Compete or Cooperate? The Effect of Patent Protection on Licensing Decisions of Inventors

Honggi Lee*
The Fuqua School of Business
Duke University

Abstract

How does patent protection affect licensing decisions of inventors? While broader patent scope can increase the value of an invention as well as lower expropriation risk and transfer costs, it can also increase the value of self-commercialization by raising the inventor's market power. Accordingly, theoretical predictions around the effect of patent scope on licensing are ambiguous, and empirical results have been mixed and usually confined to specific settings. In this study, I present more precise and generalizable findings about the relationship between patent scope and licensing propensity of inventors. Additionally, I explore various invention and firm characteristics that might enhance or limit the patent scope-licensing relationship. To do so, I employ a large sample of licensed inventions over 20-year period across multiple technology areas and exploit an exogenous variation in patent scope. The results show that reduced patent scope leads to a substantial decline in licensing propensity of inventors. Furthermore, the effect is stronger for high-quality, novel, and science-based inventions as well as for inventions generated by small and young inventors. The effect is also stronger for inventions in biotechnology than it is for inventions in other technology areas. Through this study, I contribute to our understanding of the ways in which inventors commercialize specific types of inventions as well as the mechanisms underlying inventors' decisions to compete or cooperate with incumbents and potential rivals.

Key words: commercialization, invention, licensing, patent

* Address: The Fuqua School of Business, 100 Fuqua Dr., Durham, NC 27708. Email: honggi.lee@duke.edu.

1. Introduction

For inventors, decisions regarding commercialization mode (e.g. self-commercialization, licensing, joint venture) can have crucial implications for the returns they can appropriate from their inventions (Teece, 1986). Yet, these decisions can have a broader impact on existing market power and structure (Schumpeter, 1943). Accordingly, some studies have highlighted industries (e.g. computers) where incumbents are overthrown by inventors who decide to enter a product market by commercializing their own inventions (Christensen, 1997; Henderson and Clark, 1990), while others have documented industries (e.g. biotechnology) where cooperation between inventors and incumbents is more common, often arguing that such cooperation is likely to reinforce the existing market structure so long as markets for technology and ideas are efficient enough (Gans et al., 2000, 2008; Hegde and Luo, 2017). Given such far-reaching implications, studies have put much emphasis on understanding the factors that influence inventors' commercialization strategy in relation to the decision to compete (by entering the market directly) or cooperate (through licensing or other means of co-development) with incumbents and potential rivals (Arora and Ceccagnoli, 2006; Arora and Fosfuri, 2003; Gambardella and Giarratana, 2013; Gans et al., 2000, 2008; Gans and Stern, 2003; Hegde and Luo, 2017; Hsu, 2006; Shane, 2001).

In this study, I explore the extent to which patents, a legal instrument that can increase the value of an invention and lower expropriation threats and certain transfer costs, facilitate inventors' use of licensing as a commercialization channel. Given their characteristics, it might seem that broader patents would promote inter-firm cooperation through licensing. However, it is not clear whether simply providing broader patents would increase licensing propensity because they also increase the value of self-commercialization by raising the inventor's market power (Arora and Ceccagnoli, 2006; Gans et al., 2000). Additionally, studies have shown that the effectiveness of patents is likely to vary across technology areas, firms, and even inventions (Anand and Khanna, 2000; Cohen et al., 2000; Kline and Rosenberg, 1986; Teece, 1986). Thus, theoretical predictions around the effect of broader patents on licensing are ambiguous (Arora and Ceccagnoli, 2006). Empirical results have been mixed as well, with some studies finding a positive relationship between patents and licensing while others finding no significant or even negative relationship (Fosfuri, 2004; Gans et al., 2000; Nerkar and Shane, 2007; Shane, 2001). Also, these studies have often been confined to specific industries (e.g. biotechnology) or settings (e.g. small to large firm licensing) potentially limiting the generalizability of their findings (Cassiman and Veugelers, 2002; Fosfuri, 2004; Gans et al., 2000, 2008; Smith, 2001).

I contribute to this debate by presenting more precise and generalizable findings about the relationship between patent scope and licensing decisions of inventors. Furthermore, drawing from prior studies that characterize various determinants of licensing (e.g., industry, firm size, knowledge base of an invention), I uncover specific conditions under which the patent scope-licensing relationship might be enhanced or limited. To do so, I employ a large sample of licensed inventions across multiple technology areas combined with USPTO patents. I also exploit an exogenous variation in "effective patent scope", defined as the inventor's ownership rights over an invention delineated by the extent to which the invention is similar to prior-art inventions of rivals, which affect the inventor's ability to exercise market power as well as bargaining positions in licensing negotiations.[1]

To determine whether there exists a prior-art invention similar to a focal invention, I train a machine learning algorithm to predict whether the claims of the focal patent are similar to the claims of its prior-art patents (patents whose priority date, i.e. the first time a patent application is filed for a specific invention, is earlier than the priority date of the focal patent). Then, for each focal patent, I identify the first similar prior-art patent that is disclosed to the public after the priority date of the focal patent and designate it as a "priority disclosure". Finally, the publication of a priority disclosure is used as an event that unexpectedly reduces the effective scope of the focal patent. By employing a difference-in-differences approach, I exploit this exogenous variation in patent scope to examine the effect of reduced patent scope on licensing propensity of inventors.[2]

The sample consists of inventions with USPTO patents whose priority years are between 1990 and 2010. I aggregate patents to their "simple patent family" level defined by the European Patent Office (EPO) and designate each aggregation as an "invention" to avoid double counting (EPO, 2019).[3] The treatment group consists of inventions whose patent scope has been reduced due to an unexpected publication of a priority disclosure. To control for the underlying propensity of an invention to be licensed out, a control group is constructed by matching each treated invention to a control invention (i.e. an invention whose patent scope is not reduced) on priority year, earliest publication year, international patent classification (IPC) code, forward citation quartile, and firm-level invention stock quartile. For each invention, the sample is expanded to include up to five years

[1] Legally, the scope of a patent is defined based primarily on its claims (Lemley and Shapiro, 2006), which in turn are granted based on the doctrine of enablement (Merges, 2006) pertaining to disclosure of useful knowledge in the specifications. To illustrate what a broad patent scope implies, Merges and Nelson (1990) state that "the broader the scope, the larger the number of competing products and processes that will infringe the patent." Some studies also point out that patent scope as defined in claims is often not definitive and is limited by existence of other similar patents (Gans et al., 2000; Lemley and Shapiro, 2005). Grounded in a similar logic, "effective patent scope" takes into account whether there exists a prior-art patent on an invention that is similar to the focal invention.

[2] In Section 7, I test the assumption that these events are unexpected because inventors do not have sufficient knowledge about the existence of a similar prior-art invention until the prior-art invention is published by a patent issuing authority.

[3] According to EPO, "a simple patent family is a collection of patent documents that are considered to cover a single invention. The technical content covered by the applications is considered to be identical. Members of a simple patent family will all have exactly the same priorities." Different types of patent applications that might be grouped into the same invention include provisional, foreign, continuation, divisional, and continuation-in-part (Arora et al., 2018).

before and after the publication of its priority disclosure with a dummy dependent variable taking a value of 1 if an invention is licensed out in a given year.

The empirical results show that, on average, a reduction in patent scope leads to inventors refraining from licensing their inventions. However, there is a wide variation across technology areas as well as firm and invention characteristics. The effect of reduced patent scope on licensing propensity for high-quality inventions is about 3.5 times stronger than it is for low-quality inventions. For novel inventions, the effect is around 3.7 times stronger than it is for non-novel inventions, and, for science-based inventions, the effect is around 1.5 times stronger than it is for non-science-based inventions. For inventions generated by large and mature inventors, reduced patent scope leads to around 68% and 88% increase in licensing propensity relative to inventions generated by small and young inventors, respectively. The results also show that the effect is the strongest for inventions in biotechnology, followed by those in communications and chemicals. The main findings are robust to alternative specifications and controls.

This study makes several contributions to innovation and entrepreneurship research. First, using a large sample of licensed inventions across multiple technology areas, this study provides insights into how a particular dimension of patents, i.e. patent scope, influences the licensing propensity of inventors. The findings show that reduced patent scope on average leads to a substantial decline in licensing propensity of inventors. While prior studies have examined the relationship between patents and licensing, they have often been confined to specific industries (e.g. biotechnology) or settings (e.g. small-to-large firm licensing), and finding an exogenous variation in patent scope or in other dimensions of patent effectiveness has been a challenge.

The findings from this study also provide insights into how firms profit from the specific types of inventions that they generate. While prior studies have shown that one of the reasons firms patent their inventions is to license and that the effect of patent scope is likely to vary across different types of inventions (Anand and Khanna, 2000; Cohen et al., 2000; Teece, 1986), a systematic examination of the types of inventions that might be more responsive to changes in patent scope in terms of licensing has not been undertaken. The findings from this study show that invention characteristics do indeed matter as the patent scope-licensing relationship is likely to be enhanced by high-quality, novel, and science-based inventions. This finding is consistent with the notion that high-quality inventions are much more likely to be licensed than low-quality inventions (Astebro, 2003; Gambardella et al., 2008; Harhoff et al., 1999) and that valuable applications of novel and science-based inventions tend to reside outside of the boundaries of the inventor's operations (Arora and Gambardella, 1994; Bresnahan and Gambardella, 1998; Nelson, 1959).

Furthermore, the findings are consistent with the notion that small and young inventors, including individual inventors, rely more heavily on licensing as a commercialization strategy than large and mature inventors. To the extent that inventor size and age reflect ownership of (or access to) downstream complementary assets required for commercialization, the findings imply that inventors without adequate access to complementary assets respond more sensitively to changes in patent scope. This is especially so because the downstream assets are often specialized and difficult to acquire (Arora and Ceccagnoli, 2006; Rothaermel, 2001; Teece, 1986) while broader patent scope increases the value of self-commercialization (Arora and Ceccagnoli, 2006; Teece, 1986). This finding is consistent with prior studies showing that the ease of access to complementary assets is an important determinant of the channel through which an inventor commercializes an invention (Arora and Ceccagnoli, 2006; Gambardella et al., 2007; Gans et al., 2000).

More broadly, the findings from this study contribute to our understanding of the mechanisms underlying inter-firm cooperation and competition. Prior studies have highlighted various factors at industry and firm levels as determinants of whether inventors and incumbents compete or cooperate. For instance, studies have shown that the strength of patents, existence of intermediaries, and cost of product market entry are important determinants of whether inventors and incumbents cooperate (Dushnitsky and Lenox, 2005; Gans et al., 2000; Gans and Stern, 2003; Hegde and Luo, 2017; Hsu, 2006; Shane, 2001). Others have shown that market competition influences the inventor's licensing decisions when the upstream and downstream markets are tightly linked (Arora and Fosfuri, 2003; Fosfuri, 2006; Gambardella and Giarratana, 2013). At the firm level, studies suggest that capabilities (Henderson and Clark, 1990), incentives (Christensen, 1997; Marx et al., 2014), top managers (Eesley et al., 2014), and access to downstream resources (Arora and Ceccagnoli, 2006; Gans et al., 2000; Teece, 1986) are likely to affect the compete-cooperate dynamic. This study sheds light on another important dimension, i.e. invention characteristics, that influences inventors' decision to compete or cooperate with incumbents and potential rivals. The findings show that inventors generating high-quality, novel, and science-based inventions are less likely than other inventors to pursue a cooperative commercialization strategy as patent scope narrows.

This paper is organized as follows. Section 2 presents a simple conceptual framework and hypotheses around the factors that moderate the patent scope-licensing relationship. Section 3 describes the methodology employed to test the hypotheses. Section 4 presents validation of the main measures used in the study, and Section 5 describes the sample. Sections 6 and 7 present non-parametric evidence and estimation results. Section 8 concludes with main insights from this study and avenues for future studies.

2. Conceptual framework and hypotheses

To guide the empirical analysis undertaken in this study, this section presents a simple conceptual framework focusing on invention and firm characteristics that might enhance or limit the relationship between patent scope and licensing.

Consider invention i, which in the context of this study, is a new idea or technology disembodied from a downstream product but can be traded and integrated into a single or multiple products (Arora et al., 2001; Fleming, 2001).[4]

Let λ_i be effective patent scope, which delineates ownership claims of an inventor over invention i in part based on the extent to which the invention is similar to prior-art inventions of rivals. While claims issued on a patent prescribe the extent to which its inventor can exclude others from using his invention, the existence of a patent on a highly similar prior-art invention from another inventor in effect limits such ability (Gans et al., 2008; Lemley and Shapiro, 2005; Merges and Nelson, 1990).[5] In essence, a broad patent leads to a large market power because it prevents rivals from imitating the patented invention, or, at the least, because it raises rivals' imitation costs and time (Levin et al., 1987; Mansfield, 1986; Teece, 1981).[6]

Let v_i be the inherent, technical quality of invention i. That is, v_i reflects how well an invention performs a function and in turn limits the value of the invention.[7] As Kline and Rosenberg (1986) discuss, while closely intertwined, technical value is different from commercial value, which is tied to consumers' willingness to pay for an invention.

Let $\tilde{q}$ and q be efficiencies of a licensee and a licensor, respectively, in commercializing invention i. "Efficiency" is determined by the resources and capabilities that a firm possesses (e.g. marketing, distribution network) as well as its market size. For instance, a firm with a well-established distribution network would be able to commercialize an invention at a lower cost than a firm with no such network. Additionally, a firm with a large scale would be able to sell more units of an invention and at the same time spread the costs of the invention across more units, thereby extracting higher returns from the invention (Cohen and Klepper, 1996).

Let T be transaction costs and any other costs associated with trading an invention. For instance, T includes costs associated with searching for a potential licensee (Gans et al., 2000; Hegde and Luo,

[4] For example, OLED ("organic light-emitting diode") technology is an invention that was introduced by inventors at Eastman Kodak in 1987. Since its introduction, OLED technology has been integrated into various products such as televisions and smart phones to improve display quality.

[5] Prior studies have documented patents issued on similar inventions often do exist (Lemley and Shapiro, 2005; Thompson and Kuhn, 2017). Furthermore, an interview with an expert with knowledge about the USPTO practices on patent claims and patent examination also confirms that issuance of patents on highly similar inventions is not rare.

[6] Broader patent scope can also lower expropriation risk and certain kinds of transactions costs by providing legal protection and clarifying ownership claims to the invention (Gans et al., 2000, 2008). However, for the purposes of explicating the interaction effect of patent scope and invention or firm characteristics, its effect on costs associated with trading inventions is not explicitly represented in the framework.

[7] For example, a computer CPU ("Central Processing Unit"), which is directly tied to the overall speed of a computer, is traditionally evaluated based on how quickly and accurately it can process a given set of instructions and more recently how much parallel processing it can carry out. These attributes determine the quality and in turn the technical value of a CPU. When a product that integrates an invention is composed of components that interact with the invention, the quality of the invention also depends on the interoperability of the invention with other components.

2017; Hellmann, 2007; Lamoreaux and Sokoloff, 2002), obtaining information about an invention (Anton and Yao, 1994; Arora, 1996; Arrow, 1962), and determining the commercial value of an invention (Arora and Gambardella, 2010; Kline and Rosenberg, 1986; Rosenberg, 1998). T limits the extent to which an inventor commercializes his invention through licensing (Agrawal et al., 2015; Shapiro, 1985).

Licensing occurs when there is a licensee that is more efficient than the licensor in commercializing an invention and the difference in their efficiencies would be greater than the costs to be incurred in trading the invention (Arora et al., 2001; Bresnahan and Gambardella, 1998; Stigler, 1951). Efficiency is further limited by patent scope and the innate technical quality of an invention. Thus, inventing firms commercialize their inventions through licensing when the following condition is met:

$$\lambda v\tilde{q} - \lambda vq > T, \text{ or } \lambda v(\tilde{q} - q) > T$$

2.1. Invention characteristics

Inventions can be assessed based on specific attributes that differentiate them from other inventions. Drawing on prior studies that have characterized inventions based on various attributes, I examine the effect of patent scope on licensing for inventions based on their quality (Trajtenberg, 1990; Hall et al., 2005; Lanjouw and Schankerman, 2004), novelty (Arora and Cohen, 2017; Fleming, 2001), and knowledge base, i.e. the degree of general and abstract knowledge that an invention embodies (Arora and Gambardella, 1994; Bresnahan and Gambardella, 1998).

2.1.1. Quality. An invention can be assessed based on its inherent technical quality (Gambardella et al., 2008; Hall et al., 2005; Lanjouw and Schankerman, 2004; Trajtenberg, 1990), which reflects how well an invention performs a given function or solves a given problem.[8] Quality is an important dimension to explore when examining the patent scope-licensing relationship because most patented inventions are not likely to be licensed, except those at the top of the quality distribution, and the quality distribution of patented inventions is highly skewed to the right, leaving only a small subset in the high-quality category at risk of being licensed (Astebro, 2003; Gambardella, 2013; Gambardella et al., 2007; Harhoff et al., 1999). Thus, the effect of patent scope on licensing would be stronger for high-quality inventions than for low-quality inventions because $\tilde{q} - q$ is higher for high-quality inventions than for low-quality inventions.

Hypothesis 1. *The effect of reduced patent scope on licensing propensity is stronger for high-quality inventions than for low-quality inventions.*

[8] Because inherent technical quality of an invention is difficult to observe and measure, prior studies have often inferred invention quality using observable features of an invention such as patent attributes (Lanjouw and Schankerman, 2004) or stock market valuation (Hall et al., 2005). Also, other studies have also used similar concepts such as "usefulness" (Fleming, 2001) or "impact" (Rosenkopf and Nerkar, 2001) to describe an invention.

2.1.2. Novelty. An invention can be assessed based on its novelty, which, in the context of this study, measures how different an invention is from previous inventions. For a novel invention, due to its unfamiliarity, there is high uncertainty around its applicability and commercial success (Kline and Rosenberg, 1986; Rosenberg, 1998). Thus, it is more likely that a valuable application for the novel invention resides outside of the inventor's limited range of applications than inside (Nelson, 1959).

On the other hand, an incremental invention, which makes a small improvement on a previous invention, tends to be targeted at a specific, well-understood application. Accordingly, prior studies have shown that incremental inventions are likely to reinforce an existing capability or enhance certain aspects of an existing product (Abernathy and Clark, 1985; Christensen, 1997; Henderson and Clark, 1990; Tushman and Anderson, 1986). Thus, the inventor is likely to be better positioned to commercialize an incremental invention than a novel invention. In other words, $\tilde{q} - q$ is likely to be higher for novel inventions than for incremental inventions.

Hypothesis 2. *The effect of reduced patent scope on licensing propensity is stronger for novel inventions than for non-novel inventions.*

2.1.3. Knowledge base. Inventions can draw on different types of knowledge base ranging from general and abstract to experimental (i.e. trial and error). "General and abstract" knowledge discussed in this study is defined as in Arora and Gambardella (1994) to imply knowledge that can be represented in symbolic or formulaic terms and provides a common-thread understanding across different technical disciplines.[9]

Mechanisms underlying an invention that draws on general and abstract knowledge can be understood more clearly by a potential licensee because the nature of the knowledge enables inventors to communicate the mechanisms more clearly (Winter, 1998). Beyond clarifying the mechanisms underlying an invention, general and abstract knowledge can also enhance the understanding of how the invention is in fact related to seemingly unrelated applications (Bresnahan and Gambardella, 1998; Fleming, 2007).

On the other hand, "experimental" knowledge is knowledge obtained from trial-and-error procedures. A lack of general understanding about a problem often necessitates inventors to conduct costly experiments that tend to produce local solutions (Arora and Gambardella, 1994; Mowery and Rosenberg, 1982; Vincenti et al., 1990). Because of the local nature of experimental knowledge, it is difficult to communicate the mechanisms underlying an invention that draws on experimental knowledge (Arora and Gambardella, 1994; Von Hippel, 1994; Winter, 1998).

[9] While "science-based" does not always equate to "general and abstract" as defined in this study, I use "science-based" and "general and abstract" interchangeably throughout this study.

Therefore, inventions drawing on general and abstract knowledge can help inventors to identify a wider range of potential downstream applications and in turn find a licensee who can extract high returns from the invention. In other words, $\tilde{q} - q$ would be greater for inventions drawing on general and abstract knowledge than those drawing on more experimental knowledge.

Hypothesis 3. *The effect of reduced patent scope on licensing propensity is stronger for inventions that draw on general and abstract knowledge than those that do not.*

2.2. Firm characteristics

To successfully commercialize an invention, downstream complementary assets such as manufacturing, marketing, and distribution network are required. These assets can either lower production costs or increase the willingness to pay, or both. However, these downstream assets are often difficult to access through arms-length market contracts as they tend to be specialized to particular uses rather than being generic (Teece, 1986). Imitating these assets is also difficult because they are often built through interactions among different functions within a firm or across firms over an extended period of time (Arora and Ceccagnoli, 2006; Teece, 1992).

Large firms are more likely to own such complementary assets than small firms, implying that q is larger for large firms than small firms. Thus, large firms should be more efficient in exploiting a "larger market" afforded through a broad patent scope than small firms. In other words, $\tilde{q} - q$ would be larger for small firms than for large firms.

Hypothesis 4. *The effect of reduced patent scope on licensing propensity is weaker for large firms than for small firms.*

3. Methodology

The methodology used in this study relies on determining the degree to which focal inventions are similar to their prior-art inventions and the sequence in which the patent applications of the inventions are filed and disclosed to the public.[10] For each focal invention, the information on invention similarity along with priority and earliest publication date sequence is used to identify a prior-art patent on a highly similar invention that is disclosed to the public after the priority date of the focal invention ("priority disclosure"). Given that a prior-art patent would have a priority in claiming ownership over an invention, this disclosure event effectively reduces the scope of the focal patent. Thus, the publication of a priority disclosure is used to examine the effect of reduced patent scope on licensing propensity.

[10] As previously pointed out, because this study uses EPO's simple family designation to group patents into inventions, unless specified otherwise, the patent application and publication dates referred to in this study are priority and earliest publication dates rather than application and publication dates of individual patents. Priority date is the date of the first patent application within a simple patent family, which consists of related patents whose underlying invention is considered to be the same. The earliest publication date is the first time a patent or a patent application from a simple patent family is disclosed to the public.

The following sections provide the essential details of the steps taken to identify a *priority disclosure* for each focal invention in the sample.

3.1. Similarity between focal and prior-art inventions

To determine whether there exists a prior-art invention that is similar to a focal invention, a textual similarity score is calculated between a focal patent and each of its prior-art patents by comparing the claims text of the focal and its prior-art patents. Then, using this textual similarity measure and other patent-specific measures as predictors, a machine learning algorithm is trained to predict whether the claims text of the focal patent and its prior-art patent are similar enough for a patent examiner to issue a rejection letter, i.e. a letter from a patent examiner indicating which claims of a patent application do not meet novelty and non-obviousness requirements based on claims of prior-art patents.[11] (Appendix A provides more technical details on how the similarity measure is derived.)

First, for each patent, every word is extracted from the claims section of the patent document and stemmed to reduce ambiguity. Stop words along with punctuation are also removed in the process. An important step is reverting abbreviated words and acronyms back to their original states. Reversion is performed by mapping abbreviations to their original words or phrases using a pre-loaded library. The remaining abbreviations not mapped in the pre-loaded library are checked and manually reverted back to their original words or phrases. Each word stem is then weighted by term frequency-inverse document frequency (tf-idf) ratio to account for its specificity (Salton and Buckley, 1988). The idea is that a word that is used more frequently across all documents is not likely to be a technical word tied to a specific concept or invention but is likely to be a general word that is used broadly. These normalized and weighted word stems are used to construct a vector associated with each patent document.

Given the extracted word stems, a "technical distance" between each pair of the word stems is calculated. A major challenge in measuring similarity between any two words is that similar ideas can be expressed using different words. This challenge is likely to be more severe for patent documents because inventors and their patent agents often describe their inventions in different terms from those used in prior-art patent documents in order to demonstrate novelty, leading to an artificial distance between two different words bearing a similar meaning. To address this challenge, word-pair distances are computed using examiner rejection letters that relate conceptually similar patent documents and claims to reject patent applications for a lack of novelty or for obviousness. These distances account for the likelihood that two distinct words might refer to a similar technical concept.

[11] While a rejection letter can contain other reasons for a rejection of a claim, only the novelty and non-obviousness reasons are used to ensure that the claims of a focal patent application are similar to the claims of a prior-art patent.

Using the word-specific weights and the distances between word stem pairs, a similarity score between a pair of patent documents is computed. More specifically, to derive a textual similarity score between two patent documents, a cosine similarity score between the two vectors is calculated while taking into account term frequency-inverse document frequency ratio for each word stem and word stem-pair distances. The higher the score between the weighted vectors of two patent documents, the more textually similar the documents are.

Lastly, the textual similarity scores between focal patents and their prior-art patents are used along with patent-specific measures to train a machine learning algorithm to predict whether the claims of a focal patent and a prior-art patent would be deemed highly similar by a patent examiner. More specifically, I train a random forest algorithm (Breiman, 2001) with a dummy variable *Similar* as an outcome taking a value of 1 if a patent examiner finds the claims of a focal patent to be similar enough to those of a prior-art patent to issue a rejection letter. The predictors used in the model include priority and publication years, different types of citations (forward, backward, and NPL citations), patent family size, number of claims, patent stock, and technology area.

Figure 1 presents a receiver operator characteristics (ROC) curve that plots true positive rate against false positive rate of the predicted outcomes using a test set, i.e. a sample not used to train the algorithm but set aside to test the prediction rate of the algorithm. The area under the ROC curve, which measures the performance of the classifier, is approximately 0.87 indicating that, given a randomly chosen patent pair deemed similar by a patent examiner (a rejection) and a randomly chosen patent pair not deemed similar by a patent examiner (non-rejection), the algorithm is able to predict with 87% accuracy which patent pair is more likely to be a rejection.

3.2. Priority disclosures

To determine which of the predicted similar prior-art patents would unexpectedly reduce the scope of the focal patent, prior-art patents (which by definition have an earlier priority date than the focal patent) that are published after the priority date of the focal patent are identified. Among these similar prior-art patents, the first one disclosed after the priority date of the focal patent is designated as "priority disclosure" and its publication is used as an event that unexpectedly reduces the scope of the focal patent.

Figure 2 illustrates a situation that gives rise to a priority disclosure. The figure shows the priority and the earliest publication dates of USPTO patents 7386594 ("System and method related to generating an email campaign") and 7774408 ("Methods, systems, and emails to link emails to matters and organizations") which the machine learning algorithm predicts to be similar. The horizontal line running left to right represents time. The single vertical line above the horizontal line represents the priority date of patent 7774408 (a focal patent), and the vertical lines below the horizontal line represent the priority date and the earliest publication date of patent 7386594 (a prior-art patent).

To determine whether the invention covered by patent 7386594 is a priority disclosure of the invention covered by patent 7774408, the priority dates and the earliest publication dates of the patents are compared. Comparing the dates reveals that the priority date of patent 7386594 (April 25, 2000) falls before the priority date of patent 7774408 (April 23, 2001). It further shows that the earliest publication date of US 7386594 (November 1, 2001) comes after the priority date of patent 7774408 (April 23, 2001). Given this date sequence, patent 7386594 is designated as the priority disclosure of patent 7774408. (Patent 7386594 is the first similar prior-art patent disclosed after the priority date of patent 7774408.)

The illustrated sequence of dates for the focal and the prior-art patents gives rise to a situation in which the focal inventor finds out about a similar prior-art invention only after the first patent application for his invention is filed.

4. Validation of measures

The new similarity measure used in this study is validated by examining 1) the likelihood that two patents predicted to be similar are citation pairs and 2) the share of prior-art patents that are predicted to be similar for a patent that bears a new IPC (International Patent classification). To the extent that patent citation reflects common knowledge shared between citing and cited patents, the first validation would show the degree to which the new measure captures the common knowledge shared between a focal patent and its prior-art patent, and to the extent that a new IPC reflects novelty (Fleming, 2001), the second validation would show the degree to which the new measure captures novelty of an invention.

4.1. Citation pairs

Prior studies have characterized forward citations as an indication that either the citing patent builds on the knowledge contained in the cited patent or at the least that the patent pairs share some common knowledge (Arora et al., 2018; Jaffe et al., 1993). Thus, to the extent that patent citations reflect knowledge shared by two patents, comparing the share of predicted similar patents that are citation pairs with the share of predicted non-similar patents that are citation pairs would show the degree to which the predicted similarity captures the common knowledge shared between a focal invention and its prior-art invention.

Figure A1 presents the fraction of patent pairs that are linked by a citation among predicted similar patent pairs and among predicted non-similar patent pairs. The sample consists of USPTO patents with application years between 1990 and 2010 and top fifty prior-art patents for each focal patent in terms of the textual similarity scores (close to 112 million patent pairs). The figure shows that predicted similar patent pairs are around seven times as likely to be a citation pair as predicted non-similar patent pairs (2.19% for predicted similar patent pairs and 0.31% for predicted non-similar patent pairs), providing evidence that the new similarity measure captures the common knowledge shared between a focal invention and its prior-art invention.[12] [13]

4.2. New IPCs

The machine learning algorithm used in this study is trained to predict whether a patent examiner would judge that a focal patent application is similar to a prior-art patent and, as a result, fails to pass the novelty and non-obviousness requirements. Thus, a focal patent with few or no predicted similar prior-art patents would be considered more novel than a focal patent with many predicted similar prior-art patents. This idea is tested using international patent classification (IPC) maintained by World Intellectual Property Organization (WIPO). WIPO periodically and retrospectively reclassifies patents into newly created classes to reflect that the inventions covered by the patents were novel at the time they were invented. Thus, to the extent that a reclassified patent reflects novelty of an invention, it is expected that a patent with few or no predicted similar prior-art patents would be more likely to be a reclassified patent.

Figure A2 presents the fraction of *non-novel* focal patents that are reclassified into a new IPC versus the fraction of *novel* focal patents that are reclassified into a new IPC, where a *novel* patent is a patent in the bottom half of the similar prior-art patent share distribution and a *non-novel* patent is a patent in the top half of the similar prior-art share distribution. The sample consists of USPTO patents with application years between 1990 and 2010 along with the share of top fifty prior-art patents in terms of textual similarity that are predicted to be similar for each focal patent. The figure shows that patents in the bottom half of the similar prior-art patent share distribution are over twice as likely to be reclassified into a new IPC as those in the top half of the distribution (0.082% for those in the bottom half and 0.037% for those in the top half), providing evidence that the new similarity measure reflects invention novelty.[14] [15]

[12] Among the citation pairs, 91.69% are predicted to be similar, while 60.76% of the non-citation pairs are predicted not to be similar.

[13] Table A2 presents estimation results showing the relationship between a dummy variable indicating a citation pair and a dummy variable indicating that a focal and prior-art patent pair is predicted similar by the machine learning algorithm. The specification controls for key patent attributes, forward, backward, and NPL citations and patent stock as well as priority and earliest publication years, and patent fixed effects. The results show that a similar focal and prior-art patent pair is over one and a half times more likely than non-similar pairs to be a citation pair.

[14] Among the patents reclassified into a new IPC, 38.67% are in the bottom half of the similar prior-art patent share distribution, while 25.68% of patents not reclassified into a new IPC are in the bottom half of the similar prior-art patent share distribution.

[15] Table A3 presents estimation results showing the relationship between a dummy variable for a patent that is reclassified into a new IPC and the share of top fifty prior-art patents for each focal patent that are not predicted to be similar to the focal patent. The specification controls for key patent attributes as well as priority and earliest

5. Data

The sample used in this study is constructed by combining the 2016 version of the European Patent Office's Worldwide Patent Statistical Database (PatStat) with the patent licensing database from ktMINE. It is further supplemented with additional patent data from PatentsView and Thomson Innovation.[16]

The sample consists of patents with priority years falling between 1990 and 2010 that experience an unexpected reduction in patent scope (treatment group) and control patents matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and invention stock quartile (control group).[17] To remove double counting, patents are aggregated to the invention level using the simple family hierarchy defined by the European Patent Office and each family is referred to as an "invention". That is, when multiple patents cover a single invention, they are grouped together and counted only once. The resulting sample consists of around 1.3 million inventions in the treatment group and the equal number in the control group. 2,072 inventions (corresponding to 4,029 patents) are licensed out by 862 inventors.

Finally, in order to examine the effect of reduced patent scope before and after priority disclosures (i.e. patent scope reducing events), for each invention, the sample is expanded to include up to five years before and after the publication of its priority disclosure, resulting in 17,674,697 observations at the invention-year level.

5.1. Main variables

This section provides details on how the key variables and technology areas are defined. Table A1 reports an expanded list of variables and their definitions.

5.1.1. Dependent variable. The dependent variable for this study is a dummy variable indicating whether an invention was licensed out in a given year (i.e. takes a value of 1 if an invention was licensed out in a given year and 0 if it was not). Given that some of the inventions are licensed out multiple times within the pre- and post-treatment window, only the first instance of the licensing agreement is included in constructing the variable.[18]

---

publication years, 4-digit IPC, and patent assignee fixed effects. The results show that 10% increase in the share of non-similar prior-art patents is associated with around 6% increase at the mean in the probability of being reclassified into a new patent IPC.

[16] ktMINE maintains a database of publicly available patent licensing agreements. They collect information from SEC filings and news articles. Fosfuri et al. (2012) indicate that ktMINE's licensing database is more comprehensive than other commercial and non-commercial databases.

[17] The matching is done with replacement, by allowing an invention to be chosen multiple times as a control.

[18] To identify the year in which an invention was licensed out, patent families of the patents in the ktMINE database are identified by matching the patents to PatStat's Applications database (tls201), which contains a simple patent family identifier. Then, using the patent licensing date in the ktMINE database, only the first instance of licensing agreements for each invention is kept and merged with the main database described in the previous section.

5.1.2. Invention quality. The main measure of invention quality used in this study is the number of forward citations that an invention receives (Trajtenberg, 1990; Gambardella et al., 2008; Hall et al., 2005; Lanjouw and Schankerman, 2004). To account for truncation for patents published in the latter part of the years in the sample, only the first five years after the publication of a patent is used to count forward citations. Also, to account for the fact that the main analyses of this study are done at the invention (or simple patent family) level, patent-level forward citations are aggregated up to the invention level by keeping only the first instance of forward citations coming from patents in the same citing patent family.

5.1.3. Invention novelty. The measure of novelty used in this study is based on the component recombination measure from (Fleming, 2001). To construct this measure, all IPC subclasses associated with each of the focal patents are extracted from PatStat's IPC database. Then, for each patent, its IPC combination is compared to the combinations of IPCs that have been used in previously filed patents to determine the number of times the IPC combination has been used. More formally, the following is implemented: $C_i = \sum_1^{i-1} D$, where C_i is the IPC combination used in patent i and D is a dummy indicating whether a previously filed patent uses the same IPC combination as invention i.

5.1.4. General and abstract knowledge. The type of knowledge base examined in this study is general and abstract knowledge underlying inventions. The main measure of the degree of general and abstract knowledge underlying an invention is based on textual similarity between focal patents and university publications. First, each focal patent is compared to all university publications that were published in a scientific journal before the priority date of the focal patent to derive a document similarity score.[19] Then, a count variable is created to capture how many times each focal patent appears as one of the top one hundred textually similar patents across the publications. For instance, if focal patent F1 appears as one of the top one hundred textually similar patents for university publications P1 and P2, then the value assigned to F1 would be 2. In essence, this measure is the degree of research-based knowledge that an invention draws from.

5.1.5. Firm size. Prior studies have shown that ownership or access to downstream complementary assets (e.g. marketing, production facility, distribution network) enables inventing firms to be more efficient in commercializing their inventions (Teece, 1986; Arora and Ceccagnoli, 2006; Gans et al., 2000). Thus, the effect of patent scope on licensing would be less pronounced for firms with ownership of downstream complementary assets. Also, to the extent that such ownership is associated with firm size, firm size would interact with patent scope in a similar way.

[19] The same algorithm used to calculate patent document similarity is used without the part pertaining to patent examiner rejection letters.

In this study, *Firm size* is proxied by the number of inventions issued to a firm in the sample period. To build this measure, patent assignees file is downloaded from PatentsView, a database maintained by the USPTO and further cleaned to remove misspellings and ambiguity in firm names. Then, the number of inventions (i.e. patent families) issued to each assignee in the sample period is recorded.

Furthermore, given that prior studies have shown that business unit size is significantly associated with R&D (Cohen and Klepper, 1996), an alternative measure of *Firm size* is constructed as the number of inventions issued to a firm in a given technology area during the sample period.

5.1.6. Technology area. To examine the extent to which patent scope-licensing relationship varies across technology areas, patents are grouped into biotechnology, chemicals, communications, computers, electrical machines, mechanicals, medical technology, and pharmaceuticals based on PatStat's technology fields, which group IPC's representing similar technologies. Communications include basic communication process, digital communication, and communications. Chemicals include basic materials chemistry, chemical engineering, food chemistry, macromolecular chemistry and polymers, and organic fine chemistry. Computers include computer technology and semiconductors. Mechanicals include handling, machine tools, and mechanical elements. Biotechnology, electrical machines, medical technology, and pharmaceuticals do not have sub-fields as defined by PatStat.[20]

6. Non-parametric evidence

Table 1 presents summary statics for the main variables in the treatment group. Panel A of Table 1 includes only the licensed inventions while Panel B includes all inventions.

Comparing forward citations, a measure of invention quality, across the two tables shows that the licensed inventions receive many more citations than inventions in the full sample. The average number of forward citations is 18.8 with a standard deviation of 31.7 in the full sample whereas the average is 61.1 with a standard deviation of 117.5. This observation is consistent with the view that high quality inventions are more likely to be licensed than low quality inventions.

Comparing IPC combination familiarity, a measure of invention novelty, provides evidence consistent with the notion that downstream applications for novel inventions (i.e. those that are novel to the world including the inventor) are more likely to be outside of the inventor's operations than inside. For inventions in the full sample, the mean of IPC combination familiarity is 149.7 with a standard deviation of 832.0, and, for licensed inventions, the mean is 38.8 with a standard deviation of 280.5. (A lower value indicates higher novelty.)

[20] Figure A3 shows the distribution of licensed inventions across technology areas. Pharmaceuticals and medical technologies have the most active licensing activity followed by biotechnology, chemicals, and computers in the order listed. The category "Other" includes civil engineering, consumer goods, furniture, measurements, etc.

The number of university publications that a patent is close to is also higher for licensed inventions (mean of 141.6 and standard deviation of 453.5) than for inventions in the full sample (mean of 63.7 and standard deviation of 273.7). This is consistent with the view that an invention drawing on general and abstract knowledge is likely to help the inventor find a more efficient licensee than an invention that does not draw on such knowledge.

Table 2 presents mean comparisons of main variables used in the study using only the licensed inventions (Panel A) and using the full sample (Panel B). The invention-level averages are derived from the entire sample period, while firm-level averages are derived from pre-treatment periods. The comparisons show that the measures generally have similar distributions for the treatment and control groups. For the licensed inventions (Panel A), comparisons of invention-level variables show that the differences in means are between 10%-20% of the standard deviation of the corresponding variables, and comparisons for the firm-level variables do not show any statistically significant differences.

Figure 3 shows licensing patterns before and after the publication of a priority disclosure for the treated and control groups. X-axis represents years relative to the publication of a priority disclosure, and Y-axis represents the average number of licensing agreements entered into per 1,000 patents. The figure shows that the trend lines track similar paths until around the arrival of a priority disclosure, but diverge thereafter, with licensing for the treated inventions increasing at a significantly slower rate than for the control inventions. This pattern is further explored through parametric analysis in the following sections.[21]

Figures 4a-d compare changes in licensing propensity of treated inventions relative to control inventions after a priority disclosure is published (i.e. patent scope is reduced for the treated inventions) and across different invention and firm characteristics. Figure 4a compares changes in licensing propensity between high- and low-quality inventions. It shows that, after patent scope is reduced, licensing propensity for high-quality inventions goes down by 0.001 whereas licensing propensity for low-quality inventions does not change (both relative to controls), a result consistent with Hypothesis 1.

Figures 4b-4d further show patterns consistent with Hypotheses 2 through 4. They show that, after patent scope is reduced, licensing propensity goes down more for novel, science-based, and small-inventor inventions than for non-novel, non-science-based, and large-inventor inventions, respectively.

[21] Figure A4 shows licensing lags in years from the priority year of the focal inventions. The sample includes inventions that are licensed within ten years from the priority year of each invention. Licensing activity appears to peak between years 2 and 4, and half of the inventions get licensed by around year 3, largely consistent with findings in prior studies (Gans et al., 2008; Hegde and Luo, 2017).

7. Estimation results

This study examines the effect of reduced patent scope on licensing propensity of inventors as well as factors that might enhance or limit this relationship at the invention and firm levels and across technology areas. To test the relationship, this study employ a difference-in-differences approach at the invention-year level by exploiting an exogenous variation in patent scope with the pre- and post-treatment window set to five years.[22] The sample consists of about 1.3 million inventions whose priority years range from 1990 to 2010 and an equal number of controls matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and invention stock quartile.

7.1. The effect of patent scope on licensing

Prior studies have argued that broader patent scope is likely to lower transaction costs for licensing. For instance, by providing legal protection against expropriation, a broad patent allows the inventor to share relevant knowledge about his invention with a potential licensee and as a result decreases the potential licensee's cost of obtaining the necessary knowledge to assess the value of the invention (Arrow, 1962). Furthermore, by clarifying ownership claims, a broad patent can help licensing parties to avoid certain disagreements during negotiations (Gans et al., 2008). As a result, a broad patent might have a positive effect on licensing propensity.

At the same time, however, a broad patent can also increase the market power of the inventor, raising the inventor's expected returns from self-commercializing an invention. At its core, a broad patent allows the inventor to exclude others from using his patented invention. Furthermore, it is possible that a broad patent affords the inventor the extra time required to refine his invention without the threat of imitation and even provide the inventor enough time to acquire the complementary assets required for commercialization (Teece, 1986). Given these opposing forces, it is not clear whether a broad patent would increase the licensing propensity of inventors.[23]

Thus, in the baseline analysis, the following difference-in-differences specification is employed at the invention-year level to examine the effect of patent scope and licensing propensity of inventors:

$$Licensed_{i,t} = \beta_1 Treated_i \times Post_{i,t} + \beta_2 Post_{i,t} + \mathbf{Z}_{i,t}\boldsymbol{\gamma} + \boldsymbol{\sigma}_i + \boldsymbol{\tau}_t + \epsilon_{i,t}$$

Licensed is a dummy variable taking a value of 1 if an invention i is licensed out in year t restricted to five years before and five years after the publication of a priority disclosure. *Treated*

[22] The results are robust to restricting the pre- and post-treatment window to three years or one year.

[23] While some studies point out this tension (Arora and Ceccagnoli, 2006), others have argued that under specific conditions, the relationship between patent scope and licensing propensity could be positive (Gans et al., 2000) or even negative (Galasso and Schankerman, 2014). Empirical findings have also been mixed and often confined to specific settings, with some studies finding a positive relationship between patent scope and licensing (Anand and Khanna, 2000; Gans et al., 2000; Smith, 2001) and others finding the opposite or no significant relationship (Branstetter et al., 2006; Cassiman and Veugelers, 2002; Fink, 2005; Fosfuri, 2004; Shane, 2001).

is a dummy variable taking a value of 1 for invention i if the invention has a priority disclosure (i.e. patent scope is reduced), and 0 otherwise. $Post$ is a dummy variable taking a value of 1 for invention i in year t if t is within five years after the publication of a priority disclosure, excluding the year of the disclosure, and 0 if t within five years prior to the publication of a priority disclosure, including the year of the disclosure. $\mathbf{Z}_{i,t}$ is a vector of controls at the invention- and firm-year levels, including invention quality, firm size, the number of similar prior-art and non-prior-art inventions. $\boldsymbol{\sigma}_i$ and $\boldsymbol{\tau}_t$ represent invention and year fixed effects. $\epsilon_{i,t}$ is the *iid* error term. Standard errors are clustered at the invention level.

The coefficient of interest is β_1, which captures the differential change in licensing propensity between treated inventions whose patent scope has been reduced and control inventions whose patent scope has not been reduced. For instance, $\hat{\beta}_1 < 0$ would be consistent with the prediction that an inventor would be less likely to license out his invention when the scope of a patent on his invention is narrowed.

Table 3 presents the baseline results on the effect of reduced patent scope on licensing propensity of inventing firms. Column 1 includes no controls but has invention and year fixed effects. The results show that a reduction in patent scope on average leads to around 125% decline in licensing propensity for treated inventions relative to control inventions. To mitigate the potential concern that the effect of reduced patent scope (triggered by the publication of a priority disclosure) is confounded with effects of other unobserved events, robustness tests are conducted using smaller pre- and post-treatment time windows (i.e. three-year pre- and post-treatment time window, and one-year pre- and post-treatment time window), and the results are reported in Columns 1 and 2 of Table A4. The results are robust to these changes in the pre- and post-treatment time window.[24]

Column 2 includes a control for invention quality, proxied by forward citation, which prior studies have shown to be associated with the likelihood that an invention gets licensed (Astebro, 2003; Gambardella et al., 2008; Harhoff et al., 1999). While the coefficient invention quality is positive, it is not statistically significant.

Column 3 adds controls for firm characteristics, i.e. firm size, firm age, and firm licensing experience, which can influence the likelihood that an invention is licensed. For instance, access to downstream complementary assets required for commercialization, which large firms (proxied by invention stock) are more likely to have than small firms, would encourage firms to self-commercialize rather than license. The results show that, once these controls are added, reduced patent scope

[24] Columns 9 and 10 of Table A4 also report results using only the inventions from assignees that have ever licensed during the sample period and using only the licensed inventions. The results are robust to these different sample selections.

leads to over 200% decrease in licensing propensity for the treated inventions relative to control inventions, a substantial increase in the magnitude of the effect.

Furthermore, inventors located in prominent technology centers such as Silicon Valley, are likely to have more established licensing networks and contacts with intermediaries (e.g. venture capitals) than those not located in these regions with concentrated inventing and licensing activities (Cunningham, 2017; Gans et al., 2008; Saxenian, 1990). Thus, to account for having access to such networks that might influence licensing, Column 4 includes shares of patent inventors located in Silicon Valley, Route 128, or Canada for an inventor in a given year. The coefficient is positive and statistically significant for *Silicon Valley inventors* and for *Canada inventors*, but negative and statistically significant for *Route 128 inventors*. The results continue to show that reduced patent scope leads to over 200% decline in licensing propensity.

Prior studies have shown that technology market competition is likely to lead to an increase in licensing activity for technology producers (Arora and Fosfuri, 2003; Fosfuri, 2006). Thus, Column 5 controls for competition in the technology market for each of the focal inventions by adding similar prior-art and non-prior-art inventions. As expected, similar prior-art and non-prior-art inventions both have a positive relationship with licensing propensity. The results continue to show that reduced patent scope has a string effect on licensing as it leads to over 200% decline in licensing propensity relative to controls.[25]

A key assumption of the empirical setup in this study is that inventors working on a specific invention do not know about similar inventions generated by other inventors until the similar inventions are disclosed to the public by a patent issuing authority. If inventors know about similar inventions that will be disclosed to the public in the future, it is possible that they try to license their inventions prior to the disclosure of the similar inventions to avoid losing profits from licensing their inventions. This situation could lead to an artificial increase in licensing activity before and in turn a decline after the disclosure of a similar invention. To test for this possibility, the main analysis is conducted again using a sample that consists of focal and prior-art invention pairs whose inventors are located in different states and a separate sample where the inventions are assigned to different 4-digit IPCs.

Columns 6 and 7 show that the results are robust to using a sample where inventors of a focal invention and a priority disclosure are located in different states and a separate sample where a focal invention and a priority disclosure are assigned to different 4-digit IPCs. The results show that reduced patent scope leads to around 220% decline in licensing propensity relative to

[25] Table A8 presents a relationship between patent scope reduction and licensing propensity of inventions (Panel A) and patents (Panel B). The results are consistent with the main findings reported in this section. That is, there is a positive association between patent scope and licensing propensity.

control inventions when using a sample with inventors located in different states, and 186% decline relative to control inventions when using a sample with inventions assigned to different 4-digit IPCs. The magnitudes of the effects are similar to the effects found Column 5, mitigating potential concerns that inventors somehow know about similar inventions generated by other inventors before the inventions are disclosed to the public and respond to such knowledge by licensing out their inventions before the similar inventions are disclosed.

7.2. Heterogeneous effects of patent scope on licensing

The following difference-in-differences specification is employed to explore firm- and invention-level factors that might limit or enhance the patent scope-licensing relationship:

$$Licensed_{i,t} = \beta_1 C_i \times Treated_i \times Post_{i,t} + \beta_1 Treated_i \times Post_{i,t} + \beta_2 Post_{i,t} + \mathbf{Z}_{i,t}\boldsymbol{\gamma} + \boldsymbol{\sigma}_i + \boldsymbol{\tau}_t + \epsilon_{i,t}$$

In addition to the variables described for the baseline specification, C_i denotes invention and firm characteristics. The coefficient of interest is β_1 where $\hat{\beta}_1 < 0$ would indicate a stronger effect of patent scope on licensing for higher values of C_i. For instance, when examining how the effect of reduced patent scope on licensing varies across invention quality, $\hat{\beta}_1 < 0$ would indicate that the effect is stronger for high-quality inventions than for low-quality inventions.

7.2.1. Invention characteristics. Table 4 presents results on how the relationship between patent scope and licensing varies across invention characteristics. Columns 1 and 2 examine the effect of reduced patent scope on licensing propensity across different levels of invention quality. Column 1 interacts a continuous measure of invention quality (i.e. number of forward citations received by an invention within five years from its earliest publication year) with $Treated \times Post$, and Column 2 interacts a dummy variable indicating a high-quality invention with $Treated \times Post$, where the dummy variable takes a value of 1 for inventions in the top half of the forward citation distribution, and a value of 0 for those at the bottom half. The results show that the effect of reduced patent on licensing propensity is stronger for high-quality inventions than for low-quality inventions. In particular, Column 2 shows that the effect is over 3.5 times stronger for high-quality inventions than for low-quality inventions. These results confirm Hypothesis 1 and are consistent with the findings from prior studies showing that high-quality inventions are much more likely to be licensed than low-quality inventions.

To test the hypothesis that the effect of reduced patent scope on licensing is stronger for novel inventions than for non-novel inventions (Hypothesis 2), Columns 3 interacts a continuous measure of novelty (i.e. number of times a specific IPC combination has been used prior to the priority date of the focal invention) with $Treated \times Post$, and Column 4 interacts a dummy variable indicating a novel invention with $Treated \times Post$, where the dummy variable takes a value of 1 for inventions in the top half of the IPC combination usage count distribution, and a value of 0 for those in the bottom half. The results from Colum 4 show that the effect of reduced patent scope on licensing propensity of novel inventions is over 3.5 times stronger than it is for non-novel inventions. These results confirm Hypothesis 2 and is consistent with the notion that valuable application for a novel invention is more likely to be found outside of the inventor's boundaries than inside.

Columns 5 and 6 test the hypothesis that the effect of reduced patent scope on licensing propensity would be stronger for inventions drawing on general and abstract knowledge (i.e. science-based inventions) than for inventions drawing on more applied knowledge. Column 5 interacts a continuous measure capturing the degree of general and abstract knowledge embodied in an invention (i.e. number of times an invention is textually similar to university publications published prior to the priority date of the invention) with $Treated \times Post$, and Column 6 interacts a dummy variable indicating a high degree of general and abstract knowledge embodied in an invention with $Treated \times Post$, where the dummy variable takes a value of 1 for inventions in the top half of the close university publication count distribution, and a value of 0 for those in the bottom half. Consistent with Hypothesis 3, the results from Column 6 show that the effect of reduced patent scope on licensing is around 1.5 times stronger for inventions drawing on general and abstract knowledge than for inventions drawing on more applied knowledge.

7.2.2. Inventor characteristics. Inventor size is an important dimension that is likely to moderate the patent scope-licensing relationship. To the extent that inventor size implies the inventor's ability to access specialized downstream assets, inventor size should facilitate commercialization of inventions through internal resources. For small inventors, their reliance on patent scope would be more essential because they are less likely to have alternative ways to capture value from their inventions. Accordingly, this section explores the moderating effect of inventor size on the patent scope-licensing relationship (Hypothesis 4).[26]

Table 5 presents the estimation results on the effect of patent scope on licensing for small and large inventors. Column 1 interacts a continuous measure of inventor size (the size of inventor's invention stock) with $Treated \times Post$, and Column 2 interacts a dummy variable indicating a large inventor with $Treated \times Post$, where the dummy variable takes a value of 1 for an invention from an inventor in the top half of the invention count distribution, and a value of 0 for an invention from an inventor in the bottom half. Consistent with Hypothesis 4, the positive and statistically significant coefficients in both Columns 1 and 2 show that the effect of reduced patent scope on

[26] While inventor size would be better measured using sales, due to a lack of information on sales for many of the inventors in the sample over the twenty-year period, invention stock is used as a proxy for inventor size. In a test at the end of this section, inventor size is alternatively proxied by a threshold for the number of employees set by the USPTO to define "small entities" and "micro entities".

licensing propensity is weaker for inventions from large inventors than it is for inventions from small inventors. Specifically, Column 2 shows that the effect leads to around 68% increase in licensing propensity for inventions from large inventors relative to those from small inventors.[27]

Prior studies have found that business unit size is highly correlated with R&D (Cohen and Klepper, 1996), and thus Columns 3 and 4 use invention stock at the firm-technology area level to proxy business unit size and examine the moderating effect of business unit size on the patent scope-licensing relationship. Column 3 interacts a continuous measure of business unit size with $Treated \times Post$, and Column 4 interacts a dummy variable indicating a large business unit with $Treated \times Post$, where the dummy variable takes a value of 1 for inventions generated by business units in the top half of the invention stock distribution, and a value of 0 for inventions generated by business units in the bottom half. Consistent with the results found using invention stock at the firm level, the results show that reduced patent scope leads to around 63% increase in licensing propensity for inventions generated by large business units relative to those generated by small business units.[28]

Lastly, Column 5 reports results from an analysis comparing the effect of reduced patent scope on licensing across individual and non-individual inventors.[29] Given that individual inventors are likely to be among the smallest of inventors (individuals and firms combined), it is expected that the effect of reduced patent scope would be stronger for individual inventors than for non-individual inventors.

To test this prediction, a dummy variable is constructed and is assigned a value of 1 for a non-individual inventor and 0 for an individual inventor and is interacted with $Treated \times Post$. Consistent with the prediction, the results show that reduced patent scope leads to around 41% increase in licensing propensity for non-individual inventors relative to individual inventors.

[27] Columns 1 and 2 of Table A5 report results using *Firm age* as an alternative measure proxying for access to downstream resources. The results continue to hold and show that reduced patent scope leads to around 88% increase in licensing propensity for inventions generated by mature inventors relative to inventions generated by young inventors.

[28] To mitigate the potential concern that invention stock is not a good proxy for inventor size, the effect of reduced patent scope on licensing propensity is examined using inventor size defined by the USPTO for the purposes of maintenance fee payments. USPTO designates an inventor as "small entities" if it has up to 500 employees and as "micro entity" if it has up to 500 employees, has not filed more than four patent applications, and does not have an income exceeding three times the amount of the U.S. median household income. A dummy variable is constructed and assigned a value of 1 if the maintenance fee information indicates that an invention's owner did not claim a small or micro entity status, or a value of 0 if the owner did claim such status. In Column 3 of Table A5, the dummy variable is interacted with $Treated \times Post$. The results continue to hold, showing that the effect of reduced patent scope leads to around 67% increase in licensing propensity for inventions generated by large inventors relative to inventions generated by small inventors.

[29] Individual inventors are identified based on the DWPI assignee code field in the Thomson Innovation's patent database and the assignee type code in PatentsView. Additional cleaning and manual designation of individually owned patents are performed as required.

7.3. Technology areas

Prior studies have shown that the effectiveness of patents might vary across industries (Anand and Khanna, 2000) and that different inventors might use patents for different purposes (Cohen et al., 2000; Somaya, 2012). For instance, Cohen et al. (2000) highlight that, in complex product industries where a typical product embodies many patantable technologies owned by multiple firms, one of the main reasons that firms patent is to gain bargaining power in negotiations with other firms. On the other hand, in discrete product industries where a typical product is composed of a small number of patentable technologies owned by few firms or a single firm, patents are often used to prevent rivals from patenting substitute inventions. The importance of the composition of patent ownership in affecting licensing has also been emphasized by other studies (Galasso and Schankerman, 2010).

To understand how the effect of patent scope on licensing varies across technology areas, the patent scope-licensing relationship is examined separately for inventions in each of the technology areas explored in this study: biotechnology, chemicals, communications, computers, electrical machines, mechanicals, medical technology, and pharmaceuticals. Table 6 presents the results on how the effect of patent scope on licensing varies across technology areas. The results show that the effect is the strongest in biotechnology, followed by communications and chemicals.

7.4. Robustness tests

This section presents several robustness tests to mitigate concerns relating to sample selection, control matching, and other potential biases. The results from these tests show that the main findings of the study are robust to these additional tests.

7.4.1. Cross-licensing. Prior studies have documented that inventors enter into cross-licensing agreements often when each inventor has patent rights to inventions that the other inventor needs. Basically, a cross-license grants each inventor involved in the agreement the right to use the other inventor's patented inventions. However, there is still a wide variation in the specific terms of cross-licensing agreements, and the cross-licensing rates are likely to vary across industries. For instance, prior studies have found that cross-licensing agreements are likely to be more prevalent in complex product markets (Cohen et al., 2000; Giuri and Torrisi, 2010). Importantly, cross-licensing agreements might be limited to already-patented inventions or include future inventions (Shapiro, 2000), and they can arise as a result of a settlement of a court litigation (Anand and Khanna, 2000).

Column 3 of Table A4 reports results of the analysis after excluding cross-licensing agreement from the sample. It shows that the main results continue to hold even after excluding inventions associated with cross-licensing agreements, with reduced patent scope leading to over 160% decline in licensing propensity relative to control inventions.

7.4.2. For-profit and non-profit organizations. Prior studies have shown that licensing patterns of non-profit organizations such as universities and medical schools are likely to be different from for-profit organizations. This difference could be a result of university licensing relying less on patent protection than licensing by for-profit organizations or could be a result of different objectives that non-profit organizations have compared to for-profit organizations (Jensen and Thursby, 2001; Thursby et al., 2001).[30]

Columns 4 and 5 on Table A4 present results on the effect of reduced patent scope on licensing propensity for for-profit organizations and for non-profit organizations (such as universities, government entities, medical schools, and public research centers), respectively.[31] As expected, the effect of reduced patent scope on licensing is weaker for non-profit organizations (157% decline in licensing propensity for treated inventions relative to control inventions) than it is for for-profit organizations (223% decline in licensing propensity for treated inventions relative to control inventions). These results are consistent with the notion that universities and other public research institutes do not rely on patents as heavily as for-profit organizations do, and also the notion that licensing by non-profit organizations is driven less by profit motives than it is for for-profit organizations.

7.4.3. American Inventor's Protection Act (AIPA). One of the provisions of American Inventor's Protection Act (AIPA) of 1999, which took effect in November of 2000, was that patent applications be disclosed to the public eighteen months from their priority date (the earliest application date among patents that relate to the same invention, i.e. simple patent family) rather than when the patent is issued. Therefore, in the sample explored in this study, a priority disclosure (i.e. a prior-art invention that reduces the scope of the focal patent) whose priority date falls before AIPA is likely to be an issued patent while a priority disclosure whose priority date comes after AIPA is likely to be a patent application before a patent is issued. Thus, the effect of a priority disclosure before and after AIPA might be different.

Columns 6 and 7 on Table A4 present results on the effect of reduced patent scope on licensing propensity for inventions whose priority disclosure has a priority date comes before AIPA and the effect on licensing propensity for inventions whose priority disclosure has a priority date coming after AIPA, respectively. Consistent with the main findings, the results show that the effect of reduced patent scope is stronger for treated inventions than for control inventions for both pre- and post-AIPA time periods. (Reduced patent scope leads to around 161% decline in licensing

[30] Thursby et al. (2001) find in a survey-based study that almost half of university licensing occurs when an invention is only a proof of concept and that universities are driven less by profits as their goals include making broader economic contributions.

[31] I use a combination of string search and PatentsView's assignee type field to identify public organizations.

propensity relative to controls during the pre-AIPA period, and the effect of reduced patent scope on licensing propensity is driven entirely by treated inventions in the post-AIPA period).

7.4.4. Publication and issuance of focal patents. Hegde and Luo (2017) argue that publication of the focal inventions by credible intermediaries such as the USPTO promotes licensing by reducing search costs for both potential licensors and licensees. Also, Gans et al. (2008) argue that the notice of patent allowance spurs licensing by lowering information asymmetries, search costs, and expropriation risks.[32] Thus, a publication or an issuance of a patent for the invention invention is likely to influence licensing patterns of inventors.[33]

Column 8 on Table A4 presents the results from excluding inventions for years that come after the earliest publication year of of the focal invention. While the effect is weaker than it is in the baseline results (166% decline in licensing propensity relative to controls), it remains negative and large (81% decline in licensing propensity relative to controls).

7.4.5. Patent-level analysis. As previously described, this study aggregates patents into inventions using EPO's definition of simple patent family. In doing so, various patent level measures are also aggregated to the inventions level. For instance, textual similarity score between a focal patent and a prior-art patent is aggregated up to the invention level by choosing the highest score between any two patents, each belonging to the patent family of the focal and the prior-art patents.

To mitigate concerns that aggregation of the sample and the patent-level measures potentially introduce a systematic bias that lead to spurious results, the main analysis of the study is replicated using a patent-level sample in Table A6. Column 5, which controls for 5-year forward citations, firm characteristics, and competition, shows that reduced patent scope leads to over 150% reduction in licensing propensity for treated inventions relative to control inventions. These results provide support for the main findings in the study.

7.4.6. Propensity score matching. To address the potential concern that the manual matching method used in this study is not sufficient to pair each focal invention with a closely resembling prior-art invention, an alternative method is used to check the robustness of the results. To do so, each focal invention is first matched on priority year, earliest publication year, and 4-digit IPC. Then, within the cells created by matching on these three characteristics, propensity scores are used to match on five-year forward citations (proxying for quality) and firm invention stock (proxying for firm size).

[32] The publication date and issuance date of a patent would be the same in most cases pre-AIPA, but the dates can diverge after the eighteen-month rule from AIPA went into effect in November of 2000.

[33] While the potential bias introduced by these events are likely to be conservative, I try to untangle the effects of these events for completeness and any potentially unexpected effects these events might have on the analysis in this study.

Table A7 presents the baseline results using propensity score matching to select controls. The results (Column 5) show that, after controlling for invention quality, firm characteristics, and competition, reduced patent scope leads to around 148% decline in licensing propensity for treated inventions relative to control inventions.

8. Conclusion

Understanding the factors that influence inventors' decision to compete or cooperate is a key to understanding when existing market structure might be disrupted or reinforced. While several studies have explored how patents (or other forms of intellectual property rights) influence such decisions, they have often been confined to specific industries and settings. Using a large sample of licensed inventions across multiple technology areas, this study presents more generalizable findings about how inventors react to unexpected reduced patent scope with regard to licensing. The findings show that on average inventors refrain from licensing out their inventions when patent scope is reduced.

Also, this study provides insights into the extent to which the patent scope-licensing relationship is moderated by invention characteristics. While prior studies have emphasized various industry and firm characteristics that limit or enhance the patent scope-licensing relationship, the role of invention characteristics has not been closely examined. The findings from this study show that invention characteristics do indeed matter in moderating the effect of patent scope on licensing propensity. In particular, inventors generating high-quality, novel, and science-based inventions are found to be more responsive to changes in patent scope in terms of licensing.

Furthermore, this study presents results that support prior studies emphasizing the importance of downstream complementary assets in commercializing an invention. The results are consistent with the view that small inventors that lack other avenues through which they can capture value from their inventions rely more heavily on broader patents when they license their inventions. The results are also consistent with the idea that downstream assets required for commercialization are often difficult to acquire because they are developed over an extended period of time through close interactions among relevant parties. Given that large and mature inventors are more likely to have access to the necessary downstream assets for commercialization, they tend to be less reliant on broader patent scope for licensing.

Lastly, this study makes a methodological improvement over prior studies, which have often struggled to find an exogenous variation in patent scope (or other dimensions of patent effectiveness) to deal with potential endogeneity issues. To examine the patent scope-licensing relationship, this study exploits an exogenous variation in patent scope triggered by unexpected disclosure of a highly similar prior-art invention that effectively reduces the scope of the focal patent.

There are several limitations to this study as well as potential avenues for future studies. Most notably, the current empirical setup cannot distinguish the channels through which patent scope operates in influencing the licensing strategies of inventors. More specifically, while broader patent scope promotes licensing, it might do so by increasing the value of an invention or by lowering transaction costs. From the current setup, it is not clear which channel (or channels) patent scope operates through. A future study should try to isolate the specific channels through which patent scope influences licensing and in turn inventors' decision to either compete or cooperate.

Another interesting avenue for future research would be to explore what happens to inventions that are not licensed. While this study provides insights into potential conditions under which inventors refrain from licensing out their inventions, it does not go further to show what eventually happens to those inventions that are not licensed. It is possible that these inventions are developed internally and later commercialized. However, it is also possible that they never get commercialized or monetize in some other way. Having insights into what eventually happens to these inventions would provide deeper understanding of inventors' behaviors with respect to inter-firm cooperation and competition.

References

Abernathy WJ, Clark KB (1985) Innovation: Mapping the winds of creative destruction. *Research policy* 14(1):3–22.

Agrawal A, Cockburn I, Zhang L (2015) Deals not done: Sources of failure in the market for ideas. *Strategic management journal* 36(7):976–986.

Anand BN, Khanna T (2000) The structure of licensing contracts. *The Journal of Industrial Economics* 48(1):103–135.

Anton JJ, Yao DA (1994) Expropriation and inventions: Appropriable rents in the absence of property rights. *The American Economic Review* 190–209.

Arora A (1996) Contracting for tacit knowledge: the provision of technical services in technology licensing contracts. *Journal of Development Economics* 50(2):233–256.

Arora A, Belenzon S, Lee H (2018) Reversed citations and the localization of knowledge spillovers. *Journal of Economic Geography* 18(3):495–521.

Arora A, Ceccagnoli M (2006) Patent protection, complementary assets, and firms incentives for technology licensing. *Management Science* 52(2):293–308.

Arora A, Cohen W (2017) Firm size, innovation and the effectiveness of dod support to corporate rd. *DoD Report* .

Arora A, Fosfuri A (2003) Licensing the market for technology. *Journal of Economic Behavior & Organization* 52(2):277–295.

Arora A, Fosfuri A, Gambardella A (2001) Markets for technology and their implications for corporate strategy. *Industrial and corporate change* 10(2):419–451.

Arora A, Gambardella A (1994) The changing technology of technological change: general and abstract knowledge and the division of innovative labour. *Research policy* 23(5):523–532.

Arora A, Gambardella A (2010) The market for technology. *Handbook of the Economics of Innovation*, volume 1, 641–678 (Elsevier).

Arrow KJ (1962) Economic welfare and the allocation of resources for invention. *The Rand Corporation* .

Astebro T (2003) The return to independent invention: Evidence of unrealistic optimism, risk seeking or skewness loving? *The Economic Journal* 113(484):226–239.

Branstetter LG, Fisman R, Foley CF (2006) Do Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from U. S. Firm-Level Panel Data*. *The Quarterly Journal of Economics* 121(1):321–349, ISSN 0033-5533, URL `http://dx.doi.org/10.1093/qje/121.1.321`.

Breiman L (2001) Random forests. *Machine learning* 45(1):5–32.

Bresnahan T, Gambardella A (1998) 10 the division of inventive labor and the extent of the market. *General purpose technologies and economic growth* 253.

Cassiman B, Veugelers R (2002) R&d cooperation and spillovers: some empirical evidence from belgium. *American Economic Review* 92(4):1169–1184.

Christensen CM (1997) The innovator's dilemma: When new technologies cause great firms to fail harvard business school press. *Boston, MA* .

Cohen WM, Klepper S (1996) A reprise of size and r & d. *The Economic Journal* 106(437):925–951.

Cohen WM, Nelson RR, Walsh JP (2000) Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not). Technical report, National Bureau of Economic Research.

Cunningham C (2017) When does novelty pay? *Academy of Management Proceedings*, volume 2017, 16957 (Academy of Management Briarcliff Manor, NY 10510).

Dushnitsky G, Lenox MJ (2005) When do firms undertake r&d by investing in new ventures? *Strategic Management Journal* 26(10):947–965.

Eesley CE, Hsu DH, Roberts EB (2014) The contingent effects of top management teams on venture performance: Aligning founding team composition with innovation strategy and commercialization environment. *Strategic Management Journal* 35(12):1798–1817.

EPO (2019) Docdb simple patent family. `https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families/docdb.html`, accessed: 2019-07-31.

Fink C (2005) Intellectual property rights and us and german international transactions in manufacturing industries. *Intellectual Property Rights and Development: Lessons from Recent Economic Research* 75–110.

Fleming L (2001) Recombinant uncertainty in technological search. *Management science* 47(1):117–132.

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

Fosfuri A (2004) Determinants of international activity: evidence from the chemical processing industry. *Research Policy* 33(10):1599–1614.

Fosfuri A (2006) The licensing dilemma: understanding the determinants of the rate of technology licensing. *Strategic Management Journal* 27(12):1141–1158.

Fosfuri A, Helmers C, Roux C (2012) Are joint patents collusive? evidence from the us and europe .

Galasso A, Schankerman M (2010) Patent thickets, courts, and the market for innovation. *The RAND journal of economics* 41(3):472–503.

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

Gambardella A (2013) The economic value of patented inventions: Thoughts and some open questions. *International Journal of Industrial Organization* 31(5):626–633.

Gambardella A, Giarratana MS (2013) General technological capabilities, product market fragmentation, and markets for technology. *Research Policy* 42(2):315–325.

Gambardella A, Giuri P, Luzzi A (2007) The market for patents in europe. *Research Policy* 36(8):1163–1183.

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

Gans JS, Hsu DH, Stern S (2000) When does start-up innovation spur the gale of creative destruction? Technical report, National bureau of economic research.

Gans JS, Hsu DH, Stern S (2008) The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. *Management science* 54(5):982–997.

Gans JS, Stern S (2003) The product market and the market for ideas: commercialization strategies for technology entrepreneurs. *Research policy* 32(2):333–350.

Giuri P, Torrisi S (2010) Cross-licensing, cumulative inventions, and strategic patenting. *Draft paper for the 5th Annual Conference of the EPIP Association, Maastricht*.

Hall BH, Jaffe A, Trajtenberg M (2005) Market value and patent citations. *RAND Journal of economics* 16–38.

Harhoff D, Narin F, Scherer FM, Vopel K (1999) Citation frequency and the value of patented inventions. *Review of Economics and statistics* 81(3):511–515.

Hegde D, Luo H (2017) Patent publication and the market for ideas. *Management Science* 64(2):652–672.

Hellmann T (2007) The role of patents for bridging the science to market gap. *Journal of Economic Behavior & Organization* 63(4):624–647.

Henderson RM, Clark KB (1990) Architectural innovation: The reconfiguration of existing. *Administrative science quarterly* 35(1):9–30.

Hsu DH (2006) Venture capitalists and cooperative start-up commercialization strategy. *Management Science* 52(2):204–219.

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics* 108(3):577–598.

Jensen R, Thursby M (2001) Proofs and prototypes for sale: The licensing of university inventions. *American Economic Review* 91(1):240–259.

Kline SJ, Rosenberg N (1986) An overview of innovation. the positive sum strategy: Harnessing technology for economic growth. *The National Academy of Science, USA* .

Lamoreaux NR, Sokoloff KL (2002) Intermediaries in the us market for technology, 1870-1920. Technical report, National Bureau of Economic Research.

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

Lemley MA, Shapiro C (2005) Probabilistic patents. *Journal of Economic Perspectives* 19(2):75–98.

Lemley MA, Shapiro C (2006) Patent holdup and royalty stacking. *Tex. L. Rev.* 85:1991.

Levin RC, Klevorick AK, Nelson RR, Winter SG, Gilbert R, Griliches Z (1987) Appropriating the returns from industrial research and development. *Brookings papers on economic activity* 1987(3):783–831.

Mansfield E (1986) Patents and innovation: an empirical study. *Management science* 32(2):173–181.

Marx M, Gans JS, Hsu DH (2014) Dynamic commercialization strategies for disruptive technologies: Evidence from the speech recognition industry. *Management Science* 60(12):3103–3123.

Merges RP (2006) Software and patent scope: A report from the middle innings. *Tex. L. Rev.* 85:1627.

Merges RP, Nelson RR (1990) On the complex economics of patent scope. *Columbia Law Review* 90(4):839–916.

Mowery DC, Rosenberg N (1982) The commercial aircraft industry. *Government and Technological Progress, Pergamon Press, New York* .

Nelson RR (1959) The simple economics of basic scientific research. *Journal of political economy* 67(3):297–306.

Nerkar A, Shane S (2007) Determinants of invention commercialization: An empirical examination of academically sourced inventions. *Strategic Management Journal* 28(11):1155–1166.

Rosenberg N (1998) Uncertainty and technological change. *The economic impact of knowledge* 17–34.

Rosenkopf L, Nerkar A (2001) Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal* 22(4):287–306.

Rothaermel FT (2001) Incumbent's advantage through exploiting complementary assets via interfirm cooperation. *Strategic management journal* 22(6-7):687–699.

Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. *Information processing & management* 24(5):513–523.

Saxenian A (1990) Regional networks and the resurgence of silicon valley. *California management review* 33(1):89–112.

Schumpeter JA (1943) *Capitalism, Socialism and Democracy* (New York: Harper Row).

Shane S (2001) Technological opportunities and new firm creation. *Management science* 47(2):205–220.

Shapiro C (1985) Patent licensing and r & d rivalry. *The American Economic Review* 75(2):25–30.

Shapiro C (2000) Navigating the patent thicket: Cross licenses, patent pools, and standard setting. *Innovation policy and the economy* 1:119–150.

Smith PJ (2001) How do foreign patent rights affect us exports, affiliate sales, and licenses? *Journal of International Economics* 55(2):411–439.

Somaya D (2012) Patent strategy and management: An integrative review and research agenda. *Journal of management* 38(4):1084–1114.

Stigler GJ (1951) The division of labor is limited by the extent of the market. *Journal of political economy* 59(3):185–193.

Teece DJ (1981) The market for know-how and the efficient international transfer of technology. *The Annals of the American Academy of Political and Social Science* 458(1):81–96.

Teece DJ (1986) Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research policy* 15(6):285–305.

Teece DJ (1992) Competition, cooperation, and innovation: Organizational arrangements for regimes of rapid technological progress. *Journal of economic behavior & organization* 18(1):1–25.

Thompson N, Kuhn JM (2017) Does winning a patent race lead to more follow-on innovation? *Available at SSRN 2899088* .

Thursby JG, Jensen R, Thursby MC (2001) Objectives, characteristics and outcomes of university licensing: A survey of major us universities. *The journal of Technology transfer* 26(1-2):59–72.

Trajtenberg M (1990) A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics* 172–187.

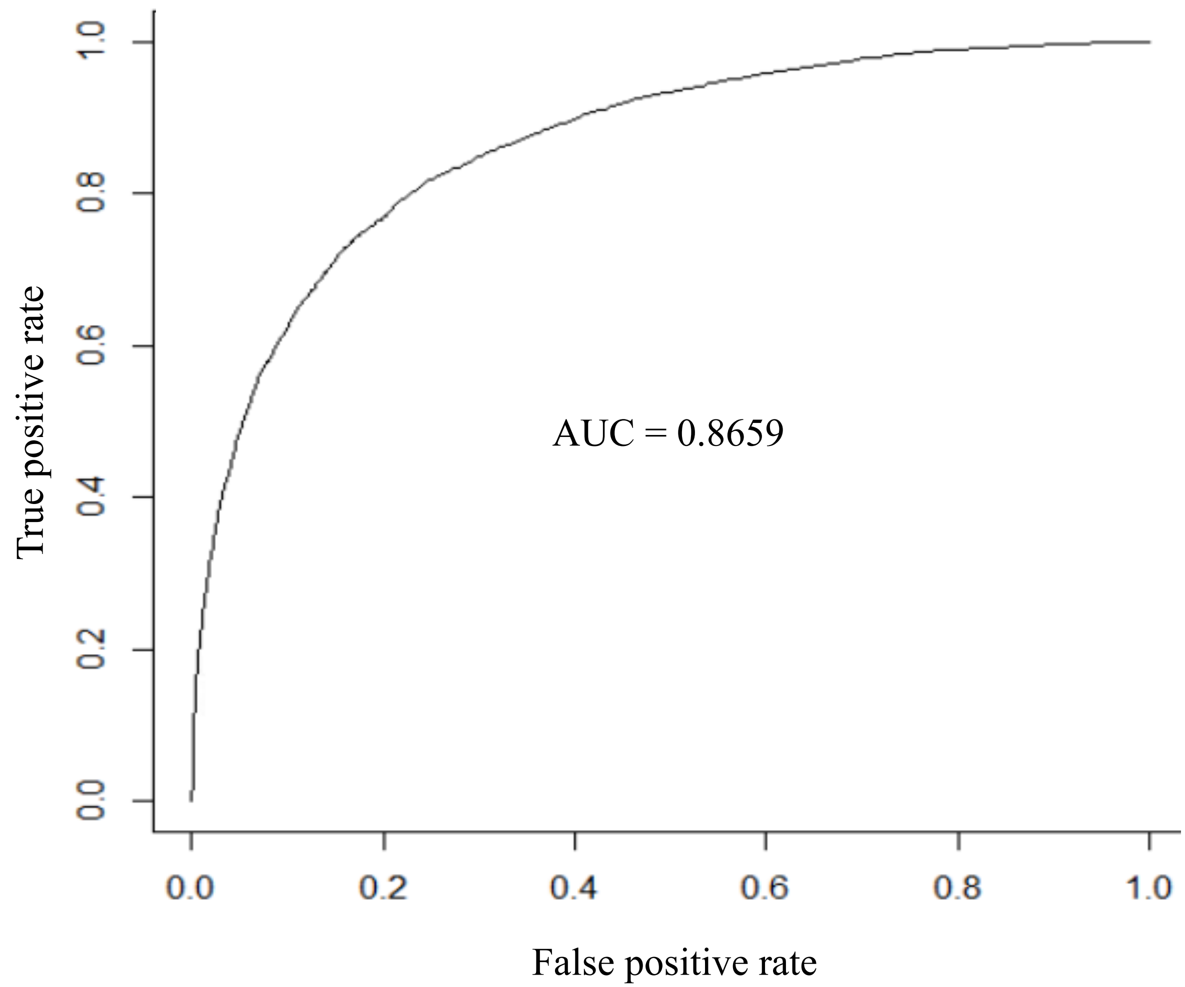
Tushman ML, Anderson P (1986) Technological discontinuities and organizational environments. *Administrative science quarterly* 439–465.

Vincenti WG, et al. (1990) *What engineers know and how they know it*, volume 141 (Baltimore: Johns Hopkins University Press).

Von Hippel E (1994) sticky information and the locus of problem solving: implications for innovation. *Management science* 40(4):429–439.

Winter S (1998) Knowledge and competence as strategic assets. *The strategic management of intellectual capital* 187:37.

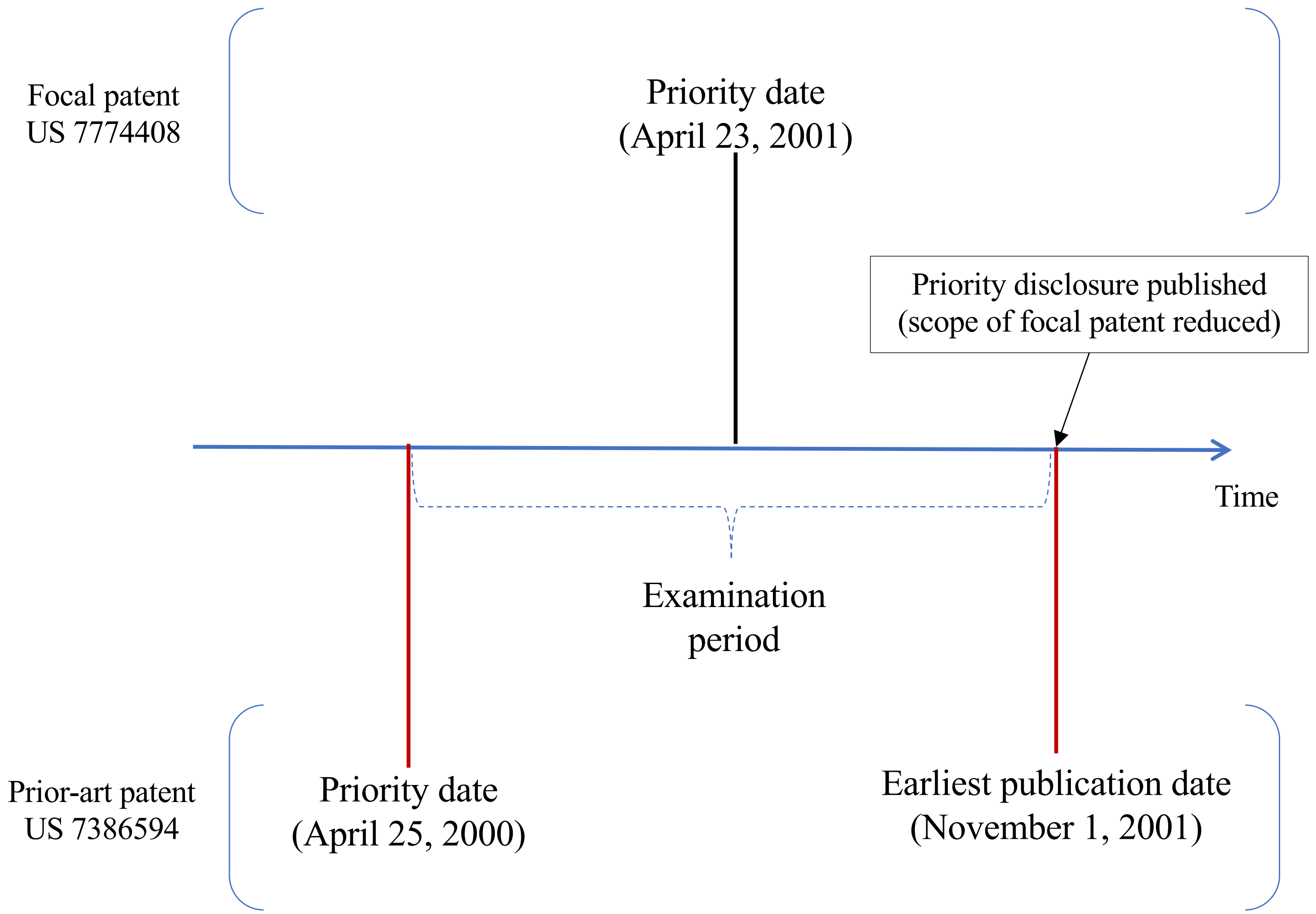
Figure 1. Receiver Operator Characteristics (ROC) Curve



Notes: The figure presents receiver operator characteristics (ROC) curve for the machine learning (random forest) algorithm employed to predict patent similarity. The Area under the ROC Curve (AUC) indicates the performance of the algorithm in predicting whether a pair of a focal and a prior-art patents is similar.

Figure 2. Priority Disclosure

USPTO Patent 7386594: System and method related to generating an email campaign
USPTO Patent 7774408: Methods, systems, and emails to link emails to matters and organizations



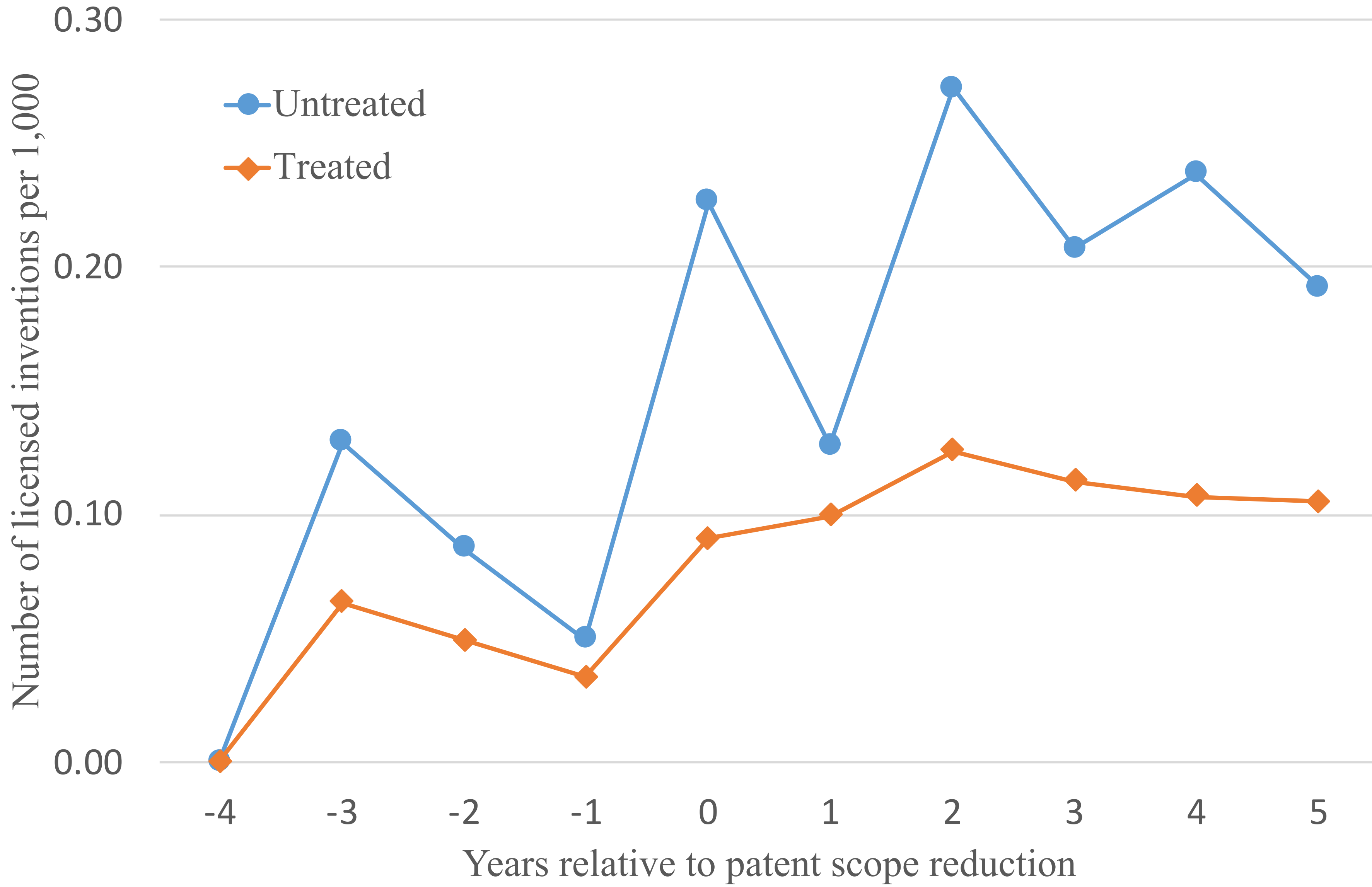
Priority date:
the earliest application date of the patents that relate to the same invention.

Earliest publication date:
the date on which the first patent from a patent family is published.

Examination period:
Time period during which a patent application is examined for patent issuance and the prior-art invention is not disclosed to the public.

Notes: The figure illustrates the sequence of priority date and publication dates of focal and prior-art patents that give rise to a priority disclosure, i.e. an event that effectively reduces the scope of the focal patent.

Figure 3. Parallel Trends between the Treatment and Control Group



Notes: The figure presents licensing patterns before and after the publication of a similar priority disclosure (i.e. patent scope reduction) for the treatment and control groups. X-axis represents years relative to the publication of a similar priority disclosure, and Y-axis represents the average number of licensed patents per 1,000 patents for both the treatment and control groups.

Figure 4. Changes in Licensing Propensity after Patent Scope Reduction

Notes: The figure presents changes in licensing propensity of treated inventions relative to control inventions after a priority disclosure is published (i.e. patent scope is reduced) across different invention and firm characteristics. High- and low-quality inventions are those in the top and bottom half of the forward citation distribution. Novel and non-novel inventions are those in the top and bottom half of IPC combination use count distribution. Science- and non-science-based inventions are those in the top and bottom half of close university publication count distribution. Large- and small-firm inventions are those in the top and bottom half of invention stock distribution at the firm level.

Table 1. Summary Statistics for Main Variables (Treated Inventions)

Panel A: Only Licensed Inventions						
				Distribution		
VARIBALES	No. Obs.	Mean	Std. Dev.	10th	50th	90th
Priority year (focal invention)	1,527	1,997.0	3.8	1,992	1,997	2,002
Earliest publication year (focal invention)	1,527	1,999.0	3.8	1,994	1,999	2,004
Priority to disclosure lag	1,527	0.769	0.691	0	1	1
Similar priority disclosures	1,527	4.5	3.8	1	3	9
Dummy for a citation	1,527	0.042	0.2	0	0	0
Forward citations	1,527	61.1	117.5	8	31	141
Citations to priority disclosures	1,527	62.8	71.2	16	42.5	125.2
IPC combination familiarity	1,527	38.8	280.5	0	0	60
Close university publications	1,527	141.6	453.5	0	19	310
Dummy for a licensed invention	1,527	0.6	0.495	0	1	1
Inventor size (Invention stock)	667	518.4	2591.0	1	13	496

Panel B: All Inventions						
				Distribution		
VARIBALES	No. Obs.	Mean	Std. Dev.	10th	50th	90th
Priority year (focal invention)	1,304,886	2,001.0	5.0	1,994	2,002	2,007
Earliest publication year (focal invention)	1,304,886	2,003.0	5.1	1,995	2,003	2,009
Priority to disclosure lag	1,304,886	0.777	0.618	0	1	1
Similar priority disclosures	1,304,886	4.2	3.6	1	3	9
Dummy for a citation	1,304,886	0.029	0.168	0	0	0
Forward citations	1,304,886	18.8	31.7	2	10	41
Citations to priority disclosures	1,304,886	32.8	37.5	8	22	67
IPC combination familiarity	1,304,886	149.7	832.0	0	0	211
Close university publications	1,304,886	63.7	273.7	0	8	128
Dummy for a licensed invention	1,304,886	0.001	0.026	0	0	0
Inventor size (Invention stock)	122,842	18.1	259.1	1	2	14

Notes: The table presents summary statistics for main variables used in this study. Only the treated inventions are included. Definitions are on Table A1.

Table 2. Mean Comparisons for Main Variables

Panel A: Only Licensed Inventions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean comp.	Treated			Untreated		
VARIBALES	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Priority disclosures	4.5**	1,527	4.5	3.8	545	0.0	0.0
Similar prior-art inventions	0.157**	1,527	0.172	0.9	545	0.015	0.191
Similar non-prior-art inventions	0.647**	1,527	2.2	4.1	545	1.6	3.4
Forward citations	18.0**	1,527	61.1	117.5	545	43.1	73.1
Backward citations	2.6	1,527	39.1	53.1	545	36.4	48.2
NPL citations	-7.1	1,527	47.5	155.2	545	54.6	159.9
IPC combination familiarity	0.116	1,527	38.8	280.5	545	38.7	236.3
Close university publications	-86.3**	1,527	141.6	453.5	545	227.9	952.6
Inventor size (Invention stock)	200.6	667	525.1	2841.5	358	324.4	1589.8
Inventor age (first patent application year)	-0.326	667	13.3	13.1	358	13.6	12.8
Inventor licensing experience	-0.112	667	0.738	0.995	358	0.849	0.860
Share of Silicon Valley inventors	0.0079	667	0.101	0.238	358	0.093	0.231
Share of Route 128 inventors	-0.0004	667	0.020	0.091	358	0.020	0.087
Share of Canada inventors	0.0038	667	0.039	0.172	358	0.035	0.163

Panel B: All Inventions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean comp.	Treated			Untreated		
VARIBALES	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Priority disclosures	4.2**	1,304,886	4.2	3.6	1,304,886	0.000	0.000
Similar prior-art inventions	0.265**	1,304,886	0.271	1.2	1,304,886	0.006	0.153
Similar non-prior-art inventions	1.2**	1,304,886	2.6	4.9	1,304,886	1.354	3.671
Forward citations	3.4**	1,304,886	18.8	31.7	1,304,886	15.5	28.0
Backward citations	5.3**	1,304,886	21.6	29.2	1,304,886	16.4	22.8
NPL citations	3.4**	1,304,886	7.8	40.1	1,304,886	4.5	32.6
IPC combination familiarity	3.4**	1,304,886	149.7	832.0	1,304,886	146.3	724.5
Close university publications	4.0**	1,304,886	63.7	273.7	1,304,886	59.7	284.7
Inventor size (Invention stock)	-10.7**	122,842	18.0	262.1	75,917	28.7	350.5
Inventor age (first patent application year)	-1.361**	122,842	6.7	7.9	75,917	8.1	8.9
Inventor licensing experience	-0.002**	122,842	0.005	0.103	75,917	0.007	0.108
Share of Silicon Valley inventors	0.0035**	122,842	0.048	0.194	75,917	0.045	0.182
Share of Route 128 inventors	0.0004**	122,842	0.006	0.068	75,917	0.006	0.063
Share of Canada inventors	0.0017**	122,842	0.041	0.190	75,917	0.040	0.184

Notes: The table presents mean comparison between treated inventions and their matched inventions. Panel A presents mean comparisons for main variables using only the licensed inventions, and Panel B presents mean comparisons for main variables using all inventions. ** p<0.01, * p<0.05

Table 3. The Effect of Reduced Patent Scope on Licensing

Dependent variable:	*Dummy for a licensed invention* (Invention-year level)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	5-year window	Invention quality	Firm characteristics	Tech centers	Competition	Different state	Different IPC (4-digit)
Treated × *Post*	-0.000137**	-0.000137**	-0.000181**	-0.000185**	-0.000179**	-0.000188**	-0.000285**
	(0.000012)	(0.000012)	(0.000012)	(0.000012)	(0.000012)	(0.000018)	(0.000034)
Post dummy	0.000109**	0.000109**	0.000078**	0.000083**	0.000089**	0.000086**	0.000153**
	(0.000012)	(0.000012)	(0.000012)	(0.000012)	(0.000012)	(0.000017)	(0.000032)
log(1+*Forward citations*)		0.000003	0.000010	0.000008	0.000008	0.000028**	0.000023
		(0.000007)	(0.000006)	(0.000006)	(0.000006)	(0.000009)	(0.000017)
log(1+*Inventor size*)			-0.000123**	-0.000169**	-0.000167**	-0.000156**	-0.000259**
			(0.000021)	(0.000022)	(0.000022)	(0.000028)	(0.000054)
log(1+*Inventor age*)			0.001261**	0.001246**	0.001242**	0.001244**	0.001920**
			(0.000038)	(0.000038)	(0.000038)	(0.000046)	(0.000098)
log(1+*Inventor licensing experience*)			0.007294**	0.007308**	0.007310**	0.008399**	0.009053**
			(0.000165)	(0.000166)	(0.000166)	(0.000236)	(0.000342)
log(1+*Silicon valley inventors*)				0.000080**	0.000079**	0.000098**	0.000115**
				(0.000004)	(0.000004)	(0.000006)	(0.000010)
log(1+*Route 128 inventors*)				-0.000029**	-0.000030**	-0.000063**	-0.000094**
				(0.000004)	(0.000004)	(0.000007)	(0.000012)
log(1+*Canada inventors*)				0.000032**	0.000032**	0.000032**	0.000061**
				(0.000004)	(0.000004)	(0.000007)	(0.000013)
log(1+*Similar prior-art inventions*)					0.000055**	0.000096**	0.000146**
					(0.000008)	(0.000012)	(0.000019)
log(1+*Similar non-prior-art inventions*)					0.000038**	0.000053**	0.000029*
					(0.000005)	(0.000008)	(0.000014)
Invention FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.00007	0.00007	0.00007	0.00007	0.00007	0.00008	0.00013
Licensed inventions	2,072	2,072	2,058	2,058	2,058	1,429	883
Observations	17,674,697	17,674,697	17,663,813	17,663,813	17,663,813	10,048,215	4,037,863
R-squared	0.300	0.300	0.303	0.303	0.303	0.302	0.285

Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors. The sample consists of inventions whose priority years range from 1990 to 2010 and that experience a reduction in patent scope and controls that were matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and patent stock quartile. *Treated* is a dummy that takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table 4. The Effect of Reduced Patent Scope on Licensing - Invention Characteristics

Dependent variable:	*Dummy for a licensed invention* (Invention-year level)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Invention quality		Invention novelty		Invention knowledge base	
VARIABLES	Forward citation Continuous	Top half of citation distribution	combination use count Continuous	Top half of IPC comb. use distribution	Proximity to scientific publications	Top half of proximity distribution
Treated × *Post* ×						
log(1+*Forward citations*)	-0.000103**					
	(0.000014)					
Dummy for a high quality invention		-0.000280**				
		(0.000027)				
log(1+*IPC combination usage count*)			-0.000237**			
			(0.000051)			
Dummy for a novel invention				-0.000181**		
				(0.000023)		
log(1+*Proximity to univ science*)					-0.000062**	
					(0.000008)	
Dummy for a science-based invention						-0.000159**
						(0.000025)
Treated × *Post*	-0.000083**	-0.000078**	0.002188**	-0.000049**	-0.000040*	-0.000104**
	(0.000020)	(0.000010)	(0.000501)	(0.000017)	(0.000018)	(0.000014)
Post	-0.000195**	-0.000071**	-0.003476**	-0.000123**	-0.000192**	-0.000074**
	(0.000011)	(0.000010)	(0.000271)	(0.000015)	(0.000019)	(0.000015)
Post ×						
log(1+*Forward citations*)	0.000198**					
	(0.000009)					
Dummy for a high quality invention		0.000405**				
		(0.000023)				
log(1+*IPC combination usage count*)			0.000357**			
			(0.000028)			
Dummy for a novel invention				0.000290**		
				(0.000018)		
log(1+*Proximity to univ science*)					0.000118**	
					(0.000007)	
Dummy for a science-based invention						0.000324**
						(0.000021)
log(1+*Forward citations*)			0.000008	0.000007	0.000008	0.000008
			(0.000006)	(0.000006)	(0.000006)	(0.000006)
log(1+*Inventor size*)	-0.000202**	-0.000190**	-0.000167**	-0.000166**	-0.000170**	-0.000171**
	(0.000022)	(0.000022)	(0.000022)	(0.000022)	(0.000022)	(0.000022)
log(1+*Inventor age*)	0.001210**	0.001233**	0.001242**	0.001249**	0.001237**	0.001241**
	(0.000037)	(0.000037)	(0.000038)	(0.000038)	(0.000038)	(0.000038)
log(1+*Inventor licensing experience*)	0.007315**	0.007313**	0.007310**	0.007313**	0.007303**	0.007303**
	(0.000166)	(0.000166)	(0.000166)	(0.000166)	(0.000166)	(0.000166)
log(1+*Silicon valley inventors*)	0.000078**	0.000079**	0.000079**	0.000077**	0.000082**	0.000082**
	-0.000004	-0.000004	(0.000004)	(0.000004)	(0.000004)	(0.000004)
log(1+*Route 128 inventors*)	-0.000034**	-0.000032**	-0.000030**	-0.000032**	-0.000028**	-0.000029**
	(0.000004)	(0.000004)	(0.000004)	(0.000004)	(0.000004)	(0.000004)
log(1+*Canada inventors*)	0.000037**	0.000035**	0.000032**	0.000032**	0.000033**	0.000033**
	(0.000004)	(0.000004)	(0.000004)	(0.000004)	(0.000004)	(0.000004)
log(1+*Similar prior-art inventions*)	0.000058**	0.000057**	0.000056**	0.000055**	0.000056**	0.000056**
	(0.000008)	(0.000008)	(0.000008)	(0.000008)	(0.000008)	(0.000008)
log(1+*Similar non-prior-art inventions*)	0.000036**	0.000037**	0.000038**	0.000037**	0.000042**	0.000041**
	(0.000005)	(0.000005)	(0.000005)	(0.000005)	(0.000005)	(0.000005)
Invention FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007
Licensed inventions	2,058	2,058	2,058	2,058	2,058	2,058
Observations	17,663,813	17,663,813	17,663,813	17,663,813	17,663,813	17,663,813
R-squared	0.303	0.303	0.303	0.303	0.303	0.303

Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors for different invention characteristics. The sample consists of inventions whose priority years range from 1990 to 2010 and that experience a reduction in patent scope and controls that were matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and patent stock quartile.*Dummy for a high quality invention* takes 1 if an invention is in the upper half of the forward citation distribution. *Dummy for a novel invention* takes 1 if an invention is in the upper half of IPC / IPC combination use count. *Dummy for a science-based invention* takes 1 if an invention is in the upper half of close university publication count distribution. *Treated* is a dummy that takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table 5. The Effect of Reduced Patent Scope on Licensing - Firm Characteristics

Dependent variable:	*Dummy for a licensed invention* (Invention-year level)				
	(1)	(2)	(3)	(4)	(5)
	Firm level patent stock		Firm-tech area level patent stock		Indiv. inventor
VARIABLES	Continuous	Dummy	Continuous	Dummy	Dummy
Treated × *Post* ×					
log(1+*Patent count*)	0.000046**		0.000024**		
	(0.000005)		(0.000005)		
Dummy for a large inventor		0.000365**		0.000261**	
		(0.000097)		(0.000056)	
Dummy for a non-individual inventor					0.000126**
					(0.000045)
Treated × *Post*	-0.000462**	-0.000536**	-0.000294**	-0.000413**	-0.000306**
	(0.000047)	(0.000096)	(0.000037)	(0.000055)	(0.000043)
Post	0.001152**	-0.000569**	0.000672**	0.000393**	-0.000483**
	(0.000045)	(0.000095)	(0.000031)	(0.000044)	(0.000043)
Post ×					
log(1+*Patent count*)	-0.000154**		-0.000102**		
	(0.000005)		(0.000004)		
Dummy for a large inventor		0.000672**		-0.000350**	
		(0.000095)		(0.000045)	
Dummy for a non-individual inventor					0.000620**
					(0.000045)
log(1+*Forward citations*)	0.000008	0.000005	0.000007	0.000006	0.000007
	(0.000007)	(0.000006)	(0.000007)	(0.000006)	(0.000006)
log(1+*Inventor size*)					-0.000262**
					(0.000024)
log(1+*Inventor age*)	0.000594**	0.001465**	0.000850**	0.001163**	0.001472**
	(0.000035)	(0.000046)	(0.000035)	(0.000037)	(0.000046)
log(1+*Inventor licensing experience*)	0.007456**	0.007327**	0.007401**	0.007309**	0.007308**
	(0.000169)	(0.000166)	(0.000168)	(0.000166)	(0.000166)
log(1+*Silicon valley inventors*)	0.000064**	0.000068**	0.000064**	0.000068**	0.000086**
	(0.000004)	(0.000004)	(0.000004)	(0.000004)	(0.000004)
log(1+*Route 128 inventors*)	-0.000039**	-0.000034**	-0.000038**	-0.000034**	-0.000029**
	(0.000005)	(0.000005)	(0.000005)	(0.000005)	(0.000004)
log(1+*Canada inventors*)	0.000034**	0.000028**	0.000032**	0.000029**	0.000034**
	(0.000004)	(0.000004)	(0.000004)	(0.000004)	(0.000004)
log(1+*Similar prior-art inventions*)	0.000036**	0.000066**	0.000042**	0.000059**	0.000058**
	(0.000008)	(0.000008)	(0.000008)	(0.000008)	(0.000008)
log(1+*Similar non-prior-art inventions*)	0.000030**	0.000036**	0.000031**	0.000037**	0.000039**
	(0.000005)	(0.000005)	(0.000005)	(0.000005)	(0.000005)
Invention FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.00007	0.00007	0.00007	0.00007	0.00007
Licensed inventions	2,058	2,058	2,058	2,058	2,058
Observations	17,663,813	17,663,813	17,663,813	17,663,813	17,663,813
R-squared	0.303	0.303	0.303	0.303	0.303

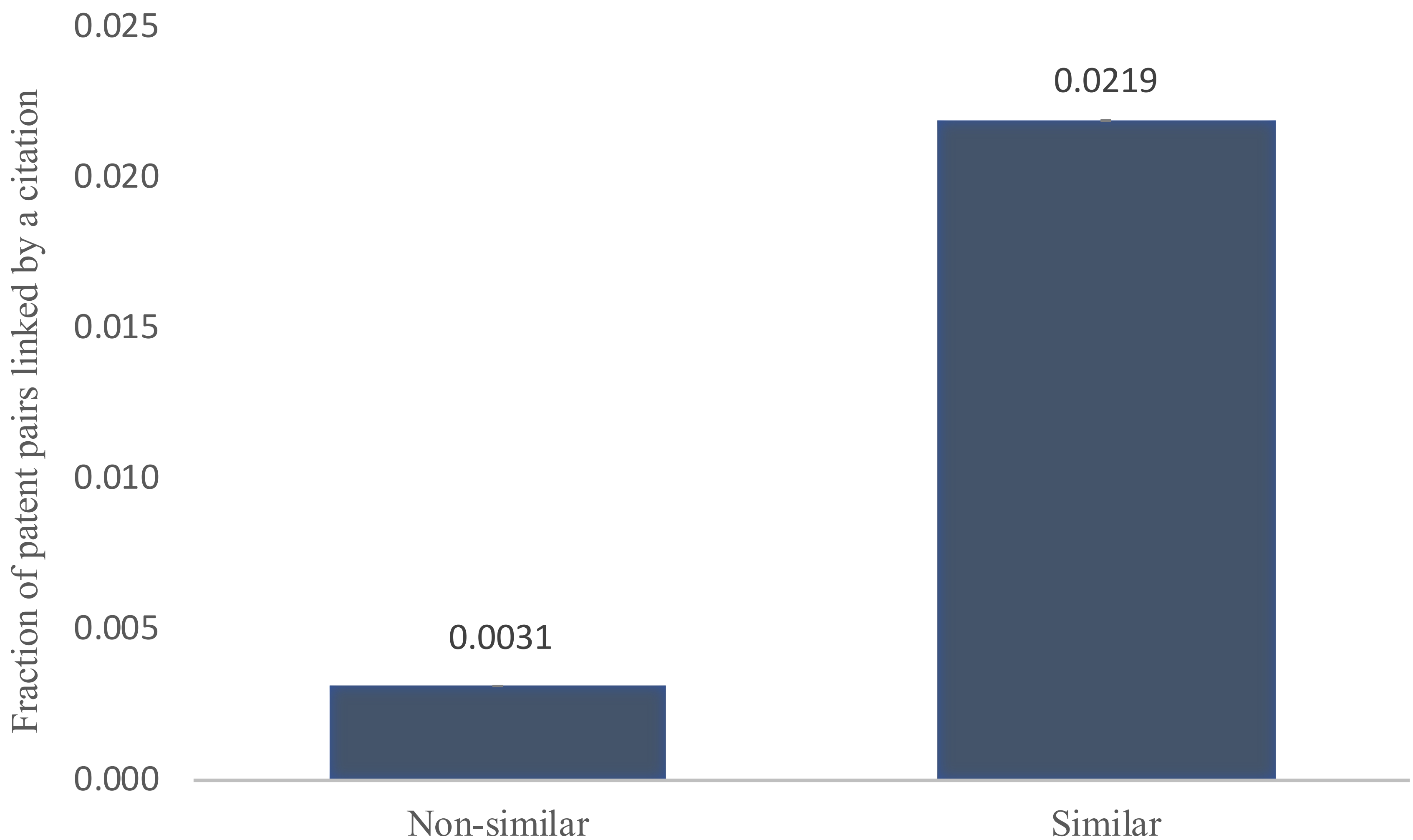
Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors. The sample consists of inventions whose priority years range from 1990 to 2010 and that experience a reduction in patent scope and controls that were matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and patent stock quartile. *Dummy for a large inventor* takes 1 if a firm is in the upper half of the invention count distribution at the firm level prior to the treatment year. (An alternative measure uses invention count at the firm-technology area level.) *Dummy for a non-individual inventor* takes 1 if an inventor is not an individual and 0 otherwise. *Treated* is a dummy that takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table 6. The Effect of Reduced Patent Scope on Licensing - Technology Areas

Dependent variable:	*Dummy for a licensed invention* (Invention-year level)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Biotechnology	Chemicals	Communication	Computer	Electr. Machines	Mechanicals	Medical Tech.	Pharmaceuticals
Treated × Post	-0.000818**	-0.000135*	-0.000392**	-0.000002	-0.000328**	0.000000	-0.000405**	0.000340
	(0.000230)	(0.000052)	(0.000039)	(0.000016)	(0.000058)	(0.000021)	(0.000114)	(0.000223)
Post	-0.000336	0.000150**	0.000353**	0.000067**	0.000393**	-0.000004	-0.000562**	-0.000021
	(0.000174)	(0.000049)	(0.000039)	(0.000016)	(0.000069)	(0.000020)	(0.000127)	(0.000218)
log(1+*Forward citations*)	0.000764**	-0.000012	-0.000128**	0.000012	0.000045*	0.000023*	0.000060	0.000205
	(0.000132)	(0.000033)	(0.000020)	(0.000008)	(0.000021)	(0.000010)	(0.000045)	(0.000135)
log(1+*Inventor size*)	0.003996**	-0.000209	-0.000257**	-0.000164**	0.000013	0.000131**	-0.001365**	-0.002101**
	(0.000707)	(0.000112)	(0.000071)	(0.000033)	(0.000078)	(0.000040)	(0.000141)	(0.000437)
log(1+*Inventor age*)	0.005040**	0.002484**	0.001548**	0.001002**	0.001591**	0.000247**	0.003603**	0.008163**
	(0.000515)	(0.000271)	(0.000130)	(0.000094)	(0.000155)	(0.000043)	(0.000227)	(0.000724)
log(1+*Inventor licensing experience*)	0.034355**	0.009688**	0.006254**	0.001662**	0.009381**	0.005477**	0.037655**	0.026153**
	(0.002513)	(0.000864)	(0.000410)	(0.000133)	(0.000785)	(0.000769)	(0.001831)	(0.001814)
log(1+*Silicon valley inventors*)	0.000885**	-0.000026	0.000055**	0.000022**	0.000046**	0.000013*	0.000842**	0.000117
	(0.000098)	(0.000026)	(0.000010)	(0.000005)	(0.000009)	(0.000006)	(0.000056)	(0.000097)
log(1+*Route 128 inventors*)	0.000474**	0.000061*	-0.000073**	-0.000022**	0.000126**	-0.000022**	0.001036**	-0.000620**
	(0.000177)	(0.000029)	(0.000009)	(0.000005)	(0.000017)	(0.000005)	(0.000077)	(0.000115)
log(1+*Canada inventors*)	-0.001143**	0.000034*	0.000098**	0.000011**	-0.000008	0.000000	-0.000557**	0.000702**
	(0.000176)	(0.000017)	(0.000012)	(0.000003)	(0.000010)	(0.000004)	(0.000058)	(0.000133)
log(1+*Similar prior-art inventions*)	0.000412**	0.000036	-0.000023	0.000040**	0.000059*	-0.000015	-0.000780**	0.001871**
	(0.000109)	(0.000032)	(0.000018)	(0.000008)	(0.000025)	(0.000014)	(0.000102)	(0.000194)
log(1+*Similar non-prior-art inventions*)	0.000083	-0.000006	-0.000001	0.000003	0.000069**	0.000001	0.000473**	0.001472**
	(0.000084)	(0.000023)	(0.000009)	(0.000005)	(0.000013)	(0.000010)	(0.000049)	(0.000173)
Invention FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.00037	0.00011	0.00010	0.00002	0.00013	0.00005	0.00013	0.00018
Licensed inventions	290	184	152	159	112	97	361	366
Observations	313,976	1,135,530	2,195,155	4,385,404	1,159,723	1,839,804	733,817	327,158
R-squared	0.279	0.284	0.300	0.276	0.360	0.458	0.374	0.247

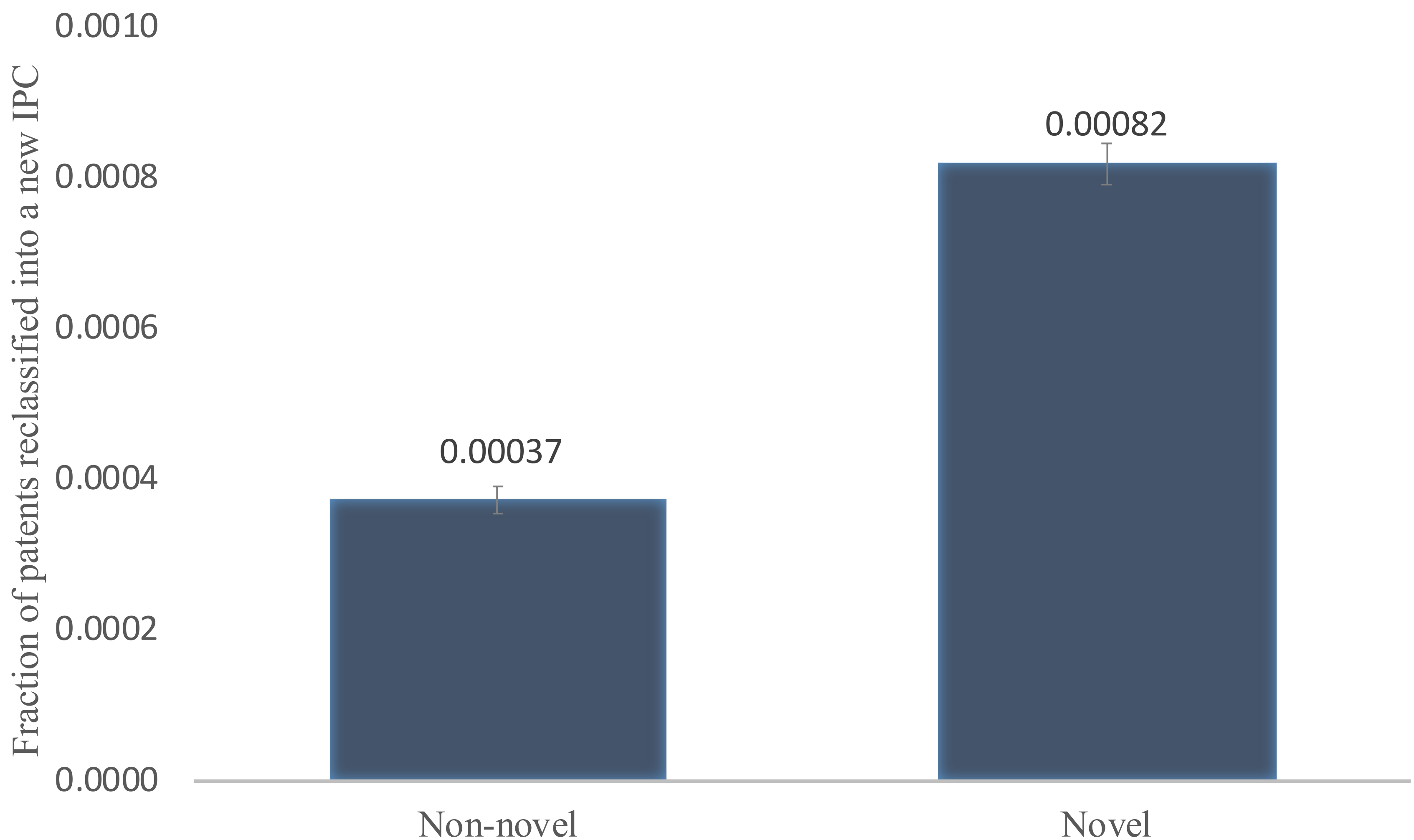
Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors. The sample consists of all inventions whose priority years range from 1990 to 2010 and that experience a reduction in patent scope and controls that were matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and patent stock quartile. *Treated* is a dummy that takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Figure A1. Validation of the Similarity Measure - Patent pairs linked by a citation



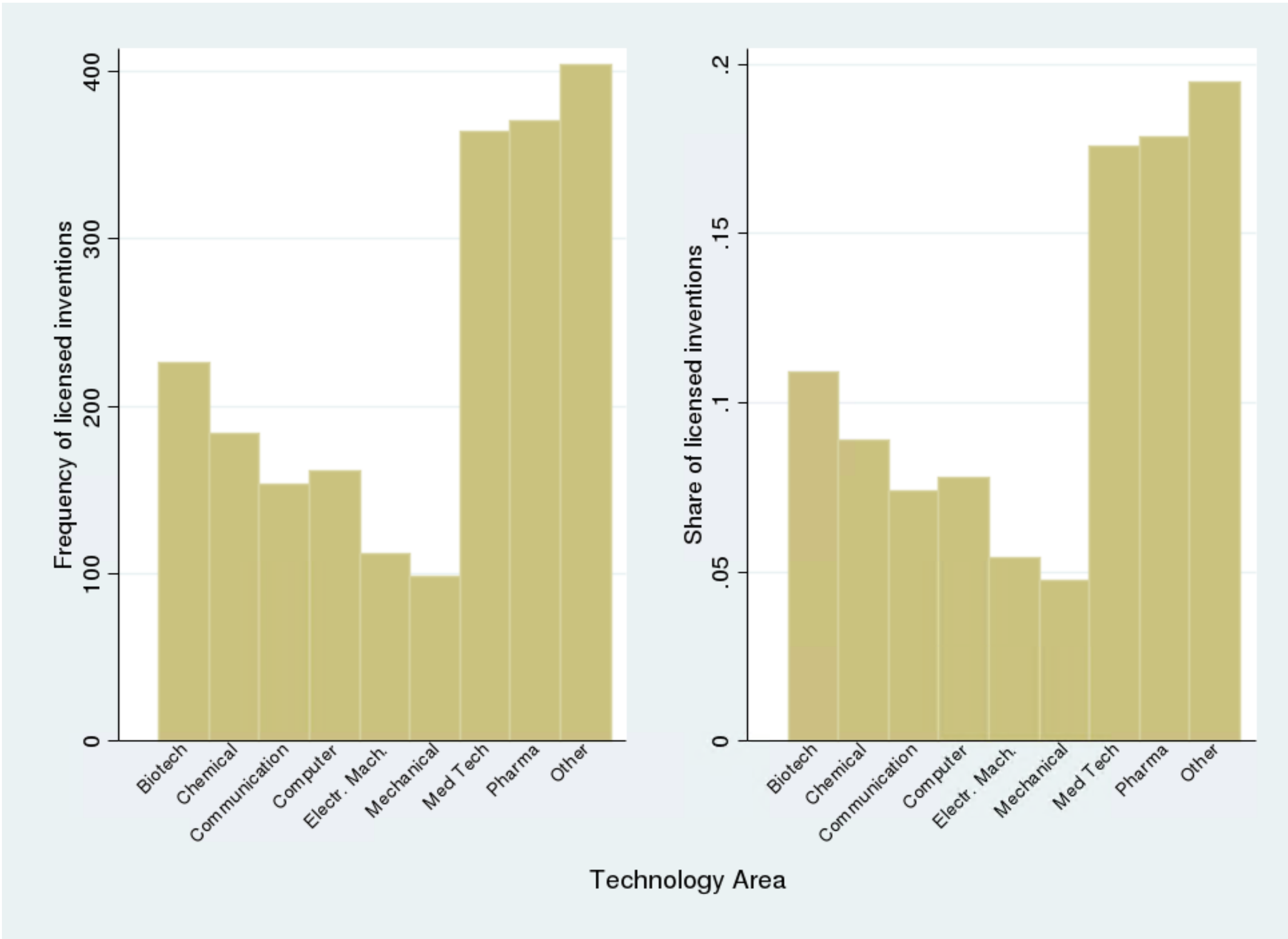
Notes: The figure presents the comparison of the fraction of patent pairs that are linked by a citation between similar patent pairs and non-similar patent pairs. The sample consists of USPTO patents with application years between 1990 and 2010 and top fifty prior-art patents for each focal patent in terms of the textual similarity scores (close to 143 million patent pairs).

Figure A2. Validation of the Similarity Measure - Patents reclassified into a new IPC



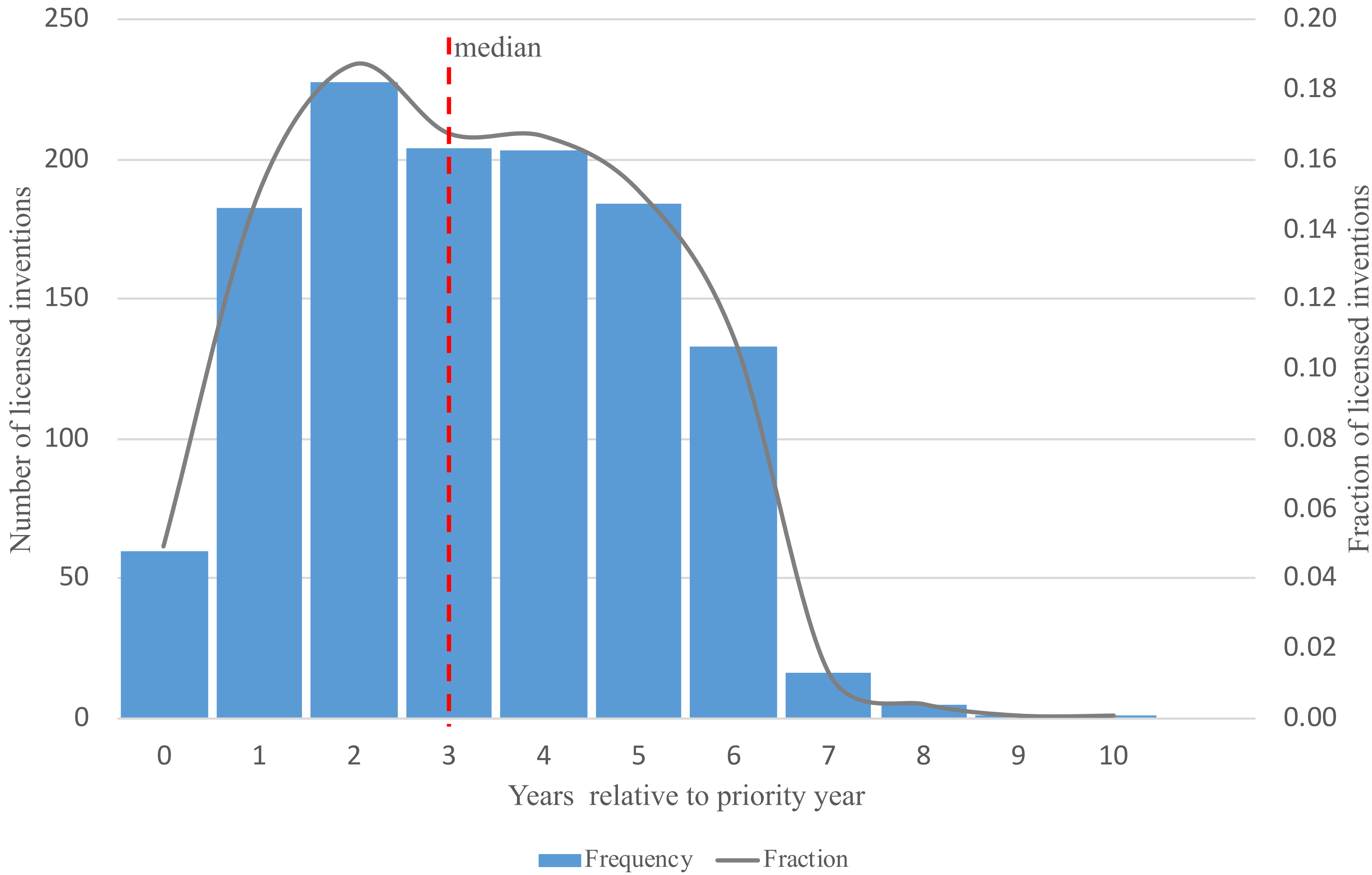
Notes: The figure presents the fraction of non-novel focal patents that are reclassified into a new IPC versus the fraction of *novel* focal patents that are reclassified into a new IPC, where a *novel* patent is defined as a patent in the bottom half of the similar prior-art share distribution and a *non-novel* patent is defined as a patent in the top half of the similar prior-art share distribution. The sample consists of USPTO patents with application years between 1990 and 2010 along with the share of top fifty prior-art patents in terms of textual similarity that are predicted to be similar for each focal patent.

Figure A3. Licensed Inventions across Technology Areas



Notes: The figure presents both number and share of inventions in specified technology areas defined based on PatStat's Technology Field database.

Figure A4. Licensing Lag from Priority Year of the Focal Inventions



Notes: The figure presents licensing lags in years from the priority year of the focal invention for the licensed inventions in the sample.

Table A1. Definitions of Main Variables

Variable	Definition
Invention (Patent)	
Priority date (year)	Date (year) in which the first patent application from a patent family was filed with a patent issuing authority
Earliest publication date (year)	Date (year) in which the first patent application or patent from a patent family was disclosed to the public by a patent issuing authority
Priority to disclosure lag	Lag in years between the priority year and the treatment year (i.e. publication of similar priority disclosure)
Similar priority disclosures	Number of priority disclosures that are similar to the focal inventions
Similar prior-art inventions	Prior-art inventions that are similar to a focal invention
Similar non-prior-art inventions	Inventions that come after a focal invention and are similar to the focal inventio
Dummy for a citation	Indicator that the focal invention and the similar priority disclosure is linked by a forward citation
Forward citations	Number of times a focal invention is cited by inventions that come after the focal invention's publication
Backward citations	Number of prior-art inventions that a focal invention cites
NPL citations	Number of non-patent literature that a focal invention cites
Citations to priority disclosures	Number of forward citations that similar priority disclosures paired with a focal invention receives
IPC combination familiarity	Number of times an IPC or a IPC combination has been used prior to the generation of the focal invention
Close university publications	Number of univeristy publications similar to a focal invention and published prior to the focal invention
Dummy for a licensed invention	Indicator taking a value of 1 if an invention is licensed out in a given year
Inventor	
Inventor size	Number of inventions that a focal inventor generates prior to the treatment year of an invention
Inventor age	Treatment year of an invention minus the first year that the focal inventor generates its first patented invention
Inventor licensing experience	Number of licensing agreeemnts that a focal inventor has entereed into prior to the treatment year of an invention
Share of Silicon Valley inventors	Fraction of inventions for a focal firm in a given year whose inventor addresses include a location in Silicon Valley
Share of Route 128 inventors	Fraction of inventions for a focal firm in a given year whose inventor addresses include a location in Route 128
Share of Canada inventors	Fraction of inventions for a focal firm in a given year whose inventor addresses include a location in Canada

Table A2. Validating Invention Similarity - Citations

Dependent variable:	*Citation dummy*			
	(1)	(2)	(3)	(4)
Dummy for a similar patent pair	0.021**	0.021**	0.022**	0.022**
	(0.000)	(0.000)	(0.000)	(0.000)
log(1+*Forward citations*)		0.002**	0.002**	
		(0.000)	(0.000)	
log(1+*Backwoard citations*)		0.009**	0.009**	
		(0.000)	(0.000)	
log(1+*NPL citations*)		-0.001**	-0.001**	
		(0.000)	(0.000)	
log(1+*Patent stock*)			-0.000**	
			(0.000)	
Application year FE (focal)	Yes	Yes	Yes	Yes
Publication year FE (focal)	No	No	Yes	Yes
4-digit IPC FE (focal)	Yes	Yes	Yes	No
Patent (focal)	No	No	No	Yes
SE Cluster	Patent level	Patent level	Patent level	Patent level
Dep Mean	0.01	0.01	0.01	0.01
Std Deviation	0.12	0.12	0.12	0.12
Observations	111,795,127	111,795,127	110,935,605	111,794,941
R-squared	0.010	0.014	0.014	0.053

Notes : The table presents a relationship between a dummy variable indicating whether a pairwise patent pair (i.e. focal and its prior art patents) is predicted to be similar and a dummy variable indicating whether the focal patent cites the prior-art patent. The sample consists of focal and prior-art patent pairs, with the application year of the focal patents ranging from 1990 to 2010. *Citation dummy* is a dummy variable taking a value of 1 if a focal patent cites the prior-art invention that it is paired with. *Dummy for a similar invention pair* takes a value of 1 if a focal patent and a paired prior-art patent are predicted by the machine learning algorithm to be similar. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table A3. Validating Invention Similarity - New IPC

Dependent variable:	*Dummy for a new IPC*			
	(1)	(2)	(3)	(4)
Share of non-similar prior-art patents	0.00037**	0.00036**	0.00035**	0.00034**
	(0.00010)	(0.00010)	(0.00010)	(0.00011)
log(1+*Forward citations*)		0.00004*	0.00004*	0.00008**
		(0.00002)	(0.00002)	(0.00002)
log(1+*Backwoard citations*)		-0.00003	-0.00003	-0.00001
		(0.00002)	(0.00002)	(0.00003)
log(1+*NPL citations*)		0.00004*	0.00004*	0.00007**
		(0.00002)	(0.00002)	(0.00002)
log(1+*Patent stock*)			-0.00000	
			(0.00001)	
Application year FE	Yes	Yes	Yes	Yes
Publication year FE	Yes	Yes	Yes	Yes
4-digit IPC FE	Yes	Yes	Yes	Yes
Assignee FE	No	No	No	Yes
SE Cluster	Assignee	Assignee	Assignee	Assignee
Dep Mean	0.001	0.001	0.001	0.001
Std Deviation	0.024	0.024	0.024	0.024
Observations	2,224,432	2,224,432	2,224,432	2,066,956
R-squared	0.004	0.004	0.004	0.050

Notes : The table presents a relationship between a dummy variable indicating whether a focal patent is in the bottom half of the similr prior-art share distribution and a dummy variable indicating whether the focal patent is reclassified into a new IPC. The sample consists of focal and prior-art patent pairs. The application year of the focal patents ranges from 1990 to 2010. *Dummy for a new IPC* is a dummy variable taking 1 if a focal invention was assigned to a new IPC. *Share of non-similar prior-art patents* is the fraction of top 50 prior-art patents in terms of textual similarity score that are not deemed to be similar to the focal patent. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table A4. Robustness Tests - Subsample Analysis

Dependent variable:	*Dummy for a licensed invention* (Invention-year level)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	3-year window	1-year window	Excl. cross-licensing	For-profit organizations	Non-profit organizations	Years ≤ 2001	Years > 2001	Priority disc year ≤ Focal pub year	Only assignees ever licensed	Only licensed inventions
Treated × Post	-0.000146**	-0.000063**	-0.000144**	-0.000185**	-0.000141**	-0.000268**	-0.000085**	-0.000110**	-0.002597**	-0.031814**
	(0.000013)	(0.000019)	(0.000012)	(0.000012)	(0.000044)	(0.000022)	(0.000010)	(0.000021)	(0.000223)	(0.004228)
Post	0.000024	-0.000064**	0.000087**	0.000083**	0.000090*	0.000166**	-0.000012	-0.000146**	0.002099**	0.014443**
	(0.000014)	(0.000018)	(0.000012)	(0.000012)	(0.000044)	(0.000021)	(0.000013)	(0.000022)	(0.000222)	(0.003179)
log(1+*Forward citations*)	-0.000019*	-0.000003	0.000019**	-0.000010	0.000094**	0.000012	0.000002	0.000003	0.000154	-0.007218**
	(0.000008)	(0.000016)	(0.000006)	(0.000006)	(0.000025)	(0.000010)	(0.000006)	(0.000014)	(0.000114)	(0.001393)
log(1+*Inventor size*)	-0.000217**	0.000077	-0.000153**	-0.000358**	-0.000428**	-0.000217**	-0.000078**	-0.000664**	-0.021697**	-0.009257*
	(0.000031)	(0.000097)	(0.000021)	(0.000028)	(0.000075)	(0.000033)	(0.000014)	(0.000067)	(0.000943)	(0.004462)
log(1+*Inventor age*)	0.001249**	0.001020**	0.001205**	0.001817**	0.001237**	0.001639**	0.000608**	0.001575**	0.103522**	0.108194**
	(0.000048)	(0.000125)	(0.000037)	(0.000068)	(0.000066)	(0.000055)	(0.000043)	(0.000102)	(0.003147)	(0.003731)
log(1+*Inventor licensing experience*)	0.008521**	0.008855**	0.006297**	0.005784**	0.022924**	0.007309**	0.007904**	0.011075**	0.034537**	0.690581**
	(0.000248)	(0.000746)	(0.000155)	(0.000154)	(0.000993)	(0.000186)	(0.000394)	(0.000498)	(0.000759)	(0.015891)
log(1+*Silicon valley inventors*)	0.000053**	0.000018	0.000061**	0.000080**	0.000171**	0.000133**	0.000028**	0.000089**	0.002902**	0.013354**
	(0.000005)	(0.000012)	(0.000004)	(0.000004)	(0.000029)	(0.000007)	(0.000003)	(0.000011)	(0.000101)	(0.001953)
log(1+*Route 128 inventors*)	-0.000028**	-0.000038*	-0.000018**	-0.000020**	-0.000104**	-0.000134**	0.000149**	-0.000107**	-0.000509**	-0.029159**
	(0.000006)	(0.000017)	(0.000004)	(0.000004)	(0.000039)	(0.000008)	(0.000009)	(0.000012)	(0.000070)	(0.003728)
log(1+*Canada inventors*)	0.000054**	0.000053**	0.000024**	0.000032**	0.000099**	0.000063**	0.000005	0.000106**	0.000881**	-0.032456**
	(0.000007)	(0.000013)	(0.000004)	(0.000004)	(0.000034)	(0.000008)	(0.000003)	(0.000011)	(0.000091)	(0.003720)
log(1+*Similar prior-art inventions*)	0.000039**	0.000032*	0.000050**	0.000016*	0.000206**	0.000155**	-0.000027**	0.000130**	0.000572**	0.006510**
	(0.000008)	(0.000016)	(0.000008)	(0.000008)	(0.000029)	(0.000012)	(0.000008)	(0.000018)	(0.000135)	(0.001787)
log(1+*Similar non-prior-art inventions*)	0.000042**	-0.000017	0.000055**	0.000029**	0.000014	0.000028**	0.000042**	0.000092**	0.000047	-0.000437
	(0.000005)	(0.000014)	(0.000005)	(0.000005)	(0.000020)	(0.000009)	(0.000005)	(0.000012)	(0.000088)	(0.001435)
Invention FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.000066	0.000091	0.000053	0.000048	0.000164	0.000114	0.000013	0.000065	0.001194	0.019932
Licensed inventions	2,057	2,007	1,960	1,333	770	1,783	280	2,058	2,058	2,058
Observations	12,410,249	5,215,737	17,659,588	14,939,346	2,724,467	9,378,186	8,285,627	6,997,961	961,579	52,681
R-squared	0.371	0.792	0.314	0.298	0.322	0.276	0.406	0.451	0.332	0.628

Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors. The sample consists of inventions whose priority years range from 1990 to 2010 and that experience a reduction in patent scope and controls that were matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and patent stock quartile. *Treated* is a dummy takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** $p<0.01$, * $p<0.05$

Table A5. Robustness Tests - Firm Characteristics

Dependent variable:	*Dummy for a licensed invention* (Invention-year level)		
	(1)	(2)	(3)
	Firm age		Maintenance fee
VARIABLES	Continuous	Dummy	Dummy
Treated × *Post* ×			
log(1+*Inventor age*)	0.000141**		
	(0.000024)		
Dummy for a mature inventor		0.000566**	
		(0.000042)	
Dummy for a large firm (maint. fee)			0.000292**
			(0.000060)
Treated × *Post*	-0.000684**	-0.000643**	-0.000425**
	(0.000096)	(0.000041)	(0.000059)
Post	0.004008**	0.000968**	0.000384**
	(0.000120)	(0.000041)	(0.000048)
Post ×			
log(1+*Inventor age*)	-0.001068**		
	(0.000031)		
Dummy for a mature inventor		-0.001066**	
		(0.000041)	
Dummy for a large firm (maint. fee)			-0.000346**
			(0.000049)
log(1+*Forward citations*)	0.000016*	0.000012	0.000009
	(0.000007)	(0.000007)	(0.000006)
log(1+*Inventor size*)	-0.000368**	-0.000171**	-0.000134**
	(0.000024)	(0.000022)	(0.000023)
log(1+*Inventor age*)			0.001141**
			(0.000041)
log(1+*Inventor licensing experience*)	0.007398**	0.007045**	0.007240**
	(0.000167)	(0.000159)	(0.000165)
log(1+*Silicon valley inventors*)	0.000099**	0.000087**	0.000074**
	(0.000004)	(0.000004)	(0.000004)
log(1+*Route 128 inventors*)	-0.000029**	-0.000034**	-0.000032**
	(0.000004)	(0.000005)	(0.000004)
log(1+*Canada inventors*)	0.000037**	0.000028**	0.000032**
	(0.000004)	(0.000004)	(0.000004)
log(1+*Similar prior-art inventions*)	0.000028**	0.000039**	0.000050**
	(0.000008)	(0.000008)	(0.000008)
log(1+*Similar non-prior-art inventions*)	0.000046**	0.000050**	0.000034**
	(0.000005)	(0.000005)	(0.000005)
Invention FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.00007	0.00007	0.00007
Licensed inventions	2058	2072	2045
Observations	17,663,813	17,674,697	17,662,772
R-squared	0.303	0.303	0.30356

Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors. The sample consists of inventions whose priority years range from 1990 to 2010 and that experience a reduction in patent scope and controls that were matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and patent stock quartile. *Dummy for a mature inventor* takes 1 if firm age is in the upper half of firm age distribution prior to the treatment year and 0 if it is in the bottom half. *Dummy for a large firm* (based on maintenance fee threshold) takes 1 if an inventor does not claim "small entity" or "micro entity" status and 0 otherwise. *Treated* is a dummy that takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table A6. Robustness Tests - Patent Level

Dependent variable:	*Dummy for a licensed patent* (Patent-year level)				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	5-year window	Invention quality	Firm characteristics	Tech centers	Competition
Treated × Post	-0.000229**	-0.000231**	-0.000361**	-0.000361**	-0.000322**
	(0.000020)	(0.000020)	(0.000021)	(0.000021)	(0.000020)
Post	0.000272**	0.000260**	0.000222**	0.000224**	0.000214**
	(0.000018)	(0.000018)	(0.000018)	(0.000018)	(0.000019)
log(1+*Forward citations*)		0.000077**	0.000053**	0.000054**	0.000056**
		(0.000010)	(0.000010)	(0.000010)	(0.000010)
log(1+*Inventor size*)			0.000582**	0.000575**	0.000585**
			(0.000040)	(0.000040)	(0.000041)
log(1+*Inventor age*)			0.001732**	0.001730**	0.001740**
			(0.000049)	(0.000049)	(0.000050)
log(1+*Inventor licensing experience*)			0.009627**	0.009643**	0.009648**
			(0.000206)	(0.000206)	(0.000206)
log(1+*Silicon valley inventors*)				0.000015*	0.000014*
				(0.000007)	(0.000007)
log(1+*Route 128 inventors*)				-0.000151**	-0.000152**
				(0.000009)	(0.000009)
log(1+*Canada inventors*)				0.000105**	0.000105**
				(0.000008)	(0.000008)
log(1+*Similar prior-art inventions*)					0.000101**
					(0.000010)
log(1+*Similar non-prior-art inventions*)					0.000032**
					(0.000006)
Invention FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.000188	0.000188	0.000188	0.000188	0.000188
Licensed inventions	1,938	1,938	1,931	1,931	1,931
Observations	13,477,225	13,477,225	13,470,533	13,470,533	13,470,533
R-squared	0.363	0.363	0.367	0.367	0.367

Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors. The sample consists of patents whose priority years range from 1990 to 2010 and that experience a reduction in patent scope and controls that were matched on priority year, earliest publication year, 4-digit IPC, forward citation quartile, and patent stock quartile. *Treated* is a dummy that takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table A7. Robustness Tests - Propensity Score Matching

Dependent variable:	*Dummy for a licensed invention* (Invention-year level)				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	5-year window	Invention quality	characteristics	Tech centers	Competition
Treated × *Post*	-0.000150**	-0.000152**	-0.000134**	-0.000134**	-0.000132**
	(0.000013)	(0.000013)	(0.000013)	(0.000013)	(0.000013)
Post	0.000170**	0.000161**	0.000090**	0.000090**	0.000089**
	(0.000013)	(0.000013)	(0.000013)	(0.000013)	(0.000013)
log(1+*Forward citations*)		0.000038**	0.000032**	0.000032**	0.000031**
		(0.000007)	(0.000006)	(0.000006)	(0.000007)
log(1+*Inventor size*)			0.000104**	0.000104**	0.000103**
			(0.000023)	(0.000023)	(0.000023)
log(1+*Inventor age*)			0.001033**	0.001034**	0.001028**
			(0.000035)	(0.000035)	(0.000035)
log(1+*Inventor licensing experience*)			0.006598**	0.006603**	0.006603**
			(0.000171)	(0.000171)	(0.000171)
log(1+*Silicon valley inventors*)				0.000005	0.000005
				(0.000004)	(0.000004)
log(1+*Route 128 inventors*)				-0.000065**	-0.000066**
				(0.000005)	(0.000005)
log(1+*Canada inventors*)				0.000029**	0.000030**
				(0.000004)	(0.000004)
log(1+*Similar prior-art inventions*)					0.000034**
					(0.000008)
log(1+*Similar non-prior-art inventions*)					0.000036**
					(0.000005)
Invention FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV (pre-treatment)	0.000055	0.000055	0.000054	0.000054	0.000054
Licensed inventions	1,601	1,601	1,593	1,593	1,593
Observations	15,836,340	15,836,340	15,827,674	15,827,674	15,827,674
R-squared	0.301	0.301	0.305	0.305	0.305

Notes: The table presents estimation results on the effect of reduced patent scope on licensing propensity of inventors. The controls were identified by through propensity score matching using forward citations and inventor patent stock after doing an exact match on priority year, earliest publication year, and 4-digit IPC. The sample consists of all inventionswhose priority years range from 1990 to 2010 and that experience a reduction in patent scope. *Treated* is a dummy that takes 1 if a focal invention experiences a reduction in patent scope (i.e. publication of similar priority disclosure) and 0 for a matched control invention. *Post* is a dummy that takes 1 for post-treatment years and 0 for pre-treatment years. Robust standard errors in parentheses. ** p<0.01, * p<0.05

Table A8. Relationship between Patent Scope and Licensing

Panel A: Invention Level

Dependent variable:	*Dummy for a licensed invention* (Invention level)				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Baseline	Invention characteristics	Firm characteristics	Technology centers	Competition
Dummy for an invention with a priority disclosure	-0.00017*	-0.00018**	-0.00024**	-0.00024**	-0.00022**
	(0.00007)	(0.00007)	(0.00007)	(0.00007)	(0.00007)
log(1+*Forward citations*)		0.00049**	0.00045**	0.00046**	0.00046**
		(0.00007)	(0.00007)	(0.00006)	(0.00006)
log(1+*Inventor size*)			0.00028	0.00002	0.00001
			(0.00034)	(0.00035)	(0.00035)
log(1+*Inventing age*)			-0.00329**	-0.00308**	-0.00310**
			(0.00048)	(0.00042)	(0.00042)
log(1+*Inventor licensing experience*)			-0.00842**	-0.00871**	-0.00870**
			(0.00299)	(0.00293)	(0.00293)
log(1+*Silicon Valley inventors*)				0.00028	0.00028
				(0.00019)	(0.00019)
log(1+*Route 128 inventors*)				0.00093	0.00092
				(0.00051)	(0.00051)
log(1+*Canada inventors*)				-0.00006	-0.00005
				(0.00031)	(0.00031)
log(1+*Similar prior-art patents*)					-0.00013**
					(0.00004)
log(1+*Similar non-prior-art patents*)					0.00010
					(0.00006)
Firm FE	Yes	Yes	Yes	Yes	Yes
4-digit IPC FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
SE Cluster	Invention level	Invention level	Invention level	Invention level	Invention level
Sample Mean of DV	0.0012	0.0012	0.0012	0.0012	0.0012
Licensed Inventions	2,154	2,154	2,154	2,154	2,154
Observations	1,790,167	1,790,167	1,790,167	1,790,167	1,790,167
R-squared	0.31	0.31	0.31	0.31	0.31

Panel B: Patent Level

Dependent variable:	*Dummy for a licensed patent* (patent level)				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Baseline	Invention characteristics	Firm characteristics	Technology centers	Competition
Dummy for a patent with a priority disclosure	-0.00035**	-0.00037**	-0.00044**	-0.00045**	-0.00024**
	(0.00012)	(0.00012)	(0.00013)	(0.00012)	(0.00009)
log(1+*Forward citations*)		0.00041**	0.00034**	0.00035**	0.00037**
		(0.00007)	(0.00007)	(0.00007)	(0.00007)
log(1+*Inventor size*)			-0.00004	-0.00023	-0.00023
			(0.00062)	(0.00060)	(0.00060)
log(1+*Inventing age*)			-0.00385**	-0.00365**	-0.00370**
			(0.00068)	(0.00058)	(0.00058)
log(1+*Inventor licensing experience*)			-0.00942**	-0.00981**	-0.00984**
			(0.00351)	(0.00344)	(0.00344)
log(1+*Silicon Valley inventors*)				0.00021	0.00022
				(0.00023)	(0.00023)
log(1+*Route 128 inventors*)				0.00136*	0.00136*
				(0.00060)	(0.00060)
log(1+*Canada inventors*)				-0.00017	-0.00017
				(0.00043)	(0.00043)
log(1+*Similar prior-art patents*)					-0.00043**
					(0.00010)
log(1+*Similar non-prior-art patents*)					-0.00014
					(0.00010)
Firm FE	Yes	Yes	Yes	Yes	Yes
4-digit IPC FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
SE Cluster	Patent level	Patent level	Patent level	Patent level	Patent level
Sample Mean of DV	0.0020	0.0020	0.0020	0.0020	0.0020
Licensed Patents	4,275	4,275	4,275	4,275	4,275
Observations	2,123,592	2,123,592	2,123,592	2,123,592	2,123,592
R-squared	0.40	0.40	0.40	0.40	0.40

Notes: The table presents the relationship between patent scope and licensing propensity at the invention (Panel A) and patent (Panel B) levels. The sample used in Panel A consists of all inventions, both those that experience a patent scope reduction and those that do not, with priority years ranging from 1990 to 2010. The sample used in Panel B consists of all patents, obth those that experience a patent scope reduction and those that do not, with application years ranging from 1990 to 2010. Robust standard errors in parentheses. *Dummy for an invention with a priority disclosure* takes a value of 1 if an invention experiences a patent scope reduction and 0 otherwise. *Dummy for a patent with a priority disclosure* takes a value of 1 if a patent experiences a scope reduction and 0 otherwise. ** p<0.01, * p<0.05

Appendix A: Invention Similarity Measure

This section describes how invention similarity measure is derived using a machine learning algorithm and textual similarity scores.

Step 1: Bag of words

As the first step to calculating the proximity between two patent documents, the bag-of-words approach is used to extract all words from the claims text of all USPTO patent documents. For each patent, a vector of all word stems is created. Each word stem is weighted by the inverse of its frequency in the complete patent corpus. More formally, for each word, an inverse frequency index is created as follows:

$$I_i = N_i \times (1 - \frac{p_i}{P})$$

N_i is the number of times ith word stem in the word stem vector appears throughout the claims section of the USPTO patents. p_i is the number of patent documents that contain the ith word stem, and P is the number of patents issued by USPTO. Each item in the index represents the weight assigned to extracted word stems according to their specificity across all USPTO patent documents.

An important part of the word stemming process is mapping acronyms and technical concepts. For example, the acronym RAM refers to Random Access Memory. Thus, in the textual comparison algorithm, when the sequence of words Random Access Memory appears, it is collapsed into RAM. Acronyms appear in capital letters on patent documents. All words with at least two capital letters are retained and manually search is conducted to find their meaning. To mitigate cases where multiple meanings exist for a given acronym, acronym-meaning match is performed at the four-digit IPC level. (Chemical compounds also appear in capital letters, but they are unchanged.)

Step 2: Distance between words

Similar ideas might be described using different text. Thus, a major challenge is how to compute the "technical distance" between two words, that is how to calculate the likelihood that two different words describe the same technical concept. This challenge is likely to be more severe for patents because inventors and their patent agents have a strong incentive to describe their inventions differently from the descriptions of prior art patents, leading to an artificial textual "distance" between the invention and most relevant prior-art inventions.

To address this challenge, a dictionary is developed that aims to measure the probability that two distinct words refer to the same technical concept. For this purpose, words used in patent documents deemed to be technically similar by human experts (i.e. patent examiners) are identified. A measure of technical distance between two given words is computed using prior-art patents referenced by examiners in rejecting patent applications for a lack of novelty or obviousness.

To create the technical distance between two words the following steps are followed. First, a random sample of about 150,000 non-final rejection letters is extracted from the USPTO's Public PAIR (Patent Application Information Retrieval) system. Only non-final rejection letters with rejections pertaining to novelty and non-obviousness as outlined in 35 U.S.C. 102 and 35 U.S.C. 103 of the USPTO's Manuals for Patent Examining Procedure. For each patent with a non-final rejection, the claims text of the original patent application associated with that rejection is extracted along with the text of the prior-art patents cited as the reason for the rejection ("rejection prior-art patents").[1]

Second, all relevant word stems are extracted from the claims section of the focal patent application and corresponding prior-art patents listed by the patent examiner as the basis for a rejection.[2] At the end of this step, all relevant word stems are extracted from the rejected applications and prior-art patents listed on non-final rejection letters. Next, the proximity between each pair of the word stems is calculated based on their co-occurrence. To account for the baseline tendency of two word stems to co-occur across two documents, for each rejected application and rejection prior-art patent pair, a control pair is constructed by linking the rejected application with a control patent that was not cited as a reason for the rejection but is in the same 4-digit IPC (International Patent Classification) and has the same application year as the rejection prior-art patent. Proximity between a pair of word stems is calculated as the ratio of the number of times the pair appears in the rejected application and rejection prior-art patent to the number of times it appears in the rejected application and the control prior-art patent. More precisely, proximity between two word stems is calculated as:

---

[1] In cases where multiple rejections are associated with the same application, the relevant (modified) application claims are extracted for each rejection.

[2] Original applications rather than the final patent documents because claims can change through patent examination process and thus using the original applications allows us to compare the relevant set of claims between the applications and rejection prior-art patents.

$$Proximity_{w1,w2} = \frac{(A \bigcup R)_{w1,w2}}{(A \bigcup C)_{w1,w2}}$$

$A \cup R_{w1,w2}$ is the number of times the words *w1* and *w2* co-occur within the focal application *A* and rejection prior-art patent *R*. $A \cup C_{w1,w2}$ is the number of times the words *w1* and *w2* co-occur in the focal application *A* and control patent *C*. Because the same word stem pair *w1* and *w2* can co-occur in more than one application and rejection prior-art patent pair, the proximity scores between *w1* and *w2* are averaged across all application and rejection prior-art patent pairs, denoted by $\bar{P}_{w1,w2}$.

Step 3: Textual overlap between documents

The final step of our algorithm is to construct a similarity score between a pair of patent documents based on their words and the "technical distance" between these words from Step 2. To derive the textual proximity between two patent documents, a vector of words is created for each patent document with their corresponding weights (i.e. inverse frequency) as described in step 1. Then, the cosine proximity score is calculated between word vectors *W1* and *W2*, where each vector consists of *n* elements while taking into account the average word pair proximity, $\bar{P}_{w1,w2}$ calculated in step 2:

$$PS_{W1,W2} = \frac{\sum_{i=1}^{i=n} W1_i \times W2_i \times \bar{P}_{w1_i,w2_i}}{\sqrt{\sum_{i=1}^{i=n} W1_i^2}\sqrt{\sum_{i=1}^{i=n} W2_i^2}}$$

Finally, the proximity score $PS_{W1,W2}$ is normalized to be between 0 and 1 by dividing it by $\max(PS_{W1i,W2i})$. 1 indicates the highest similarity, and 0 indicates the lowest similarity between two documents.

Step 4: Machine learning algorithm

Finally, using textual similarity scores between the focal and prior-art patent pairs and patent-specific attributes, a random forest algorithm (Breiman 2001) is trained to predict whether a patent examiner would judge the claims of a focal patent to be highly similar to the claims of a prior-art patent. To do so, textual similarity scores for a randomly chosen 41,000 rejection patent pairs (i.e. a pair of a focal patent application under examination and a prior-art patent used as a reason for rejecting the focal patent's claims[3]) are computed and key attributes for each of the

[3] As explained in the previous section, only the rejections pertaining to novelty and non-obviousness as outlined in 35 U.S.C. 102 and 35 U.S.C. 103 of the USPTO's Manuals for Patent Examining Procedure are used. 41,000 is the number after dropping rejection patent pairs pertaining to reasons other than novelty and non-obviousness.

patents are added.[4] (The attributes include priority and earliest publication years, different types of citations, IPC, and assignees.) Then, for each focal patent, a control prior-art patent, which has not been judged to be similar to the focal patent, is found by matching on priority year, earliest publication year, and 4-digit IPC code. Lastly, the sample is split into two parts, where one of the subsamples is designated as a training dataset used to train the machine learning algorithm and the other is designated as the test dataset used to test the performance of the trained algorithm.

A random forest model is trained using the training dataset, where the dependent variable is a dummy indicating whether the claims of a focal patent application and its paired prior-art patent were judged by a patent examiner to be similar enough to merit a rejection letter. The predictors used in the model include a textual similarity score for each rejection patent pair and each control patent pair along with priority and publication years, different types of citations (forward, backward, and NPL citations), patent family size, number of claims, patent stock of the assignee, and technology areas for both the focal and the prior-art patent.

Figure 1 presents a receiver operator characteristics (ROC) curve that plots true positive rate against false positive rate of the predicted outcomes using the test dataset. The area under the ROC curve, which measures the performance of the classifier, is approximately 0.87 indicating that, given a randomly chosen patent pair deemed similar by a patent examiner (a rejection) and a randomly chosen patent pair not deemed similar by a patent examiner (non-rejection), the algorithm is able to predict with 87% accuracy which patent pair is more likely to be a rejection. Both mean decrease in prediction accuracy and mean decrease in Gini Index (a node purity measure that indicates the degree to which a predictor contributes to the homogeneity of the nodes) indicate that textual similarity score, forward citations of the prior-art patent, and technology area of the focal patent are strong predictors.

[4] The same algorithm used to calculate textual similarity scores outlined in Steps 1 through 3 is used.