NBER WORKING PAPER SERIES

PATENT QUALITY: TOWARDS A SYSTEMATIC FRAMEWORK FOR ANALYSIS AND MEASUREMENT

Kyle W. Higham
Gaétan de Rassenfosse
Adam B. Jaffe

Working Paper 27598
http://www.nber.org/papers/w27598

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2020

Patent Quality: Towards a Systematic Framework for Analysis and Measurement
Kyle W. Higham, Gaétan de Rassenfosse, and Adam B. Jaffe
NBER Working Paper No. 27598
September 2020
JEL No. O31,O34

ABSTRACT

The 'quality' of novel technological innovations is extremely variable, and the ability to measure innovation quality is essential to sensible, evidence-based policy. Patents, an often vital precursor to a commercialised innovation, share this heterogeneous quality distribution. A pertinent question then arises: How should we define and measure patent quality? Accepting that different stakeholders have different views of this concept, we take a multi-dimensional view of patent quality in this work. We first test the consistency of popular post-grant outcomes that are often used as patent quality measures. Finding these measures to be generally inconsistent, we then use a raft of patent indicators that are defined at the time of grant to dissect the characteristics associated with different post-grant outcomes. We find broad disagreement in the relative importance of individual characteristics between outcomes and, further, significant variation of the same across technologies within outcomes. We conclude that measurement of patent quality is highly sensitive to both stakeholder viewpoint and technology type. Our findings bear implications for scholarly research using patent data as well as for policy discussions about patent quality.

Kyle W. Higham
Institute of Innovation Research,
Hitotsubashi University
2-1 Naka, Kunitachi
Tokyo 186-8603
Japan
higham@iir.hit-u.ac.jp

Gaétan de Rassenfosse
Ecole polytechnique fédérale de Lausanne
College of Management of Technology
Odyssea Station 5
CH-1015 Lausanne
gaetan.derassenfosse@epfl.ch

Adam B. Jaffe
Brandeis University
188 Brookline Avenue
Apartment 26A
Boston, MA 02215
and Queensland University of Technology
and also NBER
adam.jaffe@motu.org.nz

1 Introduction

In order to study the sources and effects of innovation, or to formulate innovation policy, it is necessary to measure the rate of innovation across a variety of social, geographical, and technological dimensions. Measuring the rate of innovation is an inherently challenging task because the magnitude or significance of innovations varies enormously (Scherer and Harhoff, 2000; Fleming, 2007). Hence any effort to quantify the extent or rate of innovation in a given context requires some way to measure the relative significance of different innovations.

Patents are frequently used as an indicator of the rate of invention, which is a crucial precursor to innovation (Roberts, 1988). In the context of patents, the inherent variability of innovation significance rears its head in the form of enormous heterogeneity in patent importance, value or quality. As an early, accessible, and rich data source, availability of patent data has kick-started a long line of fruitful research that has led to the development of a variety of different metrics or indicators of the importance, impact, value or quality of individual patents (see Section 2).

We use the generic term 'quality' to encompass the distinct and diverse concepts of importance, impact, value, or significance that have been associated with different metrics applied to patents. In addition to semantic practicality, one reason for this simplification is that all of the concepts that come under the quality umbrella are so stakeholder-dependent (Khanna, 2019) that these various terms are not specific enough to accurately convey useful information. For example, a patent that is extremely valuable to the owner may actively harm economic activity in the society-at-large. On the other hand, a groundbreaking patent may be quickly built upon or around by competitors, reducing the useful life and value of the patent in the long-term but providing great societal benefits in the form of new technologies or competitive markets. The tension between different stakeholders of an effective patent system often forces some political and philosophical framing before any we make any empirical measurements for a particular application. As such, we refer to quality in the abstract and any specific indicators by name to avoid forcing meaning onto them.

To give ourselves the best opportunity to uncover empirically useful results, we develop a systematic framework for analyzing how different quantitative indicators or metrics relate to the different dimensions of quality and each other. Our goal is *not* to determine the 'best' quality metric; indeed, one of our messages is that because 'quality' is an intrinsically multidimensional concept, there cannot be a single best metric. Instead, we suggest that analyzing the distinct dimensions of quality in this systematic way provides the most meaningful way to understand what can be known about the quality-related characteristics of individual inventions based on patent data. Further, researchers have often associated specific metrics with specific dimensions of quality; e.g., forward citations are interpreted as capturing technological impact. However, these interpretations cannot be derived from first principles and, at times, tread dangerously close to becoming definitional; e.g., a patent with high technological impact *is defined as* one with a large number of forward citations. This vagueness becomes more problematic as one continues to peruse the literature and finds, for example, forward citations being used as a proxy for economic value. It is not incorrect, per se, to use the same metrics as proxies for different patent properties, as long as we acknowledge that there is no single interpretation for them. Accordingly, our approach is to examine the empirical relationships among and between ex post metrics and ex ante characteristics to try to determine the extent to which such interpretations have real empirical content.

We first look directly at the relationships between some commonly used and easily accessible post-grant patent quality indicators, referred to hereafter as *outcomes*. Finding that these outcomes are generally not in agreement as to which patents are 'high-quality,' we examine how patent characteristics established at grant are associated with each outcome. In doing so, we observe significant but (in light of the conflicting results in the empirical literature) expected differences between antecedents of these outcomes. This result adds additional color to extant qualitative explanations of patent quality, as seen through the eyes of different stakeholders (e.g., Guerrini, 2013; Khanna, 2019). Additionally, these results suggest that composite indicators that simply average different quality measures, while useful in some contexts, may hurt attempts to answer specific research questions if the outcomes that are combined reflect opposing views of 'quality.' Our empirical strategy also attempts to capture the true extent of heterogeneity across *technologies*, with regard to the measurement of patent quality. Our empirical results highlight patent characteristics that are particularly sensitive to the technology field. The relationships between science dependence and ex-post outcomes, for example, appear to be quite technology-dependent. A literature review also reveals that science dependence is likely very sensitive to the specific operationalization as well. We do not believe this indicator is unique in these aspects and suggest that extreme caution and comprehensive robustness testing are required before drawing definitive conclusions from such indicators—both in terms of metric construction and field of application.

Further, many patent characteristics, from the properties of the underlying invention to the applicant's drafting strategies, may hurt the perception of quality for some technologies more than others, even when seen through the eyes of the same stakeholder. For example, the impact of a (perhaps intentionally) prolonged prosecution process on the value of a patent may depend on how the assignee intends to use it, which may, in turn, depend on its technological nature. Indicators that are strongly related to the interplay between technology types and industry-specific patent strategy norms, such as grant lag, claim counts, or family size, should be treated with particular care.

Finally, and most importantly, we urge an abundance of caution when attempting to measure patent quality. This conclusion is not a wholly negative one at which to arrive, however. Although we find even the most basic and well-established patent quality outcomes generally do not agree on what 'high-quality' means, and that significant technological heterogeneity exists in the antecedents of these outcomes, we do not conclude that all attempts to measure patent quality are futile. The conclusion is simply that a sensible one-size-fits-all approach does not exist. Research questions should be well-defined, and the appropriate patent quality dimension and technology areas selected (or controlled for) to obtain the best results. Patentable inventions include everything from semiconductors to dog toys, and from pharmaceuticals to turbines—situations where informative results can be obtained after applying quality measures equally across technological space are, predictably, few and far between. In sum, patented inventions are myriad and complex, as are the policies that govern and encourage them, and average effects across the universe of patents may not be as informative as we have given them credit for.

We organize the remainder of this article as follows. Section 2 provides an overview of related work and rationale for the departures from this literature in the current work. Section 3 briefly describes data-set construction; however, due to the varied sources and empirical relevance of these data, specifics are discussed in Sections 4 and 5. In the former of these sections, we test the overlap between four common post-grant outcomes related to patent quality, while the latter relates those outcomes to the patents' intrinsic properties. Section 6 concludes and relates the measurement of patent quality to the overarching goals and social value of patent systems.

2 Measurement of Patent Quality

Despite patent quality being of great interest to policymakers and researchers for decades, surprisingly few works take a comparative view of this concept. Of course, comprehensive articles concerning patent quality do exist (van Pottelsberghe, 2011; van Zeebroeck and van Pottelsberghe, 2011b; Squicciarini et al., 2013), but there are only a small number of thorough analyses of the consistency of the different aspects of patent quality. The most notable of those that do, and the most similar to our analysis here, were provided by van Zeebroeck and van Pottelsberghe (2011a) and van Zeebroeck (2011). These studies test the consistency of various quality indicators using Spearman rank correlations.

However, neither of these works aim to examine the relationships between these metrics beyond their correlations. Indeed, one may expect significant levels of heterogeneity in these relationships across different types of patentable inventions. Further, one can exploit easily-observed or -calculated ex-ante characteristics (those established at or near grant) to understand the antecedents of these heterogeneities. To capture both the general relationships and the heterogeneities in these relationships, we leverage the United States Patent and Trademark Office's (USPTO) docketing of each incoming patent application (and eventual grant) to a single *art unit*. Art units are discussed in more detail in Section 4.3.

The variables we have selected for the current analysis, and the justifications for these choices, can be found in Section 4. There are two conspicuous omissions from our set of patent-level outcomes, both of which were excluded for lack of data. The first is legal outcomes, such as involvement in infringement suits or legal disputes concerning the validity of the patent or one of its claims. Despite concerns about rising rates of litigation (Bessen and Meurer, 2013), these events remain quite rare relative to the universe of patents and are concentrated in specific technology sectors (Lanjouw and Schankerman, 2001; Helmers et al., 2018). These facts rule out the use of these events for the kind of analyses presented here (where patents are grouped by distinct technology types) due to their small sample size. Despite this, it is important to note that litigation data are becoming more accessible with time (Marco et al., 2017; Miller, 2018), and the antecedents of patent disputes continue to be studied in depth by others (e.g., Mann and Underweiser, 2012; Love et al., 2019).

The second is evidence of commercialization, for which it is not easy to obtain data. Patenting without commercialization (or follow-on innovation resulting in commercialization) has a limited societal impact, and so information on the 'real-world' status of patented inventions is extremely valuable from a public policy viewpoint. Current commercialisation evidence based on virtual patent marking websites (e.g., de Rassenfosse, 2018) or surveys (e.g., Giuri et al., 2007; Webster and Jensen, 2011) are simply not sufficient on the fine-grain level required here.

These omissions are symptomatic of a well-known phenomenon in the patent quality literature: most patents are simply not worth very much (Scherer and Harhoff, 2000; Silverberg and Verspagen, 2007). Patents are filed very early in the invention process (Cotropia, 2009), and many do not transition into fully-fledged, functional, and useful products. Of those that do, many are not worth an expensive legal battle to keep in force. Some of the most explicit expressions of patent quality are rare events. However, rare events comprise the tip of an iceberg of inventive activity and represent inventions with sufficient private and technological value to be rendered visible to the wider world. There is a place for the analysis of these events in particular contexts and technological domains, but their inclusion is not compatible with the broad view we aim for in this work.

Apart from these omissions, there are a number of more abstract ways of thinking about patent quality that are very difficult to measure directly (if at all), but remain useful for framing concerns about patent quality or the contribution of patent systems to societal well-being and technological progress. We present two examples. First, the quality of a patent may not be reflective of the quality of the underlying invention that it protects. As such, the degree to which the patent scope overlaps or eclipses the scope of the invention can be a measure of quality in its own right. A valid patent that is much broader than the invention it covers may be considered particularly high quality from several perspectives. Of course, measuring any disparities between the scope of a patent and the invention it covers is only feasible on a case-by-case basis, if at all. Second, the economic 'quality' of a patent may be thought of as its net benefit to society (Guellec and van Pottelsberghe, 2007). This requires a comparison between the benefits realised by the existence of a patent on an invention and the counterfactuals to this scenario: either the invention would not exist without the potential for patent protection or the invention would exist but without a patent. The mutual exclusivity of these scenarios means that simultaneous comparison is impossible for any individual invention, even if societal benefit was well-defined and measurable. However, indirect estimations of the levels of follow-on innovation via such a comparison have been calculated through invalidation events (See, e.g, Galasso and Schankerman, 2015).

3 Data and Methodological Roadmap

The data used in the present work can be categorized by the way we use them. We broadly delineate these categories as data that:

- Define our focal cohort;
- Constitute ex-post outcomes of the patents in the focal cohort;
- Allow us to measure semantic differences between patents in the focal cohort.
- Are used to construct ex-ante patent indicators characterizing the intrinsic properties of patents in the focal cohort;

As far as the focal cohort is concerned, we select the cohort of patents granted by the United States Patent and Trademark Office (USPTO) in the years 2001–2004. There are several reasons for this choice. First, USPTO data are easily accessible and often augmented in interesting ways by researchers working both inside and outside the USPTO, resulting in an extensive literature base upon which to draw information and inspiration. Second, patent indicators, including ours, are dependent on citation data for information about the nature of backward-looking knowledge inputs and forward-looking impact measures. To maximize the information content of our indicators, our cohort begins in the same complete calendar year (2001) that the USPTO started explicitly distinguishing examiner citations and applicant citations. Each of these parties has different reasons for adding citations, and so we believe it is worth separating these citations where possible. Third, we want enough time after the final patent in the cohort for specific outcomes to manifest themselves in the data. Most pertinently, the final maintenance payment (or non-payment) needs to be observed and made available by the USPTO—it occurs twelve years (at the latest) after the grant of the patent. We chose the year 2004 as the final year of the cohort for this reason.

All other categories of data incorporate information from a diverse set of sources, some of which are external to the USPTO. Rather than introducing all these sources simultaneously, we introduce these data at the points where they become empirically relevant. Specifically, ex-post outcomes are described in Section 4.1, semantic data are described in Section 4.4 (and in more detail in Appendix A), and ex-ante indicators are summarized in Table 1 with more detail added in Appendix B.

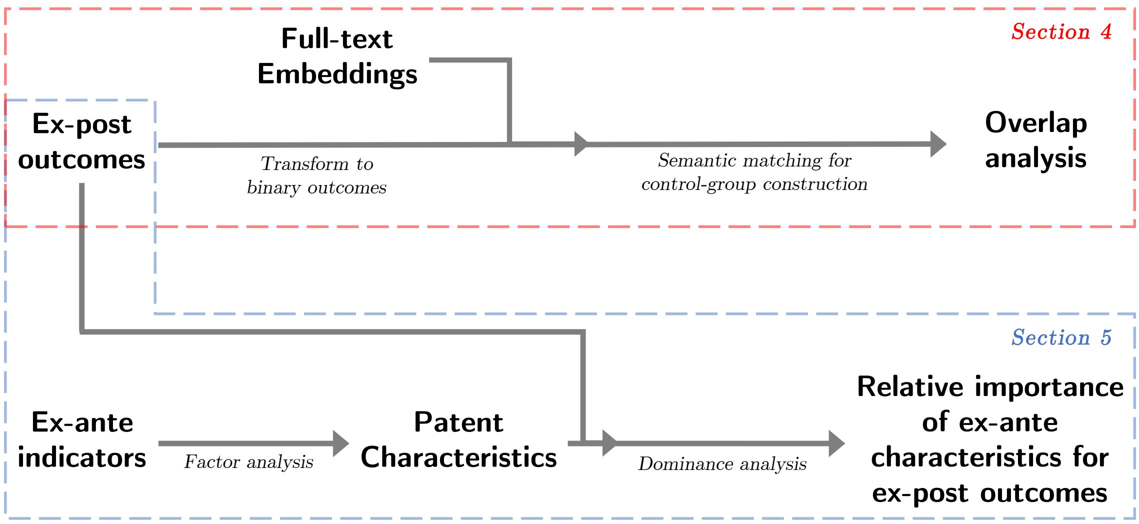

Figure 1: Analytical pipelines for the two empirical sections of this article.

In this work we make use of a wide range of data sources and empirical methods. To make our analytical pipeline explicit, Figure 1 outlines the various inputs, methodologies, and outputs of the next two sections.

4 Consistency of Patent Quality Measures

We start our analysis with a simple test of outcome overlap to examine the consistency of the selected ex-post outcomes. For the current work, we have selected three commonly-used ex-post patent quality outcomes: the anomalous stock-market reaction to grant events (Kogan et al., 2017), renewals, and forward citations (split into examiner and applicant citations). We chose these particular indicators for a number of reasons:

- They are all widely used as indicators of patent quality on the individual-patent level;
- They are available for a significant fraction of all patents;
- They afford simplicity in presentation and interpretation of results;
- If even these most simple, common, and accessible metrics show significant disagreement in their definitions of what constitutes a high-quality patent, then it is highly likely that disagreement will be observed for any pair of quality metrics one could choose.

The next section describes these three outcomes in detail.

4.1 Ex-Post Outcome Variables

4.1.1 Anomalous Stock Market Returns

In their popular paper, Kogan et al. (2017) looked at fluctuations in the stock prices of publicly-traded companies in the days after they were granted a patent. These data were made publicly available, and many have used them as an indicator of patent financial value, hereafter referred to as KPSS values after the authors of the associated paper. While this approach is certainly an interesting way of measuring patent value, it is still unclear what precisely these measured values mean in the real world, how accurate they are for different types of patents, or whether they hold much information for patents not of high ex-ante value.[1] Due to the skewness of the observed value distribution, we take the log of these values to achieve an approximately normally-distributed outcome,[2] which we then use for all analyses involving this variable.

4.1.2 Patent Renewal

To keep a patent in force, patentees must pay renewal fees (referred to as maintenance fees by the USPTO). At the USPTO, these fees are due at 3.5, 7.5, and 11.5 years after patent grant, and will extend the life of the patent to 8, 12, and 17 years respectively. Patents that are not renewed at all will fall into the public domain four years after grant.

Patent renewal is generally considered to be an indicator of patent value to the firm holding the patent (Lanjouw et al., 1998). This notion is sensible, as one would expect a firm to pay to keep a patent in force only when the expected returns to this state exceed this cost, implying the patent has a present value to the firm of at least that of the maintenance fee (Pakes and Schankerman, 1984). In a world where many patents appear to be close to worthless (Scherer and Harhoff, 2000; Silverberg and Verspagen, 2007), renewal information is valuable.

4.1.3 Forward Citations

Forward citations are those citations that are *received* from subsequently granted patents and (in this work) are counted ten years after grant. These citations are made for numerous reasons by inventors, attorneys, and examiners. Generally, they reflect either a disclosure regarding knowledge of prior art (in the case of inventors and attorneys) or justification for the limiting

[1] The authors show a power-law relationship between their patent values and forward citations; however, this trend is non-existent for patents with less than the mean number of citations.

[2] In fact, we observe a slightly bimodal distribution of values after this transformation, which may suggest a lower limit in patent value detectable via this method, below which only noise might be observed.

of claims in the patent application under consideration (in the case of examiners). If we think about patents as defining a technological space from which the patentee has the right to exclude others from entering, citations act to delimit this space (Jaffe and de Rassenfosse, 2017). In this geometric analogy, therefore, a patent that receives many citations likely occupies a newly discovered frontier of this space—attracting many new neighbors that attempt to capture some of this lucrative technological real estate.

In the economic literature, forward citation counts were one of the first invention-level metrics available to measure what might be termed 'technological importance' (Carpenter et al., 1981; Narin et al., 1987; Griliches, 1990). Due to their availability and presumed information content, the use of forward citations as indicators of patent quality has become well-established (See, e.g., Albert et al., 1991; Trajtenberg et al., 1997; Benson and Magee, 2015). Jaffe and de Rassenfosse (2017) offer a comprehensive review of the use of patent citations and their relationship to observable outcomes of interest. With our choice of cohort, we can split these citations into two distinct groups: examiner and applicant citations. Applicant citations are those submitted by the patentee as part of their duty of candor and reflect the applicant's knowledge of existing research. By contrast, examiner citations generally represent IP-limiting prior art cited during the examination process (Cotropia et al., 2013). The mechanisms that generate these two types of citations are, therefore, quite distinct and worth comparing.

To isolate the more informative citations, we remove self-citations from our analysis of forward citations. While forward self-citations may be the most direct way to observe follow-on innovation *within* a firm (Hall et al., 2005; Moser et al., 2018), the number of patents that are applied for in the course of an extensive research program is largely a matter of firm strategy (Hegde et al., 2009). Before omission, self-citations constituted 10% of all forward citations to patents in our cohort occurring within ten years of grant.

4.2 Conversion to Binary Quality Indicators

In order to fairly compare the different data types in which the above outcomes are encoded, we convert all outcomes to binary indicators, where a 1 is consistent with extant notions of high patent quality or value. As the outcomes in question are already known to be very noisy in their ability to accurately measure these notions of high quality (Lanjouw and Schankerman, 2004; van Zeebroeck, 2011), reducing them to binary indicators does not remove as much information as one might expect a priori. Then, we can directly compare these outcomes to check their overlap relative to chance.

We need to set a threshold for each outcome to separate the 'high-quality' patents from the 'low-quality' patents. We set this threshold at the 90th percentile of total applicant forward citations, examiner forward citations, and KPSS values. 'High-quality' as renewal to full term is, in fact, a very low bar, with about half of the patents in our cohort meeting this requirement. Our solution to this problem is presented in Section 4.5. In brief, we focus on non-renewed patents rather than patents renewed to full term.

The overlap between these binary outcomes may differ by technology type, and there may exist technological heterogeneity within predefined classifications. To obtain the fairest and most comprehensive comparison between the outcomes as possible, we take additional measures before any analysis; by first splitting the cohort by art unit, then by implementing semantic matching to control for any remaining intra-unit technological variation (see Section 4.4).

4.3 On the Use of Art Units

We will study the consistency of patent quality measures at the level of art units. Art units are groups of examiners with similar and particular expertise to examine a narrow set of technology types. Each patent is only granted by one art unit, and working at the art unit level avoids the problems associated with multiple classifications such as CPCs or IPCs. Grouping patents by art unit, we obtain a set of non-overlapping blocks that we analyze individually. We can then look at the distribution of outcomes across these art units for any analysis we choose. This method reduces our reliance on average effect sizes and exposes the variation in effect sizes across distinct technology groupings.

Art units have additional advantages as well, such as a more uniform size distribution than other classification systems (in terms of quantity of patents in each), and the fact that the art unit a patent is sent to is determined by the knowledge required to examine the patent. Conducting the analysis at the art unit level allows us to observe a distribution of results across a wide range of different technologies.

Yet. the size distribution of art units is not perfectly uniform by any means, so we implement a minimum size threshold. The most restrictive data we use across all analyses are KPSS values, which, after filtering out patents with undefined ex-ante indicators (see Section 5.1), are only defined for 35% of patents in our cohort. Because we want to use the same set of art units for all analyses, we must set the size threshold at a point that ensures a sufficient sample size for any analyses involving KPSS values. As such, all art units we consider here have at least 200 patents with defined KPSS values. While this selection may bias our set of art units towards those that commonly grant patents to publicly traded companies, it still covers a wide range of technologies. This threshold excluded 46% of all art units existing in the time frame we consider. However, of those excluded, 65% have fewer than 50 patents granted in total (including those without KPSS values), 31% have fewer than ten, and the *patents* in the excluded art units only constitute 5% of the cohort.[3] Of the 249 art units we include, the mean size is 2420 patents, and the median is 2441 patents. Only six art units granted more than 5000 patents of the patents remaining in our cohort, reflecting the natural size limitation of examiner teams at the patent office.

4.4 Semantic Patent Matching

Even after grouping patents by their art units, some within-group technological heterogeneity remains. In particular, art units, by their examiner-based construction, group technologies by the skills required to examine the applications and not necessarily by technological function. Of course, these concepts often overlap, but we do not want to assume that this is the case for every art unit. Therefore, to boost the signal-to-noise ratio, we match every 'high quality' case patent with a semantically similar control patent. This way, for each outcome, we can compare the frequency of another outcome in the high-quality group with its frequency in the lower-quality control group without being as concerned about bias due to differences in technological profile between the groups. Many measures of semantic similarity exist, and many of these have been tested on patent text, from simple Jaccard indices (Arts et al., 2018) to unsupervised machine learning techniques (Magerman et al., 2010). A review of some of these different methods for use with patent data can be found in van Looy and Magerman (2019).

For our purposes, we make use of the patent embeddings constructed by Google,[4] wherein patent full-texts are embedded in a low-dimensional vector-space as described in Appendix A. However, the purpose for which we use natural language processing here is not to extract semantic patterns or interpret the output of such semantic analysis. Rather, we only aim to be able to measure the similarity of any pair of documents. As this does not require an

[3]The two largest art units excluded (which were both about twice the size of the third-largest) were the units that handled exercise devices and animal-related inventions often associated with hobbyists (fishing, beekeeping, horse riding, etc.), where, respectively, only 5% and 7% of patents granted were to publicly traded companies.

[4]`https://tinyurl.com/googlepatentdata`

understanding of exactly *how* documents are similar, we do not have to sacrifice the effectiveness of matching in exchange for the interpretability of the raw output (such as imposed priors or distributional assumptions (Deerwester et al., 1990; Blei et al., 2003)). However, the resultant matching is still manually checked to ensure the calculated pairwise similarity measures roughly align with our intuition.

4.5 Control Group Construction

To form our control groups, we match each case patent with a semantically similar non-case patent.[5] Mechanically, for each patent, we first find and order (from closest to furthest) the 50 most similar patents granted in the same year by the same art unit, as calculated via the cosine distance between the embedding vectors. For each case patent (e.g., a patent in the top 10% most cited by applicants), we run through these similar patents to find the most similar patent that is both non-case (e.g., not in the top 10% most cited by applicants) and does not already fulfill this criterion for a previous patent (i.e., a non-case patent cannot be a control for multiple case patents). This matching results in equal-sized sets of case and control patents for each art unit. So we include only those art units with a sufficiently large set of case patents, a minimum threshold is put in place at 20 case patents for this overlap analysis.

In the case of patent renewals, there exists a problem with this approach: about half of all patents granted by the USPTO in 2001–2004 are kept in force for the full term (17 years). Further, the half that is kept in force may have a very different technological profile than the half that is not, and so a one-to-one match between the groups is not possible. To remedy this issue, therefore, we ask the inverse question: what are the characteristics of patents that are *not* renewed at the first maintenance payment (3.5 years after grant)? As this is a much smaller group, some 13% of patents, we can match *these* patents to those that were kept in force for their full term. Note that while the definition of high-quality remains intuitive (renewed to full term), the matching direction (low to high quality) is the opposite to that of the other patent quality outcomes in this analysis.

Unlike the percentile threshold of the other binary outcomes, we cannot automatically guarantee sufficient sample size for this overlap analysis in the case of renewals and place a threshold

[5]Note that for all patent quality indicators considered here, the matching described will likely bias the case group toward patents in specific technological subgroups—those which are almost always renewed, have universally higher levels of citations, or have a lower ex-ante chance of grant (generating a larger anomalous stock market reaction and thus KPSS value). However, our matching strategy will mitigate the technological differences between the case and control groups that *are* included.

of at least 20 non-renewed patents—some smaller art units may have an extremely high renewal rate, so the number of non-renewed patents available to match could be very low. Likewise, in the other matching direction, if fewer patents are renewed to full term than those that are never renewed, then matching will be incomplete. We remove any art units in this category, alongside any falling below the threshold above. These restrictions (which apply only the 'renewal' outcome) result in the omission of six art units from the overlap comparison with both citation outcomes, and 26 omissions from the overlap comparison to KPSS values.

4.6 Overlap of Patent Quality Measures: Results

We first present a measure of *overlap* between pairs of outcomes relative to chance. The calculation for this overlap is exactly the Pearson correlation coefficient which, in the case of two binary variables (i.e., high-quality dummies for each outcome variable), is equivalent to the Matthews correlation coefficient (Matthews, 1975).

With the current case-control method, the number of case and control patents are equal for a particular outcome. Therefore, while one outcome in the overlap calculation is observed 50% of the time by construction, the other may be relatively rare. The correlation coefficient takes this feature into account (Chicco and Jurman, 2020). Additionally, the case and control patents are different for every outcome,[6] so the overlap matrix for any given art unit is slightly asymmetric. Figure 2 shows the distributions of overlap between a focal outcome (defining case/control patents) and each of the other outcomes.

These results display several interesting patterns. The strongest (and perhaps most expected) level of agreement we see between outcome pairs is between the number of application citations and the number of examiner citations. This agreement is interesting because this implies that there is no strong substitution effect when applicants submit many potential citations in their knowledge disclosure statement. The nature of the technology appears to be driving both sets of citations higher simultaneously—patents with many applicant citations generally have many examiner citations.

In contrast, there is a striking *lack* of agreement between the KPSS values and the other outcomes. It makes sense that renewal is the most closely related to this measure, given that it is also an indicator of financial values. However, our results here stand somewhat at odds to

[6] In the case of overlaps with KPSS value, we remove patents with undefined value (i.e., patents assigned to non-listed companies) before the matching procedure.

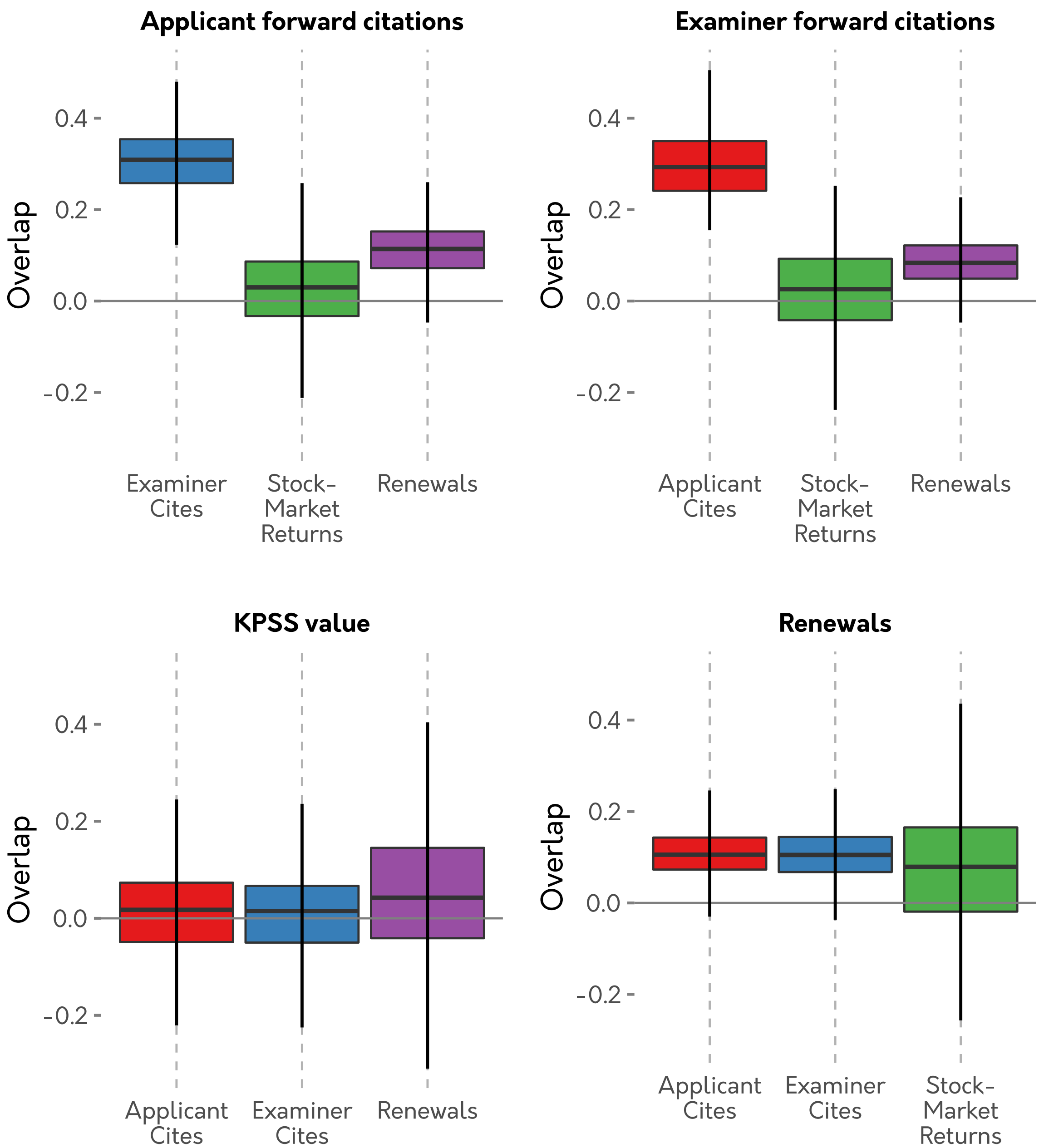

Figure 2: Distributions of overlap of high-quality outcome measures across art units, with a high-quality patent defined as being: *top-left*: in 90th percentile of applicant forward citation count after 10 years, *top-right*: in 90th percentile of examiner forward citation count after 10 years, *bottom-left*: in 90th percentile of KPSS values, and *bottom-right*: renewed to full term (17 years). Positive values indicate the overlap between outcomes within an art unit is greater than chance, negative is less than chance, and zero is that expected by chance alone. Box represents lower and upper quartiles of the results across art units, with median noted inside the box and whiskers extending to 10th and 90th percentiles.

the relationship between citations and the KPSS values found in the work from which we source these data (Kogan et al., 2017). We only look at the top decile (by art unit) in this work for each outcome, which is where we may expect the relationship to be most robust—indeed, Kogan et al. (2017) only report a strong positive relationship between the financial value measure and citations for patents with an above-average number of citations. However, the authors also describe a significant weakening of this relationship as controls for firm size are introduced. This finding is consistent with what we observe here given that the most prolific firms (which is presumably related to their size) patent most top-decile patents.[7] Of course, this does not mean that either citations or KPSS value must be a poor measure of patent quality; this result could point to nearly orthogonal dimensions of patent quality, a much more exciting prospect.

Finally, it is particularly interesting that renewals (i.e., patent lifetimes) appear to contain a small but consistently positive amount of overlap with all other indicators. This relationship may reflect the time patent quality can take to reveal itself; a patent renewed to full term will likely be both financially valuable and technologically impactful. The KPSS value, because it is measured so early in the patent's life, is a very noisy measure of financial expectations, but only when this is paired with revealed technological usefulness is renewal a sensible decision.

This simple analysis shows us the level of disagreement between popular ex-post measures of patent quality. To investigate these relationships further, we look into the intrinsic characteristics of patents reflected in each of these outcomes to gain insight into their possible antecedents.

5 Patent Quality & Intrinsic Character

While technological importance and the financial value of a new invention are generally revealed over time, patents do not change after grant and therefore have some intrinsic characteristics that precede these qualities.[8] Controlling for narrow technology categorizations, we can leverage these intrinsic characteristics to see how they may be associated with different outcomes. To do this, we collect ex-ante indicators that may add information about an invention, in terms of both its technological traits and views on its ex-ante value or importance, that may influence

[7] Among patents granted to publicly traded firms, we find the 1.2% most prolific firms were granted half of all patents in our cohort. However, among patents in the top KPSS value decile, this number drops to 0.6% of firms accounting for half of the top-decile patents.

[8] Post-grant amendments are very rare. They occur as a result of a validity challenge, see, e.g., Chien et al. (2018).

drafting or filing strategies. These indicators constitute the input into a factor analysis to identify intrinsic components of patent character, which are then used as independent variables in a dominance analysis to tease out their relative association with the outcomes presented in Section 4.1. Dominance analysis, explained in greater details below, is a method of measuring relative contributions of independent variables in a regression.

5.1 Ex-Ante Indicator Variables

To capture the intrinsic characteristics of a patent, we rely exclusively on ex-ante indicator variables; that is, those defined at (or very close to) the grant of the patent. There are two primary rationales for this choice. The first is that external factors unrelated to the patent or invention's intrinsic characteristics may influence the perceived quality of a patent over time. By temporally separating these intrinsic characteristics from the post-grant outcomes, we hope to obtain a more 'pure' signal regarding the relationships between these two sets of variables (at the cost of explanatory power). The second rationale is that the ability to relate characteristics of a patent *at grant*, systematically, to ex-post dimensions of quality is valuable information from a technology management perspective. If we were to include ex-post indicators into our analysis, the use of our framework for the characterization of newer patents (for any purpose) would be delayed until the requisite ex-post information became available. Appendix B provides a full descriptions of all indicator variables included in the factor analysis, while Table 1 lists them and defines the abbreviations that we use for the remainder of this document.

We transform all indicators before further analysis in order to reduce the influence of extreme values—we add 1 and take the logarithm such that zero is the minimum value for every indicator.[9,10]

The specific ex-ante variables included in our analysis from this point forward, while numerous relative to the existing literature on patent quality, are necessarily somewhat arbitrary. However, the decision around what to include depends primarily on two factors: data availability and the breadth of patent characteristics captured. That is, providing the variable is measurable and defined for the vast majority of patents, we have endeavored to include a range of ex-ante variables that capture distinct (but not necessarily unrelated) characteristics of a patent or invention. The fact that this set is incomplete and arbitrary by nature was an

[9] Some indicators have 1 as a minimum by default; these have 1 subtracted from their raw values such that all indicators have a theoretical minimum of zero before this transformation. This is indicated in Table 1.

[10] Originality (BOR in Table 1) does not require a log transformation to achieve approximate normality.

Variable	Abbreviation
Backward citations to patents, applicant added	BCA
Backward citations to patents, examiner added	BCE
Backward self-citations to patents, applicant added	BSA
Backward self-citations to patents, examiner added	BSE
Backward citations to foreign patents	BFP
Backward citations to scientific lit. (front page) (Marx & Fuegi, 2019)	BNP
Backward citations to scientific lit. (in-text) (Bryan et al., 2019)	BIT
Originality (Trajtenberg et al., 1997)	BOR
Backward-citations' pedigree (Higham et al., 2019)	BPE
Average age of backward citations	BAG
# independent claims -1	CIN
# dependent claims per independent claim	CDE
# words in first claim (Kuhn & Thompson, 2019)	CFW
# CPC technology class memberships -1	TCM
Grant lag	LAG
Non-final rejections	NFR
Geographic Family size (# countries) -1	GPF

Table 1: List of the variables used in this work as raw measures of intrinsic patent indicators and their abbreviations.

important consideration in our chosen methods, as discussed in Sections 5.2 and 5.3.

Some of the ex-ante variables we use are not defined for a small minority of patents, but we consider them sufficiently meaningful to include. Specifically, we can extract much information about a patent by looking at the patents that it cites. Some constructed variables (such as the average age of cited patents) are simply not defined when no patents are cited. Weighing the information content of these indicators against the number of patents excluded, we decided it was worth the inclusion of these variables. A total of 4.8% of our cohort are excluded from all analyses, with 3.3% due to a lack of backward citations. A further 1.5% of patents are excluded due to patents with exclusively negative-time citation lags (see Appendix B for details).

5.2 Extracting Intrinsic Patent Characteristics

We conduct a factor analysis to construct quantitative measures of different ex-ante patent characteristics that can be uncovered with the selected indicators. In prior work, this method has both been successfully applied to the construction of composite quality indices (Lanjouw and Schankerman, 2004), the interpretation of patent quality indicators (van Zeebroeck, 2011), and models of patent citation dynamics (Higham et al., 2019), the latter building on a simple preferential attachment model as a description of forward citation accrual (Higham et al., 2017a,b). We favored factor analysis over principal component analysis (another popular dimensionality reduction technique) because we do not want to assume that the lower-dimensional space can capture *all* variance among the original set of indicators. After having extracted a lower-dimensional space, we use the relationship between the indicator variables and the extracted factors to assign a position to each patent within this space. We performed this factor analysis on the whole cohort to gain a broad picture of this space of characteristics that is not technology-specific.

The dimensionality reduction is conducted by selecting factors based on a scree test (Cattell, 1966). This test is a graphical method whereby factors are ordered by variance explained, and the factor retention threshold is chosen to be the point at which this quantity decreases dramatically for additional factors. While somewhat subjective, this method is appropriate for our purposes, as our goal is to find naturally occurring and interpretable components of intrinsic (ex-ante) patent character, hereafter referred to as *characteristics*. Later, we use this information as input into the dominance analysis described in Section 5.3 in order to determine the characteristics that are generally associated with each ex-post patent quality measure—our goal is not to predict these outcomes, so we keep the number of retained factors small for the sake of interpretability. After discarding factors as described above, we apply an oblique rotation to the factor loading matrix in order to obtain the human-interpretable results shown in Figure 3. This rotation allows factors to become correlated if this provides a more optimal result (in terms of a subjective, but explicit, notion of interpretability). Appendix C contains a description of this rotation process and the objective function used in the optimization process. Even with an oblique rotation, the factor analysis produces factors that appear to be mostly independent of one another, as evidenced by the mostly small values displayed in the covariance matrix in Table 2. We note that while the factors are mathematically defined in the same way for all patents, the differences between covariance matrices for different technologies could be substantial.

Figure 3 shows the results of the factor analysis, after which we normalize each factor to unit length and variance. Most of these factors correspond to intuitive notions of different patent characteristics, such as science dependence and interdisciplinarity. Our interpretations for all of these factors can be found in Table 3. In particular, we describe what we believe a large factor score on a particular patent means, given the ex-ante indicator variables that contribute the most to each factor. The shorthand for each characteristic will be used to refer to each factor and will be written in italics.

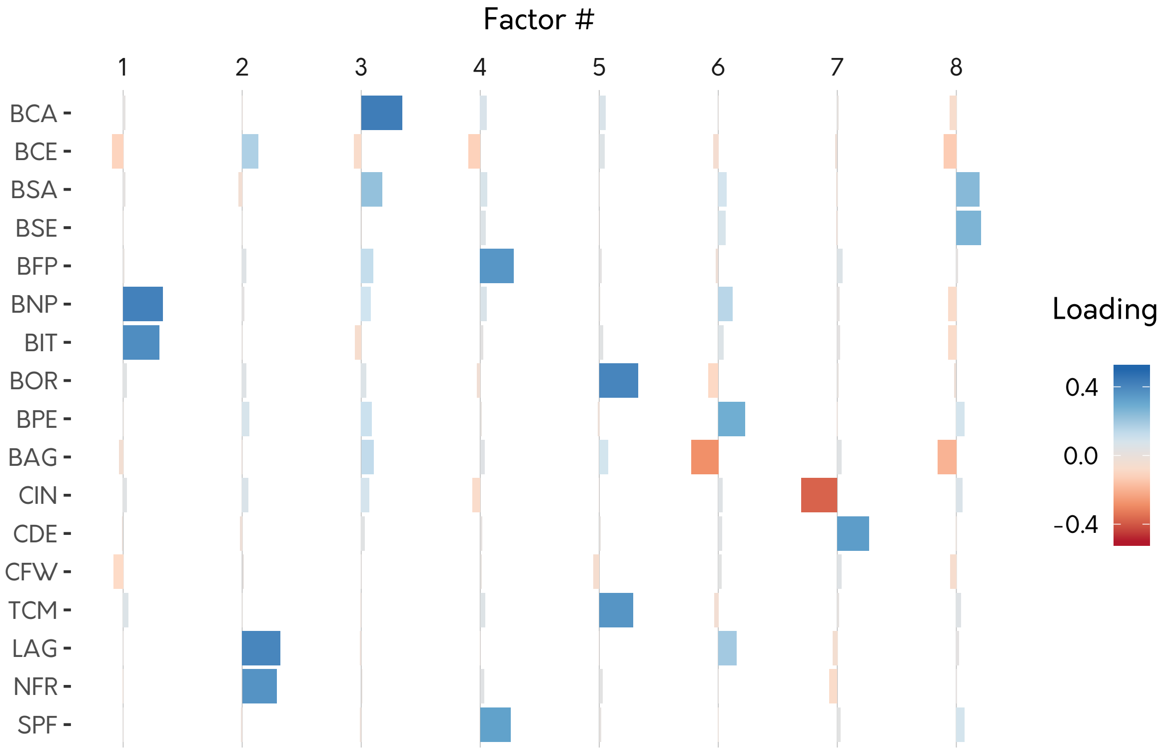

Figure 3: Results of the exploratory factor analysis on ex-ante measurable input variables, after an oblique rotation to increase interpretability. 'Loading' in this context refers to the relative contribution of each indicator to each factor. Factors in order of eigenvalue, an input-variable-dependent quantity that should not be read into.

2	**3**	**4**	**5**	**6**	**7**	**8**	**Factor**
0.10	0.33	0.38	0.25	0.46	0.07	0.07	**1**
	0.02	-0.01	0.14	0.52	-0.44	0.11	**2**
		0.52	0.35	0.21	0.10	0.14	**3**
			0.23	0.17	0.36	0.26	**4**
				-0.21	0.16	-0.13	**5**
					-0.20	0.57	**6**
						-0.20	**7**

Table 2: Covariance matrix of factors after an oblique rotation. All factors are normalised to unit variance and zero mean, so table is also the correlation matrix.

5.3 Dominance Analysis

The goal of dominance analysis is to sensibly decompose the explanatory power of a set of correlated independent variables with respect to some dependent variable. This method is most useful when one is concerned about the *relative explanatory power* of a particular variable relative to others, and not in the *direction* of the relationship between this variable and the outcome variable, which is indeed our intention here. After all, we have deliberately sacrificed explanatory power by excluding ex-post variables from the factor analysis. For this reason, we are only interested in obtaining a comprehensive picture of the parts that *are* explained.

Factor #	Characteristic shorthand	Large loadings	Interpretation of a high factor score
1	Science	BNP, BIT	Many citations to scientific and technical articles, likely a large basic science dependence.
2	Firm Tenacity[a]	LAG, NFR, BCE	Long grant lag and many non-final rejections justified with prior art cited by examiners.
3	Knowledge Synthesis	BCA, BSA, BAG	Many citations made by applicant to established patents, both to the applicants own patents and those of others. Perhaps indicative of crowded technological fields.
4	International	BFP, GPF	Many citations to foreign patents, while the family includes patents granted in many different countries.
5	Interdisciplinary	BOR, TCM	Cites patents from many different 4-digit CPC technology classes while also belonging to many itself.
6	Cutting Edge	BPE, -BAG, LAG, BNP	Cites patents that are new but well-cited compared to similar patents of the same age. Long grant lags and many citations to the scientific or technical literature suggest patent is difficult to examine.
7	Narrowness	-CIN, CDE	Few independent claims, and many dependent claims per independent claim. Could reflect a dense patent space.
8	Ongoing Research	BSE, BSA, -BAG	Many self-citations to new patents from both applicants and examiners, likely part of an ongoing research program.

Table 3: Interpretation of factor analysis results and mapping to preliminary 'characteristics' that may be indicated by a large score on each factor. Large loadings column lists variables with loadings > 0.15, listed in order of absolute magnitude (negative sign indicates negative loading). The first column displays the factor numbers that correspond with those in Figure 3. [a] 'Firm Tenacity' may be attributed in small part to intentional delays by the applicant throughout the prosecution process (Zahringer et al., 2018). However, the evidence for this behaviour is much more limited for applications to the USPTO than for other patent offices (Palangkaraya et al., 2008; Henkel and Jell, 2010; Mejer and de la Potterie, 2011; Zhang et al., 2020). The relationships between the main components of tenacity (LAG & NFR) and private economic value are discussed in Appendices B.3.3 & B.3.5.

We use the most simple version of dominance analysis, as proposed by Budescu (1993) (see also Azen and Budescu, 2003). Once a set of independent variables is constructed (such as the factors extracted in Section 5.2), this method is designed to measure the average contribution of any particular variable to a pre-defined goodness-of-fit measure, given the contributions of the other variables in the same regression against some dependent variable. In practice, we conduct our dominance analysis as follows:

1. Define a quantity to capture explanatory power (usually R^2 values or similar[11]);
2. Select a particular focal independent variable;
3. Construct the set of all subsets of the other independent variables;
4. For each subset, regress on the outcome variable with and without the addition of the focal variable. Calculate the explanatory power gained with the addition of the focal variable;
5. Average these gains across all subsets;
6. Repeat for every independent variable.

In short, paraphrasing Budescu (1993), this method averages, across all subsets, the difference between the fit of the models that include a particular variable and the models that exclude it. After we compute these averages, they are normalized to reflect the approximate percentage of the explanatory power they contribute—we are not interested in goodness of fit per se; we are only interested in understanding the different contributions to the explained variance. For this reason, we do not report specific R^2 or pseudo-R^2 values but note that these are generally in the 0.1–0.4 range and strongly depend on the outcome-art unit pair.[12]

For this method, converting outcomes to binary quality indicators (as done in Section 4) does not simplify the analysis. As such, we select a regression model for each outcome with their natural data type in mind—ordinary least squares (OLS) for (log) KPSS value, Poisson for both forward citation counts,[13] and ordered logit for renewals.

For our purposes, with the independent variables being the characteristics described in Table 3 and the outcome variables being those described in Section 4.1, we conducted this analysis for each art unit individually with grant-year dummies to account for any systematic temporal changes occurring over the four-year period considered (2001–2004) that may affect our selected

[11] All non-OLS models use McFadden's pseudo R^2 (McFadden, 1973).

[12] We additionally note that even regressions resulting in very low R^2 values are not necessarily misspecified, and the model may still possess predictive power. For example, a low-R^2 model that nonetheless has a 50% chance of identifying a patent in the top 20% with respect to some outcome measure clearly has utility. More generally, a model can have high precision (in the classification sense) without a high R^2, if one is willing to accept many false negatives.

[13] A negative binomial model was only computationally feasible using a heuristic version of dominance analysis (Johnson, 2000) and results were qualitatively identical to the Poisson model. However, this method cannot be applied to regressions on categorical data such as ordinal logit models (which we use to model renewal events). We opted to avoid a mixture of dominance analysis methods that could render a comparison between outcomes less convincing. Further, the Poisson count model is robust for the mean under misspecification (whereas the negative binomial model may introduce bias) (Gourieroux et al., 1984). This property is ideal for our empirical strategy as the dominance analysis runs independently for each art unit; using a Poisson model means that we do not have to assume universal distributional properties for all art units. For these reasons, we continue to use a Poisson model for forward citation data.

outcomes.[14] We present the results in the same way as for the overlap analysis in Section 4, as a distribution over art units.

One advantage of dominance analysis, for our purposes, is the lowered sensitivity of this method to the semi-arbitrariness of the independent variables we include. Even if we now introduce an extremely predictive new independent variable into our analysis, many of the subsets on which the dominance analysis is calculated (for any particular focal characteristic) would not include this new variable. In this way, the variables we currently include will still be recognized as having valuable explanatory information, and the new results will have obvious qualitative similarities to the old.

Figures 4 and 5 show the results of the dominance analysis, conducted for each art unit meeting the criteria laid out in Section 4.3. The sign of the relationships between the dependent variable and the characteristics in the full model (i.e., the regression with all characteristics included) is averaged over all art units to indicate the level of agreement in the relationship's direction. The usefulness of these results lies in their descriptions of the kinds of patents that are likely to be deemed high (or low) quality when using a particular outcome measure.[15] In other words, this method depicts the biases that a researcher may wish to anticipate and control for when using patent quality measures for multiple technology fields. For example, KPSS values have clear biases against inventions granted in multiple countries, which may be more common for particular technologies, larger firms, or non-U.S.-based assignees.

We observe, for all outcomes, that there exists a significant amount of heterogeneity in their relationships with characteristics (as defined in Table 3) across art units. This variation is, perhaps, intuitive; we do not expect a high score on a particular characteristic to be consistently associated with a particular outcome for all technologies. As such, we note that the most inconsistent and variable relationships (e.g., between *Science* or *Knowledge Synthesis* and Applicant Forward Citations) could be the most interesting to dissect along technological divisions. We also show the results for three technologically distinct art units (1648, 2122, and 2828, covering inventions related to viruses, artificial intelligence, and lasers respectively), in Figures 6 and 7, against the background of the distribution of results for all art units.

However, the differences between these common patent quality measures are more striking. While examiner and applicant citations show obvious similarities, these outcomes contrast

[14] Note that while the results presented do include these dummies, omitting them does not change the results.

[15] It is worth emphasizing that the relationships observed in this work are not necessarily causal.

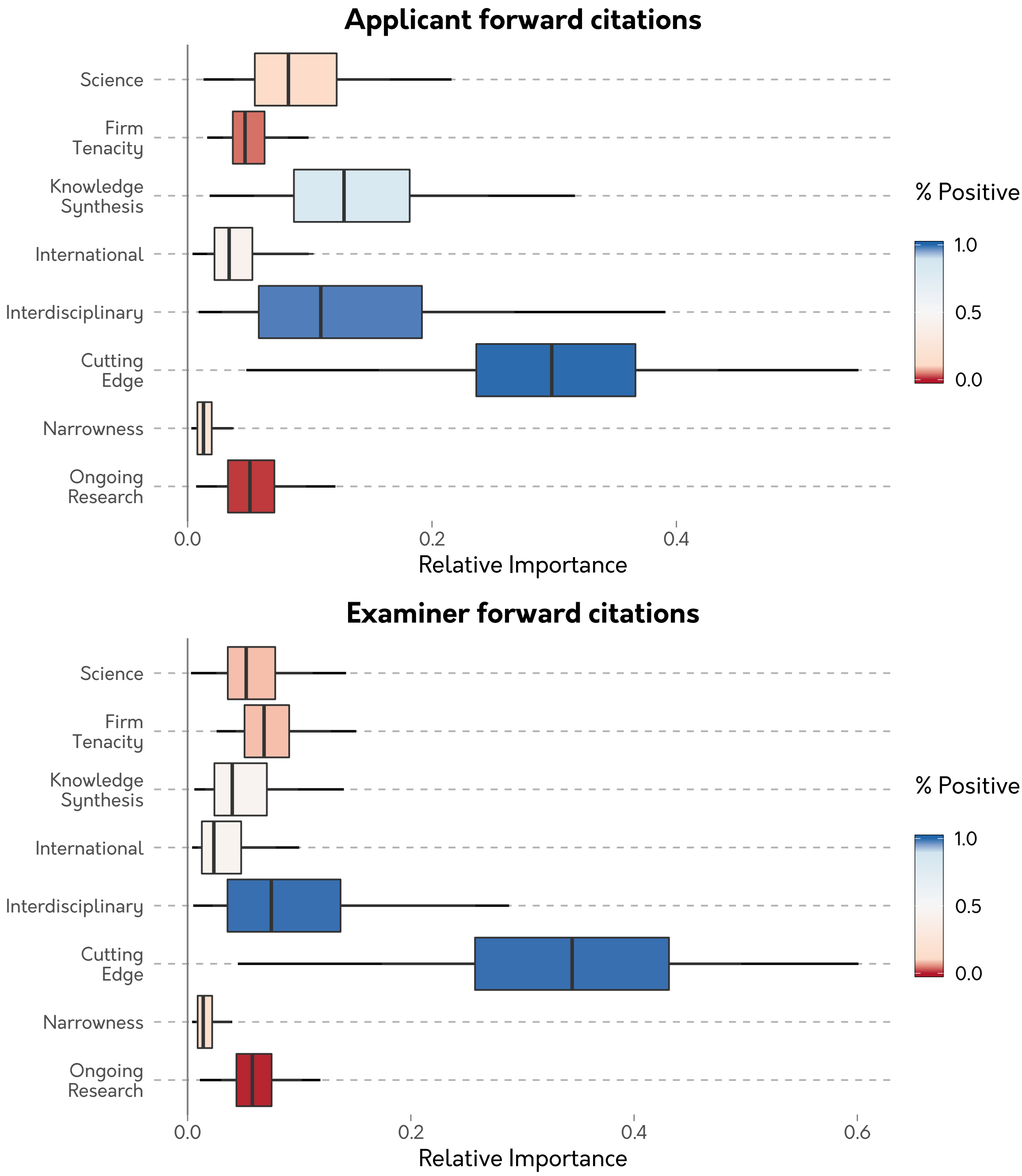

Figure 4: Relative importance calculated via a dominance analysis of patent characteristics on: ***Top***: forward applicant citations after 10 years, and ***Bottom***: examiner citations after 10 years. Box represents lower and upper quartiles of the distribution of results of dominance analysis across art units, with the median noted inside the box and whiskers extending to 10th and 90th percentiles. Color indicates the percentage of art units with positive-signed relationships between each characteristic and the outcome in the full model (separate from the dominance analysis).

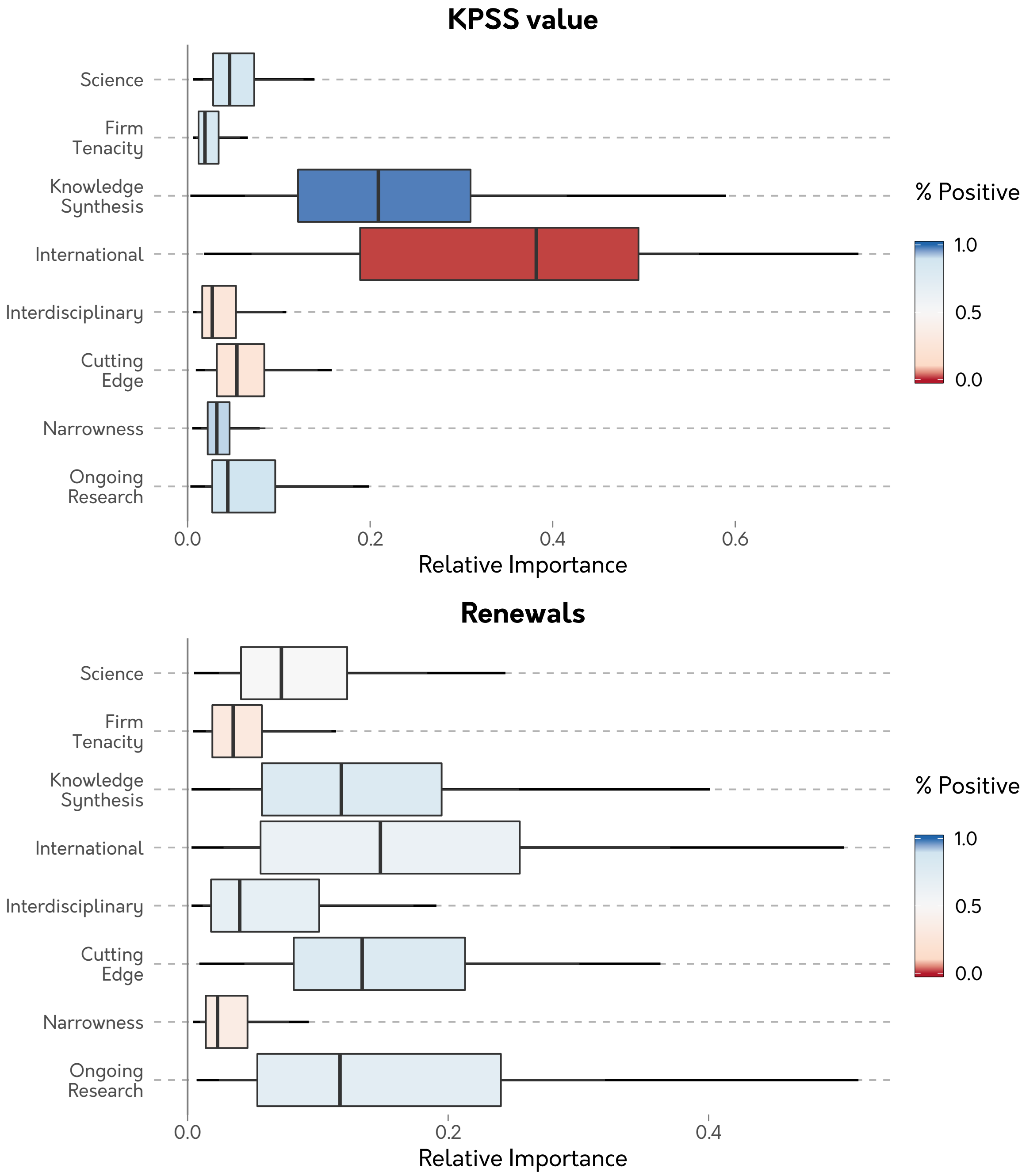

Figure 5: Relative importance calculated via a dominance analysis of patent characteristics on: ***Top***: KPSS value, and ***Bottom***: renewal to full term relative to patents abandoned before a single maintenance payment. Box represents lower and upper quartiles of the distribution of results of dominance analysis across art units, with the median noted inside the box and whiskers extending to 10th and 90th percentiles. Color indicates the percentage of art units with positive-signed relationships between each characteristic and the outcome in the full model (separate from the dominance analysis).

against the results obtained for the more financially-orientated outcomes in Figure 5. The lifetime of patents, as evidenced by renewal data, is the outcome that is least consistently related to ex-ante characteristics. Renewal is a decision made by patent holders, while the other outcomes are the result of forces mostly outside of their control. As such, we may expect that renewal has a stronger relationship to assignee characteristics than technological ones (Bessen, 2008).

What becomes clear through this analysis is the potential pitfalls of composite patent quality measures. The obvious exemplar here is the relationship between *International* and KPSS value, aspects of both of which can be related to the financial value of a patent (Harhoff et al., 2003; Fischer and Leidinger, 2014; Kogan et al., 2017) and, as such, have been used directly as proxies for this value (Yang et al., 2015; Poege et al., 2019; Feng and Jaravel, 2020). Due to the generally negative relationship between these two metrics shown in Figure 5,[16] a composite measure where both were included could be empirically problematic. On the level of technology categories, we can see that in the case of lasers, *Science* is associated with poorer long-term citation outcomes but a longer patent lifetime, while the exact opposite is true for *Interdisciplinarity.* Including both forward citations and patent lifetime in a composite metric would, again, *reduce* the information about the influence of both science-dependence and interdisciplinarity on patent quality in this high-tech sector, while also reinforcing other influences such as that of the 'Cutting Edge' characteristic. Generalizing from these examples, it is easy to demonstrate how composite indicators, while potentially reducing noise, can water-down or eliminate the information content of the constituent metrics for particular use cases.

Concerning specific outcome-characteristic relationships, there are almost no individual results that are surprising. The shear spread in the relationships between characteristics and outcomes across technologies may have been dismissed as simple variance in the past. However, the nature of the spread here is quite different—each data point represents the average difference in explanatory power between 128 pairs of regressions for a single art unit containing about 900 patents (on average). Therefore, while there will undoubtedly be some noise in the output of this process, the spread shown in Figures 4 and 5 is *not* analogous to the standard error that could be calculated by aggregating all art units and regressing an outcome on the

[16]Patent values obtained by anomalous stock market returns are a combination of the perceived value of the patent *and* the level of surprise when the patent is granted. To obtain a financial value for each patent, Kogan et al. (2017) scaled the measured value by the average probability of grant across all patents to account for this 'surprise.' If, however, an equivalent patent is granted first at a foreign office, which will generally have a larger required inventive step for patentablity (de Rassenfosse et al., 2016), the US grant is less surprising. That is, the firm stock price already has this information built-in by the time the US equivalent is granted, and the patent is thus designated a lower value to reflect the lower level of surprise.

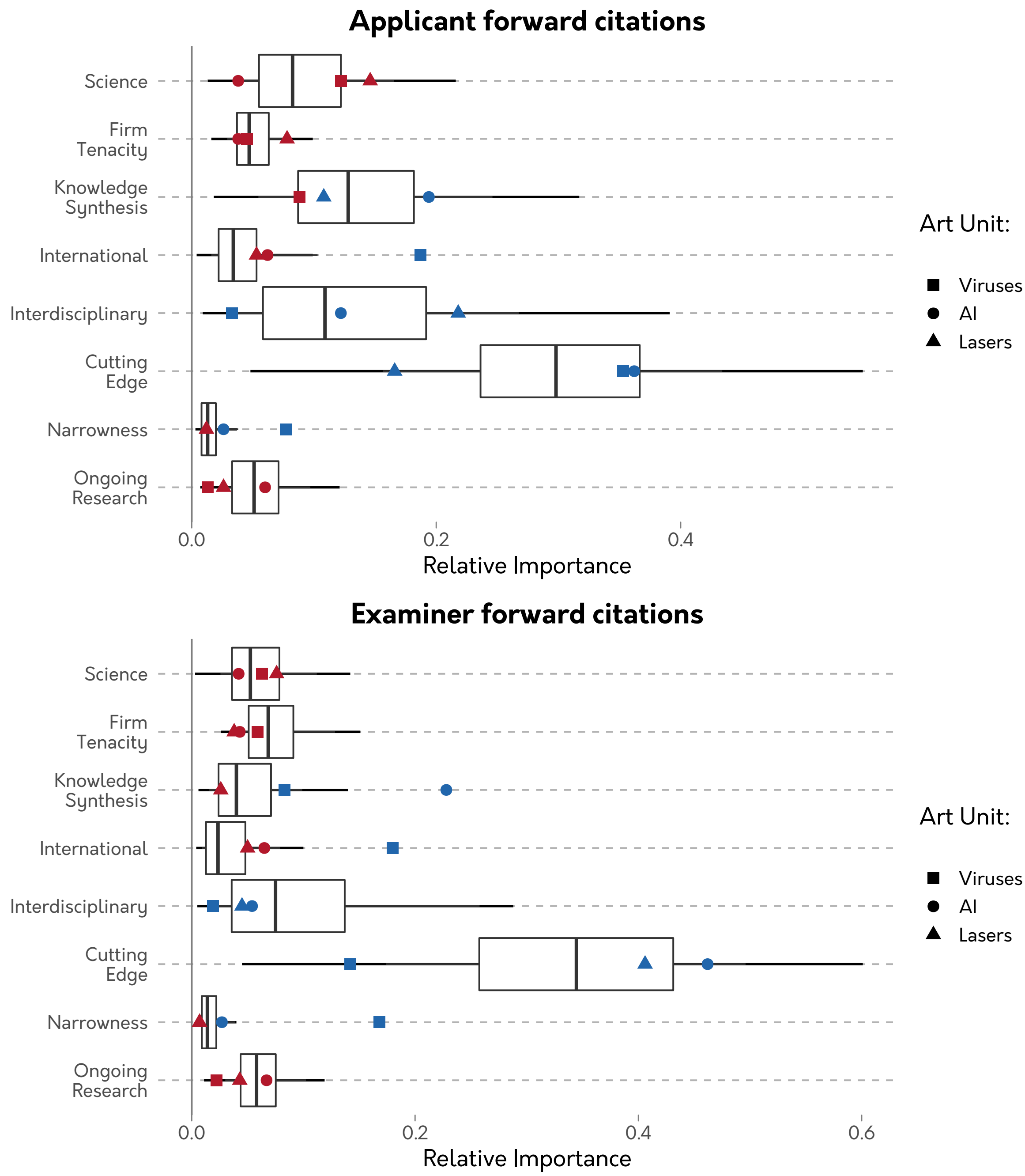

Figure 6: Relative importance of patent characteristics calculated for particular art units with respect to: ***Top***: forward applicant citations after 10 years, and ***Bottom***: examiner citations after 10 years. Color represents sign of the relationship in the full model, and symbols are overlaid on the distribution across all art units where the box represents lower and upper quartiles, with median noted inside the box and whiskers extending to 10th and 90th percentiles.

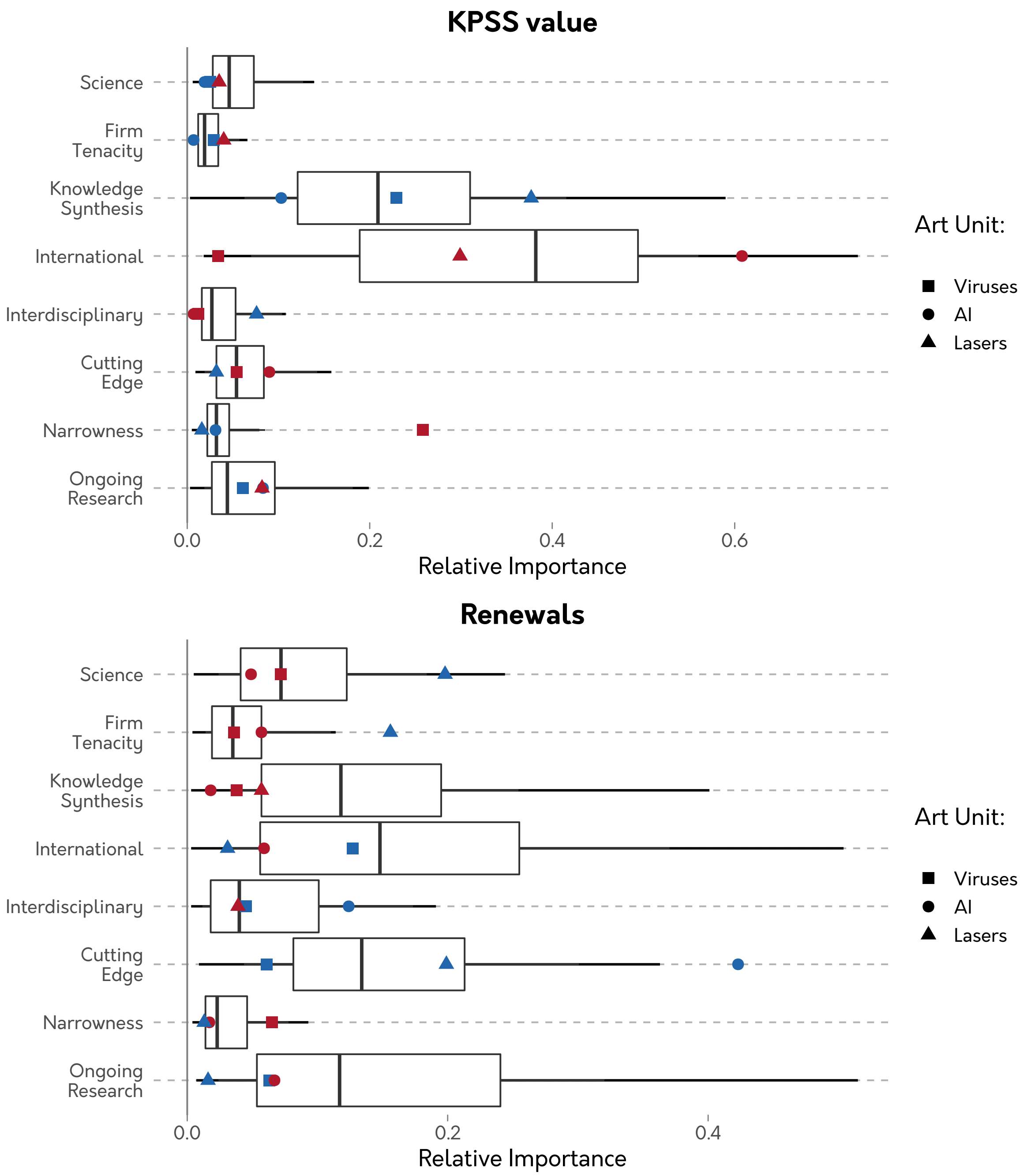

Figure 7: Relative importance of patent characteristics calculated for particular art units with respect to: ***Top***: KPSS value, and ***Bottom***: renewal to full term relative to patents abandoned before a single maintenance payment. Color represents sign of the relationship in the full model, and symbols are overlaid on the distribution across all art units where the box represents lower and upper quartiles, with median noted inside the box and whiskers extending to 10th and 90th percentiles.

characteristics. In addition, the usual presentation of these relationships may be biased towards particular technology types simply because of the skewed distribution of patent volume across these technologies (Lafond and Kim, 2019), which (by construction) is not a concern here.

Relationships between ex-ante indicators and the outcome variables examined here have been explored in the past, though often for a single technology field. To the extent that we can compare this work to previous research, we find broad agreement. Answering questions about specific relationships is not the goal of this work; however, Table 4 presents an overview of some prominent work relating to the most common outcome used, forward citations (applicant and examiner citations combined), that can easily be compared to the current results. The characteristics constructed in this work are (naturally generated) composites of ex-ante indicators, but it is generally straightforward to identify the characteristic in which any particular indicator has the most weight. For example, prior work relating international-ness to forward citations generally uses the simple family size—while not a perfect match to any of our characteristics, this metric is most closely related to the *International* characteristic. While some subjectivity is involved in this exercise, it is a good sanity check for our method. The only characteristics for which we find significant disagreement with past work are *Science* and *Narrowness*. Science dependence is a very technology- and operationalization-dependent quantity (consistent with the moderate spread we see in this variable across technologies). Narrowness is defined somewhat unusually here—it decreases with the number of independent claims and increases with the number of dependent claims *per independent claim*.

6 Discussion and Conclusions

Patent quality exists in the eyes of the beholder. Patent holders, courts, society-at-large, and all other stakeholders in the patent system have different ideas about what makes a good patent. While restricted to ex-post patent quality outcomes available at scale, we provide evidence that these outcomes are generally not in agreement about what makes a 'high-quality' patent and, further, that ex-ante characteristics of patents themselves (both intrinsic and extrinsic to the underlying invention) also differ significantly in their relationship to these outcomes. We also observe these heterogeneities within ex-post outcomes—between different technology groupings as defined by art units.

While not previously presented in the comparative form shown here, these conclusions are

	Sign of Relationship with Forward Citations		
Characterisic	+	∼	−
Science	Fleming and Sorenson (2004); Jung and Lee (2016)	Falk and Train (2017)	Sapsalis et al. (2006)
Firm Tenacity			
Knowledge Synthesis	Fleming and Sorenson (2004); Sapsalis et al. (2006); Sapsalis and van Pottelsberghe (2007); Petruzzelli et al. (2015); Falk and Train (2017)		
International	Sapsalis and van Pottelsberghe (2007)	Sapsalis et al. (2006)	
Interdisciplinary	Lerner (1994); Fleming and Sorenson (2004); Jung and Lee (2016); Mukherjee et al. (2017); Falk and Train (2017)	Petruzzelli et al. (2015)	
Cutting Edge	Fleming and Sorenson (2004); Jung and Lee (2016); Mukherjee et al. (2017)		
Narrowness	Petruzzelli et al. (2015); Jung and Lee (2016); Falk and Train (2017)		
Ongoing Research		Jung and Lee (2016)	Sapsalis et al. (2006); Sapsalis and van Pottelsberghe (2007)

Table 4: A tabular summary of previous work looking at specific relationships between one or more of the ex-ante indicators included in the current work and forward citations in aggregate. Shades indicate the level of agreement of art units as examined in this work, with darker shading representing higher consistency; where results for applicant cites and examiner cites (shown in Figure 4) have been subjectively aggregated.

both intuitive and consistent with previous work on particular aspects of patent quality. Our contribution here is to compare each of these outcomes directly and systematically to explicitly demonstrate both the quantitative and qualitative differences between outcomes and technologies with respect to our ability to measure patent quality. In contrast to prior work on this topic (e.g., van Zeebroeck and van Pottelsberghe, 2011a; Squicciarini et al., 2013), we do not offer a composite indicator of patent quality. In fact, the inconsistency of ex-post patent quality outcomes suggests that each outcome contains information that other outcomes do not. As such, merely averaging these measures across all stakeholder viewpoints may dilute the useful information that can be leveraged to answer specific questions.

This point leads us directly to a discussion about the interaction between patent quality and policy. Patent quality is a complex, multidimensional concept, and policies looking to improve patent quality must consider this complexity. For example, while we do not examine patent validity empirically in this work, a 'high-quality' patent must be enforceable. That is, patent owners need to have some confidence that their patent is valid; otherwise, they will not seek a patent, and all the supposed benefits of the system are moot. At the same time, if policymakers are attempting to put laws in place to reduce the incidence of invalid patents, this will not necessarily be in the public's best interest. An extreme example, as noted by Guerrini (2013), is a law that makes it impossible to challenge a patent's validity. This law would decrease the incidence of patent invalidity to zero immediately—this does not mean patent quality has increased, and, in fact, the effectiveness of the patent system (with respect to its purpose) would likely decrease. Of course, any policy that makes challenging a patent more expensive or unlikely to succeed will decrease invalidity rates, disproportionately hurt small businesses, and incentivize the patenting of less novel and more obvious inventions.

Further, policies that strongly increase quality in one dimension for one set of technologies may decrease quality as measured via another dimension for a different set of technologies. A policy that encourages, for example, broad diffusion of knowledge in the semiconductor industry, could be counterproductive in the biotechnology industry. The results presented here suggest that there is no panacea for improving any specific aspect of patent quality. If we accept this, then very targeted policy changes (including non-patent innovation policies) are likely the most efficient way to address many common criticisms of the patent system (Hemel and Ouellette, 2018), which may, in turn, require system-wide structural changes to practically implement (de Rassenfosse and Higham, 2019).

Our method's primary limitation is the inability to easily incorporate rare event data such as patent litigation outcomes or commercialization of the underlying invention, which could be argued to be more direct or informative indicators of patent quality. This restriction is simply a sample size problem, and broader technology classes or cohort sizes (i.e., more than four years) may provide a remedy that would allow a fair comparison.

However, there is a more general limitation that afflicts all research on patent quality: high-quality patents may not be associated with a high-quality invention, regardless of how we define these concepts. One of the primary goals of the patent system, as a government institution, is to serve (and presumably benefit) society. If the content of a patent does not eventually find its way to the market, then that patent is economic dead-weight, and society would have been better served by a defensive publication, academic article, or other contribution to the intellectual commons. At their core, patent systems are a governmental policy response to the risks associated with the development of new inventions and are accompanied by the assumption that these inventions will confer a net benefit to society. If we were to dichotomize patent quality, therefore, we suggest that a high-quality patent is one that facilitates the production of such inventions where they otherwise would not exist. Measuring quality in this fashion is a challenge, but recent policy changes such as virtual patent marking have made this task more achievable in recent years (de Rassenfosse and Higham, 2020).

Lastly, the unavoidably arbitrary selection of metrics, both ex-post and ex-ante, led us to adopt the flexible and robust empirical methods described above. This allowed us to construct a framework that is generalizable to research of an exploratory nature, particularly in situations where variables may be added or removed throughout the research process. For future work that addresses more specific questions, where there exist theoretical bases for particular relationships between indicators, alternative analysis frameworks such as a structural equation model (Jöreskog and Goldberger, 1975; Pakes and Griliches, 1980) may be more appropriate.

Acknowledgements

The authors would like to thank participants at the Oxford Invention Novelty workshop, the 4th Leuven Summer School on Data and Algorithms for ST&I studies, 11th Hitotsubashi University Institute of Innovation Research Summer School for their excellent feedback, as well as Dominique Guellec, Bronwyn Hall, and Ling Zhou for their insightful comments.

References

M. B. Albert, D. Avery, F. Narin, and P. McAllister. Direct validation of citation counts as indicators of industrially important patents. *Research Policy*, 20(3):251–259, 1991.

S. Arts, B. Cassiman, and J. C. Gomez. Text matching to measure patent similarity. *Strategic Management Journal*, 39(1):62–84, 2018.

R. Azen and D. V. Budescu. The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, 8(2):129, 2003.

C. L. Benson and C. L. Magee. Quantitative determination of technological improvement from patent data. *PLOS ONE*, 10(4), 2015.

J. Bessen. The value of US patents by owner and patent characteristics. *Research Policy*, 37 (5):932–945, 2008.

J. Bessen and M. J. Meurer. The patent litigation explosion. *Loy. U. Chi. LJ*, 45:401, 2013.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

K. A. Bryan, Y. Ozcan, and B. Sampat. In-text patent citations: A user's guide. *Research Policy*, 49(4):103946, 2020.

D. V. Budescu. Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114(3):542, 1993.

M. P. Carpenter, F. Narin, and P. Woolf. Citation rates to technologically important patents. *World Patent Information*, 3(4):160–163, 1981.

J. B. Carroll. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18(1):23–38, 1953.

J. B. Carroll. Biquartimin criterion for rotation to oblique simple structure in factor analysis. *Science*, 1957.

R. B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2): 245–276, 1966.

D. Chicco and G. Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1):6, 2020.

C. Chien, C. Helmers, and A. Spigarelli. Inter partes review and the design of post-grant patent reviews. *Berkeley Tech. LJ*, 33:817, 2018.

A. B. Costello and J. Osborne. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1):7, 2005.

C. A. Cotropia. The folly of early filing in patent law. *Hastings LJ*, 61:65, 2009.

C. A. Cotropia, M. A. Lemley, and B. Sampat. Do applicant patent citations matter? *Research Policy*, 42(4):844–854, 2013.

K. Cremers. Determinants of patent litigation in Germany. Discussion Paper 04-72, Zentrum für Europäische Wirtschaftsforschung, May 2004.

D. Czarnitzki, B. H. Hall, and R. Oriani. The market valuation of knowledge assets in US and European firms. In *The Management of Intellectual Property*, pages 111–131. Edward Elgar Publishing, 2006.

G. de Rassenfosse. Notice failure revisited: Evidence on the use of virtual patent marking. Technical report, National Bureau of Economic Research, 2018.

G. de Rassenfosse and K. Higham. Decentralising the patent system. *Available at SSRN 3446337*, 2019.

G. de Rassenfosse and K. Higham. Wanted: a standard for virtual patent marking. *Journal of Intellectual Property Law & Practice*, 15(7):544–553, 06 2020. ISSN 1747-1532.

G. de Rassenfosse and A. B. Jaffe. Are patent fees effective at weeding out low-quality patents? *Journal of Economics and Management Strategy*, 27:134–148, 2018. doi: 10.1111/jems.12219.

G. de Rassenfosse, H. Dernis, and G. Boedt. An introduction to the patstat database with example queries. *Australian Economic Review*, 47(3):395–408, 2014.

G. de Rassenfosse, W. E. Griffiths, A. B. Jaffe, and E. Webster. Low-quality patents in the eye of the beholder: Evidence from multiple examiners. Technical report, National Bureau of Economic Research, 2016.

S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6): 391–407, 1990.

N. Falk and K. Train. Patent valuation with forecasts of forward citations. *Journal of Business Valuation and Economic Loss Analysis*, 12(1):101–121, 2017.

J. Feng and X. Jaravel. Crafting intellectual property rights: Implications for patent assertion entities, litigation, and innovation. *American Economic Journal: Applied Economics*, 12(1): 140–81, 2020.

T. Fischer and J. Leidinger. Testing patent value indicators on directly observed patent value–An empirical analysis of Ocean Tomo patent auctions. *Research Policy*, 43(3):519–529, 2014.

L. Fleming. Breakthroughs and the "long tail" of innovation. *MIT Sloan Management Review*, 49(1):69, 2007.

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

A. Galasso and M. Schankerman. Patents and cumulative innovation: Causal evidence from the courts. *The Quarterly Journal of Economics*, 130(1):317–369, 2015.

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

P. Giuri, M. Mariani, S. Brusoni, G. Crespi, D. Francoz, A. Gambardella, W. Garcia-Fontes, A. Geuna, R. Gonzales, D. Harhoff, et al. Inventors and invention processes in Europe: Results from the PatVal-EU survey. *Research Policy*, 36(8):1107–1127, 2007.

C. Gourieroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica: Journal of the Econometric Society*, pages 701–720, 1984.

Z. Griliches. Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 28(4):1661–1707, 1990.

D. Guellec and B. van Pottelsberghe. *The economics of the European patent system: IP policy for innovation and competition*. Oxford University Press, 2007.

C. J. Guerrini. Defining patent quality. *Fordham L. Rev.*, 82:3091, 2013.

B. H. Hall, A. B. Jaffe, and M. Trajtenberg. The NBER patent citation data file: Lessons, insights and methodological tools. Working Paper 8498, National Bureau of Economic Research, October 2001.

B. H. Hall, A. Jaffe, and M. Trajtenberg. Market value and patent citations. *RAND Journal of Economics*, 36(1):16–38, 2005.

D. Harhoff and S. Wagner. The duration of patent examination at the European Patent Office. *Management Science*, 55:1969–1984, 2009.

D. Harhoff, F. M. Scherer, and K. Vopel. Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8):1343–1363, 2003.

D. Hegde, D. C. Mowery, and S. J. Graham. Pioneering inventors or thicket builders: Which us firms use continuations in patenting? *Management Science*, 55(7):1214–1226, 2009.

C. Helmers et al. *The economic analysis of patent litigation data*, volume 48. WIPO, 2018.

D. J. Hemel and L. Ouellette. Innovation policy pluralism. *Yale LJ*, 128:544, 2018.

J. Henkel and F. Jell. Patent pending–why faster isn't always better. *Available at SSRN 1738912*, 2010.

K. W. Higham, M. Governale, A. Jaffe, and U. Zülicke. Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. *Physical Review E*, 95(4):042309, 2017a.

K. W. Higham, M. Governale, A. Jaffe, and U. Zülicke. Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields. *Journal of Informetrics*, 11(4):1190–1200, 2017b.

K. W. Higham, M. Governale, A. Jaffe, and U. Zülicke. Ex-ante measure of patent quality reveals intrinsic fitness for citation-network growth. *Physical Review E*, 99(6):060301, 2019.

A. B. Jaffe and G. de Rassenfosse. Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology*, 68 (6):1360–1374, 2017.

A. B. Jaffe, M. Trajtenberg, and M. S. Fogarty. Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, 90(2):215–218, May 2000.

J. W. Johnson. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35(1):1–19, 2000.

K. G. Jöreskog and A. S. Goldberger. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *journal of the American Statistical Association*, 70 (351a):631–639, 1975.

H. J. Jung and J. Lee. The quest for originality: A new typology of knowledge search and breakthrough inventions. *Academy of Management Journal*, 59(5):1725–1753, 2016.

H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23 (3):187–200, 1958.

N. Khanna. Patent quality: Does one size fit all? *4iP Council, January*, 2019.

L. Kogan, D. Papanikolaou, A. Seru, and N. Stoffman. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712, 2017.

J. M. Kuhn and N. C. Thompson. How to measure and draw causal inferences with patent scope. *International Journal of the Economics of Business*, 26(1):5–38, 2019.

F. Lafond and D. Kim. Long-run dynamics of the us patent classification system. *Journal of Evolutionary Economics*, 29(2):631–664, 2019.

Y.-H. Lai and H.-C. Che. Evaluating patents using damage awards of infringement lawsuits: A case study. *Journal of Engineering and Technology Management*, 26(3):167–180, 2009a.

Y.-H. Lai and H.-C. Che. Modeling patent legal value by extension neural network. *Expert Systems with Applications*, 36(7):10520–10528, 2009b.

J. O. Lanjouw and M. Schankerman. Characteristics of patent litigation: A window on competition. *RAND Journal of Economics*, pages 129–151, 2001.

J. O. Lanjouw and M. Schankerman. Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114(495):441–465, 2004.

J. O. Lanjouw, A. Pakes, and J. Putnam. How to count patents and value intellectual property: The uses of patent renewal and application data. *The Journal of Industrial Economics*, 46 (4):405–432, 1998.

J. Lerner. The importance of patent scope: An empirical analysis. *RAND Journal of Economics*, 25:319–333, 1994.

B. J. Love, S. P. Miller, and S. Ambwani. Determinants of patent quality: Evidence from inter partes review proceedings. *U. Colo. L. Rev.*, 90:67, 2019.

T. Magerman, B. van Looy, and X. Song. Exploring the feasibility and accuracy of latent semantic analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics*, 82(2):289–306, 2010.

R. J. Mann and M. Underweiser. A new look at patent quality: Relating patent prosecution to validity. *Journal of Empirical Legal Studies*, 9(1):1–32, 2012.

A. C. Marco, A. Tesfayesus, and A. A. Toole. Patent litigation data from US district court electronic records (1963-2015). Technical Report 2017-06, 2017.

A. C. Marco, J. D. Sarnoff, and A. Charles. Patent claims and patent scope. *Research Policy*, 48(9):103790, 2019.

M. Marx and A. Fuegi. Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal*, 2020.

B. W. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, 1975.

D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142, 1973.

M. Mejer and B. v. P. de la Potterie. Patent backlogs at uspto and epo: systemic failure vs deliberate delays. *World Patent Information*, 33(2):122–127, 2011.

S. P. Miller. Who's suing us: Decoding patent plaintiffs since 2000 with the Stanford NPE litigation dataset. *Stan. Tech. L. Rev.*, 21:235, 2018.

P. Moser, J. Ohmstedt, and P. W. Rhode. Patent citations—an analysis of quality differences and citing practices in hybrid corn. *Management Science*, 64(4):1926–1940, 2018.

S. Mukherjee, D. M. Romero, B. Jones, and B. Uzzi. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances*, 3:e1601315, 2017. doi: 10.1126/sciadv.1601315.

S. A. Mulaik. *Foundations of factor analysis.* CRC Press, 2009.

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

F. Narin, E. Noma, and R. Perry. Patents as indicators of corporate technological strength. *Research Policy*, 16(2-4):143–155, 1987.

A. Pakes and Z. Griliches. Patents and r&d at the firm level: A first report. *Economics letters*, 5(4):377–381, 1980.

A. Pakes and M. Schankerman. The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In *R&D, patents, and productivity*, pages 73–88. University of Chicago Press, 1984.

A. Palangkaraya, P. H. Jensen, and E. Webster. Applicant behaviour in patent examination request lags. *Economics Letters*, 101(3):243–245, 2008.

A. M. Petruzzelli, D. Rotolo, and V. Albino. Determinants of patent citations in biotechnology: An analysis of patent influence across the industrial and organizational boundaries. *Technological Forecasting and Social Change*, 91:208–221, 2015.

F. Poege, D. Harhoff, F. Gaessler, and S. Baruffaldi. Science quality and the value of inventions. *Science Advances*, 5(12):eaay7323, 2019.

P. Régibeau and K. Rockett. Innovation cycles and learning at the patent office: Does the early patent get the delay? *The Journal of Industrial Economics*, 58:222–246, 2010. doi: 10.1111/j.1467-6451.2010.00418.x.

E. B. Roberts. What we've learned: Managing invention and innovation. *Research-Technology Management*, 31(1):11–29, 1988.

E. Sapsalis and B. van Pottelsberghe. The institutional sources of knowledge and the value of academic patents. *Economics of Innovation and New Technology*, 16(2):139–157, 2007.

E. Sapsalis, B. van Pottelsberghe, and R. Navon. Academic versus industry patenting: An in-depth analysis of what determines patent value. *Research Policy*, 35(10):1631–1645, 2006.

F. M. Scherer and D. Harhoff. Technology policy for a world of skew-distributed outcomes. *Research Policy*, 29(4-5):559–566, 2000.

S. Scotchmer. Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives*, 5(1):29–41, 1991.

G. Silverberg and B. Verspagen. The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance. *Journal of Econometrics*, 139(2):318–339, 2007.

M. Squicciarini, H. Dernis, and C. Criscuolo. Measuring patent quality: Indicators of technological and economic value. Working Paper 2013/03, OECD Science, Technology and Industry Working Papers, 2013.

M. Trajtenberg, R. Henderson, and A. Jaffe. University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, 5(1):19–50, 1997.

B. van Looy and T. Magerman. Using text mining algorithms for patent documents and publications. In *Springer Handbook of Science and Technology Indicators*, pages 929–956. Springer, 2019.

B. van Pottelsberghe. The quality factor in patent systems. *Industrial and Corporate Change*, 20(6):1755–1793, 2011.

N. van Zeebroeck. The puzzle of patent value indicators. *Economics of Innovation and New Technology*, 20(1):33–62, 2011.

N. van Zeebroeck and B. van Pottelsberghe. Filing strategies and patent value. *Economics of Innovation and New Technology*, 20(6):539–561, 2011a.

N. van Zeebroeck and B. van Pottelsberghe. The vulnerability of patent value determinants. *Economics of Innovation and New Technology*, 20(3):283–308, 2011b.

E. Webster and P. H. Jensen. Do patents matter for commercialization? *The Journal of Law and Economics*, 54(2):431–453, 2011.

J. Weston, S. Bengio, and N. Usunier. WSABIE: Scaling up to large vocabulary image annotation. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

G.-C. Yang, G. Li, C.-Y. Li, Y.-H. Zhao, J. Zhang, T. Liu, D.-Z. Chen, and M.-H. Huang. Using the comprehensive patent citation network (cpc) to evaluate patent value. *Scientometrics*, 105(3):1319–1346, 2015.

K. Zahringer, C. Kolympiris, and N. Kalaitzandonakes. Time to patent at the uspto: the case of emerging entrepreneurial firms. *The Journal of Technology Transfer*, 43(4):923–952, 2018.

G. Zhang, L. Xiong, H. Duan, and D. Huang. Obtaining certainty vs. creating uncertainty: Does firms' patent filing strategy work as expected? *Technological Forecasting and Social Change*, 160:120234, 2020.

Appendices

A Semantic Embedding of Patents

To order to effectively use as much of the information available in the full-text patent specifications as possible (as opposed to a keyword-based approach), we can embed patent specifications in a lower-dimensional (continuous) vector space, and compute the distance between patents in this space. This kind of natural language processing can be very computationally expensive; however, the team at Google Patents have generously made their document embeddings public for research use. These document embeddings are the result of a bag-of-words approach, trained to predict Cooperative Patent Classification (CPC) codes using the WSABIE classification model (Weston et al., 2011) on full-text patent specifications.

WSABIE jointly trains two functions to map two different sets of input data into the same lower-dimensional space.[17] In the current case, one set is the full-text bag-of-words for each patent, and the other is the CPC classifications for the same patent. A machine-learning process then trains each of these functions such that the bag-of-words and CPC classification for each patent are mapped into approximately the same location in the lower-dimensional space. These WSABIE-based embeddings complement our art-unit approach, as art units are *not* based on the CPC system (instead, they are still defined by the United States Patent Classification codes). We can then use the locations of these documents in this space to construct measures of

[17] This dimensionality is predefined, and in the case of the Google Patents data utilized here, this was set to 64 dimensions.

patent similarity. The embeddings are 64-component vectors, with each component restricted to the range [0,1]. As such, we can measure the cosine similarity between each pair of patents to identify the semantically similar patents.

B Indicator Variables for Factor Analysis

B.1 Backward-Citations-based Indicators

Backward citations are the citations a patent makes to related earlier patents and other sources of knowledge. They contain not only information about the citing patent, but also information on the inventions or ideas to which the citing patent is related.

From the perspective of the patent owners, the number of backward citations should be minimized for strategic reasons, as this would maximize the technological space this patent allows them to exploit (Jaffe and de Rassenfosse, 2017). Therefore, one could expect many backward citations to hurt the invention's prospects for economic-value generation. On the other hand, many backward citations could also be an indicator of a technologically complex invention reliant on or related to a wide array of technology types. To give us the best chance to observe these potential effects, we consider specific subsets and properties of backward citations separately.

Backward citations to US patents can be split along two lines: whether or not the citation is a self-citation, and whether the applicant or the examiner added it. The USPTO tag the latter, which are added for different reasons by different parties (see Jaffe and de Rassenfosse, 2017, for an overview). In short, applicants are obliged to submit known prior art material to patentability, and examiners find closely related prior art that they discover during the prosecution process that impacts the assessment of novelty or non-obviousness of the application.

Self-citations can either be inventor-centric or firm-centric in their definition. We choose to define self-citations in the latter sense (firms citing their own patents) because identifying firms is much simpler and less error-prone than identifying inventors—firms are given a USPTO assignee code while inventors are not systematically disambiguated in this way. Firm self-citations are often interpreted as the result of cumulative research (Scotchmer, 1991; Moser et al., 2018), which in turn is an indicator of a firm's willingness to invest in R&D to strengthen their competitive position in a particular technology space (Hall et al., 2005; Czarnitzki et al., 2006). Firms probably do not undertake this kind of research without the expectation that the technology, or the particular invention, will be valuable somehow. At the same time, self-citations could also be indicative of incremental improvement rather than a technological breakthrough, which may hinder impact (Sapsalis and van Pottelsberghe, 2007). As self-citations can be challenging to identify in the presence of complex and dynamic corporate structures, we take advantage of the work done by Kogan et al. (2017) to match the patent assignees to their publicly traded parent companies. We identify self-citations through a two-stage matching procedure: we first check if the cited-citing patent pairs were assigned to the same publicly-traded parent company, and if not, whether they are assigned to the same assignee as recorded by the USPTO.

The four mutually-exclusive categories of backward citations to patents are therefore[18]: *backward non-self citations to patents, applicant added* (**BCA**); *backward non-self citations to patents, examiner added* (**BCE**); *backward self-citations to patents, applicant added* (**BSA**); and *backward self-citations to patents, examiner added* (**BSE**).

The age of these backward citations to patents can be indicative of the proximity of the citing patent to the technological frontier (Benson and Magee, 2015); a large number of citations to very recent inventions could indicate that a patent is 'cutting edge.' Conversely, a patent that cites mostly older inventions may be either a late addition to an increasingly obsolete technological field or part of a generally slow-moving technological field. Perhaps intuitively, therefore, technologically influential patents have been observed to be more likely to cite new patents (Mukherjee et al., 2017). As such, we include as an indicator the *average age of backward citations* (**BAG**), which is the arithmetic mean of the ages of the patents cited by a new patent i (including self-citations) at the latter's time of application $T_i^{(a)}$. We define the age of a citation as the time between the grant of the cited patent and the application of the citing patent, thus bypassing the noise added by grant lag. To avoid mathematical difficulties, we exclude citations for which the citation lag is negative (i.e., when the application date of the citing patent is *earlier* than the grant date of the cited patent).

The relative popularity of cited patents (at the time they are cited) may also affect ex-post outcomes of the citing patent, and could be particularly informative if the cited patents are still young—this may imply that the citing patent is 'riding a wave' at the technological frontier. *Backward-citations' pedigree* (**BPE**) is a metric that is calculated based on the normalized geometric mean of the number of forward citations accrued by the cited patents listed on a particular patent document at the latter's time of application. Introduced by Higham et al.

[18] All abbreviations for indicators derived exclusively from backward citations start with the letter B.

(2019), the pedigree of patent i that has been applied for at time $T_i^{(\mathrm{a})}$ is defined as

$$\mathrm{BPE}_i = \left[\prod_{j=1}^{N_i} \frac{k_j\big(T_i^{(\mathrm{a})}\big)}{\big\langle k\big(T_i^{(\mathrm{a})}\big)\big\rangle_{S_j}} \right]^{\frac{1}{N_i}} \quad , \tag{1}$$

where N_i is the number of patents cited by patent i, and $k_j\big(T_i^{(\mathrm{a})}\big)$ denotes the number of citations the cited patent j has at $T_i^{(\mathrm{a})}$. The citation counts of cited patents j are normalized by the average number $\big\langle k\big(T_i^{(\mathrm{a})}\big)\big\rangle_{S_j}$ of citations of patents from the same CPC Section S_j of patent j granted within the same three-month period. The geometric, rather than the arithmetic, mean has been adopted because of the high skewness of citation distributions.

Originality (**BOR**) is a popular metric that quantifies the diversity of technologies that a patent cites (Trajtenberg et al., 1997). We define the originality of patent i as

$$\mathrm{BOR}_i = 1 - \sum_C \left(P_C^{(i)}\right)^2 \quad , \tag{2}$$

where $P_C^{(i)}$ is the proportion of the N_i backward citations from patent i to other patents going into a given Class C from the CPC system.[19] Information about all Classes a cited patent j is categorized into is utilized here, rather than just the first listed as was done in Ref. (Trajtenberg et al., 1997). Specifically, each cited patent j contributes to $P_C^{(i)}$ for each one of the $N_j^{(\mathrm{cl})}$ classes it is listed under with a value $(1/N_j^{(\mathrm{cl})}) \cdot (1/N_i)$. This construction ensures the relation $\sum_C P_C^{(i)} = 1$ and, thus, the reasonability of the definition (2)—BOR is limited to the half-open interval $[0, 1)$. In particular, a patent with $\mathrm{BOR}_i = 0$ will have all its backward citations to patents from one particular class. In contrast, BOR_i will be large when the citations of a given patent are distributed over many CPC Classes. In contrast to Ref. (Trajtenberg et al., 1997), we use the CPC Classes rather than the technology classes defined in the NBER patent database (Hall et al., 2001).

There are two idiosyncrasies inherent in the originality variable. First, it is sensitive to the classification system being used—a coarse system would skew the originality distribution towards zero, and vice versa for a highly specific system. Second, it is correlated with the number of backward citations because a greater number of citations creates more opportunity for technological diversity in those citations. The CPC system makes adjusting for the first problem quite straightforward; it is possible to tune the originality distribution by moving up

[19]'Classes' are the second-highest hierarchy level in the CPC, just below 'Sections.'

or down the classification hierarchy until a distribution with reasonable spread is obtained, and we find CPC Class (CPC 2-digit level) suffices. Our factor analysis addresses the second issue.

B.2 Backward Citations to Other Literature

Backward citations can also reference information that is not contained in a granted USPTO patent document. The following indicators are three such types of citations.[20]

Front-page backward citations to scientific literature (**BNP**) is the count of front-page citations that can be matched to articles in Microsoft Academic, extracted by Marx and Fuegi (2020). The vast majority of these citations originate with the applicant (as part of their duty of candor) rather than with the examiner (Cotropia et al., 2013). *In-text backward citations to non-patent literature* (**BIT**) appear in the invention description, and generally not the same as the BNP citations that appear on the front page (20-30% overlap exists). These citations generally act to illuminate the "history, usefulness, and development of the invention" (Narin and Noma, 1985) and contribute to invention disclosure by providing pertinent information that allows a person who is 'skilled in the art' to replicate and use the invention. These citations are extracted from the database associated with Ref. (Bryan et al., 2020), and correspond to those that can be matched to articles in a pre-defined set of prominent journals representing a broad range of fields. A patent with many citations of either of these types may indicate a strong scientific grounding, or otherwise build on fundamental scientific research, and could be more likely to be among the first patents granted in a new technological avenue.

Lastly, *backward citations to foreign patents* (**BFP**) cover knowledge disclosed in patents granted by a jurisdiction other than the USPTO. The vast majority of these citations are added by the 'applicant' rather than the examiner (Cotropia et al., 2013)—in reality, the foreign patents cited are generally those listed in search reports compiled by a foreign office such as the European Patent Office. As the applicant is aware of this prior art before the grant of the patent, they are obliged to pass this information to the USPTO, where they are designated 'applicant citations.'

[20] As these are still a form of backward citation, the abbreviations retain the initial B.

B.3 Variables Unrelated to Citations

B.3.1 Claims

Claim counts are a common indicator of both technological significance and economic value (van Zeebroeck and van Pottelsberghe, 2011b). This metric appears to be a robust correlate of many measurable dimensions of patent value across all technology categories (e.g., economic value in surveys (Gambardella et al., 2008), probability of litigation (Lanjouw and Schankerman, 2001), and forward-citation counts (Jaffe et al., 2000)). Also, the number of claims varies quite widely, so bias or error due to the discretization of count data is smaller than for most other available indicators.

However, the relationship between claim counts and quality is ambiguous because the effect of addition claims depends on its type- independent claims and dependent claims have very different functions (Marco et al., 2019). Independent claims are written in broad terms and serve to delineate the distinct components of an invention and stand on their own. On the other hand, dependent claims generally specify the precise embodiment of the independent claims such that the scope of each associated independent claim is well-defined. As such, we count these two claim types independently and differently: as *independent claims* (**CIN**) and *dependent claims per independent claim* (**CDE**) (de Rassenfosse and Jaffe, 2018; Marco et al., 2019). The separation of claims into these two categories is an attempt to roughly distinguish the claims that narrow technological breadth from those that broaden it, and as the total number of claims is simply the sum of the independent and dependent claims, we can only gain information via this separation. Generally, a dependent claim will narrow a single independent claim; as such, we calculate the number of dependent claims per independent claim to capture the average amount of narrowing present in each patent. Recent work has also found that independent claim length may also play an important role in determining both the scope and pendency of a patent (Kuhn and Thompson, 2019; Marco et al., 2019). Specifically, as independent claim length (and particularly the length of the first claim) increases, the more specific the claim becomes, and the narrower the patent may be, and the faster it can be granted. We include the *words in the first claim* (**CFW**) in this work, as this has been expertly validated as a measure of patent scope (Kuhn and Thompson, 2019).

B.3.2 Class Membership

Class Membership (**TCM**), which has also been (perhaps confusingly) called 'patent scope' (Lerner, 1994; Squicciarini et al., 2013), is simply a count of how many technology categories under which a patent has been classified. It constitutes a measure for technological breadth and has been linked to patent value (Lerner, 1994). For this work, we are using the CPC Classes[21] as the categories, as they represent a reasonable level of specification. If the categorization levels used for this variable are too specific (e.g., using CPC Sub-Classes), we may expect the number of classifications a patent has to become technology-dependent, and the explanatory power would weaken when attempting to detect within-category variation in quality. On the other hand, if the categories were too broad (e.g., when using the CPC Sections as categories), very few patents will belong to more than one category, and therefore the information content of this variable would be minimal.

B.3.3 Grant Lag

Grant Lag (**LAG**) is the length of time between the initial application date (not necessarily the priority date) and the eventual grant date of a patent. While a very simple metric, there are many ways this time lag may be relevant to technological influence. Recent work (Harhoff and Wagner, 2009; Régibeau and Rockett, 2010; Squicciarini et al., 2013) suggests that the grant lag is inversely proportional to patent quality, for two factors. Firstly, the claims on a patent that is marginal with respect to patentability, and therefore of lower quality, may take longer to negotiate and examine. Secondly, in some jurisdictions, companies can pay for accelerated examination of their most valuable patents. This latter phenomenon does not affect the patents in our cohort—the USPTO instituted their accelerated-examination program after these patents were granted. The former point may be relevant for us, but there is also evidence that both the number of claims and the number of backward citations are positively correlated with forward citations (Lanjouw and Schankerman, 2004), and one would expect that a more protracted negotiation process would result in increases in both of these variables. Additionally, a longer examination process requires a greater financial investment on the part of the inventor/assignee because the attorney spends a longer time negotiating with the examiner during this process. Thus, one would only expect an extended examination period to be tolerated if there was some prospect of a return on this investment. Therefore, a long grant lag may reflect a firm's confi-

[21] `cooperativepatentclassification.org`, accessed 10 July 2020

dence in the economic value of an invention, even if the invention is only marginally patentable. Finally, the complexity of an invention may also affect the length of the examination period, as the more components an invention has, or knowledge required to understand its function, the longer it may take to determine its patentability.

B.3.4 Family Size

Geographic patent family size (**GPF**) is the number of jurisdictions in which a particular invention (with the same inventors and priority date) has been patented (de Rassenfosse et al., 2014). Patentees will often seek patent protection for their inventions in multiple jurisdictions. This process can be very costly, with additional patenting fees, attorney costs, and translation costs stacking up with each new application. There are options for simplifying this process (e.g., Patent Cooperation Treaty filings), but the costs listed above generally still apply. It is intuitive, therefore, that the number of countries in which a patent is granted should indicate the confidence that the applicant has in the invention with regard to recouping all costs associated with the invention and patenting process, and thus the economic value of the patent (Cremers, 2004; Gambardella et al., 2008; Lanjouw et al., 1998; van Zeebroeck and van Pottelsberghe, 2011b). Following Sapsalis and van Pottelsberghe (2007) and Squicciarini et al. (2013), we only count one family member per jurisdiction—continuing patent applications that claim priority from an application filed in the same jurisdiction appear to be less important and may increase the noise in this indicator.

B.3.5 Non-Final Rejections

Similar to GPF, *non-final rejections* (**NFR**) are another indicator of an applicant's confidence that a patent will more than recoup the costs associated with the invention and patenting process. Non-final rejections are office actions that indicate that an examiner has decided that a patent application is not valid and invites the applicant to reply to the criticism provided or amend the patent application appropriately if they wish to continue the examination process. If the applicant wishes to continue with the process, they will generally have to pay an attorney to address the non-final rejection. Assuming that the applicant will only continue the process if they expect to recoup the additional legal fees, the number of non-final rejections (with the knowledge that the patent was eventually granted) may be an indicator of the financial value of a patent to the assignee. Non-final rejections also extend the length of the examination process, and so may be useful for teasing out useful components of the grant lag, which, as mentioned above, could reflect many different patent qualities. This indicator is particularly uncommon in studies on patent quality and as far, as we can tell, have only been used to assist in the estimation of damage awards in infringement lawsuits (Lai and Che, 2009b,a).

C Oblique Factor Rotation

Unlike in the case of orthogonal factor rotation, there is often little difference between oblique rotation strategies (Costello and Osborne, 2005), and we find the same is true for the current analysis. For this reason, we a rotation from the oblimin family, which is both popular and mathematically simple.

The oblimin criterion has its roots in the oblique counterpart to the varimax criterion (as used for the orthogonal rotation in our patent fitness paper). In words, oblimin is the minimization of the covariances of squared (communality-normalized) factor loadings between each pair of factors, summed across variables. Mathematically, the objective function is (Mulaik, 2009)

$$B = \sum_{j=1}^{r} \sum_{k=1}^{r} \left[\sum_{i=1}^{n} \frac{\lambda_{ij}^2 \lambda_{ik}^2}{h_i^4} - \frac{\gamma}{n} \left(\sum_{i=1}^{n} \frac{\lambda_{ij}^2}{h_i^2} \right) \left(\sum_{j=1}^{n} \frac{\lambda_{ik}^2}{h_i^2} \right) \right] ,$$
$$\text{with} \quad j \neq k \quad , \tag{3}$$

in the case of r factors, n variables, factor loading matrix λ, and variable communalities h_i. Parameter γ may be set to zero to obtain the 'quartimin' criterion (Carroll, 1953), to unity for the 'covarimin' criterion (Kaiser, 1958), and a compromise known as 'biquartimin' for the case $\gamma = 0.5$ (Carroll, 1957). However, this latter value does not appear to be special, and, in fact, any value in [0,1] may be substituted to suit the situation faced by the factor analyst. In general, the $\gamma = 0$ produces the most oblique solutions, while $\gamma = 1$ produces the most orthogonal. The choice of gamma depends on how much freedom one would like to give to the optimization procedure. In this work, we choose $\gamma = 0$ to maximize the obliqueness of our solution and find few highly correlated factors. We find that our results are not particularly sensitive to this choice.