes between 1 and 83 patent examiners.

Although U.S. law mandates that patent applications are initially owned by their inventors, most patent applications are assigned by the inventors to a firm prior to the issuance of a patent. For each patent that is so assigned prior to issuance we link each patent to a firm identifier via firm disambiguation data provided by Hanley (2015). Firms are commonly listed in assignment documents as a company name in text, but the same company may appear in different ways across different patent documents due to typographical errors and other inconsistencies. The firm disambiguation data resolves these inconsistencies and assigns each firm a distinct identifier.

Finally, in early 2017 we collected data from legal experts by conducting a survey of seven patent attorneys within a law firm specializing in the drafting and examination of patent applications. We selected a random sample of patents stratified on changes to patent scope according to our measure. We then asked each attorney to rate using a Likert scale the scope of each patent as a whole at the time of filing, at the time of issuance, and as a change between filing and issuance.

B. Data Coverage

We begin by selecting all patents issued between January 1, 2005 and December 31, 2012, inclusive, that were filed on or after January 1, 2001. We restrict the sample to patents issued in 2005 or later to take advantage of the improved bibliographic data available from the USPTO for patents issued on or after that year. We also limit the data to patents issued by 2012 in order to reduce rightward truncation bias in certain outcome measures. Virtually every patent issued in this range is included in the data set, with three exceptions.

First, we do not observe claim scope data for patent applications that were not published prior to issuance. Patents whose applications were unpublished include those filed before January 1, 2001. Patent applications filed before that date were not published prior to issuance, and observing the claims of these patents at the time of filing is not possible using our current methodology. However, a change to U.S. patent law caused patents filed in 2001 or later to be published by default 18 months after filing. The claims in each published patent match those submitted with the application prior to beginning patent examination. Patents whose applications were unpublished include those where the patent applicant opted out of pre-grant publication. U.S. law allows patent applicants to make this choice provided that the applicant pays a fee and agrees not to file the patent application in foreign jurisdictions. In practice, about 10% of issued patents are not published prior to issuance; we have nothing to say about these patents because we do not observe their claims at filing.

Second, we exclude from our sample each patent that claims priority to an earlier-filed patent application. Each of these "continuing" patent applications is part of an ongoing interaction between the patent applicant and the patent office. A continuing patent application is normally assigned to the same patent examiner who examined its parent. At the same time, the patent applicant can use information gleaned from the examination of the parent application to formulate the claims of the continuing application. Thus, our assumption that patent applicants are quasi-randomly assigned to patent examiners would not hold for this set of patent applications.

Third, we exclude from our main sample patents examined by the biotechnology technology center because of the way language is used in their claims. Patents examined by art units within the USPTO's biotechnology center are the most likely to use Markush language, where lists are used to make a patent broader, e.g. “a compound consisting of drug A and a drug selected from the group consisting of: drug B, drug C, and drug D”.[11] Amending a Markush claim to eliminate one of these alternatives will narrow and shorten the claim, reversing the normal relationship between claim length and scope, as discussed in Section 2.1 above. Moreover, biotechnology patent claims use sequence numbers to incorporate by reference genetic sequences listed in the technical description of the application and other idiosyncratic claim drafting phenomena that can confound measurements of claim length.

The original data set includes all patents for which we observe claim scope data, which virtually all patents issued from 2005 to 2012 that are published prior to issuance. After imposing the restrictions described above, we are left with 1,061,863 patents remaining in the final data set.

C. Summary Statistics

Table 1 provides summary statistics for the words present in the first independent claim. The data shows that patent claims tend to lengthen considerably between filing and issuance. Figure 2 plots the distribution of the number of words in the first claim of each issued patent in the data, both at the time of filing and at the time of issuance. The average patent has 130 words in the first claim at the time of filing and 181 words in the first claim at the time of issuance.

--------Insert Figure 2 and Table 1 about here--------

As discussed in Section III.A, each patent is assigned to one of about 450 art units, which each includes a number of patent examiners who specialize in the area of technology associated with the art unit. For technological and administrative reasons, there is considerable variation across different art units and within an art unit between individual examiners. For example, Figure 3 plots the number of patent examiners within art units in each USPTO Technology Center in our sample. Although most art units include fewer than 20 patent examiners, some art units include more than 60. Figure 4 plots the examiner-level average number of words added to the first independent claim for the 10 largest art unit in each Technology Center. Both the mean and the variance of the examiner-level average differ considerably, even within a Technology Center.

--------Insert Figure 3 and Figure 4 about here--------

[11] Preliminary testing in our data suggests that 27% of biotechnology patents employ Markush claim language, while 12% of chemical patents do so. In other technology centers, fewer than 2% of patents include such language in the claims.

IV. Validation

A. Our measure

We validate our measure of patent scope in two ways. First, we directly compare our measure of patent scope with patent attorney evaluations for a randomly selected sample of patents. Second, we evaluate how well our measure of scope predicts private patent value as evidenced by patent renewal rates.

We conducted a survey of seven patent attorneys within a law firm specializing in the drafting and examination of patent applications. Each patent attorney was asked three questions for each of twenty randomly selected patent applications:

1) If this patent had been issued with the claims as written in the patent publication, what is your best estimate of this patent's scope as compared to other issued patents in similar technologies?

2) Based on the claims in the patent that was issued, what is your best estimate of this patent's scope as compared to other issued patents in similar technologies?

3) Compare the first independent claim in the issued patent to the first independent claim in the patent publication. How much narrower / broader is the scope of the first independent claim in the issued patent?

The answer to each question was provided on a 0-10 Likert scale, with 0 (10) labeled as "Very Narrow" ("Very Broad") for the first two questions and "Much Narrower" ("Much Broader") for the third question. The survey instrument is shown in the Appendix, Figure A1.

Figure 5 and Figure 6 present the survey responses graphically. Figure 5 plots for each patent the change in patent scope as reported by the practitioner against our measure of the change in patent scope (normalized number of words added), while Figure 6 plots for each patent the final patent scope as reported by the practitioner (i.e. survey question 2) against our measure of the change in patent scope (i.e. survey question 3). In both plots, the survey responses suggest that our measure of patent scope is consistent with patent attorney evaluations.[12]

[12] Recall that, for the reasons discussed in the Data section, we exclude the Biotechnology Technology Center from this analysis.

--------Insert Figure 5 and Figure 6 about here--------

Table 2 presents the survey results, including six regression specifications correlating our measures with survey responses. Specifications 1, 2 and 3 correlate the number of words in the first claim with rater evaluations of the scope of each issued patent. Each additional 100 words in a patent claim correlates with a 1.3*** point reduction in patent scope according to survey respondents' 0-10 scale (Specification 1). Specification 2 shows that this number is virtually unchanged if we add in fixed effects to control for differences in means between art units. In specification 3 we further control of art-unit differences by controlling for both differences in the mean and variance between art units, which we do by creating an art-unit normalized measure of words added (a Z-score). When normalized this way, a 1 standard deviation increase in the number of words added correlates with a 1.3*** reduction in patent scope, again on the 0-10 scale, according to survey respondents.

Specifications 4, 5, and 6 show that similar effects are observed for changes to the number of words between patent filing and issuance, with a 100-word increase in claim length corresponding to a patent 1.5*** points narrower, and a 1 standard deviation increase in words added corresponding to a patent 1.2*** points narrower according to survey respondents. In addition to all of these effects being highly significant, they also have considerable explanatory power, explaining between 30 and 50% of the variance in patent scope (as judged by our expert evaluators). The survey results thus support the claim that our measures accurately measure both absolute and relative claim scope as perceived by patent attorneys.

--------Insert Table 2 about here--------

Next, we compare our measure of patent scope to private patent value as evidenced by patent holder behavior. A patent holder is more likely to pay to keep in force a patent that is more valuable, and practitioners report that broader patents are more valuable. Thus, if our measure is accurate, then patents having broader patent scope under our measure should be more likely to be renewed. Figure 7 shows that this is indeed the case. Most patents are renewed regardless of scope, but renewal rates are higher for broader patents, particularly for the more expensive second renewal fee.

--------Insert Figure 7 about here-------

We normalize our measure at the level of the art unit to account for any systematic technological and institutional differences in how inventions are claimed. While such a fine-grained approach is the most-compelling in theory, it is nevertheless also more challenging for practitioners to work with. As such, we also explore the extent to which claim length can proxy for patent scope even without such granular normalization.

	Dependent variable:					
	Survey patent scope					
	Final			Change		
	(1)	(2)	(3)	(4)	(5)	(6)
Number of words in first claim (total)	−0.013***	−0.014***				
	(0.002)	(0.002)				
Number of words in first claim (normalized)			−1.304***			
			(0.172)			
Number of words added to first claim (total)				−0.015***	−0.015***	
				(0.001)	(0.001)	
Number of words added to first claim (normalized)						−1.207***
						(0.109)
Constant	7.450***	8.326***	5.026***	4.465***	4.477***	3.808***
	(0.356)	(0.554)	(0.162)	(0.107)	(0.257)	(0.091)
Art unit fixed effects?	*No*	*Yes*	*No*	*No*	*Yes*	*No*
Observations	129	129	129	128	128	128
R^2	0.304	0.357	0.312	0.502	0.538	0.492
F Statistic	55.590***	7.327***	57.659***	127.261***	15.285***	121.909***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 3 uses different measures of final patent scope and changes to patent scope to predict the likelihood that the second renewal fee is paid. As the results show, the length of the first claim is a highly significant predictor of renewal fee payments regardless of whether claim length is normalized across all patents (Specifications 1 and 5), within USPTO technology center (Specifications 2 and 6), within USPTO art unit (Specifications 3 and 7), or within the intersection of USPTO art unit and main class (Specifications 4 and 8).

--------Insert Table 3 about here--------

This suggests the remarkable conclusion that scholars needn't limit their analyses to within art unit or within main class effects, *but can instead compare patents across the whole patent office by simply counting the words in the first independent claim.* It is not at all obvious that this pattern should be true; the words to describe the scope of a car and a solar panel needn't be at all comparable. Nevertheless, the empirical results indicate that it does seem to be broadly true.

The effect strengthens at coarser levels of analysis. This suggest that the "between" variation for technologies is even more correlated than "within" variation. Put another way, technologies that have more words in the first independent claim are also more likely to have higher abandonment rates at renewal and this effect is stronger than the relationship between words and renewal within an art unit.

Our results suggest that scholars should be able to predict renewal rates (and perhaps other commercial outcomes) by just counting the words in the first claim.[13] To facilitate others using this measure effectively, we include Table A2 in the Appendix, which can be used by empiricists to map their patent data to scope percentiles in the overall patent distribution.

B. Other measures

In Section II.B, we discussed reasons why other measures of patent scope previously proposed in the literature present significant problems. Our validation exercise provides the opportunity to test these arguments. Table 4 presents the results of ordinary least squares regression models that use the number of IPC subclasses[14] to which a patent is assigned, the number of claims in a patent, and the number of forward citations received by a patent to predict the patent's scope as reported in the survey.

--------Insert Table 4 about here--------

As shown in Table 4, the number of IPC subclasses to which a patent is assigned is *negatively* correlated with its overall claim scope, although the relationship is not significant. This is the reverse of the relationship typically assumed in the literature (see, e.g., Lerner (1994)), suggesting that IPC is a misleading measure of scope.

The number of claims in a patent is significantly and positively correlated with claim scope, but the magnitude of the coefficient, 0.065, is quite small. This implies that an increase of 15.4 claims is associated with an increase in patent scope of 1 survey point in a 0-10 Likert scale. But the average patent includes only 16.9 claims, and thus very large changes in the numbers of claims would be needed to produce any significant variation in predicted patent scope. Thus, not surprisingly, the number of claims can only explain 7% of the variance in patent scope as judged by our survey of experts. Accordingly, while the number of claims is at least correlated with patent scope, we find it to be uninformative for practical use.

We similarly find that the number of citations to a patent is not a useful measure of that patent's scope. Our survey validation exercise suggests that there is only a very slight (0.004) correlation between the number of citations and the scope of a patent and that this relationship is not statistically significant. To put that coefficient in perspective, ~250 citations would be needed to increase the

[13] We do, however, provide the important caveat that while this measure of patent scope is useful for cross-technology comparisons, one cannot build a comparable exogenous examiner-based equivalent (discussed later) for causal inference due to non-randomness in examiner assignment.

[14] The International Patent Classification System has five levels: section, class, subclass, main group, and subgroup. We count the number of unique classifications at the subclass level.

predicted scope by 1 point in the 0-10 Likert scale on the survey. In practice, the modal and median patent have *no* citations within 4 years of issuance, and even the average patent only has 1.4. So variation in citations in *not* a good measure of patent scope.

These findings, plus the theoretical issues discussed in Section II.B, suggest that none of the traditional measures of patent scope are informative for empirical studies of the effect of receiving a broader patent.

V. Causal inference

A. Context

Having introduced a measure of patent scope, there may be a natural tendency to want to run regressions of the form shown in equation (1). Unfortunately, such a regression would not produce unbiased estimates because the quality of the underlying invention is likely to induce commercial outcome changes through multiple channels, not just patent scope. For example, a good invention might lead a company to invest more in the development of the product, which would lead to better commercial outcomes, irrespective of the scope of the patent.

(1) $$CommercialOutcome = \alpha + \beta * PatentScope + \epsilon$$

A causal graph (Figure 8**)**, such as those advocated by Jordan (2003) and Pearl (2009), illustrates why this is likely to be true. As Figure 8 suggests, invention quality is likely to influence commercial outcomes both via patent scope and other factors. These alternate channels create an omitted variable problem that is very hard to control for, and which will lead to biased effect estimates if one were to try to directly analyze the effect of patent scope on commercial outcomes.

--------Insert Figure 8 about here-------

To resolve this endogeneity problem, we introduce an instrument: patent examiner *scope toughness*. In this we apply the approach suggested by Cockburn et al. (2002) and taken by Sampat and Williams (2015) in that we will focus on the role of patent examiners in the determination of a patent's scope. Recall that "often claims are lengthened through the addition of more specific language during prosecution at the request of the patent examiner." (Quinn, 2015). Such lengthened claims will, on average, cause those patents to have less scope. As we will show, some examiners are systematically tougher across all the patents that they face, while others are systematically more lenient. This variation implies that a patent evaluated by such a more stringent examiner is treated differently, not because of the characteristics of the patent or the invention it describes, but because of characteristics of the examiner. Insomuch as examiners are randomly assigned (which we will address below), this variation provides the exogenous shock to patent scope that is needed for a causal analysis.

B. Our measure: Patent Examiner Scope Toughness

We define our measure of patent examiner *Scope Toughness* as the average number of words added to the first claim in an examiner's patents (excluding the patent being evaluated), normalized by the art unit that the examiner is in. This formula is shown mathematically in equation (2), where $ExaminerScopeToughness_p$ is the examiner's toughness on patent p being evaluated by examiner *e,* in art unit *a*. where $mean(WordsAdded)_i$ represents the mean of words added by that examiner (*e*) or by all examiners in that art unit (*a*), and where the subscript ($\neq p$) indicates that the focal patent is excluded from all calculations. This exclusion is important because without it the focal patent would impact the result mechanically, inducing bias. In sum, our measure is a leave-one-out average of the words added by a particular examiner, which is then converted into a Z-score.

$$(2) \qquad ExaminerScopeToughness_p = \left(\frac{mean(WordsAdded)_e - mean(WordsAdded)_a}{stdev(WordsAdded)_a}\right)_{\neq p}$$

The motivation for normalizing within an art unit is that, as shown previously in Figure 4**,** art units exhibit considerable differences in both the mean and variance in words added. This variation probably reflects both substantive reasons (different treatment of the subject matter) and administrative differences (either derived from subject-area differences, or idiosyncratic to the examiners working there).

We find abundant evidence that patent examiner *Scope Toughness* is a strong instrument. Figure 9 and Figure 10 illustrates these in summary form, showing the number of words in the patents, at filing and issuance respectively, by the top 25% toughest patent examiners in an art unit as compared to the bottom (most lenient) 25%. While the number of words in the claims at filings for the two are almost identical, the patents assigned to the tougher examiners have their scope narrowed much more, yielding much longer claims at issuance. This can also be seen in more continuous form in Figure 11.

Table 5 shows that examiner *Scope Toughness* (i.e. word added in an examiner's other patents) is highly predictive of the words *added* to the focal patent during examination (with F-scores greater than 1,900). These results conclusively demonstrate that patent examiner *Scope Toughness* has a strong “first stage” in predicting changes in the words added to patent claims, and thus to the scope of those patents.

--------Insert Figure 9, Figure 11, Figure 11, and Table 5, about here-------

C. Random Assignment and Balance

There is substantial debate about whether the assignment of examiners to patents within the USPTO is random, or as good as random. Ultimately this is a claim about the internal operations of the patent office, which we cannot observe. Nevertheless, those who have studied it in detail, including those with visiting positions there, have concluded that the assignment process is as good as random. (Farre-Mensa, Hegde, & Ljungqvist, 2016; Lemley & Sampat, 2012; Sampat & Williams, 2015). Patent examiners themselves lend support to the view that there isn’t any selection on the applicant side, reporting that it is also nearly impossible for applicants to pick their examiners: (Key, 2015)

Question: “Is there any way for an inventor to influence who examines his or her patent?”

Answer: "No, you have absolutely no influence over who reviews an application. I don't know anyone who has been successful at getting a new examiner assigned to a case."

Nevertheless, recent work by Righi and Simcoe (2017) that evaluates the types of patents assigned to different examiners, find that “examiners specialize in particular technologies, even within relatively homogenous art units,” suggesting that assignment isn’t always random, but rather that particular types of inventions are assigned to specialists at higher-than-random rates. At the same time, they “find no evidence that certain examiners specialize in applications that have greater importance or broader claims.”

We do not take a strong point of view on this issue in general, but rather consider it in our context. If examiners are randomly assigned, then we should find that the patents assigned to tough and lenient examiners are different by no more than chance. If, however, we find that they are different, we need to evaluate by how much – since that will affect the extent of bias that could be introduced.

We test for balance between the type of patents examined by tough and lenient examiners both in means and in distributions. Table 6 shows the mean differences in application between the top 25% toughest and the bottom 25% (most lenient) examiners.[15] We find that, of average, the toughest examiners receive applications with 4 additional words in the claims. Because of our enormous sample

[15] Ideally this comparison would examine all applications by these examiners, but data availability concerns restrict the sample to issued patents.

size, this difference is highly significant in a test of difference of means, but it is also quite small, representing less than a 3% normalized difference between them.[16]

--------Insert Table 6 Error! Reference source not found. **about here--------**

The extreme closeness of these distributions at filing is supportive of the general contention of quasi-random assignment of patents to examiners. But the strong statistical significance of differences in means suggests there may be either rare instances of significant non-randomness or very subtle non-randomness more broadly, in line with the findings of Righi and Simcoe (2017). That said, in practice we would suggest that these balances are close enough to limit the potential bias introduced by the instrument across large groups of patents,[17] but that future work that uses them should be cognizant of, and test, what effect these small imbalances might have on their estimated coefficients, particularly when dealing with small samples.

D. Impact of getting a tougher examiner

Getting one's patent assigned to a tough examiner should have a number of effects. First, that difference should manifest as a more difficult negotiation process between the examiner and the applicant. Table 7 examines this, showing that for each standard deviation in examiner scope toughness negotiations take longer (a.k.a. pendency – 0.29 years***[18]), involve more back and forth communications (0.37 more office actions***), and involve the examiner being more likely to find reasons to reject the patent application because it is not novel (1.5% greater chance of a 102 rejection***) or is just an obvious extension of other work (7.6% greater chance of a 103 rejection***).

--------Insert Table 7 about here--------

More importantly, getting a tough examiner should affect the value of a patent, and thus decrease the applicant's willingness to pay to issue or maintain the protection that it provides. Table 8 shows that indeed this is the case. Patents evaluated by a one standard deviation tougher examiner are abandoned before issuance 5.4%*** more often, after which their owners are 0.3%*** less likely to pay $1,600 to

[16] Intuitively, the normalized difference corresponds to a standard deviation of the pooled data set.
[17] Importantly, this claim is one about covariate balance. If random assignment holds, then the observed covariate balance is likely to also hold for unobservables. If, however, assignment is largely non-random then such assumptions about unobservables cannot be made and our claim must be limited to the bias introduced by the observable factors. In cases where there is some randomness, but it is not complete (as seems likely given the literature), the potential bias will be between these extremes.
[18] This correlation (also shown in
Figure **14**) also suggests that looking at the pendency of an examiner might be another way of constructing an examiner toughness metric. We find that it is an acceptable metric, but has both more theoretical issues and less empirical predictiveness, and thus we focus on words added.

maintain the patent after 3.5 years, and 3.4%*** less likely to pay $3,600 to maintain it after 7.5 years.[19] The net effects on maintenance fee payments are different across different technology classes, as Figure 12 shows. Patents in Computers, Communications and Semiconductors are most likely to be abandoned if they've had a tough examiner.

--------Insert Table 8 and Figure 12 about here--------

It is perhaps more revealing to examine this by the identity of the patentee. Table 9 and Table 10 show this for the first and second renewal fees respectively, with public firms broken into the top and bottom half of assets at patent filing and private firms broken up into whether they have more or less than 50 previous patent applications (since asset data isn't available). These results show that even at the first renewal fee, both large public and large non-public firms are renewing patents less if they have less scope. This is consistent with them having more sophisticated patenting operations, and thus being able to identify a fraction of their patents that they wish to let lapse. Smaller firms and universities show no such abandonment pattern, perhaps reflecting that they either cannot or don't wish to do similarly. In contrast, by the second, more expensive renewal fee, all applicants except universities are disproportionately abandoning patents with less scope, with large firms continuing to do so more aggressively.

--------Insert Table 9 and Table 10 about here--------

One might also expect that broader patents are more likely to be cited by future patents. We find mixed results for this, with even small changes in the specification able to produce large changes in the magnitude or sign of the resulting estimates. A more detailed look at the data clarifies why this should be, since examination by a tough examiner also leads to other conflating effects: decreased likelihood of issuance, increased pendency, and decreased scope, all of which can feed back into changes in citations.[20]

E. Comparison with Sampat and Williams (2015) measure

While our inspiration for this measure came from Sampat and Williams (2015), our measure is different from theirs both mathematically and substantively. Mathematically the most import differences are our usage of words added (instead of abandonment rates) and our normalization at the

[19] Source: USPTO. USPTO Fee schedule. (http://www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule). Our panel is too short to accurately measure effects at the third maintenance fee payment.
[20] Discussed in further detail in the Appendix.

art unit level. Substantively, and more importantly, the two measures characterize patent examiners based on different sets of patents that they are evaluating.

Our measure characterizes patent examiners by how they treat the scope of valuable patents, where the question is *how broad* a patent will be granted, not *whether* one will be. Sampat and Williams' measure characterizes patent examiners by how they treat patents that *might be allowed or might not be*. Since ultimately only an applicant can abandon an application, by definition such patents must have valuations close to zero. Because these two measures consider such different types of patents, one might imagine that examiners treat them differently. Figure 13 shows that this is indeed the case, with examiners that are tough on valuable patents having notably different toughness on low-value ones. As Figure 13 shows, the two measures have some correlation, but there is also lots of variation around them. An area of particular difference is the examiners that are only moderately tough on scope (0-1 standard deviations), but have enormous differences in the share of abandonments that they cause.

--------Insert Figure 13 about here--------

In addition to comparing these measures directly, we can also ask which does better at predicting the scope changes that an examiner will impose on patents they evaluate. Table 11 shows that both measures are highly predictive (<0.1%), but that Scope Toughness predicts ~4x as much of the variance (R^2 of 3.1% vs 0.8%).

--------Insert Table 11 about here--------

It is clear that both our measure and that of Sampat and Williams (2015) have value. Theirs is likely to be most useful where the issue of patent issuance (or not) is the chief contention, for example if the existence of patents is used as a signal by a small firm trying to raise venture capital. In contrast, we contend that our measure is better suited for evaluating effects of the protection offered by patents – in particular the ability to exclude competitors. Most of the value from the patent system comes from a set of highly valuable patents. For these patents, the question isn't typically will they get a patent, but how much coverage will it provide and therefore how effective will it be in generating additional profits for the applicant. In these contexts, our measure's focus on the scope of coverage and on evaluating patent examiners by how they treat valuable patents is particularly useful. The next section presents an example of how our instrument can be applied.[21]

VI. Application: Standards Essential Patents

We demonstrate the power of examiner toughness instrument based on patent scope by looking at standards-essential patents. Technical standards can be essential for facilitating the adoption of new technology and for coordinating to produce subsequent innovations that are compatible. (Merges & Kuhn, 2009; Simcoe, 2016). They can also have substantial effects on competition and follow-on innovation. (Lampe & Moser, 2016). Thus, it is of substantial interest to understand what drives the inclusion of patents into a standard. Here we test whether a patent's scope affects whether it is declared "essential" to a technical standard. Crucially, by using our instrument this test is exogenously varying the scope of protection, but not the innovativeness or contribution from the technology underlying the patent.

To analyze this question we start with a data set, developed by Bekkers, Catalini, Martinelli, and Simcoe (2012), of 14,057 patents that have been declared standards essential by their owners. This data "provides a full overview of all disclosed IPR at standard setting organizations world-wide" and is "[b]ased on the archives of thirteen major SSOs". Most art units at the USPTO lack of patented standards in this data, so we limit our analyses to the subsample of in art units where at least one patent has been declared standards essential.

Using a direct OLS regression of the examiner toughness instrument on inclusion of patents in a standard[22] we find that for every additional standard deviation of toughness of the patent examiner, a patent is 7.1% less likely to be incorporated into a standard. Because so few patents (<.01%) are declared standards essential, we estimate the null distribution with a Fisher's Exact Test (i.e. a permutation test) to build up a distribution of outcomes under the hypothesis that patent scope has no effect on the likelihood of being declared standards essential.[23] Figure 15 plots the results of this

[21] It would be appealing to be able to apply our metric to Sampat and Williams' dataset. They found precise nulls in their estimation of the effect of patent coverage on downstream innovation. It would be interesting to ask if we considered patent examiners based on their treatment of more valuable patents (which many pharmaceutical patents surely are), whether we would get a different result. Unfortunately, the presence of Markush language and less randomness in the assignment of patent examiners (Righi and Simcoe) make this much more difficult. However, an enterprising researcher might find a way to identify these systematically and then validate our measure in the rest of biotechnology, in which case they would be able to check this.

[22] Notice, this is the intent-to-treat estimator, rather than the treatment-on-the-treated estimator. We report this one because we suspect that it will be the most useful to practitioners.

[23] We randomly permute examiner toughness across all patents while holding constant whether a patent is declared standards essential. Then, we regress the outcome (i.e. whether the patent is declared standards essential) on examiner toughness. We repeat this process 10,000 times to construct the distribution of outcomes.

analysis. We find that only 0.52% of the permutations of examiner toughness produce as large an estimated coefficient as the real one, and thus our estimate has a 0.52% significance level and conclude that patent scope plays an important role in whether a patent is essential to the use of a technological standard, *even apart from the technology underlying the patent*.

--------Insert Figure 15 about here-----

VII. Conclusion

This paper addresses a fundamental challenge for patent scholars: evaluating the effects of greater patent coverage. Traditionally such evaluations have been difficult because the scope of protection has been difficult to measure and because, even if scope were measured, omitted variable biases would preclude drawing causal inferences about the effects of patent scope. This paper introduces two new tools for patent scholars: an easy-to use way of measuring patent scope and a way to use this as an econometric instrument to produce unbiased casual estimates.

Our measure of patent scope is the number of words in a patent's first claim, with *longer* claims producing *narrower* scope. This approach is grounded in the practice and law of patenting: a product, composition, or process must exhibit *all* the elements in a claim to infringe it, and therefore adding more elements (and hence words) almost always makes it narrower. We extensively validate this measure, showing that it agrees with (i) qualitative pronouncements by practitioners, (ii) patent evaluations by patent attorneys, and (iii) the behavior of applicants.

Our validation exercise also allows us to evaluate previously proposed measures of patent scope, including: counting the number of patent classes, counting the patent's citations, and counting the number of claims in a patent. We find that they are all either misleading or uninformative. In particular the number of patent classes is *negatively* correlated with scope, exactly contrary to traditional assumptions. The citations to a patent are so weakly (and statistically insignificantly) related scope as to be of no use to patent scholars. Finally, we find that the number of claims in a patent is correlated with patent scope, but that this relationship is quite weak and explains little variation, meaning that it provides virtually no information across the patents that most scholars work with.

We also introduce a new instrument: patent examiner *scope toughness*, which provides an exogenous source of variation in patent scope that can be used for causal inference. We validate this instrument, showing that it is predictive of the behavior of patent applicants, and that it is a strong first stage when used as an instrument. In contrast to previous instruments which focus on abandonment rates on marginal patents, our instrument reflects the behavior of patent examiners on all patents, including the important patents that comprise the vast majority of the value in the patent system. Further, we show that *scope toughness* explains about 4 times more of the variation in changes in patent scope than measures based on abandonment. Finally, we provide an example of the usage of our instrument, showing that a patent that provides exogenously less scope (by being assigned to an examiner of one standard deviation more *scope toughness*) is 7.1% less likely to be declared standards essential.

It is our hope that these new measures, grounded in detailed knowledge of patent law and the functioning of the patent office, and extensively validated, will provide useful tools for innovation scholars. We provide additional tools, including the pre-calculated examiner toughness measures for causal inference, at jeffreymkuhn.com/ and www.neil-t.com/.

References

Beem, R. (2015). More Words, Less Patent Clarity. Retrieved from http://www.beemonpatents.com/2015/11/more-words-less-patent-clarity/

Bekkers, R., Catalini, C., Martinelli, A., & Simcoe, T. (2012). *Intellectual property disclosure in standards development.* Paper presented at the Conference Proceedings of the NBER conference on Standards, Patents & Innovation.

Chang, H. F. (1995). Patent scope, antitrust policy, and cumulative innovation. *The Rand Journal of Economics*, 34-57.

Cockburn, I. M., Kortum, S., & Stern, S. (2002). *Are all patent examiners equal?: The impact of characteristics on patent statistics and litigation outcomes*: National Bureau of Economic Research.

Cohen, J. E., & Lemley, M. A. (2001). Patent scope and innovation in the software industry. *California Law Review*, 1-57.

DiPietrantonio, H. J. a. W., Edmund J. (2011). Getting The Most Value From Your Patent Claims. Retrieved from https://www.ip-watch.org/2011/08/04/getting-the-most-value-from-your-patent-claims/

Farre-Mensa, J., Hegde, D., & Ljungqvist, A. (2016). *The Bright Side of Patents*. Retrieved from

Funnell, C. G. a. S., Christopher C. (2014). What is a Patent Worth? Look to the Claims.

Galasso, A., & Schankerman, M. (2014). Patents and cumulative innovation: Causal evidence from the courts. *The Quarterly Journal of Economics, 130*(1), 317-369.

Gambardella, A., Harhoff, D., & Verspagen, B. (2008). The value of European patents. *European Management Review, 5*(2), 69-84.

Gilbert, R., & Shapiro, C. (1990). Optimal patent length and breadth. *The Rand Journal of Economics*, 106-112.

Goldstein, R. Why Some Patents Are Valuable, While Others Are Worth Very Little. Retrieved from http://www.goldsteinpatentlaw.com/educationcenter/beyondthebasics/patents-valuable-others-worth-little/

Green, J. R., & Scotchmer, S. (1995). On the division of profit in sequential innovation. *The Rand Journal of Economics*, 20-33.

Hanley, D. (2015). *Innovation, Technological Interdependence, and Economic Growth.* Paper presented at the 2015 Meeting Papers.

Harhoff, D. (2016). Patent Quality and Examination in Europe. *The American Economic Review, 106*(5), 193-197.

Jordan, M. I. (2003). An introduction to probabilistic graphical models: preparation.

Key, S. (2015). A Former Patent Examiner Pulls Back the Curtain at the USPTO. Retrieved from https://www.inc.com/stephen-key/a-former-patent-examiner-pulls-back-the-curtain-at-the-uspto.html

Klemperer, P. (1990). How broad should the scope of patent protection be? *The Rand Journal of Economics*, 113-130.

Kuhn, J. M., Younge, K. A., & Marco, A. (2016). Patent Citations: An Examination of the Data Generating Process.

Kulaga, T. Business Strategies for Protecting Intellectual Property. Retrieved from http://www.knoxpatents.com/ipstrategies.html

Lampe, R., & Moser, P. (2016). Patent Pools, Competition, and Innovation—Evidence from 20 US Industries under the New Deal. *Journal of Law, Economics, and Organization, 32*(1), 1-36.

Lemley, M. A., & Sampat, B. (2012). Examiner characteristics and patent office outcomes. *Review of Economics and Statistics, 94*(3), 817-827.

Lerner, J. (1994). The importance of patent scope: an empirical analysis. *The Rand Journal of Economics*, 319-333.

Marco, A. C. (2007). The dynamics of patent citations. *Economics letters, 94*(2), 290-296.

Marco, A. C., Sarnoff, J. D., & deGrazia, C. (2016). Patent Claims and Patent Scope.

Merges, R. P., & Kuhn, J. M. (2009). An Estoppel doctrine for patented standards. *California Law Review, 97*(1), 1-50.

Merges, R. P., & Nelson, R. R. (1990). On the complex economics of patent scope. *Columbia Law Review, 90*(4), 839-916.

Pearl, J. (2009). *Causality*: Cambridge university press.

Quinn, G. (2015). For Patent Owners Patent Quality is all about Value. Retrieved from http://www.ipwatchdog.com/2015/06/24/patent-owners-patent-quality-value/id=59010/

Righi, C., & Simcoe, T. (2017). *Patent Examiner Specialization*.

Sampat, B., & Williams, H. L. (2015). *How do patents affect follow-on innovation? Evidence from the human genome*. National Bureau of Economic Research.

Scotchmer, S. (1996). Protecting early innovators: should second-generation products be patentable? *The Rand Journal of Economics*, 322-331.

Simcoe, T. (2016). *Can Standard Setting Organizations Address Patent Hold-up? Comments for the Federal Trade Commission*.

Yan, H. (2013). A Dynamic Framework for Patent Claim Construction: Insights from a Philosophical Hermeneutic Study. *Tex. Intell. Prop. LJ, 21*, 1.

Younge, K. A., & Kuhn, J. M. (2015). Patent-to-Patent Similarity: A Vector Space Model. *Available at SSRN*.

Tables and Figures

Table 1. Summary statistics for words present in the first independent claim.

Statistic	N	Mean	St. Dev.
Number of words at filing	1,061,863	130	101
Number of words at issuance	1,061,863	181	99
Increase between filing and issuance	1,061,863	51	94
Normalized increase in words	1,061,863	0	1

Table 2. Regression model correlating measures of patent scope with survey rater responses.

	Dependent variable:					
	Survey patent scope					
	Final			Change		
	(1)	(2)	(3)	(4)	(5)	(6)
Number of words in first claim (total)	-0.013^{***}	-0.014^{***}				
	(0.002)	(0.002)				
Number of words in first claim (normalized)			-1.304^{***}			
			(0.172)			
Number of words added to first claim (total)				-0.015^{***}	-0.015^{***}	
				(0.001)	(0.001)	
Number of words added to first claim (normalized)						-1.207^{***}
						(0.109)
Constant	7.450^{***}	8.326^{***}	5.026^{***}	4.465^{***}	4.477^{***}	3.808^{***}
	(0.356)	(0.554)	(0.162)	(0.107)	(0.257)	(0.091)
Art unit fixed effects?	*No*	*Yes*	*No*	*No*	*Yes*	*No*
Observations	129	129	129	128	128	128
R^2	0.304	0.357	0.312	0.502	0.538	0.492
F Statistic	55.590^{***}	7.327^{***}	57.659^{***}	127.261^{***}	15.285^{***}	121.909^{***}
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 3. Broader patents are more valuable.

	Dependent variable:							
	Second maintenance fee payment							
	Normalized words in first claim at issuance				Normalized words added to first claim			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All	-0.026^{***}				-0.006^{***}			
	(0.001)				(0.001)			
Tech Center		-0.026^{***}				-0.006^{***}		
		(0.001)				(0.001)		
Art Unit			-0.024^{***}				-0.006^{***}	
			(0.001)				(0.001)	
Art Unit x Class				-0.022^{***}				-0.007^{***}
				(0.001)				(0.001)
All	0.668^{***}	0.669^{***}	0.670^{***}	0.670^{***}	0.671^{***}	0.671^{***}	0.671^{***}	0.671^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	430,299	430,299	430,299	427,943	430,299	430,299	430,299	427,876
R^2	0.003	0.003	0.002	0.002	0.0002	0.0002	0.0002	0.0003
F Statistic	$1{,}147.923^{***}$	$1{,}123.008^{***}$	986.554^{***}	856.793^{***}	83.837^{***}	90.753^{***}	99.058^{***}	117.516^{***}
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 4. Predicting survey narrowing via other measures of patent scope.

	Dependent variable:		
	Survey patent scope (final)		
	(1)	(2)	(3)
Number of IPC classes	−0.129		
	(0.382)		
Number of claims		0.065***	
		(0.021)	
Number of citations			0.004
			(0.016)
Constant	5.183***	4.049***	5.068***
	(0.533)	(0.380)	(0.225)
Observations	107	129	129
R^2	0.001	0.073	0.0004
F Statistic	0.114	10.016***	0.050
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5. The effect of examiner toughness on words added by examiners in different USPTO technology centers and overall.

	Dependent variable:						
	Words added (Z-score)						
	Chemical	Computers	Networking	Communications	Semicond.	Miscellaneous	Mechanical
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scope Toughness (Z-score)	0.153***	0.211***	0.218***	0.227***	0.197***	0.177***	0.169***
	(0.003)	(0.009)	(0.007)	(0.004)	(0.002)	(0.002)	(0.003)
Constant	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001	−0.002
	(0.005)	(0.019)	(0.017)	(0.009)	(0.005)	(0.004)	(0.004)
Observations	131,730	90,421	53,319	160,454	341,428	133,581	150,930
R^2	0.021	0.032	0.036	0.047	0.037	0.029	0.020
F Statistic	2,823.147***	2,957.388***	1,962.915***	7,987.891***	13,006.220***	4,035.708***	3,007.234***
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors clustered by firm						

Table 6. Balance table comparing Student's T-Test results for the claims at filing by patents issued by examiners in the top 25% and bottom 25% of toughness.

Variable	Bottom 25%	Top 25%	T-Stat	Normalized Difference
Words	127.933	131.973	−15.183	0.029
Unique words	54.416	55.252	−15.409	0.030
Characters	782.411	813.661	−22.772	0.044
Verbs	11.116	11.470	−23.010	0.044
Nouns	20.202	20.504	−12.111	0.023
Adjectives	5.469	5.533	−6.720	0.013
Whereins	0.654	0.686	−10.580	0.020
Commas	4.618	4.640	−0.950	0.002
Semicolons	2.587	2.696	−15.914	0.031

Observations: 537,406

Table 7. The relationship between examiner toughness and observable patent examination outcomes.

	Dependent variable:			
	Pendency (years)	Office actions (count)	Has 102 rejection	Has 103 rejection
	(1)	(2)	(3)	(4)
Scope Toughness (Z-score)	0.291***	0.366***	0.015***	0.076***
	(0.002)	(0.003)	(0.001)	(0.001)
Constant	4.334***	2.151***	0.529***	0.712***
	(0.010)	(0.012)	(0.004)	(0.004)
Observations	377,204	332,778	332,778	332,778
R^2	0.209	0.107	0.032	0.047
F Statistic	4,741.522***	1,907.234***	519.119***	782.051***

Note: *p<0.1; **p<0.05; ***p<0.01
Models run with application year x technology center fixed effects.
Sample limited to applications filed 2005-2007 to maximize support.

Table 8. The effect of examiner toughness on the likelihood that a patent application will issue.

	Dependent variable:		
	Issuance	First renewal fee	Second renewal fee
	(1)	(2)	(3)
Scope Toughness (Z-score)	−0.054***	−0.003***	−0.034***
	(0.0003)	(0.0003)	(0.0004)
Constant	0.909**	0.872***	0.652***
	(0.380)	(0.002)	(0.002)
Observations	1,984,920	1,061,747	1,061,747
R^2	0.056	0.001	0.307
F Statistic	2,062.727***	93.357***	36,202.130***

Note: *p<0.1; **p<0.05; ***p<0.01
Application year fixed effects.

Table 9. Change in willingness to pay the <u>first</u> maintenance fee to keep their patent in force (by firm type).

	Dependent variable:				
	Likelihood of paying first renewal fee, conditional on issuance				
	Small non-public	Large non-public	Small public	Large public	University
	(1)	(2)	(3)	(4)	(5)
Scope Toughness (Z-score)	0.001	−0.006**	−0.001	−0.006**	−0.0005
	(0.001)	(0.003)	(0.002)	(0.003)	(0.003)
Constant	0.866***	0.867***	0.939***	0.865***	0.895***
	(0.001)	(0.017)	(0.007)	(0.019)	(0.008)
Observations	192,355	278,980	95,442	112,246	9,912
R^2	0.00001	0.0003	0.00001	0.0003	0.00000
F Statistic	2.422	94.026***	0.894	38.575***	0.021

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered by firm

Table 10. Change in willingness to pay the second maintenance fee to keep their patent in force (by firm type).

	Dependent variable:				
	Likelihood of paying second renewal fee, conditional on issuance				
	Small non-public	Large non-public	Small public	Large public	University
	(1)	(2)	(3)	(4)	(5)
Scope Toughness (Z-score)	−0.004**	−0.014***	−0.007**	−0.018***	−0.001
	(0.002)	(0.004)	(0.003)	(0.005)	(0.009)
Constant	0.673***	0.668***	0.806***	0.646***	0.667***
	(0.003)	(0.020)	(0.017)	(0.033)	(0.016)
Observations	111,755	155,511	55,682	59,446	5,014
R^2	0.0001	0.001	0.0003	0.001	0.00001
F Statistic	7.987***	119.467***	17.104***	72.595***	0.045

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered by firm

Table 11. Words-added as predicted by different measures of examiner toughness.

	Dependent variable:		
	Words Added (z-score)		
	(1)	(2)	(3)
Scope Toughness (Z-score)	0.192***		
	(0.001)		
Pendency Toughness (Z-score)		0.143***	
		(0.001)	
Abandonment Toughness (Z-score)			0.090***
			(0.001)
Constant	−0.001	−0.001	−0.012***
	(0.001)	(0.001)	(0.001)
Observations	1,061,863	1,061,863	937,605
R^2	0.031	0.017	0.008
Adjusted R^2	0.031	0.017	0.008
F Statistic	34,424.450***	18,865.900***	7,967.356***

Note: *p<0.1; **p<0.05; ***p<0.01
Application year fixed effects.

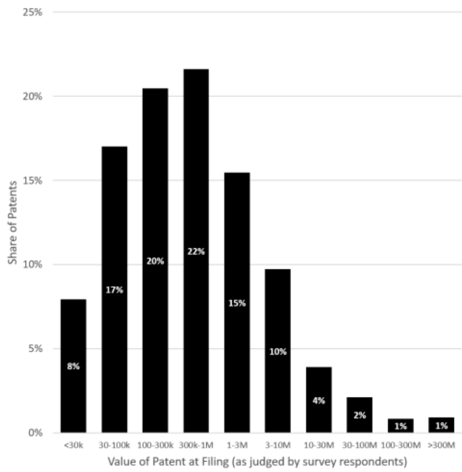

Figure 1: Patent Value at Filing (based on Gambardella et al., 2005).

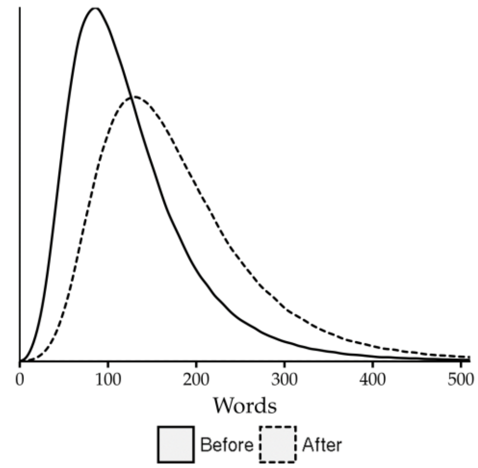

Figure 2. Number of words in first independent claim before and after prosecution.

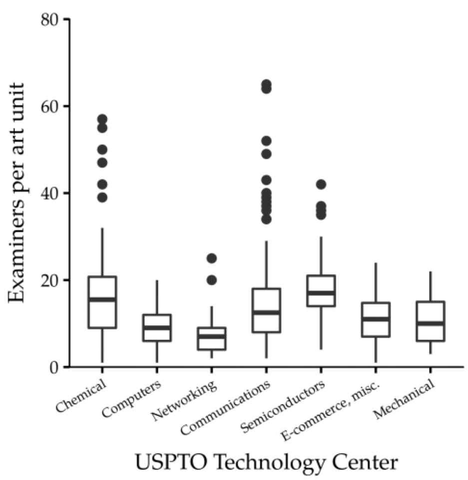

Figure 3. Number of patent examiners per USPTO art unit by Technology Center.

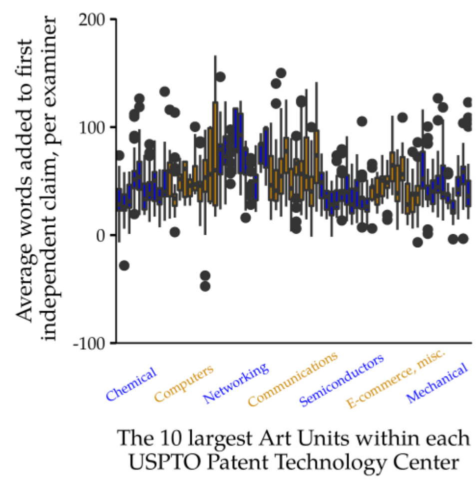

Figure 4. Average words added for examiners in the 10 largest art units within each USPTO Patent Technology Center.

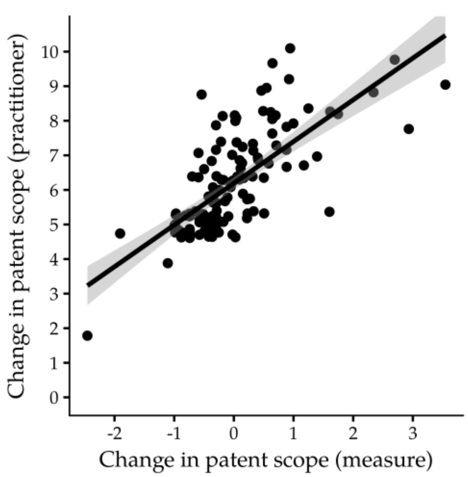

Figure 5. Survey responses as compared with measure for changes in patent scope.

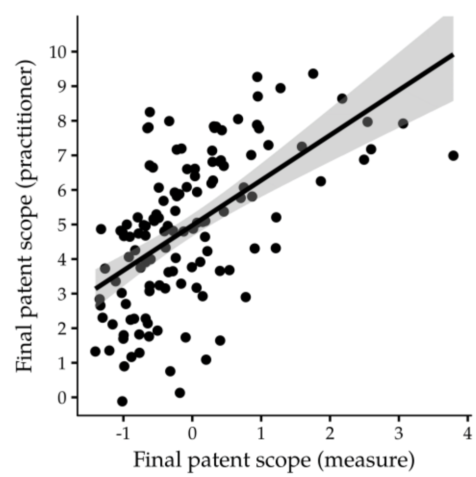

Figure 6. Survey responses as compared with measure for final patent scope.

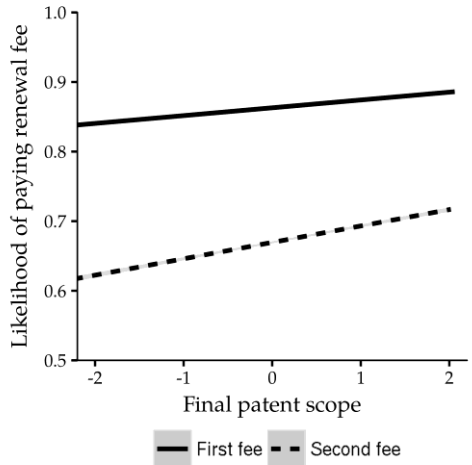

Figure 7. Likelihood of paying renewal fees as a function of patent scope.

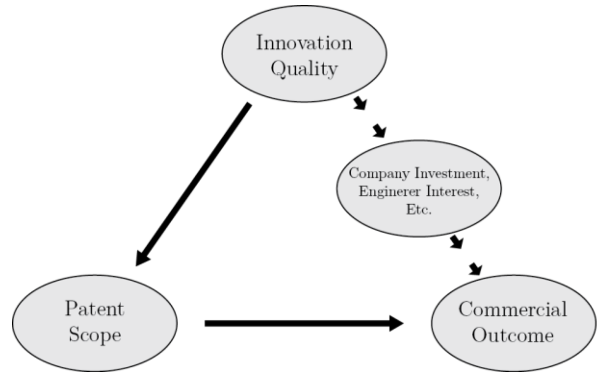

Figure 8: Causal Graph showing the endogeneity problem of inference with Scope without an instrument.

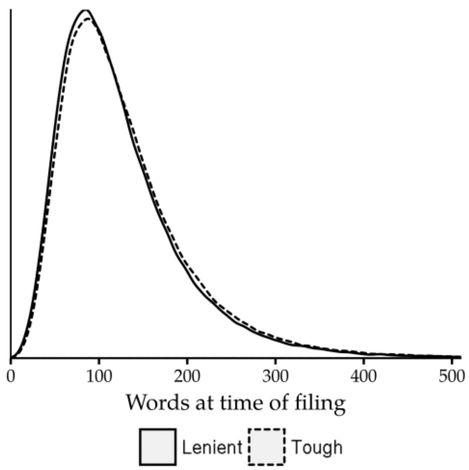

Figure 9. Average words in the first independent claim at filing by toughest and most lenient examiners.

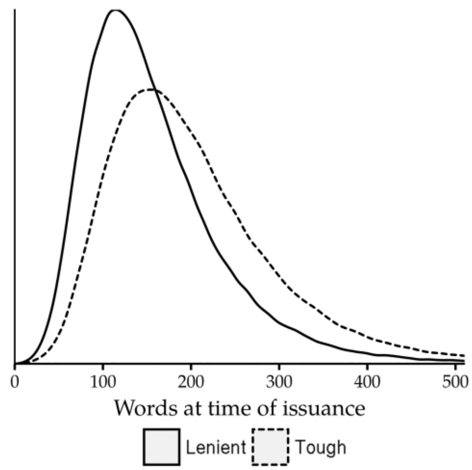

Figure 10: Average words in the first independent claim at issuance by toughest and most lenient examiners

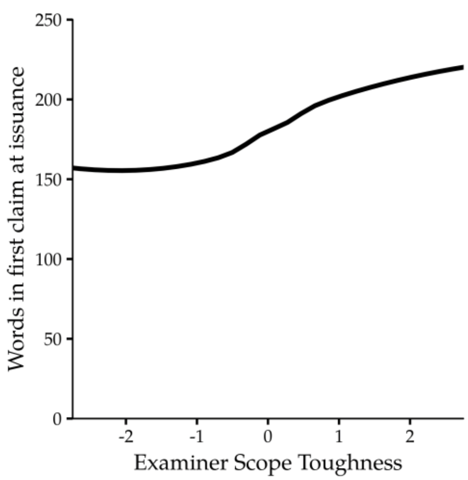

Figure 11. Number of words in the first independent claim (at issuance) as a function of examiner toughness.

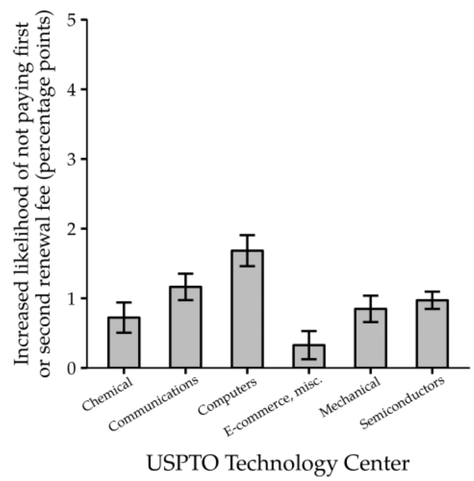

Figure 12. Increased likelihood of letting patent protection lapse (by not paying the first or second renewal fee) from getting a one standard deviation tougher examiner.

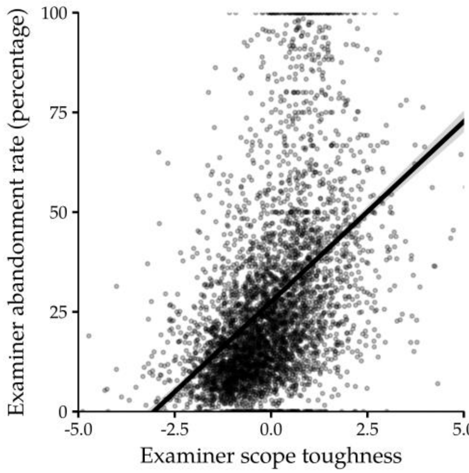

Figure 13. Scatterplot of examiner abandonment rate against examiner *scope toughness*.[24]

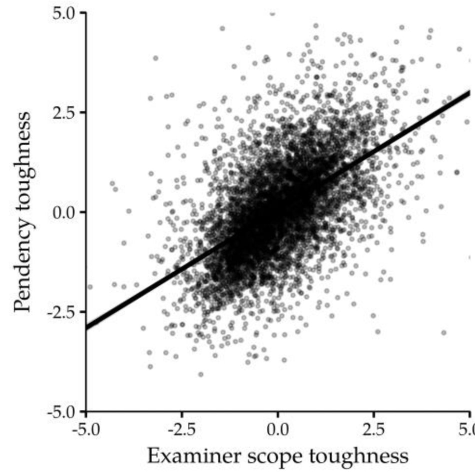

Figure 14. Scatterplot of mean examiner pendency vs. words-added toughness.

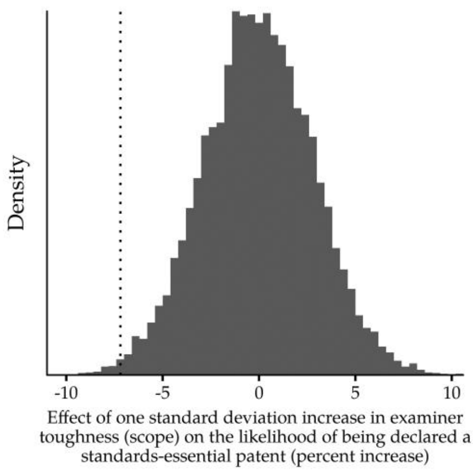

Figure 15. Permutation test of the effect of a tougher examiner on a patent being declared standards essential. Histogram is the null distribution, line at 7.1% is the observed coefficient.

[24] In this scatterplot some examiners appear to have a 100% abandonment rate, but also scope toughness, which would seem to be contradictory because scope toughness is constructed from words added to *issued* patents. This discrepancy is due to only incomplete abandonment data being available, not any error in calculation. Despite the confusion this may cause, we present the data this way since these are values that other empiricists would calculate if they used these methods in practice.

Appendix

One might also imagine that examiner toughness would affect follow-on citations. However, the expected direction and interpretation of such an effect is unclear. On one hand, a broader patent might lead to more citations due to the increased economic significance of the patent. On the other hand, a narrower patent might also lead to more citations due to increased follow-on innovation by competitors. Any interpretation is further confounded by the other effects highlighted above – patents with tougher examiners also issue later, likely impacting commercialization, other R&D activities, etc. Table A1 shows that this, with specification 3 showing that tougher examiners do lead to fewer citations from the *filing* date, but where the mechanism of this effect could be scope, pendency or many other causes. If one instead measures this effect from *issuance*, one gets the opposite effect. Again, the mechanism is unclear. The pendency difference could cause this, but so could certification effects (tougher examiners might be less likely to have their patents overturned later), etc. Overall, we conclude that there isn't a strong relationship between scope and citations, in line with our argumentation in IV.B

Table A1. Effect of Scope on Forward Citations.

	Dependent variable:			
	Forward citations (logged)			
	4 years from issuance		7 years from filing	
	(1)	(2)	(3)	(4)
Examiner Scope Toughness	0.035***		−0.009***	
	(0.001)		(0.001)	
Words at issuance (z-score)		0.013***		0.003***
		(0.001)		(0.001)
Constant	1.431***	1.422***	1.194***	1.197***
	(0.013)	(0.013)	(0.014)	(0.014)
Observations	841,851	841,851	794,364	794,364
R^2	0.049	0.047	0.030	0.030
F Statistic	472.199***	457.856***	274.224***	273.081***

Note: *p<0.1; **p<0.05; ***p<0.01
Application year by technology center fixed effects.

Table A2. Number of words in the claims of issued patents, divided by percentile.

Percentile	Total	Z-score	Percentile	Total	Z-score	Percentile	Total	Z-score	Percentile	Total	Z-score
1%	46	-1.39	26%	118	-0.64	51%	163	-0.18	76%	225	0.45
2%	55	-1.30	27%	119	-0.62	52%	165	-0.16	77%	229	0.49
3%	61	-1.23	28%	121	-0.60	53%	167	-0.14	78%	232	0.53
4%	66	-1.18	29%	123	-0.58	54%	169	-0.12	79%	236	0.57
5%	70	-1.14	30%	125	-0.57	55%	171	-0.10	80%	240	0.61
6%	73	-1.10	31%	126	-0.55	56%	173	-0.08	81%	244	0.65
7%	77	-1.07	32%	128	-0.53	57%	175	-0.06	82%	249	0.70
8%	80	-1.04	33%	130	-0.51	58%	177	-0.04	83%	253	0.74
9%	82	-1.01	34%	132	-0.49	59%	179	-0.01	84%	258	0.79
10%	85	-0.98	35%	133	-0.48	60%	181	0.01	85%	264	0.85
11%	88	-0.95	36%	135	-0.46	61%	184	0.03	86%	269	0.91
12%	90	-0.93	37%	137	-0.44	62%	186	0.05	87%	275	0.97
13%	92	-0.90	38%	139	-0.42	63%	188	0.08	88%	282	1.04
14%	94	-0.88	39%	140	-0.41	64%	191	0.10	89%	289	1.11
15%	97	-0.86	40%	142	-0.39	65%	193	0.13	90%	297	1.19
16%	99	-0.84	41%	144	-0.37	66%	196	0.15	91%	306	1.28
17%	101	-0.82	42%	146	-0.35	67%	198	0.18	92%	315	1.38
18%	103	-0.80	43%	148	-0.33	68%	201	0.21	93%	326	1.50
19%	105	-0.77	44%	149	-0.31	69%	204	0.23	94%	339	1.63
20%	107	-0.75	45%	151	-0.30	70%	207	0.26	95%	355	1.79
21%	108	-0.73	46%	153	-0.28	71%	209	0.29	96%	375	2.00
22%	110	-0.72	47%	155	-0.26	72%	212	0.32	97%	401	2.26
23%	112	-0.70	48%	157	-0.24	73%	215	0.35	98%	439	2.66
24%	114	-0.68	49%	159	-0.22	74%	218	0.39	99%	510	3.39
25%	116	-0.66	50%	161	-0.20	75%	222	0.42			

Figure A1. Patent Claim Scope Evaluation Survey Instrument

Patent Claim Scope Evaluation Survey

Thank you for participating in our patent scope survey. We greatly appreciate your help!
Please answer the following questions to the best of your ability.

Patent Number:

Patent Publication Link:

Issued Patent Link:

Section 1: All Claims

1. If this patent had been issued with the claims as written in the patent publication, what is your best estimate of this patent's scope as compared to other issued patents in similar technologies?

○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10
Very Narrow					Average					Very Broad

2. Based on the claims in the patent that was issued, what is your best estimate of this patent's scope as compared to other issued patents in similar technologies?

○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10
Very Narrow					Average					Very Broad

Section 2: First Independent Claim

1. Compare the first independent claim in the issued patent to the first independent claim in the patent publication. How much narrower / broader is the scope of the first independent claim in the issued patent?

○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10
Much Narrower					About The Same					Much Broader

○

OR **Not Applicable**

(e.g., the first claim in the patent publication is directed to a different invention than the first claim in the issued patent)