2017

Essays on patent examination and standard essential patents

BOSTON UNIVERSITY

QUESTROM SCHOOL OF BUSINESS

Dissertation

ESSAYS ON PATENT EXAMINATION AND STANDARD ESSENTIAL PATENTS

by

CESARE RIGHI

B.A., Bocconi University, 2006
M.S., Bocconi University, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader

Timothy S. Simcoe, Ph.D.
Associate Professor of Strategy & Innovation

Second Reader

Iain M. Cockburn, Ph.D.
Richard C. Shipley Professor in Management
Professor and Chair of Strategy & Innovation

Third Reader

Christian Catalini, Ph.D.
Fred Kayne (1960) Career Development Professor of Entrepreneurship
Assistant Professor of Technological Innovation, Entrepreneurship, and Strategic Management
Sloan School of Management
Massachusetts Institute of Technology

To my family.

Acknowledgments

First of all, I thank the members of my dissertation committee, Professors Timothy Simcoe, Iain Cockburn and Christian Catalini, for their help and advice on this dissertation and other research projects, and for their support during my years at Boston University. I also thank the other faculty members of the Strategy & Innovation department at Questrom School of Business for their advice and their contribution to my intellectual development.

Chapters 2 and 4 are joint works with Timothy Simcoe. During the preparation of chapter 2 I benefited from conversations with James Bessen and Michael Meurer. I also thank Kira Fabrizio, Jeffrey Furman and Olga Hawn for their suggestions on a previous version of chapter 4. The data utilized for the analysis in chapters 3 and 4 were collected in collaboration with Rudi Bekkers, Christian Catalini, Arianna Martinelli and Timothy Simcoe. I also thank Jeffrey Kuhn and Kenneth Younge for access to data used in chapter 4.

I thank Yanbo Wang, Arvids Ziedonis, Rosemarie Ziedonis, Myriam Mariani, Berardino Palazzo, Caroline Flammer and Marshall Van Alstyne for conversations and advice that contributed to my development as a doctoral student. I also thank the former and current students of the Ph.D. program at Questrom School of Business, and in particular Diego Zunino, Jax Kirtley, Marketa Sonkova, Jeremy Watson, Sina Khoshsokhan, Hye Young Kang, Rebecca Karp, Tushar Shanker, Diogo Duarte, Bingyi Chen, Thaddeus Neururer, De Chen and Albert Valenti. I also thank Senem Aydin, Navid Bazzazian, Hendrik Meder, Dominic Distel, Martin Murmann and Simone Santamaria. I would like to thank Patricia Caffrey and Marta Mcmanus for their advice, and Gregory DeFronzo for his help with the IT resources at Questrom and friendship during these years.

I thank Questrom School of Business and my dissertation committee for financial support during my studies. The last semester of my doctoral studies was funded by Boston University School of Law with funding provided by Google. I had no contacts with the company and all the views expressed in my work are solely my own.

Last but not least, I thank my family for their love and encouragement.

Cesare Righi

ESSAYS ON PATENT EXAMINATION AND STANDARD ESSENTIAL PATENTS

CESARE RIGHI

Boston University, Questrom School of Business, 2017

Major Professor: Timothy S. Simcoe, Ph.D.
Associate Professor of Strategy & Innovation

ABSTRACT

This dissertation contains three studies on the patenting process and standard essential patents. The first study analyzes the matching of patent applications to examiners at the U.S. Patent and Trademark Office. The analysis uses statistical tests originally developed to study industry agglomeration and finds strong evidence that examiners specialize in particular technologies. Specialization is more pronounced in the biotechnology and chemistry fields, and less in computers and software. Evidence of specialization becomes weaker conditioning on technology subclasses. There is no evidence that certain examiners specialize in applications that have greater importance or broader claims. Finally, the study shows that more specialized examiners have a lower grant rate and produce a larger narrowing of claim-scope during examination. The results have implications for the growing literature that exploits examiners characteristics to study the effects of patenting.

In the second study, I analyze the strategic behavior of applicants for Standard Essential Patents. Owners of these patents (and especially those that rely more on patents to generate revenues) use the mechanisms provided by the patent system to delay issuance more often than owners of similar patents. The analysis also shows

that applicants for Standard Essential Patents may delay issuance to obtain the right balance between patent breadth and strength, and that companies prolong prosecution until the standard is set, possibly to cover the standard with additional claims. Finally, I find a positive correlation between the issuance lag and the probability of patent litigation. This suggests that owners of Standard Essential Patents may delay issuance to obtain patents that are more valuable, or that longer lags are associated with failures in licensing negotiations.

The third study exploits Standard Essential Patents as a window on standardization and analyzes the direction of technical progress that builds upon compatibility standards. It uses patent citations to characterize the dispersion of cumulative inventive activity across technological areas. The overall pattern of results suggests that Standard Setting Organizations select technologies that are important in a relatively narrow technological area, and their adoption as input for following inventive activity broadens after standardization.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Whether the patent system encourages or hinders innovation and economic growth is heavily debated. A very large literature in economics analyzes costs and benefits of patents and several trade-offs involved in the design of a patent system (Arora et al., 2001; Arrow, 1962; Boldrin and Levine, 2008; Bessen and Maskin, 2009; Gans and Stern, 2003; Hall, 2007; Heller and Eisenberg, 1998; Nordhaus, 1967).

The traditional rationale for patents is to increase the appropriability of the returns to investments in innovation. Inventive activity is often affected by a market failure that justifies some form of government intervention. Innovators often do not appropriate all the returns of their investments because of knowledge spillovers and imitation by competitors. This can lead to underinvestment in innovation. Patents are designed to increase the incentives to innovate by granting a temporary monopoly on the economic exploitation of an invention. Furthermore, reducing the threat of expropriation of the original inventor and reducing the risk of imitation, patents may enable market transactions of ideas and technologies, generating gains from trade (Arora et al., 2001; Arrow, 1962; Gans and Stern, 2003; Griliches, 1992; Jaffe et al., 1993; Nelson, 1959; Nordhaus, 1967; Teece, 1986; Thompson and Fox-Kean, 2005; Wright, 1983).

Patents may also have negative effects. When innovation is cumulative, patents raise the costs for inventors that use previous research as input. Moreover, in many industries they are not an effective mechanism for appropriating the returns of research and companies often patent just for strategic reasons. Especially in industries characterized by complex technologies, in which companies need to combine multiple components protected by patent rights, this may create a patent thicket. The patent thicket not only increases transaction costs for firms that need to negotiate with many patent holders, but also raises the risk of hold up for companies that can inadvertently infringe patents of other companies. To respond to these threats, companies have developed institutions to navigate the patent thicket, such as patent pools and intellectual property policies of standard setting organizations (Boldrin and Levine, 2008; Bessen and Maskin, 2009; Cohen et al., 2000; Gallini and Scotchmer, 2002; Hall and Ziedonis, 2001; Levin et al., 1987; Rysman and Simcoe, 2008; Shapiro, 2001).

To assess the effect of the patent system on innovation and economic outcomes, it is fundamental to understand how patents are produced. An emerging literature is improving our understanding of the production of patents and its consequences (Carley et al., 2015; Cockburn et al., 2002; Frakes and Wasserman, 2014; Hall and Ziedonis, 2001; Lemley and Sampat, 2012; Ziedonis, 2004). The first two studies in this dissertation aim to contribute to this literature.

The first study focuses on a key stage of the production of patents: the matching of patent applications to patent examiners. Patents are the product of a long interaction between applicants and the patent office, and patent examiners have an important role in determining what is eventually patented and the quality of patents.

A recent stream of papers is trying to exploit variation in the characteristics of examiners to estimate the effects of patents on economic outcomes. The empirical strategies in these studies are based on the assumption that applications are randomly assigned to examiners at the United States Patent and Trademark Office (USPTO). In chapter 2, a joint project with Timothy Simcoe, we test whether this assumption holds. While other studies test whether application characteristics are correlated with examiner characteristics, in this work we address the question from a different perspective. Our approach is new and analyzes explicitly the allocation of applications to examiners (within art units, i.e. groups of examiners working on similar technologies) using statistical tests originally developed to study industry agglomeration.

We find strong evidence that examiners within art units of the USPTO specialize in particular technologies. Interestingly, there are significant differences across technological areas. Examiners in the technology centers (i.e. groups of art units) in biotechnology and chemistry seem to be more specialized than those in computers and communications. We also detect some sorting of applications based on the identity of the assignee. However, this may be due technological specialization of the inventors. Evidence of specialization becomes weaker, but does not completely disappear, if we condition on the primary classification of applications (technology class-subclass), again with differences across fields.

An important assumption for many studies that want to exploit the supposedly random assignment of applications to examiners is that important characteristics of applications are not correlated with the type of examiner they are assigned to. We use two (imperfect) proxies of the *ex ante* value of applications (scope of the first independent claim and number of patent family members) to test whether or not we observe agglomeration driven by patent value. We find no evidence that certain examiners specialize in applications that have greater importance or broader claims.

We also explore the implications of examiner specialization for the outcomes of patent examination. In the last part of the analysis we show that more specialized examiners have a lower grant rate and are more likely to narrow the claims of a patent during examination.

The results of this study have important implications for the emerging literature that relates differences in the examination process to economic outcomes. While randomization may make an instrumental variable constructed on examiner characteristics more credible, our analysis does not imply that those instruments are invalid. However, researchers willing to pursue that approach should be careful in their applications, taking into account the details of the institutional environment. Some assumptions may be more plausible than others, depending on the particular research question and setting analyzed.

The second study in this dissertation analyzes the patenting process from the perspective of the applicant, focusing on strategic behavior. In particular, in chapter 3 I analyze the prosecution strategy of the applicants for a very valuable subset of patents: Standard Essential Patents (SEPs).

The U.S. patent system provides unique opportunities to patent applicants to delay the prosecution of their patents and modify the claims of applications. Delays and changes to claims can be used by companies to cover recent technical developments and even new technologies developed by competitors. These opportunities are particularly appealing to companies involved in standardization, because they can exploit them to make their patents essential or more essential for the implementation of a standard after observing how a proposed standard evolves.

I compare SEPs with applications similar at the time of filing and likely exposed to similar examination and find that applicants for SEPs use more often mechanisms provided by the U.S. patent system to delay issuance and modify claims. This leads to longer lags between priority and issuance.

Companies involved in “upstream” activities such as pure knowledge developers, patent holding companies or producers of components use these mechanisms more aggressively. A simple explanation for this result is that upstream companies rely more on patents to generate revenues, thus have more incentives to increase the value of their patents.

I also analyze how the scope of the claims changes over time and find that owners of SEPs may delay issuance to obtain the right balance between patent breadth and strength. While SEPs and similar patents have more or less the same scope at filing, SEPs have broader scope at issuance. However, longer prosecution is associated with claim narrowing. As narrower claims are less likely to be invalidated in a court, this suggest that applicants for SEPs may want first to obtain a relatively stronger patent that is broad enough to cover a number of implementations of a standard, and then file continuations to cover additional implementations. This interpretation is reinforced by the findings that the probability of issuance increases significantly after SEP disclosure (which I use as a proxy for standardization timing) and that the latter is positively correlated with the filing of continuation applications.

Finally, I find that an increase in the lag between priority and issuance of SEPs is correlated with an increase in the probability of patent litigation. This suggests that owners of SEPs may delay issuance to obtain patents that are more valuable, or that longer lags are associated with failures in licensing negotiations.

These results help us to understand the formation of the patent thicket and the consequences in terms of holp-up, and are therefore relevant for the design of public policies and rules of standard setting organizations to limit the negative consequences of patents.

The last study in this dissertation is a joint work with Timothy Simcoe. In chapter 4 we study the relationship between technology endorsement by Standard Setting Organizations (SSOs) and the direction of inventive activity. SSOs are institutions that coordinate the collective development of new compatibility standards and are particularly important in Information and Communication Technology (ICT) industries. An important role of SSOs is to design Intellectual Property (IP) policies to partially reduce the threat of hold-up related to the inclusion of patented technology into technical standards. Chapter 4 contributes to the literature that studies the performance of SSOs, and more in general to the literature that studies the role of institutions in the process of cumulative inventive activity.

We develop a theoretical framework that combines the concepts of direct and indirect network effects with the insights of the literature on cumulative inventive activity to understand how SSOs shape the scope of inventive activity upon compatibility standards. We introduce the concepts of "deepening" (cumulative technical progress characterized by relatively low dispersion across technological areas) and "broadening" (cumulative technical progress characterized by relatively high dispersion across technological areas) and relate them to the activities of SSOs.

In the empirical analysis, we exploit the disclosure of SEPs as a window on standardization within SSOs and compare the dispersion of patent citation flows to SEPs and similar patents. In the first part of the analysis we use a measure of patent-to-patent text similarity and a new measure that takes into account the probability of inter-class citation to estimate the balance between deepening and broadening. In the second part of the analysis we separate deepening and broadening using citations from technological classes that repeatedly cite a patent as a measure of deepening and citations from new classes as a measure of broadening. The results provide evidence that both trends are occurring. The overall pattern of results suggests that SSOs select technologies that are important in a relatively narrow technological area, and their adoption as input for following inventive activity broadens after standardization. We also explore the heterogeneity across SSOs and licensing terms and find substantial differences.

The three studies in this dissertation are all related by the goal to improve our understanding of the patent system and the institutions developed to manage the trade-offs related to patents. The first two studies analyze the production of patents from different perspectives. Chapter 2 focuses on a key stage of the production of patents, the assignment of patent applications to examiners. This is an early stage that has important consequences for the outcomes of the examination process and eventually for innovation and economic outcomes. The role of patent examiners is fundamental because they determine what inventions are patented and the characteristics of patent rights. Chapter 3 focuses on the strategies of applicants during the patenting process. As patents are the product of an interaction between the patent office and the applicants, the strategies of applicants are also important to determine the characteristics of the patents produced and affect innovation and economic outcomes. The setting analyzed in this chapter is the production of SEPs, which are a subset of very valuable patents with great importance for technical progress in complex industries. In these industries, a patent thicket can be detrimental for innovation. SSOs, not only with their IP policies but also with other activities such as coordination of research efforts and certification of technical merits, are key players in those contexts. Chapter 4 studies their role in shaping the direction of inventive activities.

Chapter 2

Patent Examiner Specialization

2.1 Introduction

In 2015, the U.S. Patent and Trademark Office (USPTO) received 589,410 utility patent applications. Matching each application to a qualified examiner is a fundamental part of the examination process. This matching proceeds in two steps. First, each application is assigned to an "art unit" comprised of several examiners who specialize in a particular technology. Then the patent is assigned to an individual examiner within that art unit. Several studies have suggested that the second step in this process is more-or-less random, and then, building on an idea first proposed by Sampat and Williams (2015), used examiner characteristics as an instrumental variable for examination outcomes.[1]

We re-examine the random matching assumption, and find strong evidence of technological specialization by examiners within art units. While this does not invalidate the use of instrumental variables based on examiners' characteristics, it does imply that the exclusion restriction cannot be justified on the basis of random assignment. We show that examiner specialization is more pronounced in some technology areas (Biotechnology and Chemistry), and less in others (Computers and Software). Evidence of specialization becomes weaker, but does not completely disappear, if we condition on U.S. Patent Classification System (USPC) sub-classes. However, we

[1] Papers adopting variants on this identification strategy include Farre-Mensa et al. (2015), Feng and Jaravel (2016), Gaulé (2015), Kuhn (2016), Kuhn et al. (2016), and Sampat and Williams (2015).

find no evidence that certain examiners specialize in applications that have greater importance or broader claims. Finally, we show that more specialized examiners have a lower grant rate and produce a larger narrowing of claim-scope during the examination process.

This is the first paper to systematically test the random matching hypothesis across all of the technology areas examined by the USPTO. Our methods for detecting specialization are borrowed from the literature on industry agglomeration (Mori et al., 2005). Specifically, we compute a pair of test statistics that ask whether application characteristics (e.g. technology subclass) are less dispersed across examiners than we would expect under random assignment. These methods focus specifically on the null hypothesis of random assignment, unlike IV falsification tests that ask the slightly different question of whether examiner and application characteristics are correlated.[2] Each of our calculations is performed at the art-unit-year level, and we examine the entire distribution of p-values for various characteristics, including technology subclass, assignee, and indicators of patent value (family size) and scope (first independent claim length).

At a substantive level, our findings illustrate how the USPTO manages a tension between efficiency and fairness. One way to promote fairness is through uniform application of patentability criteria, but prior research suggests that this is difficult. Some examiners are simply tougher than others (Sampat and Williams, 2015; Kuhn et al., 2016), and experienced examiners are more lenient on average, partly because of time constraints (Lemley and Sampat, 2012; Frakes and Wasserman, 2014). Random matching provides another path to fairness, but forgoes the efficiency benefits of further technological specialization. Our analysis shows that the amount of specialization varies across art units, leading some applicants to get tougher examiners on

[2]Even if examiners are highly specialized by technology, other characteristics (e.g. propensity to grant) might be randomly distributed. This is why the IV strategy described above could still be valid.

average. But we find no evidence that particularly important applications (with large families) or broad applications (with short first independent claims) are assigned to specific examiners.

We discuss two plausible explanations for our finding that examiners are more specialized in Chemistry and Biotechnology than in the computer-related art units. One possibility is that "generalist examiners" are able to evaluate computing inventions, while more specialized skills and knowledge are required in chemistry and life sciences. Another possibility is that the USPC technology classification system works better in chemistry and biotech, so we fail to observe much of the specialization that takes place within computer-related art units. Distinguishing between these hypotheses is good topic for future research.

Finally, we find a positive correlation between specialization and a more stringent examination process, suggesting that it is easier for examiners to find relevant prior art when working in a more familiar field. Under random matching, these estimates have a causal interpretation. Alternatively, they remain important for showing how non-random matching is related to important patent examination outcomes.

The paper proceeds as follows. Section 2.2 describes how the USPTO assigns applications to examiners. Section 2.3 explains our methods and data. Section 2.4 presents results and Section 2.5 concludes.

2.2 Patent Examiner Assignment at the USPTO

When a patent application is filed, the Office of Patent Application Processing reviews the formality requirements of the application and assigns it a serial number. A contractor defines the technological classification of the application using USPC class and subclass codes.[3] Each application has at least one mandatory classification,

[3]The two main purposes of the USPC are to facilitate the retrieval of technical documents and to ease the allocation of applications to the examining personnel specialized in a particular

which is defined as a unique combination of class and subclass identifiers. The current version of the USPC has roughly 450 classes and more than 150,000 subclasses.

The USPTO has eight Technology Centers (TCs) responsible for examination of utility patent applications in broad technological areas. Each TC is comprised of several art units, or teams of patent examiners who specialize in a particular technology. Technological classifications are used to assign each new patent application to a specific art unit.[4]

Within each art unit the initial assignment of a new application is handled by a Supervisory Patent Examiner (SPE). The SPE can refine the technological classification of a new application if it is incorrect, or request that an application be transferred to another art unit.[5] But in most cases, the SPE will assign the application to an examiner within her art unit. This is the step we analyze below.

Previous research documents that SPEs have substantial discretion in examiner assignment. Some SPEs interviewed by Lemley and Sampat (2012) mention assigning applications to examiners essentially randomly within subclasses. Other SPEs give the oldest unassigned application to an examiner when she finishes the examination of another application. Cockburn et al. (2002) suggest that the degree of technological specialization varies across art units – in some art units an individual examiner is responsible for almost all applications in a specific technology class, and in others the examiners are less specialized.

While the USPTO constantly monitors the performance of art units and examiners to ensure a certain level of quality of the examination process, the assignment to a

technology. For details, see `http://www.uspto.gov/sites/default/files/patents/resources/classification/overview.pdf`. Although it was replaced by the Cooperative Patent Classification (CPC) on January 1, 2013, the USPC is the relevant classification for the entire period of our study.

[4]For the current list of classes and subclasses examined by each art unit, see `http://www.uspto.gov/patents-application-process/patent-search/understanding-patent-classifications/patent-classification`.

[5]The Manual of Patent Examining Procedure sec. 903.08 describes the rules governing assignment and transfer of applications between art units.

particular art unit and to a specific examiner can have important consequences for an application. Different practices across art units and the personal approach of each examiner can affect whether an application is eventually granted (Sampat and Williams, 2015), how quickly a decision is reached (Farre-Mensa et al., 2015), and the scope and strength of an issued patent (Kuhn et al., 2016). This variation in standards led Cockburn et al. (2002) to conclude that "there may be as many patent offices as patent examiners."

2.3 Methods and Data

We use two statistical tests originally developed to analyze industry agglomeration. In our application, patent examiners are analogous to cities, and technology subclasses (or other application characteristics) are analogous to industries. Each test might be viewed as a multivariate generalization of a t-statistic, which compares observed frequencies to the distribution under random assignment.

2.3.1 Agglomeration Test Statistics

Divergence Index

The D-index was developed by Mori et al. (2005), building on Kullback and Leibler (1951), and is based on the concept of relative entropy. Suppose we have a set of applications characterized by category $i \in \mathbf{I} = \{1, ..., I\}$, assigned to a set of examiners denoted by $r \in \mathbf{R} = \{1, ..., R\}$. In our application, the categories i may correspond to USPC subclasses, assignees or any other predetermined observable characteristic of a patent application. Under random allocation, examiner r's share of all applications from category i should equal her share of the overall population.

To formalize that idea, define n_{ir} as the number of applications in category i assigned to examiner r, and $N_i = \sum_{r=1}^{R} n_{ir}$ as the total number of applications in category i. The reference distribution $p_0 = (p_{0r} : r \in \mathbf{R})$, where $p_{0r} = \frac{\sum_{i=1}^{I} n_{ir}}{\sum_{i=1}^{I} N_i}$ measures examiner r's share of all applications, is the share we expect her to be allocated from each category under the null of random assignment.

Let p_{ir} denote the true probability that a randomly sampled application in category i is assigned to examiner r, so the distribution across examiners for the category is $p_i = (p_{ir} : r \in \mathbf{R})$. We can measure the divergence between p_i and p_0 using the relative entropy of p_i with respect to p_0, called the D-index by Mori et al. (2005):

$$D(p_i|p_0) = \sum_{r \in \mathbf{R}} p_{ir} \ln \left(\frac{p_{ir}}{p_{0r}} \right).$$

$D(p_i|p_0)$ is nonnegative, achieves its minimum at $p_i = p_0$ and its local maxima when all applications in category i are assigned to a single examiner.

To estimate the D-index, we use the observed data to estimate the probabilities p_{ir}, with $\hat{p}_{ir} = \frac{n_{ir}}{N_i}$, thus estimating:

$$D(\hat{p}_i|p_0) = \sum_{r \in \mathbf{R}} \hat{p}_{ir} \ln \left(\frac{\hat{p}_{ir}}{p_{0r}} \right). \tag{2.1}$$

These probability estimates converge to the true value exponentially fast with the increase in sample size for a given category N_i.

As shown by Mori et al. (2005) the D-index can be related to the the log likelihood ratio (λ):

$$-\frac{\ln \lambda}{N_i} = \sum_{r \in \mathbf{R}} \frac{n_{ir}}{N_i} \ln \left(\frac{\hat{p}_{ir}}{p_{0r}} \right) = D(\hat{p}_i|p_0).$$

Given that $-2\ln\lambda$ is distributed asymptotically as a chi-square with $R-1$ degrees of freedom, we can use this relationship for testing the null hypothesis $p_i = p_0$ against the alternative of a more agglomerated distribution (see Mori et al. (2005) for details).[6] In our application, the number of tests will equal the number of categories (e.g. one

[6] In practice, we compute $2N_i D(\hat{p}_i|p_0)$ and use it for a one-tailed chi-square test with $R-1$ degrees of freedom.

per technology subclass) and we examine the distribution of p-values from all of these tests conditional on a given sample-size threshold (e.g. $N_i > 20$).

Multinomial Test for Agglomeration and Dispersion

MTAD computes multinomial likelihood functions for an allocation of agents to a set of discrete locations. In our application, the agents are patent applications and locations correspond to examiners. If the likelihood of the observed data is lower (higher) than the likelihood under random choice, MTAD indicates that the agents are agglomerated (dispersed). This approach differs from the D-index because the statistic is computed for an entire art unit, and because it can detect both agglomeration and over-dispersion relative to the null of random allocation.

To provide a brief formal description of MTAD, we adapt the notation provided in Rysman and Greenstein (2005). Suppose we have R examiners, each receiving n_r applications, with $r = 1, ..., R$. The variable n_r is bounded between $\underline{n} = 0$ and $\overline{n} = \infty$ and distributed according to the discrete distribution $f(n_r)$. Each examiner can be assigned applications of c types. The unconditional probability of being assigned type c is p_c for $c = 1, ..., C$. The observed number of applications of type c assigned to examiner r is x_r^c. Define $\mathbf{x}_r$ as the vector of elements $x_r^1, ..., x_r^C$, $\mathbf{p}$ as the vector of probabilities $p_1, ..., p_C$, $\mathbf{n}$ as the $R \times 1$ vector of applications assigned to each examiner, and $\mathbf{X}$ as the $R \times C$ matrix of allocations. If examiners are assigned applications independently, the likelihood of observing outcome $\mathbf{x}_r$ for examiner r is the multinomial pdf

$$\mathcal{L}(\mathbf{x}_r, n_r, \mathbf{p}) = \binom{n_r}{x_r^1, ..., x_r^C} p_1^{x_r^1} ... \, p_C^{x_r^C}$$

and the average log-likelihood for the data is

$$l(\mathbf{X}, \mathbf{n}, \mathbf{p}) = \frac{1}{R}\sum_{r=1}^{R} \ln\Big(\mathcal{L}(\mathbf{x}_r, n_r, \mathbf{p})\Big).$$

We want to compare this log-likelihood with the value we would observe under independent random assignment. Let the random variable $l(f, \mathbf{p})$ be distributed according to the distribution $l(\mathbf{X}, \mathbf{n}, \mathbf{p})$ if $\mathbf{X}$ was *actually* drawn from a multinomial distribution and n_r was drawn from f. Then the expected log-likelihood under random allocation is given by

$$E[l(f, \mathbf{p})] = \sum_{n_r}\Bigg(\sum_{\mathbf{z}\in\Phi(n_r)} \ln \mathcal{L}(\mathbf{z}, n_r, \mathbf{p}) \times \mathcal{L}(\mathbf{z}, n_r, \mathbf{p})\Bigg) f(n_r)$$

where $\Phi(n_r)$ is the set of all possible allocations of the n_r applications. To compute $E[l(f, \mathbf{p})]$ we treat $\mathbf{p}$ as known and take f to be the empirical distribution of n_r. The MTAD test-statistic is

$$t(\mathbf{X}, \mathbf{n}, \mathbf{p}) = l(\mathbf{X}, \mathbf{n}, \mathbf{p}) - E[l(f, \mathbf{p})]. \tag{2.2}$$

A negative (positive) value of $t(\mathbf{X}, \mathbf{n}, \mathbf{p})$ signals agglomeration (dispersion) of patent applications compared to the null of random assignment. This statistic is distributed asymptotically normal and we use simulation to generate its confidence intervals.[7]

2.3.2 Data

Our main data source is the USPTO Patent Examination Research Dataset (Graham et al., 2015), which is based on information from the Public Patent Application Information Retrieval system (Public PAIR). We also use information from PATSTAT,

[7] See Rysman and Greenstein (2005) for details on the test. Timothy Simcoe developed a software module to easily perform this test in Stata, available at the following link: `https://ideas.repec.org/c/boc/bocode/s457205.html`

the USPTO Patent Assignment Dataset (Marco et al., 2015) and the Patent Claims Research Dataset (Marco et al., 2016).

We restrict our analysis to "regular" utility patent applications filed on or after the enactment of the American Inventor's Protection Act of 1999 (November 29, 2000) and before January 1st 2013, whose examiner is affiliated with one of the eight technology centers responsible for the examination of utility patent applications.[8] [9] Under the AIPA, regular utility patent applications are generally published eighteen months after the filing date.[10]

The data have several limitations. First, applications will not appear in our data if they are abandoned before publication, or if the applicant files only in the United States, requests that the application not be published and is not granted a patent. Previous research suggests that these outcomes are relatively rare. A second problem is that we do not observe whether applications are transferred from one examiner to another before or after the information on the examiner is recorded in our data.

We exclude applications filed after 2012 to avoid problems related to publication lags and a change in the USPTO technological classification scheme. We also exclude serialized continuations (continuation applications, continuations in part and divisional applications) because these applications are usually assigned to the same examiner of the original application, and would therefore lead us to overstate the extent of agglomeration. Our primary analysis sample contains 2,931,713 applications

[8]The USPTO Patent Examination Research Dataset provides the information on the examiner of record of an application as of January 24, 2015. This is the examiner as of that date for pending applications and the examiner at the time of disposal for disposed applications. The art unit is the art unit of the examiner of record for the application as of the last office action recorded for the application.

[9]We report the distribution of applications by TC in our sample in figure A1.

[10]While granted applications have always been available for public inspection after the grant date, this was not the case for rejected applications before the AIPA. Under the AIPA, inventors can avoid publication after eighteen months if they forgo foreign patent protection (Graham and Hegde, 2014, 2015). However, Graham et al. (2015) show that about 95% of the regular non-provisional utility patent applications filed between 2001 and 2012 can be found in Public PAIR.

examined by 12,389 examiners affiliated with 590 art units.

2.3.3 Variables

We focus on several application characteristics (indexed by i or c above) that may influence the assignment of applications to individual patent examiners within an art-unit-year. We use the filing date of the applications to assign them to art-unit-year cells. Grouping applications also by year helps account for possible changes in assignment practices over time and turnover in the pool of examiners. The first of these characteristics is the primary USPC subclass of the application. If patent examiners specialize in evaluating applications related to particular technologies, we expect to see agglomeration on this variable.

The identity of the applicant may also influence the allocation of applications — either directly or due to technological specialization. We measure this with the assignee of an application. Specifically, we retrieve information on the assignment of applications, identify the assignments made by the inventors to their employers before the application is docketed to an examiner, clean and standardize the assignee names and create clusters of names that are likely to belong to the same organization, to which we assign a unique identifier.[11] After completing this process, we have missing assignee data for 668,642 applications. To check the robustness of our assignee measurement, we utilize a second measure of the applicant identity: the customer number assigned by USPTO to each application. This number identifies the correspondent for application-related matters and is usually either the law firm representing the applicant or the legal department of the firm filing the application.[12]

[11] We employ an assignee name cleaning and standardization routine that builds upon Thoma et al. (2010) and the name standardization routines developed for the NBER Patent Data Project available at `https://sites.google.com/site/patentdataproject/Home/posts/namestandardizationroutinesuploaded`. Details are available upon request.

[12] Results of the customer-number analysis are similar to those for the assignee and are available upon request.

We would like to examine whether some examiners are assigned a larger share of "high value" applications. The size of a patent family is often used as a proxy for economic value of the invention because increased value leads patentees to file in more countries (Harhoff et al., 2003; Putnam, 1996). We count the number of applications in the same DOCDB patent family, with filing dates on or before the focal application date, to construct an indicator variable that equals one if a focal application is above the 95th percentile in the family size distribution (within an art-unit and filing-year). We call this variable "DOCDB Family Size."[13]

Finally, we consider whether some examiners are assigned applications seeking greater scope of protection. Kuhn et al. (2016) show that the length of the first independent claim in a patent is a good measure of patent scope. The idea behind this measure is that shorter claims provide broader scope of patent protection because every word added to the text of the claims can potentially introduce additional elements or characteristics that must be present to establish infringement. We create an indicator variable that equals one if and only if a patent application falls below the 5th percentile of the word count distribution for the first independent claim in the subsample of applications with the same filing year examined by the same art unit.[14] We call this variable "Words in 1*st* Claim."[15]

2.4 Results

This section presents evidence of patent examiner specialization, and then regression results linking specialization to examination outcomes.

[13]We test the robustness of these results using the INPADOC patent families. The results are similar to those for DOCDB patent families and are available upon request.

[14]Kuhn et al. (2016) note that this measure of scope is not suitable for the analysis of patent scope in biotechnology. So we exclude the technology center responsible for biotechnology from the analysis of this variable. We also check the robustness of the results based on the length of the first independent claim utilizing measures built upon the number of claims and independent claims. The results are similar and available upon request.

[15]We provide summary statistics for the variables used in the analysis in table A1.

2.4.1 Examiner Specialization

Figure 2·1 shows that patent examiners handle more applications from a given USPC subclass or assignee than we would expect under random allocation. Specifically, each panel shows a histogram of p-values from a sample of hypothesis tests. For the D-index (top row), we run a separate test for each art-unit-year by subclass or assignee cell containing more than 20 applications, and for the MTAD (bottom row) we run a separate test for each art-unit-year cell containing more than 50 applications.[16]

Under the null of random assignment, the p-values in Figure 2·1 should be uniformly distributed between zero and one. However, in each panel a large share of the test-statistics fall below the usual 1% statistical significance threshold, providing strong evidence of specialization. The two histograms in the left column indicate that about 35 percent of the D-index and MTAD tests for random USPC assignment have a p-value below 0.01. The two histograms in the right column show somewhat weaker evidence of specialization by assignee, with 10 to 20 percent of the p-values falling below the 1-percent threshold. The agglomeration by assignee becomes much weaker if tests are conducted within USPC subclasses (see below), suggesting that these findings are primarily a result of technological specialization of examiners and applicants. Overall, Figure 2·1 shows that the allocation of applications within art units is often far from random, and that SPEs take into account the technological classification when they assign applications to the examiner, as described in previous research (e.g. Lemley and Sampat (2012)).

Table 2.1 examines the degree of examiner specialization in different Technology Centers, and for an additional pair of application characteristics. Specifically, the table reports the share of D-index or MTAD tests that reject the null hypothesis of random allocation at a 5-percent significance level. Panel A shows that there is

[16] All of our results are robust to varying the within-cell sample size cutoffs, but going much below these thresholds leads to large numbers of uninformative tests.

evidence of examiner specialization in every technology center. However, the "Computer Architecture" and "Computer Networking" areas are less agglomerated than Biotechnology, Chemistry, Semiconductors and Mechanical Engineering. The results in Panel B are similar.

Although our data do not speak to the underlying causes of variation in examiner specialization across technology centers, there are at least two possible explanations for this pattern. First, examiners in the less agglomerated technology centers may be "generalists" who are capable of evaluating most applications within their art-unit. This would naturally lead SPEs to adopt a more random allocation process. Alternatively, patent examiners in the Computers and Communications technology centers might be just as specialized as their counterparts, but this is not apparent to us because the USPC classification system is less representative of actual differences in these fields than in the other technology centers.

The lower half of Table 2.1 examines agglomeration for a pair of dichotomous variables: "DOCDB Family Size" and "Words in 1st Claim". Both of these variables focus on extreme outcomes because we are interested in whether SPEs assign unusual applications to a specific set of examiners. The data suggest that, for the most part, they do not. There is some evidence that very large families are concentrated among a smaller set of examiners for Semiconductors and Mechanical Engineering. And there is some evidence that certain examiners specialize in broader patents (as measured by length of the first claim) in the Chemical and Materials Engineering technology center. But these effects are not large, and might easily be caused by the technological specialization observed in Panel A.

The results presented thus far beg the question of whether examiner specialization is purely technological. To explore that idea, we test for agglomeration *within* art-unit-year-USPC-subclasses to see whether conditioning on observed technological heterogeneity changes our results. There are two caveats to keep in mind. First, we cannot condition on *unobserved* technological heterogeneity. And second, many USPC subclasses receive only a few applications per year, so these tests exclude a large amount of data. However, if examiners seem to be randomly assigned within large sub-classes, we might be more comfortable that most of the specialization we observe within art-unit-years is based on technology rather than other patent characteristics.

Figure 2·2 examines agglomeration by assignee, within art-unit-years, both within and without conditioning on USPC subclass. Each panel presents a quantile-quantile plot that compares the distribution of the D-index (top row) or MTAD (bottom row) for the observed data to the distribution under simulated random assignment. The left column shows that the observed quantiles of the D-index are higher, and the observed quantiles of MTAD are lower, than the simulated quantiles under random allocation. In other words, there is strong evidence of specialization, as we saw above. However, the righthand column shows that the evidence for agglomeration is much weaker once we condition on the USPC subclass. Note how the sample size falls dramatically in the right column.[17] Although this may affect the precision of the tests, we think that the samples in our analysis are large enough to detect significant departures from random assignment.

Table 2.2 examines agglomeration within art-unit-year-USPC subclasses for individual Technology Centers. Panel A focuses on the allocation of assignees. The D-index tests reject the null of random assignment in a substantial share of tests for the Biotechnology, Chemistry and Miscellaneous technology centers. The results for MTAD are weaker. For the Computer Architecture and Computer Networking technology centers, there are too few subclasses of sufficient size to make any reliable inference, and the others seem to match examiners to applications randomly

[17]The D-index discards any assignee that does not submit more than 20 applications to a given art-unit in a particular year, which excludes the large majority of applicants. MTAD retains more data because it uses all applications filed in a year to an art unit.

conditional on technology. Although we do not have a large number of tests for the Biotechnology and Chemistry art units, the evidence of agglomeration by assignee within a art-unit-year-USPC subclass for those fields suggests to us that examiners are specializing based on unobserved differences in technology. In other words, conditioning on subclass may not ensure random assignment. Panels B and C find no evidence that SPEs in any technology center allocate "outlier applications" (in terms of family size or independent claim scope) to a specific set of examiners after conditioning on observed technological differences.

Overall, these results show that patent examiners specialize in particular technologies, even within relatively homogeneous art units. We find no evidence that certain examiners specialize in "outlier" patent applications. Moreover, much of the agglomeration by assignee disappears if we condition on USPC subclasses. However, we do find evidence of agglomeration by assignee, even within USPC subclasses, for the Biotechnology and Chemistry technology centers. This last result suggests that there may be examiner specialization based on unobserved technological differences in some art units even after conditioning on the observed USPC subclasses.[18]

2.4.2 Specialization and Examination Outcomes

As a final step in our empirical analysis, we explore the relationship between examiner specialization and patent prosecution outcomes. We focus on three outcomes: (i) whether an application is granted, (ii) the change in the number of words in the first independent claim between the published application and the granted patent, and (iii) the number of days required to process the application.[19] Our sample consists of all

[18]To complement the analysis describe in this section, we also run a set of Kolmogorov-Smirnov tests of the equality of distributions of the p-values for the tests on the real allocations and the simulations of random assignment. The results are consistent with those reported in the paper and are available upon request.

[19]It is important to note that an application is never ultimately rejected by the USPTO. If an applicant is not granted a patent, she can file a Request for Continued Examination (RCE), a continuation application or a continuation-in-part. We do not study the implications for RCE or

applications belonging to an art-unit by examiner by filing year cell containing more than 10 applications. To account for truncation, we exclude pending applications and those filed after year 2009.

Our unit of analysis is the application, and we adopt a measure of specialization that varies across both examiners and applications. Specifically, our main explanatory variable is the share of an examiner's applications (within an art unit-filing year cell) having the same USPC subclass as a focal application. To be more precise, define the set $k_{it}(j)$ of all patents (except for patent i) assigned to examiner j in year t. Let n_{jt} represent the total number of patents reviewed by examiner j in year t, and define an indicator 1_{mn} that equals one if and only if two patents (m and n) have the same USPC subclass. Our main explanatory variable can be written as:

$$Share_{ijt} = \frac{\sum_{m \in k_{it}(j)} 1_{mi}}{n_{jt} - 1}.$$

Intuitively, $Share_{ijt}$ equals the probability that a random draw from the pool of applications assigned to examiner j in year t has the same USPC subclass as the focal application.

Table 2.3 presents estimates from a series of OLS panel-data regressions that examine the correlation between $Share_{ijt}$ and prosecution outcomes. To ease interpretation, we standardize $Share_{ijt}$ and the outcome variables except the dummy for granted patents.[20] All models include art unit-by-filing year effects, and standard errors are clustered at art unit-filing year level.

Columns (1) through (3) report coefficient estimates from a within-examiner regression that has both art-unit-year and examiner fixed effects. In column (1), the coefficient on $Share_{ijt}$ is -0.03, indicating that examiners are less likely to grant patents in subclasses where they are assigned more applications. Column (2) finds no

continuation filings in this paper.

[20]Table A2 displays summary statistics for all variables used in this part of the analysis.

relationship between within-examiner specialization and the number of words added to the first claim. In column (3) we find a statistically significant but very small negative relationship between specialization and processing time: examiners are slightly faster in the subclasses they see most often.

Columns (4) through (6) report the results from a between-examiner analysis, where we regress the mean outcome for each examiner on the mean of $Share_{ijt}$ (i.e. the probability that two random draws from the pool of patents assigned to that examiner will belong to the same USPC subclass). As before, we standardize all variables except for the grant rate.

The coefficient in Column (4) indicates that a one standard deviation increase in $Share_{ijt}$ leads to a 6 percentage point drop in the grant rate. This reinforces the result in column (1) that specialized examiners are more stringent. The coefficient in column (5) also suggests that specialization leads to more stringent examination. However, the economic magnitude of this result is rather small: a one standard deviation change in $Share_{ijt}$ produces a 0.10 standard deviation change in the number of words added to the first claim. Finally, in column (6) we find a small but statistically significant positive association between specialization and the time required to process a patent examination.[21]

The overall message of Table 2.3 is that examiner specialization leads to more stringent examination, although the economic magnitudes are not dramatic. One plausible explanation for the finding is that it is easier for an examiner to find relevant prior art if she is more familiar with a given field of technology, leading to narrower claims and an increased probability that the application is abandoned. Under random assignment, these estimates are causal. We prefer a descriptive interpretation.

[21]We examined whether the results presented in Table 2.3 varied substantially across technology centers and found that they do not. Those results are available on request.

2.5 Conclusions

We study a key stage of patent prosecution: the assignment of applications to examiners. The first half of our empirical analysis focuses on characterizing the degree of examiner specialization. Using two statistical tests designed to study industry agglomeration, we find strong evidence that examiners specialize in particular technologies, even within relatively homogenous art units. The degree of specialization varies across fields. The USPTO technology centers associated with Computers and Communications exhibit relatively little specialization, while examiners in the "Biotechnology and Organic Chemistry" and "Chemical and Materials Engineering" technology centers appear highly specialized. In the latter technology centers, we find assignee agglomeration even after conditioning on USPC subclasses.

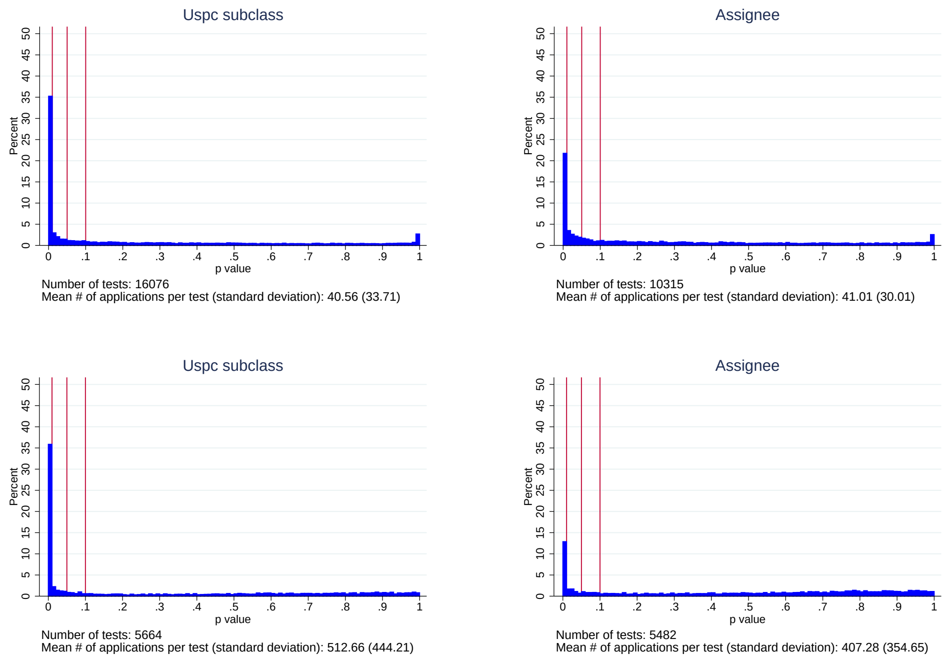
The second part of our analysis shows that more specialized examiners are more stringent on average. They have a lower grant rate, and produce a larger reduction in the scope of the first independent claim for granted patents.

It may not seem surprising that we can reject the hypothesis of random matching between applications and examiners. After all, one reason for having a patent classification system is to help route applications to appropriate examiners. However, several studies have argued that more-or-less random matching (within art-units) provides a justification for using examiner characteristics as an instrument for examination outcomes. While our findings do not invalidate this identification strategy – patent examiner characteristics might still satisfy the relevant exclusion restrictions – they do imply that we cannot rely on random assignment to justify the approach. Our findings suggest that random assignment is more plausible within USPC-subclasses, though not in every technology center.

On a more positive note, our results suggest that the USPTO's patent examination process strikes a reasonable balance between efficiency and fairness. Technological specialization is efficient. Fairness can be achieved by enforcing uniform examination standards, which is difficult, or through random assignment, which guarantees all applicants an equal shot at the more friendly examiners. Conditional on technology, examiner assignment appears relatively random in the Computing and Communications areas. And even without controlling for technology, there is no evidence that certain examiners within a given art unit handle more patents with large families or broad claims. We leave to future researchers the question of whether procedural fairness to applicants is also the best policy in terms of social welfare.

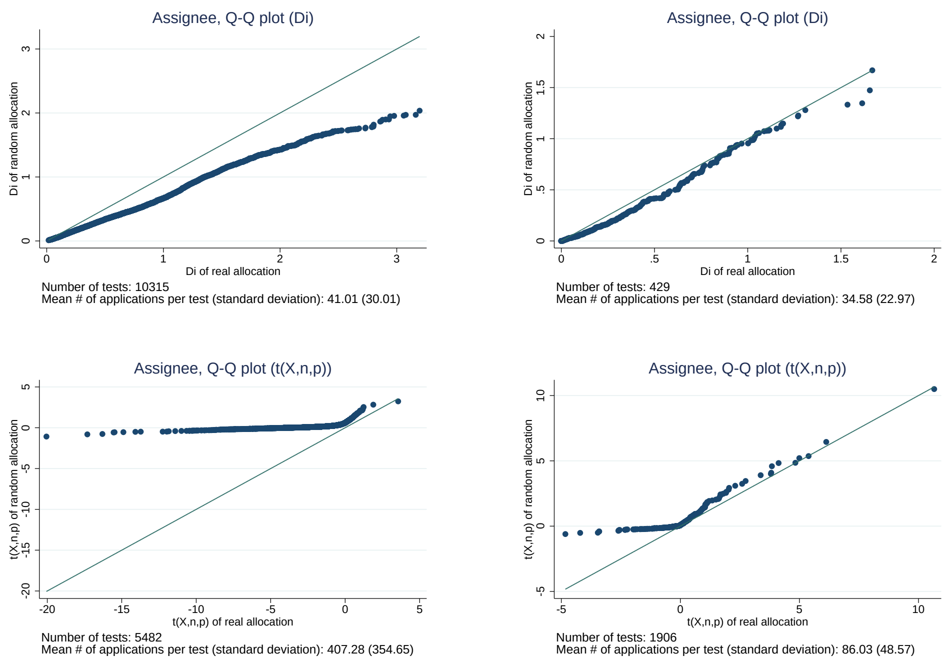
Tables and Figures

Figure 2·1: Distribution of P-values from D-index (top) and MTAD (bottom) for USPC subclass and Assignee.



Distribution of p-values of D-index and MTAD analysis for USPC subclass and Assignee codes. Tests on subsamples with more than 20 applications for D-index and 50 applications for MTAD. Vertical red lines are standard thresholds for statistical significance (0.01, 0.05 and 0.10).

Figure 2·2: Quantile-Quantile Plots of D-index (top) and MTAD (bottom) by Art-Unit-Year (left) and Art-Unit-Year-USPC-Subclass (right) for Assignee.



Each panel plots the quantiles of the D-index (top row) or MTAD statistic (bottom row) for the observed distribution (X-axis) against a simulated distribution under random assignment (Y-axis). Tests on subsamples with more than 20 applications for D-index and 50 applications for MTAD. If the observed distribution is random, the quantiles should be the same and the scatter points will fall along the 45-degree line. We observe large deviations from random assignment at the art-unit-year level, but much less evidence within art-unit-year-USPC-subclasses.

Table 2.1: D-index and MTAD Tests within Art-Unit-Years for Random Allocation by Technology Center.

Panel A: USPC subclass					
	D-index		MTAD		
Technology Center	Agg.	N	Agg.	Disp.	N
Biotechnology and Organic Chemistry	58.5	1,225	44.9	0.0	563
Chemical and Materials Engineering	82.8	1,393	80.7	0.0	726
Computer Architecture, Software, and Security	11.5	1,766	8.3	0.0	733
Computer Networking and Video Distribution	15.9	968	4.9	0.0	639
Communications	30.8	2,777	56.6	0.0	711
Semiconductors, Electrical and Optical Systems	52.5	4,335	41.8	0.8	848
Miscellaneous†	31.2	1,530	41.8	0.8	756
Mechanical Engineering, Manufacturing, Products	53.9	2,082	61.5	0.0	688
All tests	43.3	16,076	41.8	0.1	5,664

Panel B: Assignee					
	D-index		MTAD		
Technology Center	Agg.	N	Agg.	Disp.	N
Biotechnology and Organic Chemistry	66.8	253	13.5	0.0	533
Chemical and Materials Engineering	57.7	901	41.2	0.0	707
Computer Architecture, Software, and Security	12.2	1,017	0.6	0.0	720
Computer Networking and Video Distribution	15.7	516	0.6	0.0	623
Communications	17.3	1,972	11.7	0.0	686
Semiconductors, Electrical and Optical Systems	31.3	3,640	23.0	0.1	838
Miscellaneous†	40.0	875	21.7	0.0	711
Mechanical Engineering, Manufacturing, Products	51.4	1,141	29.2	0.0	664
All tests	32.1	10,315	18.1	0.0	5,482

Panel C: DOCDB Family Size					
	D-index		MTAD		
Technology Center	Agg.	N	Agg.	Disp.	N
Biotechnology and Organic Chemistry	3.4	795	6.6	0.0	560
Chemical and Materials Engineering	7.7	1,030	9.7	0.0	722
Computer Architecture, Software, and Security	1.8	902	5.3	0.0	730
Computer Networking and Video Distribution	0.8	754	3.1	0.0	637
Communications	4.4	1,034	7.2	0.0	706
Semiconductors, Electrical and Optical Systems	8.6	1,472	12.4	0.2	847
Miscellaneous†	5.5	1,172	8.9	0.3	754
Mechanical Engineering, Manufacturing, Products	8.4	1,113	11.4	0.0	686
All tests	5.5	8,272	8.3	0.1	5,642

Panel D: Words in 1st Claim					
	D-index		MTAD		
Technology Center	Agg.	N	Agg.	Disp.	N
Chemical and Materials Engineering	7.2	1,129	10.8	0.0	721
Computer Architecture, Software, and Security	0.1	895	1.0	0.0	723
Computer Networking and Video Distribution	0.0	755	0.0	0.0	627
Communications	0.5	1,052	0.9	0.0	693
Semiconductors, Electrical and Optical Systems	5.0	1,524	8.8	0.2	843
Miscellaneous†	3.4	1,194	4.9	0.0	741
Mechanical Engineering, Manufacturing, Products	3.2	1,160	4.4	0.0	679
All tests	3.1	7,709	4.6	0.0	5,027

Columns labeled "Agg." ("Disp.") report the share of tests that reject the null hypothesis of random allocation at 5% level in favor of agglomeration (dispersion). All tests are conducted within art-unit-year cells with more than 20 applications for the D-index and more than 50 applications for MTAD. † Miscellaneous = "Transportation, Construction, Electronic Commerce, Agriculture, National Security and License & Review."

Table 2.2: D-index and MTAD Tests within Art-Unit-Year-USPC-Subclasses for Random Allocation by Technology Center.

Panel A: Assignee

	D-index		MTAD		
Technology Center	Agg.	N	Agg.	Disp.	N
Biotechnology and Organic Chemistry	41.4	58	10.0	1.5	200
Chemical and Materials Engineering	17.8	73	5.4	4.7	148
Computer Architecture, Software, and Security	0.0	9	0.0	0.6	162
Computer Networking and Video Distribution	0.0	1	0.0	0.0	63
Communications	0.8	123	0.9	0.0	539
Semiconductors, Electrical and Optical Systems	0.0	92	0.0	0.2	487
Miscellaneous†	11.1	18	3.3	1.7	181
Mechanical Engineering, Manufacturing, Products	3.6	55	1.6	0.0	126
All tests	9.8	429	2.2	0.8	1,906

Panel B: DOCDB Family Size

	D-index		MTAD		
Technology Center	Agg.	N	Agg.	Disp.	N
Biotechnology and Organic Chemistry	0.0	1,107	2.6	0.0	273
Chemical and Materials Engineering	0.0	1,270	1.1	0.0	185
Computer Architecture, Software, and Security	0.0	1,608	0.0	0.0	228
Computer Networking and Video Distribution	0.0	879	0.0	0.0	78
Communications	0.0	2,609	0.8	0.0	621
Semiconductors, Electrical and Optical Systems	0.0	4,034	0.5	0.0	599
Miscellaneous†	0.0	1,427	0.7	0.0	285
Mechanical Engineering, Manufacturing, Products	0.0	1,922	0.5	0.0	208
All tests	0.0	14,856	0.8	0.0	2,477

Panel C: Words in 1st Claim

	D-index		MTAD		
Technology Center	Agg.	N	Agg.	Disp.	N
Chemical and Materials Engineering	0.0	1,270	2.2	0.0	184
Computer Architecture, Software, and Security	0.0	1,608	0.0	0.0	197
Computer Networking and Video Distribution	0.0	879	0.0	0.0	77
Communications	0.0	2,609	0.2	0.0	609
Semiconductors, Electrical and Optical Systems	0.0	4,034	1.0	0.0	573
Miscellaneous†	0.0	1,427	0.8	0.0	260
Mechanical Engineering, Manufacturing, Products	0.0	1,922	0.0	0.0	220
All tests	0.0	14,856	0.6	0.0	2,120

Columns labeled "Agg." ("Disp.") report the share of tests that reject the null hypothesis of random allocation at 5% level in favor of agglomeration (dispersion). All tests are conducted within art-unit-year cells with more than 20 applications for the D-index and more than 50 applications for MTAD. † Miscellaneous = "Transportation, Construction, Electronic Commerce, Agriculture, National Security and License & Review."

Table 2.3: Examiner Specialization and Examination Outcomes.

Model	Within Examiner			Between Examiner		
Outcome	Granted	Words	Days	Granted	Words	Days
	(1)	(2)	(3)	(4)	(5)	(6)
$Share_{ijt}$	-0.03***	0.00	-0.03***	-0.06***	0.10***	0.08***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Art-Unit-Year Effects	Y	Y	Y	Y	Y	Y
Examiner Effects	Y	Y	Y			
Observations	1,936,297	1,935,940	1,077,041	50,579	50,579	45,486
Examiners	50,579	50,579	45,486			

All models estimated with OLS. Unit of observation is a patent application. Variables $Share_{iijt}$, Words and Days are standardized. Standard errors clustered by art unit-filing year in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Chapter 3

Strategic Delays in Prosecution of Standard Essential Patents at the USPTO

3.1 Introduction

Patents that would be infringed by any implementation of a standard, or Standard Essential Patents (SEPs), are increasingly important in complex industries. Companies in the Information and Communication Technology (ICT) sector are investing significant resources to obtain them (see for example Google's acquisition of Motorola for its big patent portfolio or the auction for Nortel's patents) because they provide great benefits in terms of licensing revenues and advantages in cross-licensing negotiations. Many SEPs have been involved in important lawsuits (e.g. those between Apple and Samsung) and there are concerns about the possibility of abuse of the market power created by SEPs.

In this paper I study how company obtain SEPs. I focus on U.S. SEPs because these are the largest subset of SEPs from a single country and, more importantly, because the U.S. patent system provides unique mechanisms that applicants can exploit to increase the value of their patents. In particular, strategies that give the opportunity to the applicants to delay the issuance of their patents and to modify their claims enable companies to cover recent technical developments.

This paper addresses the question of whether applicants for SEPs pursue these strategies more intensively than other applicants, possibly to cover developments of the standardization process. Moreover, I identify the SEP applicants that rely more on patents in their business models and test whether they use those strategies more frequently. I also estimate the correlation between standardization timing and, respectively, issuance of pending applications and filings of continuation applications to see how the resolution of uncertainty on the content of a standard is associated with behavior in prosecution. Finally, I analyze the relationship between length of prosecution and litigation of SEPs.

Standardization often requires cooperation of different parties to develop a common technological platform and trigger networks effects (Farrell and Simcoe, 2012b; Katz and Shapiro, 1985; Simcoe, 2012). Participants in standardization have incentives to include their own patented technologies into a standard because technologies that could be substitutes for the focal patented technology before standardization become less attractive for standard implementers after the adoption of the standard (Lerner and Tirole, 2015). While there is evidence of significant selection on economic and technical importance, Rysman and Simcoe (2008) and Bekkers et al. (2016) show that forward patent citations significantly increase after a patented technology is included into a standard and interpret this finding as positive influence of Standard Setting Organizations (SSOs) endorsement on the adoption of the technology protected by a SEP. SEPs are also litigated more often than similar patents, suggesting that they are more valuable. Ownership of SEPs may also improve the position in a network of alliances within an industry and confer market power (Bekkers et al., 2002).

On the other hand, SEPs can be problematic. Many SEPs on a standard may increase the implementation costs because of royalty stacking. Moreover, the increase in the bargaining power of SEP owners after standardization may create serious hold-up problems. After the implementers have incurred technology-specific investments in a standard and this has become widely adopted, SEP owners may demand higher royalties for licensing their patents (Farrell et al., 2007; Lemley and Shapiro, 2007; Lerner and Tirole, 2015; Shapiro, 2001).

SEP disclosures and the cumulative number of SEPs that cover active standards are increasing over time. While the share of standards covered by declared SEPs is still relatively small, the number and the share of new standards subject to SEPs have also increased over time, especially in ICT. Furthermore, standards subject to SEPs seem to be more important, complex and innovative (Baron and Pohlmann, 2016; Baron et al., 2016; Bekkers et al., 2016).

It is therefore important to understand how companies obtain SEPs. In this study I focus on the prosecution of SEPs at the USPTO, with a focus on applicants' strategic behavior. Other studies analyze the behavior of SEP applicants and how companies may act strategically to obtain SEPs. Berger et al. (2012) analyze a sample of European SEPs related to standards developed by the European Telecommunications Standards Institute (ETSI) and find that SEPs are amended more often than matched control patents, and SEPs related to UMTS filed after the standard is set have shorter pendency. Kang and Bekkers (2015) show that owners of SEPs related to W-CDMA and LTE often file patent applications with low technical merit just before standardization meetings and negotiate their inclusion in a standard. Kang and Motohashi (2015) study the relationship between meeting attendance of patent inventors and their likelihood of obtaining SEPs. Nagaoka et al. (2009) show that a significant share of SEPs related to MPEG2, DVD and W-CDMA are filed after the standards are set and exploit continuation applications to benefit from earlier priorities. Kuhn et al. (2016) find that an outcome of prosecution, the narrowing of claims, is associated with the probability a patent is declared as essential.

This paper contributes to the literature on the relationship between patents and standards analyzing differences in the strategies of applicants for SEPs and those of similar applications in their prosecution at the USPTO. I use a large sample of SEPs from multiple important SSOs, matching them with control applications on several characteristics that are related to both the inclusion into a standard and differences in prosecution. To my knowledge, this paper is the most comprehensive study of SEP prosecution to date. Detailed data on the patenting process allow me to measure various determinants of prosecution time more directly than in previous research. I find that owners of SEPs are slower to reply to office actions and exploit some peculiar mechanisms of the U.S. patent system such as chains of continuations and provisional applications more intensively, possibly to make their patents more relevant for standards under development and thus increase their value. These differences persist when I control for application characteristics correlated with the quality of the technology and the scope of protection sought by the applicant.

Another contribution of the paper is to show with a longitudinal analysis on a relatively large sample that application disclosure, which I use as a proxy for standardization, is positively related to both issuance of U.S. SEPs and filing of continuation applications. Previous research relies on cross-sectional evidence from smaller samples of patents for a limited number of standards.

While in previous works the attention has been often on the achievement of broader patent scope, I find that SEP applicants behave in a more nuanced way. Longer prosecution may give more time to SEP owners to achieve the right balance between broad scope and patent strength. On the one hand, while SEPs and control applications have similar scope at the beginning of examination, I find that SEPs are broader when they issue. On the other hand, longer prosecution of SEPs is correlated with claim narrowing. One possible interpretation is that applicants for SEPs may want to obtain patents that are broad enough to cover multiple implementations of the standards, but not so broad that they would be invalidated in a court.

This paper also contributes to the literatures on how patents are produced and strategic patenting (Carley et al., 2015; Cohen et al., 2000; Hall and Ziedonis, 2001; Harhoff and Wagner, 2009; Lemley and Sampat, 2012; Ziedonis, 2004). I find not only differences between applicants for SEPs and those of other applications, but also that SEPs held by “upstream” organizations such as knowledge developers and producers of components have longer lags between priority and issuance than those owned by “downstream” organizations like producers of end products. Upstream SEP holders rely more intensively on the main prosecution strategies of SEP applicants, i.e. continuation applications and provisional applications, and are also slower at replying to office actions. A simple explanation for this result is that upstream organizations rely more on patents to generate revenues, and so have higher incentives to obtain more valuable patents.

Finally, this paper contributes to the literature on patent litigation and in particular on the litigation of SEPs (Bekkers et al., 2016; Cockburn et al., 2002; Galasso and Schankerman, 2010; Galasso et al., 2013; Lanjouw and Schankerman, 2001; Simcoe et al., 2009). My analysis confirms the result of Allison et al. (2004) that longer prosecution of a patent is positively correlated with litigation even in the context of SEPs, which are a subset of the population of patents with above-average value.

I provide a short description of the patenting process at the USPTO in section 3.2 and explain how applicants for SEPs can act strategically in their patent prosecution at the USPTO in section 3.3. Section 3.4 presents the methods used to study the prosecution strategies for SEPs, the relationship between standardization, patent issuance and continuation filings, and finally how I estimate the association between issuance lags from priority and litigation. Section 3.5 describes the data for the empirical analysis. I discuss the results in section 3.6. Section 3.7 concludes.

3.2 Patenting Process at the USPTO

This section describes the patenting process at the USPTO.[1] After an applicant files a patent application, a central office of the USPTO reviews its formality requirements and assigns it a serial number. A contractor assigns at least one mandatory technological classification to the application. This classification is utilized by the USPTO to assign the application to a group of examiners specialized in relatively similar technologies called art unit. The supervisory patent examiner of the art unit assigns the application to an examiner, who is responsible for determining the patentability of the invention.

The patenting process involves a substantial interaction between the examiner and the applicant. The examiner inspects the description of the invention and the claims of the application and searches for relevant prior art that may make an application unpatentable. To receive patent protection, inventions in patentable subject matter must be novel (the invention must be different from the prior art), non-obvious (the invention must be non-obvious for a person with ordinary skills in the invention's field) and useful (the invention must have a useful purpose). Also, the application must disclose the invention in sufficient detail for any person skilled in the art to make and use the claimed invention. The claims must be clear and use a definite language.

Once the examiner has determined if the application meets the patentability requirements, he sends the First Office Action on the Merits (FOAM) to the applicant. This action notifies the examiner's decision to allow or reject the claims in the application, supporting the decision with a search report of the relevant prior art. Though sometimes the examiners allow a patent application at this stage of the process, usually the FOAM is a non-final rejection.[2]

---

[1] This section draws significantly on chapter 2 and references therein, Carley et al. (2015) Cockburn et al. (2002) and Lemley and Sampat (2012).

[2] Carley et al. (2015) show that about 86% of the applications not originating from continua-

After a non-final rejection, the applicant can abandon the application, amend the claims or dispute the rejection. If the applicant submits a revised version of the application, the examiner can decide to allow the application or deliver a final rejection. However, this "final" rejection is by no means really final, because the applicant can again decide to amend the claims, request an interview with the examiner, appeal the decision of the examiner, or exploit one of the various continuation options. There are two types of continuations available to the applicants: serialized and non-serialized continuations. The first type includes continuation applications, continuations in part and divisional applications. They are treated as new applications. The second type includes Requests for Continued Examination (RCEs) and other rare types of continuations (Cotropia et al., 2013). This second type of continuations reopens the prosecution of the original application.

It is important to note that applicants can file continuations at any time while the "parent" application is pending. Divisional applications are usually filed when the examiner finds that the original application contains more than one invention, violating the principle of "unity of invention". RCEs are normally utilized after a rejection or after the notice of allowance if the applicant is not satisfied with the allowed claims. Continuation applications and continuations-in-part are utilized to obtain different claims from those in the parent application or to have another chance to obtain those that were rejected. Continuation applications cannot add additional disclosure of the invention and have the benefit of the priority of the original application. Continuations-in-part can add additional disclosure, but only the claims supported by the disclosure in the parent application have the benefits of the priority of the parent.

The use of continuations, and in particular RCEs and continuation applications,

tions receive a non-final rejection as FOAM. They also provide statistics for allowances, rejections, abandonments and appeals for the following stages of the examination process.

is particularly controversial. Some scholars argue that continuations increase the workload of the USPTO and slow down the patenting process. Continuations also provide the opportunity to “wear down” the examiners and obtain claims that were previously rejected or broad claims that the examiners would not otherwise grant, leading the USPTO to grant “low quality” patents. Moreover, they provide the opportunity to obtain claims that cover products and technologies developed before the claims are actually drafted but after the priority date, because the latter defines the relevant prior art. Companies can therefore exploit continuations of pending applications to wait for the development of competitors’ new technologies, draft new claims, and then sue infringers with patents designed to cover those technologies (Cotropia et al., 2013; Lemley and Moore, 2004).

The examination process ends with either an abandonment of the application or its issuance. Applications are abandoned for failure to reply to an office action within a specified period of time or with an express abandonment by the applicant at any point during prosecution. If the examiner concludes that the application is entitled to issue as a patent, the applicant receives a notice of allowance and the application is published as an issued patent, provided the applicant pays the fee on time and does not reopen prosecution.

There are two ways to modify a patent even after its issuance: reissue patents and reexaminations. A patent owner that finds an error in an issued patent can file an application for a reissue.[3] On the one hand, a patent owner can narrow the claims of a patent if she realizes that they are too broad and there is the risk that this would invalidate the patent. On the other hand, reissues can also be utilized to broaden the scope of the claims. If a patent owner realizes that the claims do not cover the

[3] Examples of errors that may be corrected with a reissue are typographical mistakes in the specification or the drawings, or the exclusion of an inventor from the application. On the contrary, a patent owner cannot file a reissue application for errors like intentionally failing to disclose relevant prior art.

entire scope of the disclosed invention, she can file a reissue application to broaden the scope of protection within two years from the issue date.

Requests for reexaminations can be filed either by the patent owner or by a third party at any time during the life of the patent. Third parties can file a request for reexamination to challenge the validity of a patent. Patent owners often file requests for reexamination to test the validity of their patents against newly discovered prior art to make sure the patent is "strong" before enforcing it. If necessary, the patent owner can narrow the existing claims of the patent to strengthen the patent. After reexamination, a certificate of correction containing the required changes to the claims is issued.[4]

This process offers several opportunities to the applicants to amend the claims of their patents and manipulate strategically the timing of patent issuance. These opportunities may be particularly appealing to owners of SEPs. In the next section I describe how owners of SEPs can act strategically during the prosecution of their patents at the USPTO.

3.3 Prosecution Strategies for SEPs

While delay strategies and opportunities to modify claims are available to all applicants, they may be particularly appealing for SEP applicants because of the high economic value of SEPs. Moreover, these strategies may be related to the timing of standardization. Often the drafting of standards and related patents occurs simultaneously. Applicants for SEPs can exploit various strategies enabled by the U.S. patent system to delay patent issuance and modify claims. A longer prosecution pro-

[4]The Central Reexamination Unit of the USPTO is responsible for reexaminations. Before year 2012, a patent owner could appeal an unfavorable decision of reexamination to the Board of Patent Appeals and Interferences. This was replaced with the Patent Trial and Appeal Board by the America Invents Act of 2012. This law also substituted *inter partes* reexaminations with post grant proceedings. *Ex partes* reexaminations is still available after 2012.

cess gives more opportunities to SEP holders to modify the scope of their claims to make the SEPs more relevant for the related standards and therefore more valuable. Frequently companies file patent applications that may be related to a standard just before a standardization meeting. Attending meetings inventors may acquire knowledge relevant to the standardization process and then modify the claims or file patent applications that cover possible developments of a standard. As long as the application is still pending, applicants can amend the claims, possibly to cover the scope of a standard under development (Bekkers et al., 2016; Berger et al., 2012; Kang and Bekkers, 2015; Kang and Motohashi, 2015; Nagaoka et al., 2009).

A first strategy that applicants for SEPs can exploit is to delay the start of examination without losing priority. They can do this using provisional applications and PCT applications. Provisional applications are not examined and cannot issue as patents, but the applicant for a provisional applications has 12 months to file a regular application and claim the benefits of the filing date of the provisional application. A provisional application provides a description of the invention but it does not contain claims, which are included in the regular application filed later. So, provisional applications give the opportunity to applicants for SEPs to delay the drafting of claims and observe the developments of standardization.[5]

A PCT application provides a similar opportunity. An applicant exploiting the PCT route can file a national stage entry application within 30 months from the priority date (Berger et al., 2012).[6]

A second strategy is to proceed slowly during examination with the goal of delaying the issuance of a patent. An applicant can delay the issuance of a patent at the USPTO by simply taking more time to respond to office actions. The deadline to respond to office actions is typically within 3 months from the mailing date of the office

[5] Provisional applications are available to applicants since 1995.

[6] The USA joined the Patent Cooperation Treaty in 1978.

action, but applicants can request an extension of time and postpone the deadline 3 months. Applicants for SEPs may be more likely to take more time to reply to office actions and make more requests for extension of time.

Applicants can also exploit the continuation options to delay prosecution and modify the claims. Applicants for SEPs may have higher incentives than the average applicant to reopen prosecution with an RCE after a notice of allowance to gain time and have a chance to obtain a patent with higher value than the one already allowed.

After applicants observe the final content of a standard, they can make the last amendments and then close the prosecution of their patents, eventually profiting from higher bargaining power with potential implementers. Therefore, it is likely to observe a positive association between the approval of a standard and issuance of the related patents.

Companies can also exploit serialized continuations, and especially continuation applications, to obtain SEPs or make them more valuable. Given the high economic interest at stake, applicants for SEPs have higher incentives to negotiate more aggressively with the USPTO to obtain the claims they want. This may lead to filing more continuations to have a second chance with the claims of original applications that were rejected or filing new claims that cover recent developments in standardization. Companies may also exploit continuation applications to cover a standard after major decisions regarding the design of a standard have been made. As long as an application is pending, they can file children of the application and possibly cover the developments of the standardization process with them.

Since the relevant prior art for a continuation application is determined by the priority date of the parent application, companies have incentives to keep a parent application pending and delay continuation filings until they observe the final content of a standard. This gives them the opportunity to cover the content of a standard with new claims drafted after a standard has been set, provided that these are supported by the disclosure in the parent application. Therefore, we should observe an increase in continuation application filings when a standard is approved (Lemley and Moore, 2004; Nagaoka et al., 2009).

Applicants for SEPs can exploit these strategies to obtain the right balance between claim breadth and strength. After they observe the content of the standard related to their SEPs, they can draft a set of claims to obtain a patent that is broad enough to cover multiple implementations of the standard, but not so broad that it would be easily invalidated in a court. Broad claims are generally considered more valuable because they are more difficult to invent around (Kuhn, 2016; Kuhn et al., 2016). However, overly broad claims may be easier to invalidate. They may be too broad with respect to the invention disclosed in the patent, their language may be too vague, or it may be easier to find prior art to invalidate them. Knowing the content of a related standard, companies can tailor their patents to the main implementations of the standard and obtain a patent soon after a standard is set, then exploit continuation applications to obtain additional claims to increase the value of their SEP portfolio.

SEP holders may also have higher incentives to apply for reissues and SEPs may be more likely to be involved in reexaminations than other patents. First, SEPs are more likely to be involved in lawsuits. SEP owners may want to exploit reissues and reexaminations to improve their bargaining power with potential infringers and have higher chances of winning a validity challenge. Second, SEP owners may exploit reissues to modify the claims of their patents because of important events in standardization that happen after patent issuance.[7]

Even among SEP owners, organizations with different business models may apply

[7]SEPs may also have higher reexamination rates because third parties may have higher interests to question the validity of SEPs with reexaminations.

different prosecution strategies. In particular, organizations that are more "upstream" in the value chain may have business models that rely more on the production and sale or licensing of intellectual property, as opposed to "downstream" companies for which patents are mainly necessary to obtain freedom to operate and as bargaining chips in cross-licensing negotiations (Hall and Ziedonis, 2001). Upstream organizations may have higher incentives to have stronger patents or patents that are more important for the implementation of a standard. They may apply the prosecution strategies typical of the SEPs more aggressively.

Finally, companies often try to profit from the assertion of their SEPs once their value is increased by standardization (Bekkers et al., 2016; Simcoe et al., 2009). A longer prosecution may increase the value of SEPs making them "more essential" and may be exploited to draft a stronger patent that would be less likely invalidated in a court. It can also be exploited to cover implementations of a standard and then assert the patent against those implementers unwilling to pay licensing fees to the SEP owner. Therefore, a longer prosecution is likely to be positively correlated with litigation.

3.4 Methods

In this section I describe the methods utilized to analyze the prosecution strategies of applicants for SEPs and the relationship between length of prosecution and litigation.

3.4.1 Prosecution Strategies

Suppose we have a sample of patent applications. The ideal setting to study the effect of potential standard essentiality on prosecution strategies would be an experiment in which half of these applications become at risk of becoming essential for a standard, while the others do not and are in the control group. In other words, in this experiment the "treatment" would be the random assignment of standard proposals that may cover the technology protected by the application. The comparison in mean outcomes between potentially essential applications and controls would provide an estimate of the average treatment effect of potential standard essentiality on prosecution strategies.

This experiment is not available, but we can compare the prosecution strategies of SEPs and other patent applications with observational data. However, the potential inclusion of a patented technology into a standard is not random. To compare SEP applications to similar applications, I identify "control" applications that are similar to the SEP applications with respect to a set of observable characteristics that are related to the inclusion of an application into a standard and are also related to differences in prosecution.

There is substantial technological specialization at the USPTO (see chapter 2). Differences across art units and even across examiners working in the same art unit are related to differences in prosecution (Farre-Mensa et al., 2015; Frakes and Wasserman, 2014; Kuhn et al., 2016; Lemley and Sampat, 2012; Sampat and Williams, 2015). Since SEPs are mostly concentrated in ICT-related fields (Baron and Pohlmann, 2016), it is important to take these differences into account. Moreover, small and large organizations may have different propensities to participate in standardization and to include patented technologies into a standard, and previous research shows a relationship between firm size, propensity to patent and differences in prosecution (Carley et al., 2015; Griliches, 1990; Scherer, 1983). Also, there are significant differences in prosecution between foreign and domestic applicants (Carley et al., 2015), and U.S. SEPs are less likely than the average application to have a foreign priority. Finally, different cohorts of applications may have a different probability of being included into standards and to be prosecuted differently because of changes in SSOs IP policies, in patent laws, in examination at the USPTO and in prosecution strategies of the applicants over time (e.g. Baron and Spulber (2016) and Sukhatme and Cramer (2014)).

To take into account these differences, I match each SEP application with a randomly selected regular utility patent application filed in the same year, classified in the same USPC class, assigned to the same art unit and the same examiner, with the same small entity status of the applicant (i.e. small entity or not) of the SEP application.[8] I also match on two indicator variables equal to one if the filing dates of the applications are respectively on or after June, 8, 1995 (enactment of the TRIPS) and on or after November, 29, 2000 (enactment of the American Inventor's Protection Act - AIPA) to take into account two major law changes that may have affected the prosecution strategies of applicants, and an indicator variable equal to one if the application claims the benefit of a foreign patent application to identify foreign applicants.

If one is willing to assume that the matched control applications are a valid counterfactual for the matched SEPs, the comparison of the mean outcomes between matched SEPs and matched controls would provide an estimate of the effect of potential standard essentiality on prosecution strategies. However, previous research provides evidence of substantial selection into standards based on quality of technologies and patents (Rysman and Simcoe, 2008; Bekkers et al., 2016). Valuable applications are likely prosecuted differently, so these simple comparison may have an endogeneity problem due to unobserved heterogeneity. For example, more valuable applications usually have longer prosecution (Allison et al., 2004). So the exclusion of application value from an empirical model would lead to an upward bias in the estimated correlation between SEP status and length of prosecution.

It is hard to find good measures of patent application value. To partially reduce

[8]So-called small entities receive discounts on various fees of the patenting process. The USPTO considers as small entities universities, nonprofit organizations, individual inventors and "small business concerns". The latter are businesses with no more than 500 employees or affiliates.

the endogeneity concerns related to its omission, I estimate regression models on the matched sample controlling for some observables that may be related to technical or economic value. These characteristics include the number of claims, the number of words in the first independent claim and the number of inventors. The claims of a patent define the scope of patent protection and the number of claims in a patent is positively correlated with its private value (Gambardella et al., 2008). Kuhn et al. (2016) and Kuhn (2016) advocate for the use of the length of the first independent claim as a better measure of patent scope, and show that this is negatively correlated with patent value as measured by patent sale and payment of patent maintenance fees. The number of inventors may also be correlated with the value of an invention, as patents produced by teams of inventors receive more forward patent citations, which are often utilized to measure the value of a patent (Wuchty et al., 2007).

Therefore, to estimate the correlation between SEP status and prosecution strategies, I estimate the following regression:

$$Y_i = SEP_i \alpha + X_i \beta + \varepsilon_i \tag{3.1}$$

where Y_i is the outcome of the regression (for example, the length of prosecution) for application i, SEP_i is an indicator variable equal to one for SEP applications and zero for the matched controls, X_i is a vector of control variables that includes a small entity status indicator, a foreign priority indicator, a set of filing year effects, a set of art unit effects and the natural logarithms of the number of claims, the number of words in the first independent claim of the application and the number of inventors. The coefficient of interest is α, which estimates the correlation between SEP status and prosecution strategies and outcomes. ε_i represents the error term. I also estimate similar models in which Y_i is either an application characteristic, a prosecution or post-grant outcome to compare SEPs and controls on these variables as well.

Another major threat to the identification of the effect of potential standard essentiality is reverse causality. Prosecution strategies may make an application essential for a proposed standard even if it was not essential originally or make it more likely to become essential, and some variables that I use as outcomes in the regressions are measured before a patent becomes a SEP. Moreover, many patents are declared essential after they are granted. So the causal flow may go from prosecution strategies to SEP status. Examples of this are the filing of continuation applications or RCEs to cover a standard after a standard is set, or the use of provisional applications to delay the drafting of claims. In this scenario, a regression of the use of continuation procedures or provisional applications against SEP_i would produce a coefficient α biased upwards. Ideally, I would use an instrumental variable to instrument SEP status. I leave this analysis for future work, but it is important to keep in mind this issue when interpreting the results.

Then, I discard the control group and focus on the relationship between the business model of the applicants for SEPs and prosecution strategies. To do that, I drop the SEP indicator from model 3.1 and add an indicator variable equal to one for “upstream” applicants.

3.4.2 Standardization, patent issuance and continuation filings

Suppose we observe a sample of applications at risk of becoming essential for a proposed standard under development. In the ideal experiment to estimate the relationship between standardization and patent issuance, I would treat a subsample of these SEP applications with resolution of uncertainty over the content of the related standards. The comparisons of the issuance rates after a given amount of time between treated and control SEPs would provide estimates of the effect of standardization on issuance.

I use a sample of SEPs declared essential for a standard before they issue as patents to estimate this relationship in a longitudinal analysis, exploiting variation in standardization over time. It is hard to link SEP disclosures to specific standard documents and observe when uncertainty on standard proposals is resolved, so I utilize the date of disclosure of a SEP as a proxy for standardization (Rysman and Simcoe, 2008; Simcoe et al., 2009; Bekkers et al., 2016). I match these SEPs to control applications using the same matching strategy I utilized above. Utilizing application-month as unit of observation, I estimate OLS models based on the following regression:

$$Issued_{it} = Disclosed_{it}\alpha + X_{it}\beta + \varepsilon_{it} \tag{3.2}$$

where $Issued_{it}$ is a binary variable equal to one if application i issues as patent in month t, $Disclosed_{it}$ is a binary variable equal to one since the month of disclosure of a SEP, X_{it} is a vector of control variables and ε_{it} is the error term. SEPs are at risk of issuance from their filing month to the month of issuance (and dropped from the estimation sample afterwards) or to the end of the sample period if they do not issue as patents (and censored). Depending on the model, the control variables include an indicator for the SEPs, calendar month effects, age (in months) since filing effects, art unit effects, indicators for serialized continuations, use of provisional and PCT applications, small entity status of the applicant and foreign priority, and the natural logarithms of the number of inventors, the number claims of the application and the number of words in the first independent claim of the application.

The coefficient of interest is α and measures the correlation between standardization, as proxied by disclosure, and the probability of issuance. On the one hand, if one is willing to assume that there are no unobservable time-varying factors related to both disclosure and issuance, model 3.2 estimates the effect of standardization on issuance. On the other hand, it is possible that companies disclose their SEPs when they observe signals from the examiners that they have high chances to obtain a patent. As these signals are unobservable to me, they would bias upwards my estimates of α.

To estimate the correlation between the filing of continuation applications and standardization, I use as outcome the count of continuation applications of an application i filed in month t ($CONs_{it}$). The models are based on

$$E[CONs_{it}|X_{it}] = exp(Disclosed_{it}\alpha + X_{it}\beta) \tag{3.3}$$

where $Disclosed_{it}$ is a binary variable equal to one since the month of disclosure of a SEP and X_{it} is a vector of control variables as above. I estimate these models with Poisson regressions. Applications can have continuations from their filing month to the month of issuance or abandonment or to the end of the sample period if they do not issue as patents.

3.4.3 Prosecution length and litigation

The final part of the empirical analysis studies the relationship between the length of prosecution and litigation. I use a cross-section of SEPs disclosed as essential for a standard before their issuance to estimate models based on the following equation:

$$Y_i = log(Pendency_i)\alpha + X_i\beta + \varepsilon_i \tag{3.4}$$

where Y_i is an indicator for litigated patents, $log(Pendency_i)$ is the natural logarithm of the number of days between the earliest priority date of SEP i and its issuance, X_i contains the control variables and ε_i is the error term.

The coefficient of interest is α, which estimates the correlation between the length of prosecution and litigation. We should interpret the results of these models with caution, because the estimates are only cross-sectional and vulnerable to omitted variable bias. It is hard to include in the empirical model all the characteristics of a patent that may be related to litigation and prosecution length. For example, one can imagine that more important SEPs may have a longer prosecution because the applicants may want to invest more resources in drafting their claims or may spend more time negotiating with the patent examiners to obtain the claims they want or a stronger patent. As the value of a patent is also likely correlated with the probability it is enforced, the coefficient of prosecution length may be biased upwards in a simple regression of litigation against prosecution length.

To attempt to control for the heterogeneity across patents related to prosecution length and litigation, I include in the models issue year effects, indicators for small entity status of the applicant and foreign priority, a set of USPC class effects to control for differences across technological fields, and the natural logarithms of the number of inventors, the number of claims of the patent and the number of words of the first independent claim of the patent. The next section describes the data I use in the empirical analysis.

3.5 Data

I combine data from several sources. The data on SEPs come from the Disclosed Standard Essential Patents (dSEP) Database. This database provides the SEPs disclosed to 13 major SSOs as of March 2011 and it is described in details in Bekkers et al. (2016).[9] [10] I match the SEPs with the USPTO Patent Examination Research

[9]The SSOs covered by declared SEPs database are: the American National Standards Institute (ANSI), the Alliance for Telecommunications Industry Solutions (ATIS), the Broadband Forum (BBF), the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC), the European Telecommunications Standards Institute (ETSI), the International Electrotechnical Commission (IEC), the Institute of Electrical and Electronics Engineers (IEEE), the Internet Engineering Task Force (IETF), the International Organization for Standardization (ISO), the International Telecommunication Union (ITU), the Open Mobile Alliance (OMA) and the Telecommunications Industry Association (TIA).

[10]The use of these data relies on the assumption that the patents listed in the dSEP database are at risk of being essential for the implementation of a proposed standard. The dSEP database is based on the disclosure letters submitted to SSOs by participants in the development of new standards. During the standardization process many SSOs require their members to list the patents

Dataset (Graham et al., 2015), which provides the information available in the Public Patent Application Information Retrieval system (Public PAIR) as of January 24, 2015. This is the source of information on the prosecution strategies. I supplement these data with patent citation data from PatentsView[11] and claims data from the Patent Claims Research Dataset (Marco et al., 2016). Data on patent litigation in U.S. district courts come from Thomson Innovation.[12]

The sample of SEPs contains 4,479 U.S. regular utility patents and patent applications, of which 85 are abandoned and 48 are still pending at the beginning of 2015. 881 SEPs are disclosed while they are still pending applications. Table B1 shows the descriptive statistics for this sample of SEPs. These SEPs were classified in 103 USPC classes and examined by 1,174 examiners affiliated with 353 art units. Because of the technological scope of the SSOs covered by the dSEP database, most of these SEPs cover technologies related to Computers and Communications or Electrical components (table B2 shows the distribution of SEPs by Technology Center of the USPTO). 9% of the SEPs in this sample were litigated at least once before year 2015.

To study differences in prosecution related to the business model of applicants, I exploit the information on the business model of SEP owners provided in the dSEP database and the information on the assignees from the USPTO Patent Assignment Dataset (Marco et al., 2015). I retrieve the identity of the first assignee of each SEP in my sample and match the assignees with the information on the business model from the dSEP database to create two groups of organizations: "upstream" and "downstream", which filed 725 and 2,505 SEPs respectively.[13] Table B3 provides

(or the patent applications) that would be infringed by a proposed standard.

[11]Data downloaded from `http://www.patentsview.org` on February 26, 2016 with coverage through July 21, 2015.

[12]Data downloaded on April 28, 2016.

[13]I only keep the first assignee from the filing date of the application. When there are multiple assignments on the same date, I randomly pick one. Then, I clean the assignee names for

the list of assignees that have more than 30 SEPs, along with their business model.[14]

I use the data from Public PAIR to characterize the prosecution strategies of applicants. For each patent application, I identify the number of “children” continuation applications, continuations-in-part and divisionals and whether it is a continuation, continuation-in-part or divisional of other applications. I also identify whether the application claims the benefit of a provisional or PCT application. For granted patents, I use the filing date, the earliest priority from the continuity data (serialized continuations, provisional and PCT applications) and foreign priorities to construct variables that measure the number of days between each of these dates and the issue date of the patent. I use the transaction histories of the applications to retrieve the information on the requests for extension of time, the filings of RCEs after a notice of allowance and to measure the time between the first non-final rejection of an application and the response of the applicant. For granted patents, I also identify the filings of reissues and reexaminations. In the next section, I describe the results of the analysis of the prosecution strategies, their relationship with standardization and finally with litigation.

3.6 Results

In this section I describe the results of the empirical analysis. I start with the comparison of SEPs and matched control applications. Then I analyze the relationship between standardization, SEP issuance and filing of continuation applications of the

an easier match with the dSEP database using an automated name cleaning routine that builds upon the one developed for the NBER Patent Data Project (`https://sites.google.com/site/patentdataproject/Home/posts/namestandardizationroutinesuploaded`) and Thoma et al. (2010). Details on my routine are available upon request.

[14] Upstream organizations include pure upstream knowledge developers and patent holding companies, universities and public research institutes, nation states, as well as producers of components (including semiconductors). Downstream organizations include those whose business model is classified as “Software and software-based services”, “Equipment suppliers, product vendors, system integrators”, “Measurement and instrument, test systems”, “Service providers (telecommunications, radio, television, etc.)” or “SSOs, fora and consortiam, technology promotors” in the dSEP database.

SEPs. Finally, I estimate the correlation between the length of SEP prosecution and litigation.

3.6.1 Prosecution strategies

To construct the sample of matched SEPs and control applications, I first discard applications with missing values for the variables utilized to match. I could find a good match for 4,253 of the 4,470 SEP applications without missing values for filing date, examiner, art unit or USPC class (i.e. 95%). Table 3.1 compares the prosecution strategies of matched SEPs and control applications along with other observable characteristics of the applications, prosecution outcomes and litigation, by displaying the means of the variables of interest, the p-values of t-tests for differences in means and the normalized differences between SEPs and controls.[15] [16]

Post-grant outcomes are markedly different for SEPs and control patents. Conditional on issuance, SEPs have a much higher probability of being litigated (9% against 2%) and receive a higher number of forward citations, which are often considered measures of patent value. This confirms previous findings (Bekkers et al., 2016; Rysman and Simcoe, 2008). SEPs also seem to provide broader protection than the control patents: they have shorter first claims and a higher number of claims. They are also more likely to have reissue applications and being involved in reexamination.

Consistent with the view that SEPs are more likely to be part of chains of continuations, SEPs have a number of children continuations between twice and three

[15] Table B4 compares the SEPs with all regular utility patent applications in Public PAIR.

[16] For each variable X, the normalized difference between SEPs and controls is defined as:

$$\Delta_X = \frac{\overline{X_t} - \overline{X_c}}{\sqrt{(S^2_{X,t} + S^2_{X,c})/2}}$$

where $\overline{X_t}$ is the mean of variable X for the SEPs, $\overline{X_c}$ is the mean of the variable for the controls, and $S_{X,t}$ and $S_{X,c}$ are their standard deviations. It provides a comparison of the means for the groups alternative to a t-test, because the t-statistics may be very large just because the sample is large (Imbens, 2014).

times that of the matched controls.

While the time between the filing date of the application and issuance is about the same for SEPs and control applications, the lag between the earliest priority and issuance is substantially longer for SEPs: SEPs take about 5 more months to issue as patents when the lag is computed utilizing either the priority from the continuity data (i.e. continuations, PCT applications and provisional applications) or the earliest priority (which takes into account also foreign priorities). This is consistent with a more intense use of strategies to delay the filing of the SEP application: SEPs are more likely to claim the priority of provisional applications and PCT applications, and to be serialized continuations.

Applicants for SEPs are also more likely to take actions that prolong examination or delay issuance: they are 5 days slower to respond to the first non-final rejection, file about 0.14 more requests for extension of time, accumulate about 25 more days of applicant delay for the computation of the patent term adjustment, and are 2% more likely to file an RCE after a notice of allowance.[17]

Even after matching on relevant observables, SEPs seem to be more valuable applications. SEPs have a slightly higher number of inventors. While the difference in length of the first independent claim of the application is small and not statistically significant at conventional levels, SEPs have about 2 more claims. Taken together with the results on claim length and number for the granted patents, these results support the view that patents that are narrowed less during prosecution are more likely to be included into a standard (Kuhn et al., 2016).

Table 3.2 analyzes the differences between SEP applications and matched control applications in a regression framework, controlling for differences in observables. It

[17]The U.S. patent system provides additional days of patent term to patents that experience a reduction of the patent term due to delays of the USPTO. Applications filed on or after May 29, 2000, receive additional days of patent term for various types of delays in examination. This adjustment is reduced by delays of the applicant to respond to office actions within certain deadlines.

reports a series of coefficients of an indicator variable equal to one for SEP applications, together with their standard errors and p-values. We can interpret these coefficients as the differences between SEPs and control patents in the variables of interest controlling for art unit, filing year, small entity status, foreign priority, number of inventors, number of claims and scope of the application.[18]

The results of these regressions largely confirm those of the differences in means. Conditional on being granted, SEP applications receive more citations, are broader both in terms of claims and in length of the first independent claim, and are 2% more likely to be litigated. However, the differences in the probability of reissues and reexaminations are now very small and statistically insignificant.

The regressions also confirm that SEP applications are more likely to have continuation applications, continuations-in-part and divisionals. They are also more likely to be serialized continuations and to benefit from provisional applications. However, the difference in the probability of utilizing a PCT application is now statistically insignificant.

The regressions also confirm the difference in the issuance lag when we compute it from the earliest priority: prosecution of SEPs is about 5% longer than the prosecution of the controls, *ceteris paribus*.

The differences in terms of actions taken to delay prosecution are also confirmed: applicants for SEPs take about 4% more time to reply to the first non-final rejection, file 11% more requests of extension of time, are 4% more likely to file an RCE after the notice of allowance, and accumulate almost 80% more applicant delay for the computation of the patent term adjustment. Interestingly, the USPTO is slower

[18] The sample used in these regressions is different from the sample used in table 3.1 because I exclude applications filed before the enactment of the AIPA, unpublished applications, and applications with missing values for the explanatory variables in the regressions from the potential matches. I do not include a set of indicators for the USPC classes because many of these are collinear with the art unit effects. This is due to the technological specialization of most art units in a relatively small number of technological classes.

to process SEPs applications and accumulates 63% more delay for the computation of the patent term adjustment. This may be due to their complexity or because the interaction with more aggressive applicants is more time consuming for the examiners.

To analyze differences in prosecution strategies of SEPs related to different business models, I discard the matched sample and use all the SEPs whose business model of the original assignee was classified as either upstream or downstream. Table 3.3 shows the coefficients of the upstream indicator, their standard errors and their p-values from regressions of the variables of interest against the upstream indicator and control variables.

The results show that upstream SEP owners exploit some prosecution strategies more aggressively. Upstream SEP owners are more likely to file SEPs that are continuation applications and continuations-in-part and file continuation applications of the SEPs. They also use provisional applications more often. However, SEPs of downstream organizations are more likely to claim priority to PCT applications. This result is not surprising, since the group of downstream players contains many big manufactures of end products active on a global scale.

Upstream organizations are 8% more likely to file an RCE after the notice of allowance, take 18% more days to respond to non-final rejections, request 9% more extensions of time and accumulate 70% more days of applicant delay.

Taken together, these prosecution strategies result in issuance lags from the priority date that are 13% longer for upstream players. These results are consistent with the idea that upstream organizations have higher incentives to delay issuance because patents are more relevant for their business models.[19]

In the analysis reported in table 3.4 I estimate the contribution of the main prosecution strategies to the lag between the earliest priority date and SEP issuance. I

[19] I also estimates models that pool together upstream players and unclassified assignees. Unclassified assignees are often small organizations that are likely to behave as the upstream players. I report the results in table B5. The results are consistent with those reported in the main text.

run a set of OLS regressions with the natural logarithm of the number of days from the earliest priority to the grant date as outcome variable. The main explanatory variables include (i) an indicator variable equal to one for patents from serialized continuations, (ii) an indicator variable for patents claiming priority to a provisional application, (iii) an indicator variable for patents claiming priority to a PCT application, (iv) the count of requests for extension of time between the filing date and the grant date of the patent and (v) an indicator for the filing of an RCE after the notice of allowance of the patent.

Model (1) includes only the main five explanatory variables. Model (2) controls for the number of inventors, filing year effects, art unit effects, small entity status and foreign priority. Model (3) adds the logarithms of applicant and USPTO delays as well as binary variables to control for no delays due to applicant and USPTO. Model (4) adds the measures of application claim scope. Finally, model (5) adds the indicator for upstream organizations.

Serialized continuations have longer lags. Depending on the specification, the lag for serialized continuations is between 48% and 56% longer than for other SEPs, *ceteris paribus*. SEPs claiming the benefits of provisional applications and PCT applications also have longer lags on average: the coefficients of the provisional application indicator imply an increase by 7%-26% in the lags, and those of the PCT dummy by 25%-51%, *ceteris paribus*.

Requests for extension of time also significantly delay the grant date of a patent: depending on the model, each request delays the issue date by between 7% and 15%. Similarly, filing an RCE after the notice of allowance is associated with lags longer by 9%-31%.

The prosecution of SEPs assigned to upstream organizations is 5% longer than the prosecution of SEPs assigned to downstream organizations even after controlling for the main observable strategies and application characteristics. This suggest that upstream organizations may delay prosecution using other strategies (unobservable to us), for example simply taking more time to respond to office actions.[20]

It is also interesting to test whether the scope of the claims of SEPs that spent more time in prosecution is actually modified more. Ideally, we should read the original applications, compare them to the granted SEP and assess how the scope of protection changes. In practice this is not possible, but we can test whether they are modified more. For example, Berger et al. (2012) show that SEPs in their sample have more amendments than matched controls. In the next step of the analysis, I estimate the relationship between the change in the number of words to the first independent claim between the granted patent and the published application and the time between the filing date and issuance. So I regress the change in the number of words in the first independent claim against the natural logarithm of the number of days between the filing date of the application and the issue date of the patent and a set of control variables. I estimate the models with OLS on the sample of SEPs that were filed after the enactment of the AIPA and published before issuance. I report the results in table 3.5.[21]

The model in column (1) includes only the main explanatory variable, the days between filing and issuance, and control variables for USPTO delays for the computation of the patent term in order to better measure delays due to applicant decisions. The coefficient of the filing-issuance lag is statistically significant at 1% and implies that a 1% increase in the filing-issuance lag is associated with an increase in the length of the first independent claim by 0.5 words. I add a set of filing year effects and in-

[20] I check the robustness of the OLS regressions estimating Poisson models (table B6). The results are similar to those commented in the main text. Unreported results of OLS models using only the SEPs disclosed as essential before they issue as patents are also similar.

[21] I restrict the analysis to the prosecution of the SEP between the filing date and issuance because I compare the issued patent to the published application.

dicators for serialized continuations, provisional and PCT applications, small entity status, foreign priority and control for the number of inventors in model (2). The coefficient is essentially unchanged. Adding a set of art unit effects greatly reduces the magnitude of the coefficient of interest, which now implies a 0.3 increase in claim length associated with a 1% increase in days between filing and issuance. Finally, I add controls for the initial scope of the application in column (4). The coefficient drops and is less precisely estimated, but it is still statistically significant at 10% level and implies an increase by almost 0.2 words associated with a 1% increase in time between filing and issuance.

Interpreting claim length as a measure of patent scope, the results imply that longer prosecution of a SEP is associated with claim narrowing. Given that SEPs have broader scope at the end of prosecution but are similar to matched control applications in terms of scope at filing, one possible explanation is that applicants for SEP use a longer prosecution to achieve a good balance between obtaining a sufficiently broad and strong patent. Narrower patents may be less likely to be invalidated in a court. So SEP owners may want to tailor the claims to cover development in standardization and obtain patents that are just broad enough to cover them without risking their invalidation.[22] [23]

Now I move to the results of the analysis of the relationship between standardization, patent issuance and continuation filings.

[22] I estimate these models also only on subsample of SEPs disclosed before issuance (table B7). The magnitude of the coefficient of time between filing and issue dates is even bigger in these models, though the model in column (4) is not statistically significant at conventional levels, probably for the big drop in sample size.

[23] The observed relationship may be due also to delays of the patent office. However, I partially control for this with use of control variables related to USPTO delays for the computation of the patent term.

3.6.2 Standardization, patent issuance and continuation filings

In the first part of this section I analyze the relationship between standardization and patent issuance. Taking the disclosure date as proxy for the standardization date, I focus on SEPs that are declared essential for a standard before they issue as patents and observe whether disclosure is associated with a higher probability of issuance.

Table 3.6 shows the results of the OLS models in which each observation is an application-month, the outcome is an indicator for issuance and the main explanatory variable is an indicator equal to one starting in the month of disclosure of a SEP. The models are estimated using the sample of SEP applications declared essential for a standard before their issuance and with filing date on or after the date of enactment of AIPA. This is to reduce concerns of selecting on the outcome that may arise using older applications, because before the AIPA only granted patents were published. I use a sample of SEPs and matched control applications, matched using the same strategy adopted above. Applications are at risk of issuance from their filing month to the month of issuance (and dropped from the estimation sample afterwards) or to the end of year 2014 if they do not issue as patents (and censored).[24]

Models (1)-(3) compare the matched SEPs with their matched controls. Model (1) includes only the indicator variable for SEP application disclosure, the indicator for SEPs, calendar month effects to control for changes in the grant rate of the USPTO over time and a set of age-since-filing-month effects to control for the increase in the probability of issuance over time. The coefficient of SEP disclosure is statistically significant at 1% level and implies a 0.7% increase in the probability of issuance after disclosure. This may seem a small increase in absolute terms but it is relatively

[24]The last SEPs in this sample were filed in 2011. This ensures enough time for their publication before the end of the sample period (and the publication of their matched controls, which have the same filing year). About 17% of the applications in the estimation sample are abandoned before issuance (this percentage is about 10% for the SEPs). For these applications the last month in the sample is the month of abandonment.

very large: the probability of issuance in a given month in the sample is about 1.3%. Models (2) and (3) control for characteristics of the application that are fixed at filing, such as indicator variables for serialized continuations, provisional applications, PCT applications, small entity status, foreign priority, number of inventors (model 2), length of the first independent claim and the number of claims of the application (model 3). The coefficient of disclosure in both models is still statistically significant at 1% level and very similar to the one of the previous model. It is interesting to note that the coefficient of the SEP indicator is negative and statistically significant at 1% level in all models. *Ceteris paribus*, SEPs have a longer prosecution than the controls.

Models (4)-(6) discard the control applications and are estimated on the sample of SEP applications declared essential for a standard before their issuance and with filing date on or after the date of enactment of AIPA. The specifications are the same used in models (1)-(3), but they do not include the SEP indicator. The coefficient of SEP disclosure is statistically significant at 1% level in all models and implies an increase in the probability of SEP issuance between 1% and 1.2% after disclosure.

I check the robustness of this analysis to the exclusion of ETSI SEPs (and their control applications) in table B8. Bekkers et al. (2016) show that the particular rules of this SSO lead companies to disclose a high number of SEPs compared with other SSOs, and that these SEPs may have lower importance. The results are essentially unchanged.

In the second part of this section I analyze the relationship between standardization and continuation application filings. I estimate Poisson models on samples similar to those for patent issuance. The outcome in this models is the number of continuation applications claiming priority to the focal application. The results are reported in table 3.7.

Applications are at risk of having continuation applications from their filing month until they are either issued or abandoned. I truncate the sample period at the end of year 2012 to allow enough time for the publication of applications for the construction of the outcome.

Models (1) and (2) are estimated on the matched sample. Model (1) is a very simple model that includes only the SEP disclosure dummy and the SEP indicator. The coefficient of the SEP indicator is statistically significant at 1% level and implies a number of continuations higher by 38% for the SEPs compared with the controls. Surprisingly, the coefficient of the disclosure dummy is negative but its standard error is very large. Model (2) is estimated with conditional fixed-effect Poisson regression and also includes month effects and the non-linear terms of a fourth degree polynomial of age since filing month. The coefficient of disclosure is positive but imprecisely estimated.

Models (3) and (4) are estimated with the same specifications, but discard the control applications and use all SEP applications declared essential for a standard before their issuance and with filing date on or after the date of enactment of AIPA. The coefficient of the disclosure indicator in model (3) is positive but not statistically significant at conventional levels. The coefficient of disclosure in the fixed-effects model is statistically significant at 1% level and very large: it implies an increase by 125% in the number of continuations filed after SEP disclosure.[25]

To summarize the results of these section, standardization (as proxied by disclosure of SEPs) is positively correlated with patent issuance and continuation application filings, even if the evidence for continuations is weaker. In the next section, I analyze how the lag between priority and issuance is related to litigation.

[25] The results of this analysis are robust to the exclusion of SEPs declared to ETSI. See table B9.

3.6.3 Prosecution length and litigation

In the last part of the empirical analysis I analyze the association between the length of prosecution of SEPs and litigation. Table 3.8 presents the results of linear probability models that estimate the association between length of prosecution and probability of litigation. The sample for these regressions is a cross-section of SEPs disclosed before their issue date. I limit the analysis to this sample because I want to focus on patents that are included into a standard while they are still prosecuted. The outcome is a binary indicator equal to one if a SEP is litigated before year 2015. The main explanatory variable is the natural logarithm of the number of days between the earliest priority of the patent and its issuance.

Model (1) includes only the issuance lag and issue year effects. The coefficient of the issuance lag is statistically significant at 1% and implies that a 10% increase in the issuance lag is associated with a 0.7% increase in the probability of litigation. Model (2) adds indicators for small entity status and foreign priority and controls for the number of inventors. Model (3) adds control variables for the scope of the patent. Model (4) adds a set of USPC class effects. In these three models the coefficient of the issuance lag is statistically significant at 1% and implies that a 10% increase in the issuance lag is associated with a 0.9% increase in the probability of litigation. Given that the percentage of litigated patents in the estimation sample is about 4%, this is a non-negligible increase.[26]

To summarize, SEPs with a long time period between priority and issuance are more likely to be litigated even after controlling for differences across technologies, type of applicant, time trends and proxies for patent value. This may be because

[26] I check the robustness of the results excluding SEPs disclosed to ETSI (table B10), using as outcome an indicator equal to one if the SEP is litigated within 5 years from the issue date (table B11), estimating logit models and using litigation data from an alternative source (Lex Machina). The results are qualitatively similar, even though those for the models that use the 5-year window are less precise, possibly because of the drop in sample size related to the use of the time window.

applicants exploit a longer prosecution to change the claims of their SEPs to cover a related standard and make them more valuable. On the other hand, unobserved differences across patents may also explain the results. So I suggest interpreting this result carefully.

3.7 Conclusions

In this paper I analyze the prosecution strategies of applicants for SEPs at the USPTO. The central hypothesis of the study is that applicants for SEPs have incentives to delay the issuance of their patents, especially when standards and patents are developed at the same time. Applicants for SEPs can exploit several options provided by the U.S. patent system to delay the prosecution of their patents, observe the developments of the standardization process and draft claims that cover a standard.

I compare the prosecution strategies of SEPs and very similar patents, and find that SEPs have significantly longer issuance lags from the priority date. This is obtained using more intensively peculiar instruments of the U.S. patent system, such as provisional applications, chains of continuations and filings of RCEs. Moreover, applicants for SEPs take more time to respond to office actions. The idea that SEP owners delay patent issuance to cover standards is supported by the positive correlation between standardization and patent issuance. Moreover, standardization is positively related with the filings of continuation application, which suggest a strategic use of the continuation procedure to cover standards with additional claims.

It is possible that SEP owners exploit the delays to achieve the right balance between patent breadth and patent strength in order to obtain more valuable patents: while SEPs are broader than comparable patents, longer prosecution of SEPs is positively correlated with patent narrowing. This may make their patents "more essential", i.e. tailored on the standard, and also less likely to be invalidate in a court.

SEPs are indeed litigated more often than similar patents, and longer issuance lags are positively correlated with the probability of litigation.

Another interesting result is the systematic difference between the use of certain strategies between upstream and downstream organizations. Upstream organizations, which rely more on patents, use more intensively continuation applications and provisional applications, and are also slower at replying to office actions. This leads to longer issuance lags.

While essentially descriptive in nature, this study has implications for the design of the patent system and for SSOs' IP policies. Policy makers should think carefully about the potential abuse of instruments that allow strategic delays in patent prosecution, possibly limiting their use or increasing their cost. For example, some scholars have advocated limiting continuation applications (Lemley and Moore, 2004). SSOs' licensing policies may also require to commit to license all the members of a patent family. Furthermore, SSOs may design their policies to discourage strategic delays or promote the exchange of information on claim amendments. Early disclosure has its merits, but it has limited effectiveness when claims change over time. SSOs may encourage early disclosure and complement it with requirements to timely update disclosures with changes to the scope of the claims of potentially essential patents. Cooperation between SSOs and patent offices on the amendments to claims of potentially essential patents may also help to reduce the threats of hold-up.

In this paper, I rely on the assumption that patents disclosed as SEPs are at risk of being essential for a standard. However, strategic considerations and the IP policies of SSOs may lead to over- or under-disclosure of SEPs. Further research may analyze the prosecution of patents that are known (or more likely) to be essential for standards, for example those identified by experts and patent pools. Another limitation of the study is that I can only use an indirect measure of the standardization date. Further research may exploit the link between patents and standards to obtain a better measurement of the timing of standardization. Another limitation of this study is that I do not know how the claims of SEPs are changing during prosecution and I do not have a direct way to test the idea that continuations are really filed to cover a standard. Finally, the estimates in this paper are essentially correlational. While I try to compare SEPs with similar patents, control for important application characteristics and can control in some parts of the analysis for time-invariant unobserved heterogeneity, there may be important unobservables I cannot control for. Future research may try to address the issue of causality more explicitly. Another interesting topic for further research is the study of the consequences of strategic delays in patent prosecution on the development and subsequent adoption of standards.

Tables

Table 3.1: Comparison of SEPs with matched regular utility patent applications: differences in means.

Variable	Mean control applications	Mean matched SEPs	p-value t-test	Normalized difference
Litigated [a]	0.02	0.09	0.00	0.30
Cites [a]	24.35	44.96	0.00	0.31
Granted	0.87	0.97	0.00	0.39
Abandoned	0.12	0.02	0.00	0.41
Words in patent's 1st ind. claim [a]	169.08	163.49	0.00	-0.06
Claims in patent [a]	19.94	23.64	0.00	0.21
Reissue [a]	0.00	0.02	0.00	0.15
Reexamination [a]	0.00	0.02	0.00	0.12
CON children	0.41	0.92	0.00	0.31
CIP children	0.10	0.26	0.00	0.12
DIV children	0.07	0.17	0.00	0.12
Days earliest priority-issue (all) [a]	1,527.78	1,674.21	0.00	0.18
Days earliest priority-issue (continuity) [a]	1,455.93	1,596.36	0.00	0.17
Days filing-issue [a]	1,177.95	1,177.89	1.00	0.00
Days to respond to non-final rejection [b]	105.36	110.05	0.00	0.09
Requests for extension of time	0.58	0.72	0.00	0.14
RCE after notice of allowance [c]	0.03	0.05	0.00	0.12
Applicant delays (days) [d]	49.66	74.22	0.00	0.23
USPTO delays (days) [d]	515.07	540.58	0.11	0.05
CON	0.13	0.20	0.00	0.17
CIP	0.05	0.09	0.00	0.15
DIV	0.04	0.06	0.00	0.12
Provisional application [e]	0.13	0.21	0.00	0.21
PCT application [f]	0.04	0.06	0.00	0.07
Published	0.36	0.38	0.13	0.03
Words in application's 1st ind. claim [g]	112.79	116.22	0.25	0.04
Claims in application [g]	24.29	26.60	0.00	0.12
Inventors	2.46	2.76	0.00	0.17
Small entity	0.03	0.03	1.00	0.00
Foreign priority	0.21	0.21	1.00	0.00
Filing year	1,998.74	1,998.74	1.00	0.00

SEPs and "control" patents are all regular utility patent applications and are matched on filing year, art unit, examiner, USPC class, foreign priority, small entity applicant, an indicator variable equal to one if the filing date is on or after June 8, 1995, and an indicator variable equal to one if the filing date is on or after November 29, 2000. The sample contains 4,253 SEP applications and 4,253 controls. The samples used for individual tests may be different because of variable definitions and missing values. Details on the samples for each test are available upon request. [a] Variable defined only for issued patents. [b] Variable defined only for applications with a reply to the first non-final rejection. [c] Variable defined only for applications that receive a notice of allowance. [d] Variable defined only for applications filed after May 29, 2000. [e] Provisional applications available since June 8, 1995. [f] USA joined the PCT in 1978. [g] Variable defined only for published applications.

Table 3.2: Comparison of SEPs with matched regular utility patent applications: OLS regressions.

Outcome	Coefficient SEP	Standard error SEP	p-value SEP
Litigated [a]	0.02	0.01	0.00
Cites (log) [a]	0.30	0.05	0.00
Granted	0.15	0.02	0.00
Abandoned	-0.16	0.02	0.00
Words in patent's 1st ind. claim (log) [a]	-0.04	0.01	0.00
Claims in patent (log) [a]	0.08	0.03	0.02
Reissue [a]	0.00	0.00	0.73
Reexamination [a]	0.00	0.00	0.15
CON children (log)	0.20	0.02	0.00
CIP children (log)	0.04	0.01	0.00
DIV children (log)	0.03	0.01	0.00
Days earliest priority-issue (all) (log) [a]	0.05	0.02	0.03
Days earliest priority-issue (continuity) (log) [a]	0.05	0.02	0.03
Days filing-issue (log) [a]	-0.03	0.01	0.02
Days to respond to non-final rejection (log) [b]	0.04	0.02	0.06
Requests for extension of time (log)	0.11	0.02	0.00
RCE after notice of allowance [c]	0.04	0.01	0.00
Applicant delays (days) (log)	0.79	0.09	0.00
USPTO delays (days) (log)	0.63	0.11	0.01
CON	0.06	0.02	0.00
CIP	0.03	0.01	0.03
DIV	0.03	0.01	0.03
Provisional application	0.14	0.02	0.00
PCT application	0.00	0.01	0.67

The regression models are estimated on a sample of SEPs and matched control applications filed on or after November 29, 2000, that were published as applications. The sample contains 1,551 SEP applications and 1,551 controls. The samples used for individual regressions may be different because of variable definitions and missing values. Outcomes taken in logarithms may be either the natural logarithm of the variable or the natural logarithm of one plus the variable if the latter contains zeros. Details on samples, variables and regressions are available upon request. SEP and control applications are matched on filing year, art unit, examiner, USPC class, foreign priority and small entity applicant. All regressions are estimated by OLS and include art unit effects, filing year effects, indicators for small entity status and foreign priority and control for the natural logarithms of the number of words in the first independent claim of the application, the number of claims in the application and the number of inventors. Applications with missing values for these variables are excluded from the potential matches. Robust standard errors are clustered at art unit level. [a] Variable defined only for issued patents. [b] Variable defined only for applications with a reply to the first non-final rejection. [c] Variable defined only for applications that receive a notice of allowance.

Table 3.3: Comparison of SEP applications assigned to upstream and downstream organizations: OLS regressions.

Outcome	Coefficient upstream	Standard error upstream	p-value upstream
Litigated [a]	0.01	0.01	0.64
Cites (log) [a]	0.06	0.11	0.57
Granted	0.01	0.02	0.73
Abandoned	0.00	0.02	0.79
Words in patent's 1st ind. claim [a]	0.01	0.03	0.77
Claims in patent [a]	-0.06	0.06	0.26
Reissue [a]	0.01	0.01	0.41
Reexamination [a]	-0.01	0.00	0.07
CON children (log)	0.25	0.05	0.00
CIP children (log)	0.03	0.03	0.36
DIV children (log)	0.02	0.02	0.33
Days earliest priority-issue (all) (log) [a]	0.13	0.04	0.00
Days earliest priority-issue (continuity) (log) [a]	0.13	0.04	0.00
Days filing-issue (log) [a]	0.05	0.03	0.15
Days to respond to non-final rejection (log) [b]	0.18	0.05	0.00
Requests for extension of time (log)	0.09	0.05	0.04
RCE after notice of allowance [c]	0.08	0.02	0.00
Applicant delays (days) (log)	0.70	0.16	0.00
USPTO delays (days) (log)	0.18	0.19	0.32
CON	0.06	0.03	0.07
CIP	0.05	0.02	0.03
DIV	0.00	0.02	0.97
Provisional application	0.18	0.04	0.00
PCT application	-0.06	0.02	0.00

The sample contains only regular utility SEP applications whose first assignee after the filing date is either an "upstream" or a "downstream" organization. The sample contains 946 SEP applications assigned to downstream organizations and 338 SEP applications assigned to upstream organizations. The samples used for individual regressions may be different because of variable definitions and missing values. Outcomes taken in logarithms may be either the natural logarithm of the variable or the natural logarithm of one plus the variable if the latter contains zeros. Details on samples, variables and regressions are available upon request. All regressions are estimated by OLS and include art unit effects, filing year effects, indicators for small entity status and foreign priority and control for the natural logarithms of the number of words in the first independent claim of the application, the number of claims in the application and the number of inventors. Applications with missing values for these variables are excluded from the sample. Robust standard errors are clustered at art unit level. [a] Variable defined only for issued patents. [b] Variable defined only for applications with a reply to the first non-final rejection. [c] Variable defined only for applications that receive a notice of allowance.

Table 3.4: Individual strategies and issuance lag of SEP applications: OLS regressions.

	(1)	(2)	(3)	(4)	(5)
Outcome	Issuance lag (log)	Issuance lag (log)	Issuance lag (log)	Issuance lag (log)	Issuance lag (log)
Serialized Continuation	0.48***	0.48***	0.56***	0.56***	0.48***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Provisional application	0.26***	0.13***	0.09***	0.07***	0.12***
	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
PCT application	0.51***	0.26***	0.33***	0.25***	0.33***
	(0.03)	(0.02)	(0.04)	(0.03)	(0.04)
Requests for extension of time	0.15***	0.11***	0.08***	0.07***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
RCE after NOA	0.31***	0.16***	0.10***	0.10***	0.09***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Upstream					0.05**
					(0.03)
Filing year effects	N	Y	Y	Y	Y
Art unit effects	N	Y	Y	Y	Y
Small entity	N	Y	Y	Y	Y
Foreign priority	N	Y	Y	Y	Y
Inventors	N	Y	Y	Y	Y
Applicant delay (days)	N	N	Y	Y	Y
Applicant delay (binary)	N	N	Y	Y	Y
USPTO delay (days)	N	N	Y	Y	Y
USPTO delay (binary)	N	N	Y	Y	Y
Words & Claims	N	N	N	Y	Y
Observations	4,344	4,340	1,868	1,535	1,167
R-squared	0.40	0.68	0.68	0.70	0.75

All regressions estimated by OLS on the sample of granted SEPs. Outcome in all regressions is the natural logarithm of the number of days between the earliest priority and the issue date of the SEP. The variables “Applicant delay (days)”, “USPTO delay (days)”, “Inventors”, “Words” and “Claims” are the natural logarithms of, respectively, one plus the number of days of applicant delay for the computation of the patent term adjustment, one plus the number of non-overlapping days due to delay of the USPTO for the computation of the patent term adjustment, the number of inventors, the number of words in the first independent claim of the application and the number of claims in the application. The variables “Applicant delay (binary)” and “USPTO delay (binary)” are binary indicators equal to one if the application has days of, respectively, applicant delay and USPTO delay (zero otherwise). Robust standard errors in parentheses, clustered at art unit level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 3.5: Prosecution time and claim changes.

	(1)	(2)	(3)	(4)
Outcome	Words added	Words added	Words added	Words added
Days filing-issue (log)	51.19***	52.73***	31.11***	16.56*
	(9.30)	(8.89)	(9.73)	(9.44)
USPTO delay (binary & days)	Y	Y	Y	Y
Filing year effects	N	Y	Y	Y
Serialized continuation	N	Y	Y	Y
Provisional application	N	Y	Y	Y
PCT application	N	Y	Y	Y
Small entity	N	Y	Y	Y
Foreign priority	N	Y	Y	Y
Inventors	N	Y	Y	Y
Art unit effects	N	N	Y	Y
Words & Claims	N	N	N	Y
Observations	1,533	1,533	1,533	1,533

All regressions estimated by OLS on the sample of granted SEPs filed on or after November 29, 2000, and published before issuance. Outcome in all regressions is the number of words added to the first independent claim between the published application and the issued patent. The variables "USPTO delay (days)", "Inventors", "Words" and "Claims" are the natural logarithms of, respectively, one plus the number of non-overlapping days due to delay of the USPTO for the computation of the patent term adjustment, the number of inventors, the number of words in the first independent claim of the application and the number of claims in the application. The variable "USPTO delay (binary)" is a binary indicators equal to one if the application has days of USPTO delay (zero otherwise). Robust standard errors in parentheses, clustered at art unit level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 3.6: SEP disclosure and patent issuance.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Issued	Issued	Issued	Issued	Issued	Issued
Sample	Matched sample	Matched sample	Matched sample	SEPs	SEPs	SEPs
Disclosed	0.007***	0.005***	0.005***	0.012***	0.011***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
SEP	-0.005***	-0.004***	-0.004***			
	(0.001)	(0.001)	(0.001)			
Month effects	Y	Y	Y	Y	Y	Y
Age effects	Y	Y	Y	Y	Y	Y
Art unit effects	N	Y	Y	N	Y	Y
Serialized Continuation	N	Y	Y	N	Y	Y
Provisional application	N	Y	Y	N	Y	Y
PCT application	N	Y	Y	N	Y	Y
Small entity	N	Y	Y	N	Y	Y
Foreign priority	N	Y	Y	N	Y	Y
Inventors	N	Y	Y	N	Y	Y
Words and Claims	N	N	Y	N	N	Y
Observations	61,984	61,984	61,984	42,236	42,073	39,088
Number of applications	1,060	1,060	1,060	696	694	641

All regressions estimated by OLS. Unit of observation is application-month. The matched sample contains SEPs declared before their issue date and matched control applications filed on or after November 29, 2000, that were published as applications. SEP and control applications are matched on filing year, art unit, examiner, USPC class, foreign priority and small entity applicant. Applications with missing values for variables used to match or control variables are dropped from potential matches. The SEPs sample contains only SEPs declared before their issue date with filing date on or after November 29, 2000. Applications at risk of issuance from their filing month and observed until issuance month or abandonment month or end of calendar year 2014 if they do not issue as patents. Outcome in all regressions is an indicator variable equal to one in the month of issue. The variables “Inventors”, “Words” and “Claims” are the natural logarithms of, respectively, the number of inventors, the number of words in the first independent claim of the application and the number of claims of the application. Robust standard errors are clustered at art unit level. *** p<0.01, ** p<0.05, * p<0.10.

Table 3.7: SEP disclosure and continuation application filings.

	(1)	(2)	(3)	(4)
Outcome	# of CON children	# of CON children	# of CON children	# of CON children
Sample	Matched sample	Matched sample	SEPs	SEPs
Disclosed	-0.16	0.13	0.17	0.81***
	(0.16)	(0.18)	(0.14)	(0.20)
SEP	0.32**			
	(0.13)			
Month effects	N	Y	N	Y
Age^2, Age^3 and Age^4	N	Y	N	Y
Application Fixed Effects	N	Y	N	Y
Observations	58,275	19,523	39,743	15,605
Number of applications	1,060	359	696	270

Unit of observation is application-month. All models estimated with Poisson regressions. The matched sample contains SEPs declared before their issue date and matched control applications filed on or after November 29, 2000, that were published as applications. SEP and control applications are matched on filing year, art unit, examiner, USPC class, foreign priority and small entity applicant. Applications with missing values for variables used to match or control variables are dropped from potential matches. The SEPs sample contains only SEPs declared before their issue date with filing date on or after November 29, 2000. Applications at risk of filing continuation applications from their filing month and observed until their issue month or their abandon month or the end of calendar year 2012 if they are still pending. Robust standard errors in parentheses, clustered at application level. *** p<0.01, ** p<0.05, * p<0.10.

Table 3.8: SEPs' issuance lag and litigation.

	(1)	(2)	(3)	(4)
Outcome	Litigated	Litigated	Litigated	Litigated
Issuance lag (log)	0.07***	0.09***	0.09***	0.09***
	(0.02)	(0.03)	(0.02)	(0.03)
Issue year effects	Y	Y	Y	Y
Small entity	N	Y	Y	Y
Foreign priority	N	Y	Y	Y
Inventors	N	Y	Y	Y
Words & Claims	N	N	Y	Y
USPC class effects	N	N	N	Y
Observations	686	684	680	680

All regressions estimated by OLS. Unit of observation is a SEP. Sample contains only SEPs declared before the issue date. Outcome in all regressions is an indicator variable equal to one if the patent is litigated before year 2015. The variables "Inventors", "Words" and "Claims" are the natural logarithms of, respectively, the number of inventors, the number of words in the first independent claim of the patent and the number of claims of the patent. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Chapter 4

Standard Setting Organizations and the Scope of Cumulative Inventive Activity

4.1 Introduction

Compatibility standards are very important for innovation in industries in which interoperability is fundamental. In Information and Communication Technology (ICT) Standard Setting Organizations (SSOs) often develop these standards. They have a key role in coordinating research efforts, reaching consensus on shared technological platforms and providing solutions to the threat of holdup related to the inclusion of patented technologies into a standard. Their activities promote the adoption of new standards and encourage potential inventors to build on them. However, little is known on the direction of follow-on inventive efforts.

In this paper we study the relationship between SSO endorsement of a technology and the direction of inventive activity that builds on it. While other works already try to estimate the contribution of SSOs to the *rate* of cumulative inventive activity, in this paper we focus on its *direction*. Our theoretical framework combines insights from the literature that studies standards, direct and indirect network effects with the concept of cumulative technical progress. We go beyond the study of adoption of a standardized technology to analyze its use as an input for inventive activity across technological areas. We define two types of possible directions of cumulative inventive activity: “deepening” (cumulative technical progress characterized by relatively low dispersion across technological areas) and "broadening" (cumulative technical progress characterized by relatively high dispersion across technological areas).

We use the disclosure letters provided by the public archives of thirteen major SSOs to identify pieces of technology endorsed by SSOs. Assuming that Standard Essential Patents (SEPs) disclosed to these SSOs protect technologies embodied into their standards, we analyze the dispersion in technological space of the flow of patent citations to these SEPs relative to control patents with same vintage, in the same technology class and having the same citation trend over time.

The first part of our empirical analysis uses two measures of similarity between the citing and the cited patents to estimate the balance between deepening and broadening. The estimates show that SSOs select patents that are relatively more important in a narrow technological space and increase the breadth of cumulative improvements after standardization.

The second part of the analysis tries to separate the trends in deepening and broadening utilizing citations from "new" and "old" technological classes. We find that both deepening and broadening occur, with stronger evidence for broadening.

Previous research analyzes various roles of SSOs. Lerner and Tirole (2006) develop a model that focuses on the certification role of SSOs. Chiao et al. (2007) extend this model to study the policies of SSOs and test the predictions of their model on a database of SSOs. Lemley (2002) and Baron and Spulber (2016) also study the policies of SSOs. Rysman and Simcoe (2008) show empirically that the endorsement of an SSO leads to an increase in use of a technology as an input for inventive activity, and Bekkers et al. (2016) extend this analysis taking into account different SSO policies. We combine the insights of these studies with those from the literatures on cumulative inventive activity and network effects to analyze what type of technologies are selected by SSOs and the direction of the effect of the endorsement of SSOs (Katz and Shapiro, 1985; Furman and Stern, 2011).

This study contributes to the literature on SSOs (Baron and Spulber, 2016; Chiao et al., 2007; Lerner and Tirole, 2006; Rysman and Simcoe, 2008; Simcoe, 2012). In particular, we analyze how the endorsement of SSOs is related to the direction of technical progress that builds on their standards, examining the technological dispersion of marginal and selection effects already documented in the literature.

This paper also contributes to the literature on the relationship between institutions and rate and direction of cumulative inventive activity (Aghion et al., 2008; Furman and Stern, 2011; Murray et al., 2009). Combining network effects and cumulative inventive activity, our theoretical framework distinguishes between two types of technical progress that differ in terms of dispersion in technological space. Our empirical analysis also shows considerable differences across SSOs with different IP policies and licensing terms provided by the patent owners. This suggests that institutional details are important for the direction of inventive activity.

Finally, this study contributes to the literature that uses patent data to study inventive activity (Hall et al. (2001) and the vast literature that builds on them). We introduce a simple measure of similarity between citing and cited patents based on interclass citation flows that other scholars can easily adopt and adapt to their needs. This measure weights each citation by the probability that a patent from the citing technology class cites a patent in the cited class in a given time period. The measure is easy to compute and less dependent on the technological classification adopted than existing alternatives in the literature.

We provide a description of the role of SSOs in the standardization process, explain the main factors that may shape deepening and broadening, and relate them to the selection and marginal effects of SSOs in section 4.2. We describe our empirical strategy in section 4.3 and report our results in section 4.4. Section 4.5 concludes.

4.2 What SSOs Do

Voluntary SSOs are institutions for solving industry-wide problems that have an important element of collective action, such as governing a shared technology platform. These organizations may play several different roles. Various authors have suggested that SSOs act as a certification agent (Lerner and Tirole, 2006), a forum for bargaining over design decisions (Simcoe, 2012; Farrell and Simcoe, 2012a), a cooperative R&D organization (Weiss and Sirbu, 1990) and a partial solution to hold-up problems produced by widespread adoption of patented technology (Shapiro, 2001; Lemley, 2002). One shared prediction of these diverse theoretical perspectives is that SSOs encourage coordinated adoption of key technologies.

Rysman and Simcoe (2008) provide one of the first attempts to measure the impact of SSOs on the utilization of standardized technology. They show that patents associated with an industry standard receive more citations than does a set of observably similar control patents. This difference in citation rates exists prior to standardization, and increases afterwards, leading them to conclude that SSOs produce both a selection effect (they identify and endorse promising technologies) and a marginal effect (they contribute to the widespread adoption of those technologies). Bekkers et al. (2016) confirm these results with updated data and new methods.

This paper goes beyond Rysman and Simcoe (2008) and Bekkers et al. (2016) by examining the *dispersion* in citations received by patents linked to industry standards. If we observe an increase in dispersion following SSO endorsement, it would suggest that the standards promulgated by SSOs facilitate an inter-industry division of innovative labor. If instead we observe an increase in the rate of citation, but not in the dispersion of citing technology classes, it would suggest that the marginal effect identified by Rysman and Simcoe (2008) and Bekkers et al. (2016) is produced by a set of activities that facilitate coordination *within* a specific industry or technological field.

In what follows, we refer to increased citation of standardized technology within an industry or technical field as *deepening*, and an increase in the dispersion of citing areas as *broadening*.[1] The next part of this section describes the theoretical underpinnings for these two distinct effects. It is important to note that these two effects may happen simultaneously, i.e. SSOs may positively affect *both* deepening and broadening at the same time. For part of our empirical analysis we measure the balance of these two types of developments, and then try to distinguish them.

We are also interested in the analysis of the selection of technologies for SSO standards. SSOs may select technologies that are technically important within a relatively narrow area or may pick technologies adopted more broadly. The final part of this section discusses also the selection of technologies for standards.

4.2.1 Deepening

The formulation of new standards often requires coordination of many interested parties within a single industry (Simcoe, 2012). The basic challenge is to persuade independent implementors to adopt a new standard, thus triggering demand-side economies of scale (Katz and Shapiro, 1985). Bresnahan and Greenstein (1999) coined the term "divided technical leadership" to describe this process of coordinated technology adoption within a particular industry. We use the term "deepening" to emphasize the idea of cumulative technical progress characterized by relatively low dispersion across technological areas. Though one could argue that deepening is just a particular application of *direct* network effects, the concept of deepening goes beyond that. With the concept of deepening we aim to capture both the *adoption* of

[1] Although it might seem natural to call increased dispersion a horizontal effect, and increased within-sector utilization a vertical effect, both deepening and broadening arguably have a vertical dimension when the citing patent represents a cumulative improvement upon the standarized technology.

a specific standardized technology typically analyzed in the literature that refers to direct network effects and the *cumulativeness of inventive activities* within a specific technological area that uses standardized technology as an input for further technical improvements.

The adoption of the Global System for Mobile Communications (GSM) provides a good example of how standards can promote deepening. GSM was originally developed by the European Telecommunications Standards Institute (ETSI) with the goal of designing a pan-European mobile communication network. The standard is currently implemented on 80 percent of all mobile phones, and there have been substantial improvements in price and performance of GSM technology. Later-generation standards, such as the Universal Mobile Telecommunications System (UMTS) and Long Term Evolution (LTE) build directly upon GSM. However, the use of GSM technology remains concentrated within a single large technology area – mobile telecommunications – with relatively little adoption in other devices or applications.

SSOs promote deepening and coordinate cumulative technological progress in a variety of ways. For example, the endorsement of a particular technological solution by an SSO contributes to the formation of the expectations about future adoption, leading to self-fulfilling prophecies (Katz and Shapiro, 1985; Besen and Farrell, 1994; Liebowitz and Margolis, 1994). In this respect, SSOs may also be seen as certifiers of the quality of a technology that shape users' opinion about the merits of a specific technological solution (Lerner and Tirole, 2006). The endorsement by an SSO of a new standard can also help to overcome the excessive inertia that characterizes the replacement of old generations of technical standards (Farrell and Saloner, 1985).

The activities of SSOs can also ease information flows among the members of the organizations, facilitating technological forecasting by providing a road-map for firms in the industry. Participants in the activities of SSOs can monitor the technological advances in their area, learn from other participants about recent technical improvements and anticipate the developments of the technologies in the near future.[2] Companies can use this information to invest in the right direction, starting early to develop products and services that will be related to proposed standards.

Many SSOs require the disclosure of patents that may be required to implement standards, and the provision of liberal licensing terms for standards implementers (Baron and Pohlmann, 2016; Bekkers et al., 2016; Chiao et al., 2007; Lemley, 2002; Simcoe, 2013). These policies not only reduce the cost of cumulative follow-on inventive activity (Scotchmer, 1991; Gallini and Scotchmer, 2002; Bessen and Maskin, 2009), but also reduces the risk of hold-up typical of complex industries in which a patent thicket may increase licensing costs from royalty stacking, uncertainty and transaction costs caused by negotiations with multiple parties (Cohen et al., 2000; Farrell et al., 2007; Hall and Ziedonis, 2001; Heller and Eisenberg, 1998; Lemley and Shapiro, 2007; Lerner and Tirole, 2015; Shapiro, 2001; Ziedonis, 2004).

By triggering network effects, providing a technological road-map, lowering the cost of cumulative inventive activity and reducing the risk of hold-up, the activities of SSOs may encourage investments in new technologies that build upon the standard by companies within an industry. This may lead to an increase in patent citations to patents related to the standard concentrated in the same technological area that cited the focal patents before the endorsement by an SSO. Moreover, if deepening is stronger than broadening, we should observe a decrease in the average "distance" between the standardized technologies and those building upon them.

[2] For example, Waguespack and Fleming (2009) find that simple attendance to meetings of the Internet Engineering Task Force (IEFT, the SSO that sets protocols used to run Internet) is beneficial for start-ups.

4.2.2 Broadening

In addition to promoting coordinated adoption within a single industry, standards can provide a shared interface that facilitates use of a single component technology across many different applications. For example, the three-pronged electrical plug and outlet are used by nearly all electrical devices in the USA. This type of broadening across application areas is closely related to the idea of *indirect* network effects that occur when many application developers share a common platform, and to the horizontal externalities discussed in the literature on general purpose technologies (Bresnahan and Trajtenberg, 1995; Bresnahan, 2010). Similarly to deepening, our concept of broadening goes beyond indirect network effects by combining *adoption* across many application areas and *cumulative invention* in diverse areas fostered by a standardized technology characterized by high generality.

The Wi-Fi or IEEE 802.11 standards provides a good example of broadening. The initial promoters of Wi-Fi technology within the Institute of Electrical and Electronics Engineers (IEEE) came from NCR Corporation and its partner, AT&T, who shared the goal of developing a technology standard for wireless cash registers. Since that time, Wi-Fi has been adopted in many diverse industries such as personal computers, video-game consoles, smart-phones, tablets, digital cameras, digital audio players and home appliances. Over time, the adoption of Wi-Fi in these diverse areas, and the creation of application-specific complements, has arguably produced an increase in the dispersion of the cumulative Wi-Fi improvements across different technological areas.

The basic economics of broadening is closely related to the logic of scope-economies. Cooperative development of a new standard within an SSO allows developers to share the fixed costs of innovation, lowers the cost to future developers of accessing the platform, and in some cases provides a credible commitment whereby large firms can delegate control over key technologies (Gawer and Henderson, 2007). This not only enhances division of labor and gains from specialization, but also allows the development of totally unexpected complementary products. In fact, companies from outside the technological area can adapt the standardized technology at relatively low cost and low risk of expropriation. This promotes both mix-and-match compatibility (Matutes and Regibeau, 1988) and increased experimenting with new technologies that utilize the standardized interface. Various empirical studies of cumulative invention find that a decline in the cost of accessing research inputs is associated with increased *diversity* of follow-on research lines (Murray et al., 2009; Furman and Stern, 2011).

Increasing the number and variety of complementary products can generate a positive feedback loop when network effects are important. Not only the number of producers on a platform can trigger network effects and thus attract more complementors, but it can also stimulate the production of new varieties of products. Boudreau (2012) shows empirically that the incentives to innovate for a complementor are increased by the number of producers of other products on the same platform. This can generate a virtuous cycle for both the number of new specialized innovators on the platform and their heterogeneity.

Facilitating access to and experimentation on the technology platform by specialized innovators and the increase in their heterogeneity should therefore lead to a boost in the dispersion of the cumulative improvements on the standardized technology across different technological areas. As a consequence, after a (patented) technology is endorsed by an SSO we may observe an increase in new inventive activities (and thus patent citations) on it in technologies in areas that have never built upon it before. What is more, should broadening be stronger than deepening, we may observe an increase in the technological "distance" between SEPs and patents that cite them as new experimentation in different technological areas takes place.

4.2.3 Selection and Marginal Effects

Thus far, our discussion has emphasized the idea that SSOs could have a marginal impact on technology utilization and inventive activity through either broadening or deepening effects. However, it is equally possible that SSOs could *select* technologies based on either their suitability for cumulative progress within the focal industry, or for widespread adoption across different sectors. In fact, there is no reason why SSOs could not produce both selection and marginal effects that reflect a combination of broadening and deepening.

On the one hand, SSOs may be interested in selecting unusually "broad" technologies. Especially when compatibility issues are involved, SSOs can recognize the importance of adopting technical solutions that facilitate the interactions among different technological areas. This means that they can have an interest in selecting technologies that are inherently more general and can be applied in a variety of settings for their standards. Therefore, it is possible that they select technologies that are more general and "broader" even before standardization.

On the other hand, SSOs may want to select technologies that are important in a relatively narrow area. One of the functions of SSOs is to select technologies among the potential alternatives available to firms in an industry. It is reasonable to think that they can pick above-average technologies for the formulation of a new standard. Indeed, Rysman and Simcoe (2008) show that SSOs select technologies that receive more patent citations than similar technologies even before standardization. Moreover, while SSOs have an interest in a wide adoption of their standards, participants in the standardization process are interested in their private returns to their investments, and may push for adoption of technologies for which they have complements. As companies often specialize in a relatively limited set of technologies, the selection may be focused on technologies that are important in a relatively narrow technical area.

Table 4.1 summarizes the main arguments related to deepening and broadening. Deepening and broadening are not mutually exclusive, and actually may reinforce each other. The next section of the paper describes our empirical strategy for quantifying the relative importance of each of these four different potential effects of SSOs and our approach to try to disentangle them.

4.3 Empirical Strategy

This section describes how we identify technologies endorsed by SSOs and how we test the association of this endorsement with deepening and broadening.

4.3.1 Data Sources

We use SEPs to identify technologies that are endorsed by SSOs. During the formulation of new technical standards, SSOs typically require their members to disclosure the patents that would be infringed by any implementation of the proposed standards. Some SSOs publish on-line the disclosure letters listing these patents. Bekkers et al. (2016) collected the letters made available by thirteen major SSOs and identified the declared SEPs listed on these letters. We use these declared SEPs as a window on the technologies endorsed by these SSOs, assuming that the declared SEPs are eventually essential for their standards.[3] [4] [5]

The technologies endorsed by SSOs may not only be covered by the declared SEPs, but also by other patents within the same "family". In this paper we use only patents granted by the USPTO. U.S. patents are the largest group of SEPs in all the major databases of SEPs we are aware of (Bekkers et al., 2016; Baron and Pohlmann, 2016) and we can focus on patents produced within the same institutional environment. We use an algorithm developed by Bekkers et al. (2016) to identify all the U.S. members of the "extended family" of the declared SEPs. For the sake of simplicity, for the remainder of the paper we call SEPs the members of these "extended families".[6]

We complement the information on the SEPs with information on application and patent characteristics from the Public Patent Application Information Retrieval system (Public PAIR) provided by the USPTO Patent Examination Research Dataset

[3] This database of declared SEPs is available online at `http://www.ssopatents.org/`.

[4] These declared SEPs likely include many "true" SEPs but may include also some "false positives". These false positives may be patents disclosed early during standardization that are not essential because of changes in their claims or in the proposed standards. The policies of SSOs may also affect the number of false positives. For example, the provision of strong incentives to disclose may lead companies to list many patents that are not likely to be infringed by a standard. Companies may also have incentives to over-disclose patents to improve their bargaining power in cross-licensing negotiations. Some SEPs may also be excluded from the dSEP database. Some SSOs provide the option of submitting blanket disclosures that do not list specific patents and provide general licensing commitments for the patents owned by a company. Finally, some patents may be owned by companies that do not participate in the standardization process and so are not required to file these disclosure letters. See Bekkers et al. (2016) for more details on this. Nevertheless, the results of previous research support our assumption, showing that the timing of SEP disclosure has a positive relationship with their citation rates, suggesting that disclosure is related to the inclusion of those technologies into a standard and to the subsequent adoption of the latter (Rysman and Simcoe, 2008; Bekkers et al., 2016).

[5] The SSOs covered by declared SEPs database are: the American National Standards Institute (ANSI), the Alliance for Telecommunications Industry Solutions (ATIS), the Broadband Forum (BBF), the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC), the European Telecommunications Standards Institute (ETSI), the International Electrotechnical Commission (IEC), the Institute of Electrical and Electronics Engineers (IEEE), the Internet Engineering Task Force (IETF), the International Organization for Standardization (ISO), the International Telecommunication Union (ITU), the Open Mobile Alliance (OMA) and the Telecommunications Industry Association (TIA).

[6] Specifically, we exploit the link between the dSEP database and PATSTAT to identify all the members of the DOCDB family of all the SEPs in the database, all the continuations of those patents and all the priorities of the SEPs. Then, we take all the SEPs, their family members, their continuations and their priorities and identify those that are granted by the USPTO.

(Graham et al., 2015), from PatentsView and from the Patent Claims Research Dataset (Marco et al., 2016), and information on patent litigation in U.S. district courts from Thomson Innovation.[7]

As it is common in the literature, we use forward patent citations to measure cumulative inventive activity that builds upon a patent (Hall et al., 2001; Galasso and Schankerman, 2015). While patent citations have well-known limitations, patents are an abundant source of data and have very wide coverage in terms of technological fields and over time. Previous research shows that patent citations received by patents are positively correlated with the economic value of a technology and firm value (Trajtenberg, 1990; Harhoff et al., 1999; Hall et al., 2005; Gambardella et al., 2008). Jaffe et al. (2000a) and Jaffe et al. (2000b) show that patent citations are indicators of technology and knowledge flows, and many works have used patent citations to measure knowledge spillovers and technology diffusion (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Henderson et al., 2005). On the other hand, limitations of patent data are well-known. Not all inventions are patented or even patentable; the propensity to patent varies across industries and technological areas; and even if patent citations may track spillovers and knowledge flows, they are only noisy indicators.

4.3.2 Measures of similarity between cited and citing patents

A major contribution of this paper is to take into account the technological similarity between citing and cited patents in our analysis of the dispersion of cumulative inventive activity upon standards. We use two measures of technological similarity that we descrive below. First, we propose a measure that exploits inter-class citation flows to quantify the distance between citing and cited patents in technological space. This

[7]We downloaded the data on patent litigation from Thomson Innovation on April 28, 2016 and cleaned them to identify the patents that were litigated in U.S. district courts.

measure is based on the idea of giving different weights to citations that are more or less likely to cite a patent. The second one is a measure of the similarity between the text of two patents developed by Younge and Kuhn (2016).

A natural candidate to measure the dispersion of patent citations to a focal patent is the generality index (Trajtenberg et al., 1997; Henderson et al., 1998; Hall et al., 2001; Hall and Trajtenberg, 2004; Hall, 2005).[8] However, a common critique to the generality index is that it treats all classes as equidistant in technological space, and therefore depends on the particular technological classification used. For example, for the computation of the generality index for a patent in "Multiplex communications" (USPC class 370) a citation from class "Pulse or digital communications" (USPC class 375) has the same weight of a citation from "Plant protecting and regulating compositions" (USPC class 504) even if the former two are closer in technological space, as they are related to communication technologies, while the third one is related to technologies used in agriculture. A citation to the focal patent from the latter class is thus a signal of a broader impact of the cited patent.

Also, the generality index treats all classes as they have the same size. In actuality some classes are more narrowly defined and others very wide (e.g. "Chemistry of inorganic compounds" is a single class, while we have several "Organic compounds" classes). This may result in differences in the number of patents in each class, and thus in the probability of making and receiving citations. Moreover, this has implications for the heterogeneity of technologies within classes (wider classes may be populated by patents that are very different).

A further limitation of the generality index is that it does not take into account

[8]The generality index is defined as $Generality_i = 1 - \sum_j^{n_i} s_{ij}^2$, where s_{ij} is the share of citations received by patent i from patents in technological class j, and n_i is the number of technological classes citing patent i. Generality is bounded between zero and one, and higher values suggest that the cited patent has citations more dispersed across technological fields. Hall (2005) shows how to correct this measure to take into account the number of citations received.

the timing of citations. Technologies in different classes may be more or less close to each other over time. Moreover, the generality index is computed using all the citations received by a patent, so its utilization in longitudinal analyses is limited.

To overcome these limitations, in this paper we first change the unit of analysis. While the generality index is defined at the patent level, we propose to use a citation as the unit of analysis.

Second, we use two measures of the similarity between citing and cited patents. We propose a new measure to quantify the distance in technological space between citing and cited patents, which we call "Citation Weighted Technological Similarity" (CWTS). The idea behind our measure is to weight each citation by the probability that a given class cites another class at a certain point in time.

We define the CWTS between a citing class a and a cited class b in year t as

$$CWTS_{abt} = \frac{ncites_{abt}}{ncites_{at}}$$

where $ncites_{at}$ is the number of citations made by patents filed in year t classified in class a and $ncites_{abt}$ is the number of citations made by patents filed in year t in class a to patents in class b. The intuition behind this measure is that classes that are closer in technological space should cite each other at higher rates because their technologies are more similar and so more likely to build on each other. In our analysis, a citation from patent i filed in year t and classified in a to patent j classified in b is weighted by $CWTS_{abt}$ and thus has value $\frac{ncites_{abt}}{ncites_{at}}$ instead of one.[9]

The first advantage of CWTS is that it takes into account the probability of interclass citation, giving a higher similarity score to citations that are more likely.

[9]Other measures may be computed building on this simple logic. For example, one may take the average of CWTS for all the citations received by a given patent in a given year. However, we use CWTS for an empirical analysis at the citation level and so we define CWTS in this simple form.

The second advantage is that it depends less on the specific classification utilized, because interclass citation flows should reveal differences that a mere application of a classification scheme cannot capture. The third advantage is that the similarity between classes varies over time, taking into account possible convergence or divergence between the underlying technological fields.

Of course, our measure has limitations. Most importantly, patent citations can capture only imperfectly the links between technologies that we want to measure in this study. Also, while less dependent than other measures, it still depends on the technological classification used. It is also worth noting that it is not a measure of similarity in a strict sense: it is not symmetric and the similarity between a citing class and itself is usually different from one, the maximum value of CWTS as defined above.

For illustrative purposes, we compute the CWTS between NBER technological categories (Hall et al., 2001). We report the results in table 4.2. Please note that we do not use this level of technological classification in our empirical analysis later in the paper, and that we do not take into account the timing of citations for this example.

The rows of table 4.2 represent the citing categories and the columns the cited categories. The values in each cell are the share of citations made by the citing category to the cited category. The first thing to notice is that the values on the main diagonal are the highest for each citing category. This is not surprising: technologies in an area are more likely to build on other technologies in the same area, and patent citations capture this (at least partially).[10] Second, technological areas that we may think as more similar cite each other at higher rates. For example patents in the "Chemical" category cite more often those in "Drugs & Medical" than those in any

[10] Another explanation is that examiners use the technological classification to find prior art to narrow the scope of patent claims, and they typically add citations from the same technological classes of the patent application they examine (Alcacer and Gittelman, 2006).

other category, and *vice versa*. Similarly, those in "Computers & Communications" cite more often those in "Electrical & Electronic" than those in other categories, and the other way around. In our empirical analysis, we will use a more disaggregated classification (the USPC 3-digit technological classes) and also take into account the timing of citations.

We also borrow a measure of patent-to-patent similarity from Younge and Kuhn (2016). They apply automated text analysis methods to compare the text of each pair of patents for the entire set of regular utility patents granted by the USPTO from 1976 through 2014, and assign them a similarity score. They compute this similarity score using a vector space model that uses as input the text of the technical description of the patents. They utilize a "bag of words" approach to characterize each patent as a weighted vector of words, in which weights are based on the frequency of terms within a patent and in the entire population of patents. Essentially, for a given patent, they give more weight to terms that are more frequent in its technical description and less weight to those that are more common in the entire population of patents. To measure the similarity score, they compute the cosine of the angular separation between the weighted vectors for each pair of patents. This measure of similarity is bounded between zero and one, and high values suggest two patents are very similar to each other. We use this measure to quantify the similarity between citing and cited patents in our data.[11]

We report the average text similarity by NBER category of the citations made by all the SEPs in our data and the potential control patents with the same filing year, grant year and USPC class in table 4.3.[12] The rows represent the citing categories

[11] Younge and Kuhn (2016) provide a detailed description of the construction of the measure and its validation. We thank Jeffrey Kuhn and Kenneth Younge for sharing their data on this measure with us.

[12] We restrict the analysis to this sample because we do not have data for the entire population of patents.

and the columns the cited categories. As expected, the main diagonal has the highest values, i.e. the text similarity between citing and cited patents is highest for patents in the same NBER category.

The next section describes the econometric models we use in our analysis.

4.3.3 Econometric models

Suppose we observe a sample of technologies at risk of endorsement by an SSO and that our outcome of interest measures the direction of inventive activity that builds upon them. In an experimental setting, we would randomly assign the endorsement of SSOs to a subsample of technologies, comparing them to a subsample of control technologies. The comparison in mean outcomes between "treated" and "controls"would provide an estimate of the average treatment effect of SSO endorsement on the direction of cumulative inventive activity building upon the technologies.

We cannot run this experiment. Instead, we use observational data to estimate two sets of econometric models that measure the relationship between SSO endorsement and the direction of inventive activity.

The first set of models estimates the balance between deepening and broadening. We use a sample of citations to compare the distance in technological space between citing and cited patents for declared SEPs and controls patents. We use this sample to estimate models based on

$$E[Y_{cit}|X_{cit}] = exp(Disclosed_{cit}\alpha + SEP_{ci}\beta + X_{cit}\gamma) \tag{4.1}$$

where $Disclosed_{cit}$ is a binary indicator equal to one for citations c to patent i that occur at year t greater or equal to the year of SEP disclosure and SEP_{ci} is a binary variable equal to one for the citations to the SEPs. These are the two main explanatory variables in the empirical analysis and capture the marginal and selection "effects" discussed above.

We use two outcomes Y_{cit}. The first outcome is our measure of technology similarity based on interclass citation flows, CWTS. For each citation c, we compute the CWTS between citing and cited patent using the citations in year t.[13] The second outcome is the text similarity measure developed by Younge and Kuhn (2016). For an easier interpretation of the results, we multiple CWTS and the text similarity measure by 100, so that they are bounded between 0 and 100.

A major challenge to estimate the effect of SSOs on our outcomes is that SSOs do not pick technologies randomly. The main threat is that it is hard to observe the economic and technical value of a technology, which is likely an important factor in the selection mechanisms. The direction of the possible bias is hard to establish *ex ante*, as one can imagine different scenarios under which valuable technologies contribute more to deepening or broadening. To reduce the endogeneity concerns related to the non-random nature of technology selection, first we construct a matched sample of SEPs and control patents that have similar citation trends. In particular, for each SEP in our sample we identify a control patent with the same application year, grant year, in the same 3-digit USPC class and with the same number of citations at the age of SEP disclosure minus one.[14] We discard unmatched SEPs and controls. When multiple matches are available for a single SEP, we randomly pick one control patent. Our final sample contains all the citations to these matched patents.

We also use a large set of control variables to reduce the threats related to omitted variables. Depending on the specification, X_{cit} controls for the filing year and the USPC technological class of the cited patent, whether the application of the cited patent is a serialized continuation (continuation application, continuation-in-part or

[13]As it is common practice in the literature that uses patent citations, we assign citations to years using the application year of the citing patent.

[14]Patent age is computed as the difference between grant year and calendar year as in Mehta et al. (2010).

divisional), whether it claims the benefits of a provisional application, a foreign application or a PCT application, whether the application was published and was filed by a "small entity" (i.e. an individual inventor, a nonprofit organization, a university or a small firm), the year of citation, the age of the cited patent in the year of citation, the number of backward patent and non-patent citations made by the cited patent, the number of its inventors, claims and words in the first independent claim. In some specifications we also include cited patent fixed effects and a dummy variable equal to one for patents that are litigated before or in the citation year.

To estimate the semi-elasticity of our measures of similarity between citing and cited patents with respect to our two main explanatory variables, we follow Silva and Tenreyro (2006) and estimate model 4.1 with Poisson regressions. Silva and Tenreyro (2006) show that under heteroskedasticity, the log-linear OLS models commonly utilized to estimate elasiticicities can be biased. We also estimate OLS models as robustness checks.[15]

The second set of models tries to distinguish deepening and broadening. To do that we use the matching strategy utilized to construct the citation sample, but make a panel dataset retaining also the patents without citations and aggregating the data at cited patent-year level. The empirical models in this part of the analysis are based on

$$E[Y_{it}|X_{it}] = exp(Disclosed_{it}\alpha + SEP_i\beta + X_{it}\gamma) \tag{4.2}$$

where Y_{it} is either the number of USPC classes that cite patent i in year t that already cited i in previous years (which we call "old classes"), or the number of USPC classes

[15]Our outcomes assume only non-negative values, and the distribution of the text-based measure of Younge and Kuhn (2016) includes many zeros. Elasticities estimated with log-linear models in this setting can be biased. We provide robustness checks in the appendix in which we use the natural logarithm of CWTS and the natural logarithm of one plus the text-based similarity measure as outcomes.

that cite patent i in year t that never cited i before t ("new classes"). We use the old classes to measure the deepening of inventive activity on the patent, and the new classes for broadening. $Disclosed_{it}$ is a binary variable equal to one for SEPs starting in the year of disclosure, and SEP_i is a time-invariant indicator equal to one for the SEPs. Both outcomes are nonnegative count variables, so we estimate 4.2 with Poisson models.

The coefficient α estimates the correlation between SEP disclosure and deepening or broadening. To reduce the threat of omitted variables in the pooled-cross-sectional version of model 4.2, we use the same set of control variables listed above (X_{it}). We also control for the number of classes that cite i before t, because the probability of receiving citations from classes that already cited in the past (never cited the patent before) increases (decreases) with the number of classes that cited before t. Conditional on patent fixed effects, if one is willing to assume that there are no unobservable time-varying factors related to SEP disclosure and the two trends in inventive activity, α estimates the effect of SEP disclosure on the outcomes. If companies observe factors that vary over time like changes in the technical importance of the SEPs that we cannot include in the model and may be positively related to both SSO endorsement and cumulative inventive activity, our estimates would be biased upwards.

4.3.4 Samples

The differences between SEPs in this database and other patents are widely documented in Bekkers et al. (2016) and chapter 3. In the last part of this section we provide a quick overview of our two estimation samples. Both samples are based on matched SEPs disclosed before year 2010 and their control patents and exclude data in years after 2009 to reduce truncation concerns in the citation data. The citation sample contains all the citations from regular utility patents to patents in the matched sample between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age. The sample for analysis contains 168,589 citations made between 1980 and 2009. Our two outcomes are defined for 167,707 (CWTS) and 164,930 (text-based similarity) citations respectively.[16] We report the summary statistics for this sample in table C1, and the distributions of citations by SSO and licensing terms in tables C2 and C3.

The patent sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009. We use citation data from the issue year of the patents until year 2009 (inclusive), dropping those that occur after the 20th year of age. The sample contains 6,198 patents (3,099 SEPs and 3,099 controls) for a total of 58,364 patent-year observations. Table C4 provides the summary statistics for the patents in this sample. Tables C5 and C6 report the distributions of patents by SSO and licensing terms.

In the next section we report the results of our econometric analysis.

4.4 Results

In this section we report the results of the empirical analysis. We start with the results of the analysis of the technological similarity between citing and cited patents on the citation sample and then describe the results of the analysis of citations from "old" and "new" technological classes on the patent sample.

4.4.1 Technological similarity between citing and cited patents

In the first set of models we analyze the technological similarity between citing and cited patents utilizing our measure of similarity based on interclass citation flows and the text-based measure of similarity developed by Younge and Kuhn (2016).

[16] Missing values are due to missing data for technological classes or in the data from Younge and Kuhn (2016).

Table 4.4 shows the results of a set of Poisson models that estimate the semi-elasticity of CWTS with respect to the indicators for SEP disclosure and SEP status. The unit of observation is a citation. The outcome is the CWTS between citing and cited patents. The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age.

In the first model we only include the indicators for marginal ("disclosed") and selection ("SEP") "effects". The coefficient of the selection dummy is positive and statistically significant at 1%, implying a CWTS 15% higher for SEPs compared with control patents. The coefficient of disclosure is also statistically significant at 1% but it is negative and implies a decrease by about 16% in CWTS after disclosure. The results of this very simple model suggest that SSOs select technologies that have a relatively narrower impact in technological space and then broaden their influence in more diverse technological areas. In other words, deepening dominates broadening before standardization, and broadening dominates deepening after standardization.

However, these coefficients become very small and imprecisely estimated when we add control variables in models (2) (calendar year effects, age effects, cited patent filing year effects and cited patent USPC class effects), (3) (controls for characteristics of the application of the cited patent), (4) (controls for the characteristics of the cited patent) and (5) (indicator for previous litigation of the cited patent). Model (6) discards the SEP indicator, adds cited patent effects and replaces the age effects with the nonlinear terms of a 4th degree polynomial of patent age.[17] The coefficient of the disclosure indicator is positive but very small and imprecisely estimated.[18]

[17]It is not possible to include cited patent effects, calendar year effects and age effects in the same model because of collinearity between these sets of variables (Mehta et al., 2010).

[18]Table C7 tests the robustness of these results utilizing OLS models. The outcome for these models is the natural logarithm of CWTS. The results of the simplest model confirm those reported in the main text, with coefficients that are bigger in magnitude. The results of the other models show a positive, greater and statistically significant coefficient for the selection effect.

Table 4.5 estimates models similar to those in table 4.4 utilizing the text similarity measure between citing and cited patents as outcome. In the simple model in column (1), the coefficients of SEP disclosure and SEP status are both statistically significant at 1% level; the former implies a 16% decrease in the similarity between citing and cited patents after disclosure, while the latter implies that citations to SEPs have a similarity 22% higher than citations to control patents. The inclusion of control variables in models (2), (3), (4) and (5) reduces the magnitude of these estimates but they are still statistically significant at 1%. The coefficient of the disclosure dummy in model (6) is still negative and statistically significant at 1% level, and implies a 3% decrease in the similarity of cited and citing patents after disclosure.[19] [20]

Bekkers et al. (2016) show that it is important to take into account differences across SSOs and licensing terms. In tables 4.6 and 4.7 we analyze the heterogeneity of selection and marginal effects of SSOs on the balance between deepening and broadening. The models are similar to those in tables 4.4 and 4.5, but drop the disclosure and SEP indicators and replace them with a set of interaction terms between them and a set of indicators for groups of SSOs.[21]

The results in tables 4.6 and 4.7 show that the selection effect is mainly driven

[19]Table C8 shows the results of similar OLS models whose outcome is the natural logarithm of one plus the text similarity measure. The results are similar and the coefficients are even greater in size.

[20]We also estimate the models in tables 4.4 and 4.5 excluding the citations made by the patent examiners. Citations made by the examiners may be less representative of the links between technologies we want to measure in this study. Also, Alcacer and Gittelman (2006) show that citations added by the examiners are more concentrated in terms of technological classes than those provided by the applicants, leading to possible biases in the analysis. The USPTO started to provided the information on examiner citations in year 2001, so we restrict the analysis to patents granted after year 2000 and drop the citations they receive from examiners. We report the results in tables C9 and C10. The pattern of results is consistent with the results reported in the main text, but the coefficients of interests are usually smaller in size and estimated less precisely. This may be due at least in part to the big drop in sample size.

[21]We group SSOs as in Bekkers et al. (2016). IEEE, ETSI and IETF have a relatively large number of patents, so we treat them as separate groups. We group together IEC, ISO and ITU (big international standard developing organizations, or BIG-I); CEN, CENELEC ANSI and the Broadband Forum (ANSI+); the other SSOs are assigned to the residual group OTHER.

by ETSI, which has the largest positive coefficient of the SEP status indicator in all models. The disclosure policies at ETSI, not allowing blanket disclosures, may lead companies to disclose a high number of relatively unimportant patents (Bekkers et al., 2016). The large number of ETSI SEPs thus improves the precision of the estimated coefficient, and the possible lower importance of the patents may explain their narrower influence in technological space.

Regarding the marginal effects, the results in column (1) of tables 4.6 and 4.7 show that almost all the SSO groups are related to dominance of broadening over deepening. However, they decrease in magnitude and statistical significance after we include control variables in models (2), (3), (4) and (5) and include patent fixed effects in the last model of each table. A partial exception is the coefficient of the "Other" group in table 4.7, but coefficients in table 4.6 do not confirm this result.

In tables 4.8 and 4.9 we analyze the heterogeneity of selection and marginal effects across licensing terms. The models are similar to those already described in this section, but include a set of licensing terms dummies and their interaction terms with the disclosure indicator.[22]

An interesting result of these models is that SEPs disclosed with specific licensing terms seem to drive both the deepening related to the selection effect and the broadening related to the marginal effect. The provision of specific licensing terms seems to be related to a narrow impact before standardization but to a broad impact after. While the coefficient of FREE is not statistically significant in some models, patents disclosed with royalty free commitments seem to receive citations from patents that are relatively more similar after standardization. The results for the other licensing terms are either unstable across specifications or smaller in magnitude.

[22]We group all the variations of FRAND terms in the dSEP database together. Also, we consider royalty free and non-assertion commitments as a single category. We also group together the records without licensing commitments and those whose information is missing or unclear. SEPs with specific licensing commitments (e.g. provision of a royalty rate) are the last group.

So far we estimated the balance between deepening and broadening. In the remainder of this section we try to separate these two different trends.

4.4.2 Separating deepening and broadening using citations from old and new classes

In this section we try to separate the trends in deepening and broadening related to standardization within SSOs. To do this we use two outcomes directly related to deepening and broadening. Specifically, we discard the citation sample we used in the first part of the empirical analysis and use a sample at patent-year level and construct two outcomes based on patent citations. First we focus on citations from USPC classes that already cited the focal patent before to measure deepening ("old" classes). Then, we use citations from USPC classes that have never cited the focal patent in the past as a measure of broadening ("new" classes).

The use of these two outcomes allows us to measure more directly deepening and broadening. However, we lose the benefits of our measures of similarity between citing and cited patents, that do not depend (or depend less) on the technological classification.

The sample for this analysis is a panel at patent-year level that contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009. Patent-year observations before the issue year and after the 20th year of age are excluded from estimation, as well as citations that occur after year 2009 to reduce truncation concerns in the outcomes. The first set of models uses as outcome the number of USPC classes that cite the focal patent in a given year that cited the patent in previous years. All regressions are estimated with Poisson models.

The first column of table 4.10 shows the results of a very simple model in which we use only the SEP status indicator and the SEP disclosure dummy as explanatory variables. The second model includes the cumulative number of USPC classes that cited the focal patent in previous years. The third model includes calendar year effects, age effects, filing year effects and USPC class effects. Model (4) adds control variables for the characteristics of the application of the cited patent, and model (5) adds control variables for the characteristics of the cited patent and an indicator equal to one for patents that were litigated in the past or in the current year. Finally, the last model utilizes cited patent fixed effects.

The disclosure dummy has a positive and statistically significant coefficient in the first model, but the coefficient changes sign when we control for the number of classes that cited the focal patent in the past. When we add control variables, the coefficient of disclosure becomes positive again and statistically significant in pooled-cross-sectional models (3)-(5). The coefficients of these models imply a 5%-7% increase in the number of classes that already cited the patent in the past that cite again the focal patent. In model (6) we discard the matched controls and estimate the conditional fixed-effects Poisson model using only the matched SEPs because patents that receive no citations are discarded from estimation, and these are more frequent in the matched control group. The coefficient of the disclosure dummy is still positive but it is smaller and not precisely estimated. This may be due to the decrease in sample size, but also to a positive correlation between unobserved time-invariant characteristics of the cited patents that are absorbed in the fixed effect, like technical or economic value of the underlying technology. We also estimate an unreported conditional fixed-effect Poisson model similar to the one in column (6), retaining the control patents, and the results are similar to those of the pooled cross-sectional models.[23] Taken together, these results provide weak evidence of a deepening effect of SSOs on technical progress on patented standardized technology. However, we are careful to provide a causal interpretation of the results, as SEP disclosure

[23]As a robustness check, we estimate these models excluding the citations made by the examiners. The results are less precise and the coefficients of interest are often statistically insignificant. See table C11.

and cumulative invention may be driven by factors that change over time and are unobservable to us.

A causal interpretation of the estimates described above is more credible if we observe that before disclosure the trends in the outcome for SEPs and control patents are parallel. We provide a graphical representation of the difference in these trends in figure 4·1.[24] The graph reports the coefficients of the interaction terms between a set of "years until/since disclosure" indicators and the SEP indicator from an OLS regression similar to the model estimated in column (6) of table 4.10 on the matched sample of SEPs and control patents and their 95% confidence intervals. We cannot reject that the trend in citations from old classes for SEPs and control patents is the same before disclosure, and there is an increase in the number of old classes citing the SEPs after disclosure relative to the controls, even if the coefficients are imprecisely estimated. While this may seem to support a causal interpretation of the results, we are cautious in our interpretation given the low precision of the estimates and the results of model (6) in table 4.10.

The selection dummy is not statistically significant in the simplest model in column (1), but it is positive and statistically significant at 1% level in all other models. The magnitude of the coefficient changes depending on the specification, but overall these results suggest that SEPs receive between 12% and 30% more citations from old classes than the control patents. We can conclude that SEPs are selected among those patents that are already important in a relatively narrow technological space.

Models in table 4.11 replace the disclosure and SEP indicators with a set of SSO group dummies and their interactions with the disclosure dummy. While the magnitude and sometimes the sign of the specific coefficients depend on the specification

[24]We match SEPs and control patents on the total number of citations before disclosure of the SEPs. So the trends in citations from old and new classes for SEPs and controls may be different depending on the composition of the total number of citations and how these are dispersed across classes.

used, the marginal effect observed in table 4.10 seems to be driven mostly by ANSI+, IEEE and SSOs in the OTHER category. Regarding the set of selection dummies, the most interesting set of results concerns the coefficients of the IETF dummy. In all models this is the biggest positive coefficient and it is statistically significant at 1%. This may suggest that IETF members are particularly effective at selecting very important technologies in their specific field. The coefficients for ETSI and the category OTHER are also positive and statistically significant in all models. This may be at odds with the view that many patents at ETSI have lower importance and are disclosed only because of the stringent disclosure policies of this SSO. However, we have to keep in mind that in this analysis we are measuring the number of classes citing the focal patent, and not the number of citing patents.

In table 4.12 we analyze the heterogeneity across licensing terms. FRAND and SPECIFIC terms are generally associated with an increase in the number of classes that cite repeatedly the focal patent after disclosure. Interestingly, patents disclosed without the provision of terms have a decrease in the number of classes that cite them again after disclosure, even if the magnitude of the coefficients and their statistical significance varies across models. The coefficients of the selection dummies are usually positive and statistically significant for all the licensing terms with the exception of those for the dummy for specific terms, which are estimated less precisely and even change sign with the inclusion of control variables.

To isolate the relationship between SEP status, SEP disclosure and broadening we now analyze the citations from "new" USPC classes, i.e. classes that cite the focal patent for the first time. Table 4.13 reports the results of models similar to those in table 4.10, utilizing the number of new classes citing the focal patent as outcome. While the coefficient of the disclosure indicator in models (1) and (2) is negative and statistically significant at 1%, with the inclusion of control variables in models (3), (4), (5) it becomes positive and statistically significant, implying an increase in the number of new classes citing the focal patent by almost 15% after disclosure. The result of the conditional fixed-effects Poisson model in column (6) confirm this result. The results of an unreported model similar to the one in column (6) that does not exclude the control patents are very similar to those reported. If one is willing to interpret this result causally, these estimates mean that SSOs broaden the impact of endorsed technologies as inventive inputs across technological areas.

The causal interpretation is more plausible if the pre-disclosure trends in the outcome for SEPs and control patents are parallel. We test this graphically in figure 4·2. In this figure we plot the same coefficients we show in table 4·1 and their 95% confidence intervals, but the model now has the number of new classes citing the focal patent in a given year as outcome. The SEPs seem to be on a negative trend compared with the controls before disclosure, but we cannot reject the null hypothesis that the trend in outcomes for SEPs and control patents is the same in the years just before disclosure. After disclosure, the SEPs receive more citations from new classes compared with the controls, suggesting a positive relationship between disclosure and broadening. Given these trends, if SEP disclosure is correlated with time-varying unobservables correlated with citations from new classes, this correlation seems to be negative, possibly leading to a downward bias in our estimated coefficient of disclosure.

The selection effect is also interesting. The coefficient of the SEP indicator in models (1) and (2) is positive, but becomes negative when we include the control variables in models (3)-(5), implying that SEPs receive 11% fewer citations from new classes, *ceteris paribus*.[25]

[25] We also estimate these models excluding citations made by the examiners. Table C12 shows the results. The estimates for models (1) and (2) are consistent with the first two models in table 4.13, but the coefficients of models (3), (4) and (5) are small with very large standard errors. However, the coefficient of disclosure in the conditional fixed-effects Poisson model in column (6) is very large and statistically significant at 1% level, implying an increase in citations from new classes by 73%. An unreported model similar to the one in column (6) that does not discard the control group has

In table 4.14 we analyze the heterogeneity across groups of SSOs. The selection effect seems to be driven mostly by ETSI. The ETSI indicator has the largest negative and statistically significant coefficient among the selection dummies in models (3)-(5).

Interestingly, the conditional fixed-effect Poisson model in column (6) shows that the coefficients of all the interaction terms between the disclosure dummy and the SSO group indicators are positive, statistically significant and large in magnitude, with a notable exception: ETSI. The coefficient for ETSI is negative and statistically significant at 1% level. It implies an almost 19% decrease in the number of new classes citing the focal patent after disclosure of the patent to ETSI.

In table 4.14 we analyze the heterogeneity across licensing terms. The most robust result regarding the selection effects is the negative correlation between the FRAND terms indicator and the outcome. FRAND terms seem to drive the negative selection effect we estimated before. Regarding the heterogeneity of the marginal effects, these are generally positive when we include our control variables in models (3)-(5), even if some of them have relatively large standard errors. When we include patent fixed effects, all the coefficients are positive and statistically significant at least at 5% level. Interestingly, the coefficient for FRAND disclosures is the smallest one, suggesting that the vagueness of FRAND terms may lead to a lower broadening effect compared with other terms.

4.5 Conclusions

In this paper we use patent data to analyze the direction of inventive activity that builds upon technologies endorsed by SSOs, distinguishing between deepening and broadening. We use SEPs to identify technologies selected by SSOs and compare them with similar patents having the same citation trend. We use two measures of

a coefficient of disclosure equal to 0.26, with a standard error equal to 0.15.

similarity between citing and cited patents to estimate the balance between deepening and broadening, and citations from classes that cited the patents in the past or never cited the patent before to separate these two different types of technical progress. The results show that SSOs select technologies that are important in a relatively narrow technological area, and their adoption as input for following inventive activity broadens after standardization.

The analysis of the heterogeneity across SSOs shows that cumulative invention on ETSI patents is narrower than the inventive activity upon patents selected by other SSOs. As suggested by Bekkers et al. (2016), the policies of this SSO may lead to the disclosure of patents with lower importance, with a narrower influence in technological space. Patents disclosed with FRAND commitments also have narrower cumulative inventive improvements compared with those with other types of commitments. The uncertainty related to FRAND terms may reduce the willingness of companies in diverse technological areas to experiment on the standardized technology. This interpretation is reinforced by the results for patents disclosed with specific terms, which exhibit high rates of deepening and especially broadening. Certainty about the cost of access seems to be positively related to inventive activity, and especially with experimentation in new areas.

These results are relevant for SSO policies and for the strategies of companies participating in standard setting. The differences in the direction of inventive activity related to rules and commitments should be taken into account in the evaluation of the trade-offs involved in the design of policies and strategies.

This study has several limitations. SEPs and patent citations may be noisy measures of what we want to measure. Further research may try to find other, more direct ways to measure what technologies are standardized and the cumulative inventive activity upon them. Another dimension interesting to explore would be the identity of the inventors that build upon standards. Also, while we try to compare SEPs with similar patents having similar citation trends, much more could done to move towards estimates with a causal interpretation. Another interesting direction for further research is the comparison of the different implications for cumulative inventive activities between SSOs and alternative ways to organize standardization. We leave these extensions for future work.

Tables and Figures

Table 4.1: SSOs, Deepening and Broadening: selection and marginal effects

	Deepening	**Broadening**
Selection	High-quality technologies Compatibility *within* a tech. area	General technologies Compatibility *across* tech. areas
Marginal	Cumulative innovation Direct network effects Expectations Overcoming inertia Technological forecasting Lower risk of hold-up Improvements *within* a tech. area	Cumulative innovation Indirect network effects Lower cost of access Credible commitments Specialization Unexpected new product varieties Improvements *across* tech. areas

Table 4.2: Citation Weighted Technological Similarity between NBER categories.

	Cited category					
Citing category	Chemical	Comp. & Comm.	Drugs & Medical	Electrical & Electronic	Mechanical	Others
Chemical	0.68	0.02	0.10	0.06	0.06	0.09
Computers & Communications	0.01	0.82	0.01	0.09	0.04	0.02
Drugs & Medical	0.09	0.02	0.81	0.03	0.03	0.03
Electrical & Electronic	0.05	0.11	0.02	0.72	0.06	0.04
Mechanical	0.05	0.08	0.04	0.08	0.65	0.10
Others	0.08	0.03	0.02	0.04	0.09	0.73

We compute the Citation Weighted Technological Similarity between NBER categories for illustrative purposes. NBER categories defined as in Hall et al. (2001). The Citation Weighted Technological Similarity is computed using all the citations made by granted regular utility patents in Public PAIR to regular utility patents that we can match to the NBER category data from PatentsView. The rows of the table represent the citing category and the columns the cited category.

Table 4.3: Text similarity of citations between NBER categories.

Citing category	Cited category: Chemical	Comp. & Comm.	Drugs & Medical	Electrical & Electronic	Mechanical	Others
Chemical	30.06	16.27	24.42	23.88	21.66	27.04
Computers & Communications	15.62	23.60	16.10	22.34	20.68	15.95
Drugs & Medical	18.52	11.52	26.67	16.30	14.92	11.54
Electrical & Electronic	23.76	22.83	18.75	31.95	23.45	23.39
Mechanical	21.30	20.63	24.90	21.15	31.70	23.83
Others	24.15	15.89	19.42	18.76	23.53	33.54

Each cell reports the mean text similarity (Younge and Kuhn, 2016) for the citations between NBER categories (Hall et al., 2001) in the sample of citations to SEPs and patents in the same filing year-grant year-USPC class strata. The sample contains 9,739,808 citations. The rows of the table represent the citing category and the columns the cited category.

Table 4.4: "Marginal/selection effects" of SSOs on Citation Weighted Technological Similarity between cited and citing patents: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	CWTS	CWTS	CWTS	CWTS	CWTS	CWTS
Disclosed	-0.18***	0.01	0.01	0.01	0.00	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
SEP	0.14***	0.02	0.03	0.02	0.02	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	167,707	167,707	167,387	167,363	167,363	167,179
# of cited patents	6,468	6,468	6,463	6,461	6,461	5,940

The unit of observation is a citation. All regressions are estimated with Poisson models. The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age. The outcome in all models is the Citation Weighted Technological Similarity between cited and citing patents. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent citations plus one, number of non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 4.5: "Marginal/selection effects" of SSOs on text similarity between cited and citing patents: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.
Disclosed	-0.18***	-0.06***	-0.06***	-0.06***	-0.06***	-0.03**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
SEP	0.19***	0.12***	0.12***	0.12***	0.12***	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	164,930	164,930	164,610	163,789	163,789	164,095
# of cited patents	6,478	6,478	6,473	6,445	6,445	5,887

The unit of observation is a citation. The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age. The outcome in all models is the text similarity between cited and citing patents (Younge and Kuhn, 2016). All regressions estimated with Poisson models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent citations plus one, number of non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.6: Heterogeneity of "marginal/selection effects" of SSOs on Citation Weighted Technological Similarity between cited and citing patents by SSO group: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	CWTS	CWTS	CWTS	CWTS	CWTS	CWTS
ANSI+*disclosed	-0.25***	-0.07	-0.07	-0.07	-0.07	-0.03
	(0.09)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
BIG-I*disclosed	-0.14***	0.10**	0.10**	0.10**	0.09*	0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
ETSI*disclosed	-0.14***	0.02	0.03	0.02	0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
IEEE*disclosed	-0.18***	-0.01	-0.00	-0.00	-0.01	0.04
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
IETF*disclosed	-0.15*	0.06	0.05	0.05	0.05	0.09***
	(0.07)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
OTHER*disclosed	0.02	0.03	0.02	0.01	0.00	-0.03
	(0.08)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
ANSI+	0.01	-0.05	-0.05	-0.05	-0.05	
	(0.09)	(0.06)	(0.06)	(0.06)	(0.06)	
BIG-I	0.02	-0.07*	-0.08*	-0.10**	-0.09**	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	
ETSI	0.22***	0.06***	0.07***	0.06***	0.06***	
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	
IEEE	0.14***	0.02	0.02	0.01	0.00	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	
IETF	0.00	-0.02	-0.02	-0.04	-0.04	
	(0.08)	(0.04)	(0.04)	(0.04)	(0.04)	
OTHER	-0.16**	-0.05	-0.05	-0.05	-0.05	
	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	167,707	167,707	167,387	167,363	167,363	167,179
# of cited patents	6,468	6,468	6,463	6,461	6,461	5,940

The unit of observation is a citation. The outcome in all models is the Citation Weighted Technological Similarity between cited and citing patents. All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.4. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.7: Heterogeneity of "marginal/selection effects" of SSOs on text similarity between cited and citing patents by SSO group: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.
ANSI+*disclosed	-0.20***	-0.04	-0.04	-0.05	-0.05	-0.02
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
BIG-I*disclosed	-0.05	0.07**	0.08**	0.07**	0.08**	0.00
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
ETSI*disclosed	-0.14***	-0.03*	-0.03*	-0.04**	-0.04**	-0.03*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
IEEE*disclosed	-0.17***	-0.04	-0.04	-0.03	-0.03	-0.00
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
IETF*disclosed	-0.04	0.05	0.04	0.05	0.05	0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
OTHER*disclosed	-0.34***	-0.26***	-0.27***	-0.28***	-0.28***	-0.24***
	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.04)
ANSI+	0.07	0.01	0.00	0.02	0.02	
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
BIG-I	0.07	-0.04	-0.05	-0.05	-0.05	
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	
ETSI	0.26***	0.19***	0.19***	0.19***	0.19***	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
IEEE	0.15***	0.06*	0.06*	0.06*	0.06*	
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	
IETF	0.10*	0.08	0.08	0.08	0.08	
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	
OTHER	0.07	0.05	0.05	0.05	0.05	
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	164,930	164,930	164,610	163,789	163,789	164,095
# of cited patents	6,478	6,478	6,473	6445	6445	5,887

The unit of observation is a citation. The outcome in all models is the text similarity between cited and citing patents (Younge and Kuhn, 2016). All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.5. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.8: Heterogeneity of "marginal/selection effects" of SSOs on Citation Weighted Technological Similarity of cited and citing patents by licensing terms: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	CWTS	CWTS	CWTS	CWTS	CWTS	CWTS
FRAND*disclosed	-0.17***	0.01	0.01	0.01	0.00	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
FREE*disclosed	0.06	0.16*	0.12	0.13	0.12	0.13**
	(0.12)	(0.09)	(0.08)	(0.08)	(0.08)	(0.05)
NONE*disclosed	-0.18*	0.12	0.15*	0.16*	0.15	0.08*
	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.05)
SPECIFIC*disclosed	-0.53***	-0.32***	-0.32***	-0.34***	-0.34***	-0.22**
	(0.14)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
FRAND	0.15***	0.03*	0.03*	0.03	0.03	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
FREE	-0.10	-0.05	-0.02	-0.04	-0.03	
	(0.12)	(0.09)	(0.08)	(0.08)	(0.08)	
NONE	-0.09	-0.25***	-0.27***	-0.28***	-0.28***	
	(0.10)	(0.07)	(0.08)	(0.08)	(0.08)	
SPECIFIC	0.22*	0.20**	0.20**	0.22**	0.22**	
	(0.13)	(0.09)	(0.09)	(0.09)	(0.09)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	167,707	167,707	167,387	167,363	167,363	167,179
# of cited patents	6,468	6,468	6,463	6,461	6,461	5,940

The unit of observation is a citation. The outcome in all models is the Citation Weighted Technological Similarity between cited and citing patents. All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.4. Robust standard errors in parentheses, clustered at cited patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 4.9: Heterogeneity of "marginal/selection effects" of SSOs on text similarity between cited and citing patents by licensing terms: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.
FRAND*disclosed	-0.18***	-0.07***	-0.07***	-0.07***	-0.07***	-0.03**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
FREE*disclosed	0.15*	0.20***	0.18**	0.19**	0.19**	0.02
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.06)
NONE*disclosed	-0.28***	-0.04	-0.02	-0.01	-0.01	0.03
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)
SPECIFIC*disclosed	-0.30***	-0.15	-0.16	-0.23**	-0.23**	-0.15
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)
FRAND	0.20***	0.13***	0.12***	0.13***	0.13***	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
FREE	0.05	0.04	0.05	0.05	0.05	
	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	
NONE	0.13**	-0.01	-0.02	-0.03	-0.03	
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	
SPECIFIC	0.16*	0.19**	0.20**	0.27***	0.27***	
	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	164,930	164,930	164,610	163,789	163,789	164,095
# of cited patents	6,478	6,478	6,473	6445	6445	5,887

The unit of observation is a citation. The outcome in all models is the text similarity between cited and citing patents (Younge and Kuhn, 2016). All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.5. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.10: "Marginal/selection effects" of SSOs on citations from old USPC classes (deepening): Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
Disclosed	0.10***	-0.16***	0.07***	0.05***	0.05**	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
SEP	0.04	0.26***	0.12***	0.13***	0.11***	
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	58,364	58,364	58,364	58,364	57,556	27,134
# of cited patents	6,198	6,198	6,198	6,198	6,156	2,737

The unit of observation is a patent-year. The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009. Patent-year observations before the issue year and after the 20th year of age are excluded from estimation, as well as citations that occur after year 2009. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that cited the patent in previous years. "Past classes" is the number of USPC classes that cited the focal patent before a given year. All regressions estimated with Poisson models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent citations plus one, number of non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level, with the exception of model (6), which has regular standard errors. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 4.11: Heterogeneity of "marginal/selection effects" of SSOs on citations from old USPC classes (deepening) by SSO group: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
ANSI+*disclosed	0.62***	0.12	0.24***	0.24***	0.25***	0.13*
	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)
BIG-I*disclosed	0.09	-0.19***	0.04	0.02	-0.00	0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
ETSI*disclosed	-0.08***	-0.21***	0.01	0.01	0.00	-0.01
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
IEEE*disclosed	0.35***	-0.04	0.19***	0.18***	0.20***	0.09**
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
IETF*disclosed	0.02	-0.23***	0.03	-0.01	0.01	-0.02
	(0.08)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)
OTHER*disclosed	0.37***	-0.04	0.13**	0.14**	0.13**	0.20***
	(0.09)	(0.07)	(0.06)	(0.07)	(0.07)	(0.06)
ANSI+	-0.39***	-0.06	-0.03	-0.06	-0.10	
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	
BIG-I	-0.16***	0.04	0.03	0.05	0.04	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	
ETSI	0.08***	0.33***	0.14***	0.16***	0.15***	
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	
IEEE	-0.05	0.17***	0.04	0.02	0.00	
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	
IETF	0.53***	0.57***	0.32***	0.32***	0.28***	
	(0.09)	(0.05)	(0.05)	(0.05)	(0.05)	
OTHER	0.23***	0.31***	0.14***	0.11**	0.11**	
	(0.08)	(0.06)	(0.05)	(0.05)	(0.05)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	58,364	58,364	58,364	58,364	57,556	27,134
# of cited patents	6,198	6,198	6,198	6,198	6,156	2,737

The unit of observation is a patent-year. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that cited the patent in previous years. All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.10. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.12: Heterogeneity of "marginal/selection effects" of SSOs on citations from old USPC classes (deepening) by licensing terms: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
FRAND*disclosed	0.09***	-0.15***	0.07***	0.06***	0.05***	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
FREE*disclosed	-0.11	-0.26*	-0.03	-0.05	-0.06	-0.16
	(0.17)	(0.13)	(0.10)	(0.10)	(0.09)	(0.10)
NONE*disclosed	0.02	-0.42***	-0.15*	-0.17**	-0.16*	-0.02
	(0.12)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
SPECIFIC*disclosed	0.77***	0.12	0.18*	0.19*	0.30***	0.23*
	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)
FRAND	0.03	0.26***	0.11***	0.12***	0.10***	
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	
FREE	0.49**	0.48***	0.33***	0.29***	0.28***	
	(0.20)	(0.13)	(0.09)	(0.10)	(0.09)	
NONE	0.19	0.36***	0.25***	0.26***	0.24***	
	(0.12)	(0.08)	(0.07)	(0.07)	(0.06)	
SPECIFIC	-0.31***	0.06	0.21**	0.13	0.03	
	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	58,364	58,364	58,364	58,364	57,556	27,134
# of cited patents	6,198	6,198	6,198	6,198	6,156	2,737

The unit of observation is a patent-year. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that cited the patent in previous years. All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.10. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.13: "Marginal/selection effects" of SSOs on citations from new USPC classes (broadening): Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	New Classes	New Classes	New Classes	New Classes	New Classes	New Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
Disclosed	-0.49***	-0.54***	0.14***	0.13***	0.12***	0.15***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
SEP	0.26***	0.30***	-0.11***	-0.11***	-0.12***	
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	58,364	58,364	58,364	58,364	57,556	25,687
# of cited patents	6,198	6,198	6,198	6,198	6,156	2,499

The unit of observation is a patent-year. The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009. Patent-year observations before the issue year and after the 20th year of age are excluded from estimation, as well as citations that occur after year 2009. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that never cited the patent before. "Past classes" is the number of USPC classes that cited the focal patent before a given year. All regressions estimated with Poisson models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent citations plus one, number of non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 4.14: Heterogeneity of "marginal/selection effects" of SSOs on citations from new USPC classes (broadening) by SSO group: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	New Classes	New Classes	New Classes	New Classes	New Classes	New Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
ANSI+*disclosed	-0.30***	-0.38***	0.19**	0.17**	0.18**	0.49***
	(0.09)	(0.09)	(0.08)	(0.08)	(0.09)	(0.14)
BIG-I*disclosed	-0.52***	-0.56***	-0.00	-0.02	-0.04	0.30***
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.09)
ETSI*disclosed	-0.70***	-0.72***	0.05	0.05	0.05	-0.21***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)
IEEE*disclosed	-0.39***	-0.46***	0.20***	0.20***	0.21***	0.21***
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
IETF*disclosed	-0.56***	-0.59***	0.05	0.05	0.06	0.49***
	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.12)
OTHER*disclosed	-0.21**	-0.28***	0.27***	0.28***	0.27***	0.70***
	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.12)
ANSI+	0.47***	0.52***	0.04	0.01	-0.03	
	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)	
BIG-I	0.38***	0.41***	0.07*	0.10**	0.09**	
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	
ETSI	0.08**	0.12***	-0.27***	-0.28***	-0.28***	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
IEEE	0.40***	0.43***	-0.05	-0.08	-0.10*	
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	
IETF	0.66***	0.66***	0.09	0.07	0.07	
	(0.08)	(0.08)	(0.07)	(0.06)	(0.07)	
OTHER	0.47***	0.48***	0.05	0.03	0.04	
	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	58,364	58,364	58,364	58,364	57,556	25,687
# of cited patents	6,198	6,198	6,198	6,198	6,156	2,499

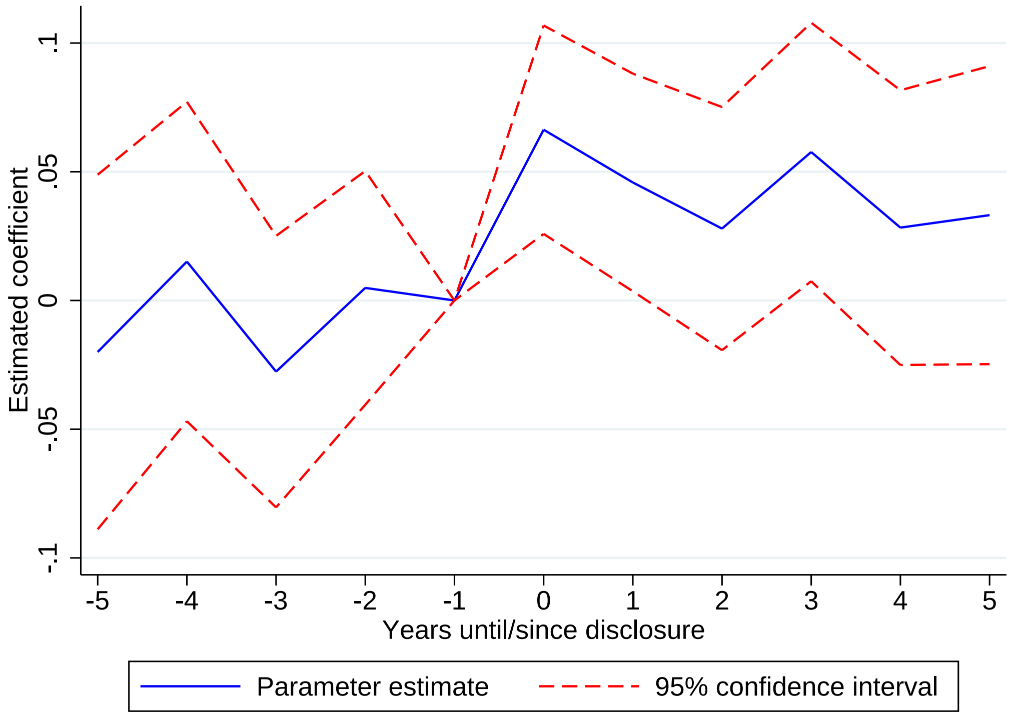
The unit of observation is a patent-year. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that never cited the patent before. All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.13. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table 4.15: Heterogeneity of "marginal/selection effects" of SSOs on citations from new USPC classes (broadening) by licensing terms: Poisson models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	New Classes	New Classes	New Classes	New Classes	New Classes	New Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
FRAND*disclosed	-0.52***	-0.57***	0.12***	0.11***	0.11***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
FREE*disclosed	-0.30	-0.33*	0.21	0.21	0.18	0.40**
	(0.19)	(0.18)	(0.15)	(0.15)	(0.15)	(0.17)
NONE*disclosed	-0.49***	-0.57***	0.07	0.05	0.02	0.45***
	(0.13)	(0.12)	(0.11)	(0.11)	(0.11)	(0.17)
SPECIFIC*disclosed	-0.01	-0.15	0.39***	0.41***	0.47***	0.48**
	(0.16)	(0.16)	(0.14)	(0.13)	(0.15)	(0.21)
FRAND	0.24***	0.28***	-0.12***	-0.13***	-0.13***	
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	
FREE	0.44**	0.43***	-0.05	-0.11	-0.11	
	(0.17)	(0.16)	(0.12)	(0.12)	(0.13)	
NONE	0.63***	0.66***	0.18**	0.15	0.16*	
	(0.11)	(0.10)	(0.09)	(0.09)	(0.09)	
SPECIFIC	0.34**	0.41***	-0.11	-0.17	-0.24*	
	(0.13)	(0.13)	(0.11)	(0.11)	(0.13)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	58,364	58,364	58,364	58,364	57,556	25,687
# of cited patents	6,198	6,198	6,198	6,198	6,156	2,499

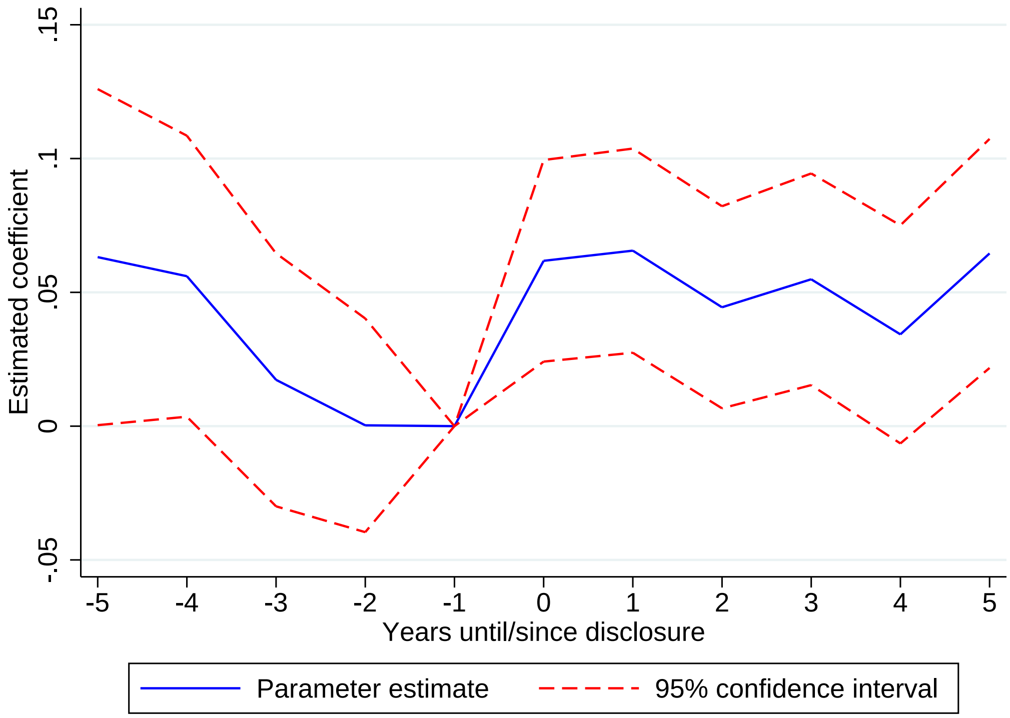
The unit of observation is a patent-year. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that never cited the patent before. All regressions estimated with Poisson models. Sample and control variables are described in the main text and in the note to table 4.13. Robust standard errors in parentheses, clustered at cited patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Figure 4·1: Pre- and post-disclosure trend in citations from old USPC classes (deepening).



This figure plots the coefficients of the interaction terms between SEP disclosure (equal to one from the year of disclosure for the SEPs, zero for the SEPs before disclosure and the matched control patents) and a set of "year until/since" disclosure indicator variables (blue solid line), together with their 95% confidence intervals (red dashed lines). The omitted category for this set of interaction terms is the interaction for the year before disclosure (year -1 in the graph). The model is estimated with an OLS regression with patent fixed effects in which the unit of observation is a patent-year and the outcome is the number of USPC classes that cite the focal patent in a given year that cited the patent in previous years. The model also includes a set of calendar year indicators, the non-linear terms of a fourth degree polynomial of age since the issue year, the number of USPC classes that cited the focal patent before a given year and an indicator equal to one from the year the focal patent is litigated for the first time in a U.S. district court. Standard errors are clustered at patent level. The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009. Patent-year observations before the issue year and after the 20th year of age are excluded from estimation, as well as citations that occur after year 2009.

Figure 4·2: Pre- and post-disclosure trend in citations from new USPC classes (broadening).



This figure plots the coefficients of the interaction terms between SEP disclosure (equal to one from the year of disclosure for the SEPs, zero for the SEPs before disclosure and the matched control patents) and a set of "year until/since" disclosure indicator variables (blue solid line), together with their 95% confidence intervals (red dashed lines). The omitted category for this set of interaction terms is the interaction for the year before disclosure (year -1 in the graph). The model is estimated with an OLS regression with patent fixed effects in which the unit of observation is a patent-year and the outcome is the number of USPC classes that cite the focal patent in a given year that never cited the patent in previous years. The model also includes a set of calendar year indicators, the non-linear terms of a fourth degree polynomial of age since the issue year, the number of USPC classes that cited the focal patent before a given year and an indicator equal to one from the year the focal patent is litigated for the first time in a U.S. district court. Standard errors are clustered at patent level. The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009. Patent-year observations before the issue year and after the 20th year of age are excluded from estimation, as well as citations that occur after year 2009.

Chapter 5

Conclusions

This dissertation studies a key stage of the production of patents, i.e. the assignment of applications to examiners, in chapter 2, the strategic behavior of applicants for SEPs in chapter 3 and uses SEPs to study the influence of SSOs on the direction of cumulative inventive activity in chapter 4.

Chapter 2 studies the matching of patent applications to examiners at the USPTO. Using statistical tests originally developed to study industry agglomeration, the analysis finds strong evidence that examiners specialize in particular technologies, even within relatively homogeneous art units. Examiner specialization is more pronounced in the biotechnology and chemistry fields, and less in computers and software. Evidence of specialization becomes weaker, but does not completely disappear, conditioning on technology sub-classes. There is no evidence that certain examiners specialize in applications that have greater importance or broader claims. Finally, the study shows that more specialized examiners have a lower grant rate and produce a larger narrowing of claim-scope during the examination process.

To measure the technological specialization we utilize the technological classes and subclass assigned to applications by the USPTO. However, technological classifications can only partially capture real technological differences across patent applications. The analysis could be extended applying recent developments in automated text analysis to measure technological similarities using the text of patent applications. This extension may help to capture technological differences currently unobserved. Another interesting area for further research would be the study of the implications of technological specialization for the outcomes of the examination process. The analysis we provided in chapter 2 is only a first attempt to understand the consequences of technological specialization. It is important to understand the implications of specialization not only for examination outcomes, but also for post-grant outcomes such as cumulative inventive activity and patent litigation.

Chapter 3 studies the strategic behavior of applicants for SEPs at the USPTO. I compare SEPs with similar applications and find that applicants for SEPs use more often mechanisms provided by the U.S. patent system to delay issuance. This leads to longer lags between priority and issuance. Companies that rely more on patents to generate revenues exploit these mechanisms more aggressively. I also analyze how the scope of the claims changes over time and find that owners of SEPs may delay issuance to obtain the right balance between patent breadth and strength. Furthermore, I find that the probability of issuance increases significantly after SEP disclosure and that the latter is positively correlated with the filing of continuation applications. These results are consistent with the idea that companies prolong the prosecution of their SEPs until the standard is set, possibly to cover the standard with additional claims. Finally, I find that a 10% increase in the lag between priority and issuance of SEPs is correlated with an increase in the probability of litigation by almost 1%. This suggests that owners of SEPs may delay issuance to obtain patents that are more valuable, or that longer lags are associated with failures in licensing negotiations.

The ideas proposed in this chapter could be utilized to understand the strategic behavior of patent applicants more generally. While SEPs are an important subsample of patents, there are other areas in which applicants may exploit the patent system strategically. Further research may generalize the analysis in chapter 3, analyzing how applicants react to the introduction of new products by competitors and exploring the consequences in terms of welfare of strategic behavior in prosecution. Another interesting topic for further research could be combining the insights of chapters 2 and 3 and try to model the interaction between examiners and applicants more formally, and explore the implications in terms of welfare of the outcomes of this interaction.

In chapter 4 we study the relationship between technology endorsement by SSOs and the the direction of inventive activity. We introduce the concepts of "deepening" and "broadening" and relate them to the activities of SSOs. In the empirical analysis, we exploit the disclosure of SEPs as a window on standardization within SSOs and compare the dispersion of patent citation flows to SEPs and similar patents. In the first part of the analysis we use a measure of patent-to-patent text similarity and a new measure that takes into account the probability of inter-class citations to estimate the balance between deepening and broadening. In the second part of the analysis we separate deepening and broadening using citations from technological classes that repeatedly cite a patent as a measure of deepening and citation from new classes as a measure of broadening. The results provide evidence that both trends are occurring. The overall pattern of results suggests that SSOs select technologies that are important in a relatively narrow technological area, and their adoption as input for following inventive activity broadens after standardization. We also explore the heterogeneity across SSOs and licensing terms and find substantial differences.

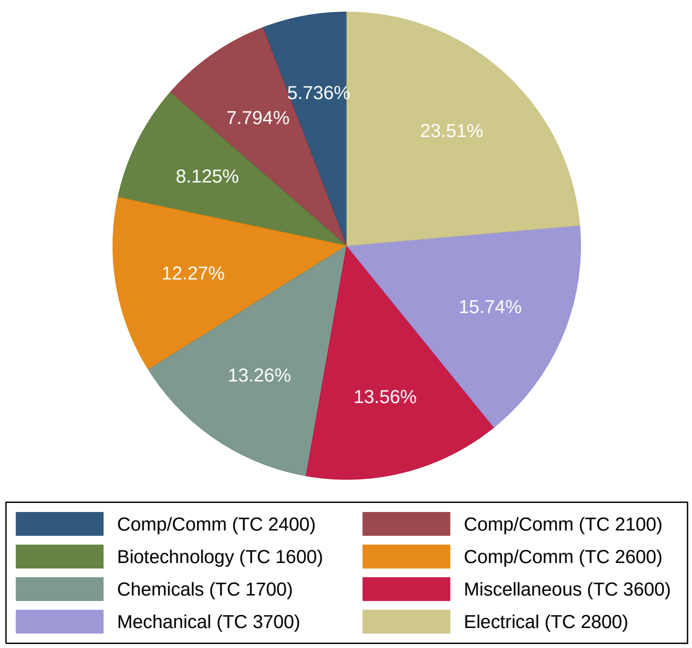
Further research in this area may study more explicitly the implications of different SSO policies on standard adoption and cumulative inventive activity. Another interesting area for further research would be the comparison of the performance of different ways to organize standardization in terms of adoption and inventive activity. More generally, further research may explore how other ways to respond to the hold-up problem affect the rate and direction of inventive activity.

I see this dissertation as my first attempt to contribute to our understanding of the patent system, the related institutions and company strategies. I look forward to further develop the work included in this dissertation, the related ideas described here and other projects that hopefully will increase our knowledge of the implications of the patent system for innovation and growth.

Appendix A

Appendix to Chapter 2

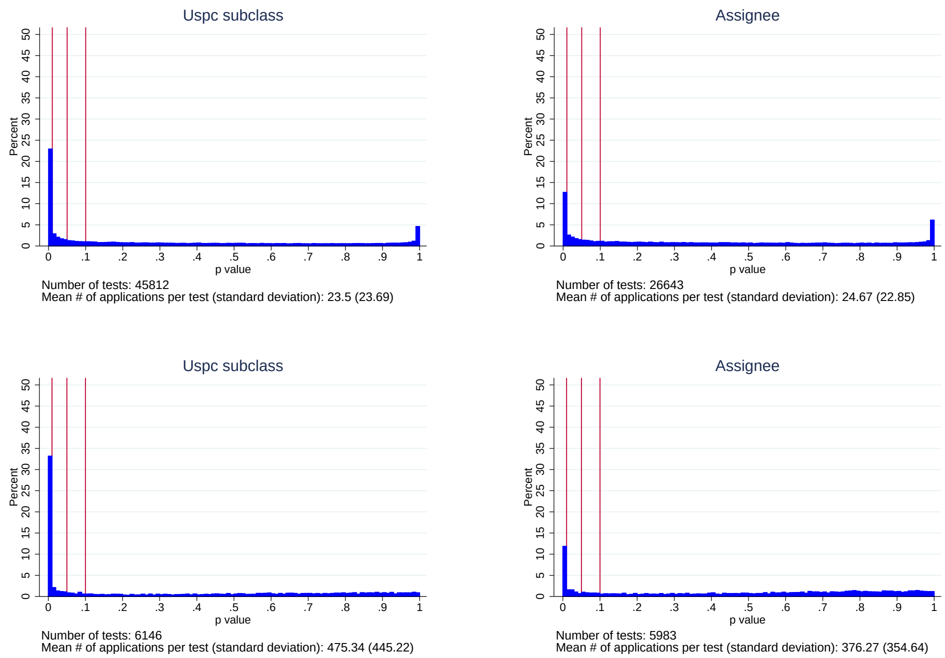
Figure A1: Distribution of applications by Technology Center.



The labels in the legend of the chart correspond to the type of technologies examined by each technology center as reported in Graham et al. (2015). The full names of the eight technology centers currently responsible for the examination of utility patent applications are:

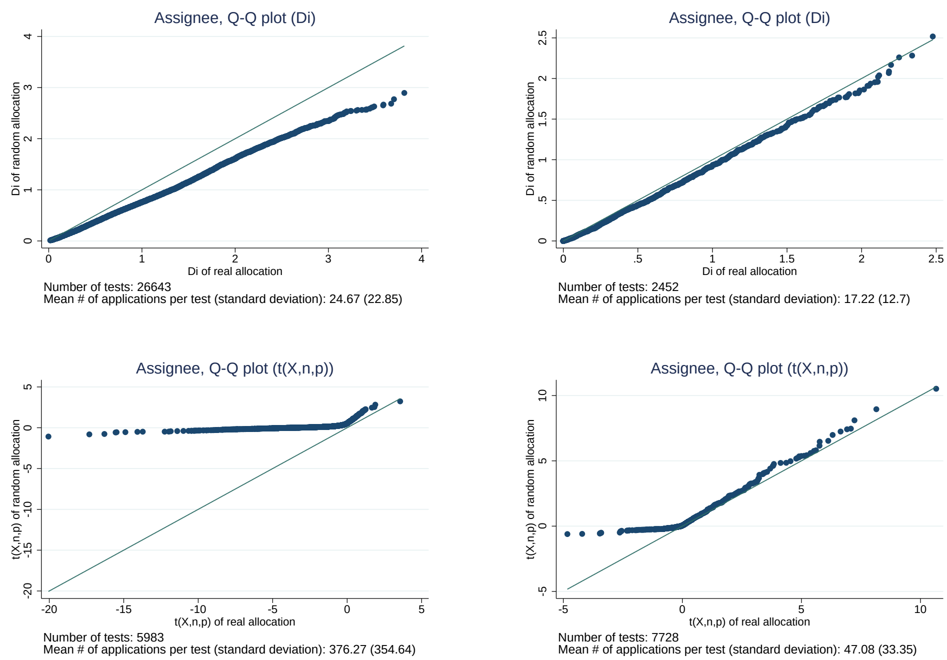
- 1600 - Biotechnology and Organic Chemistry
- 1700 - Chemical and Materials Engineering
- 2100 - Computer Architecture, Software, and Information Security
- 2400 - Computer Networks, Multiplex communication, Video Distribution, and Security
- 2600 - Communications
- 2800 - Semiconductors, Electrical and Optical Systems and Components
- 3600 - Transportation,Construction, Electronic Commerce, Agriculture, National Security and License & Review
- 3700 - Mechanical Engineering, Manufacturing, Products

Figure A2: Distribution of P-values from D-index (top) and MTAD (bottom) for USPC subclass and Assignee (lower thresholds).



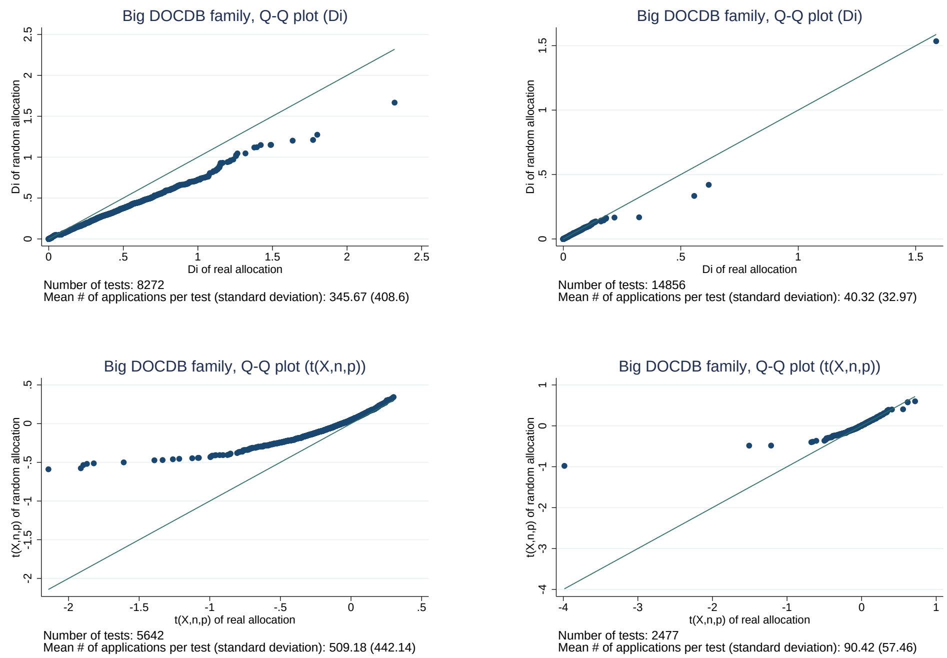
Tests on subsamples with more than 10 applications for D-index and 25 applications for MTAD. Vertical red lines are standard thresholds for statistical significance (0.01, 0.05 and 0.10).

Figure A3: Quantile-Quantile Plots of D-index (top) and MTAD (bottom) by Art-Unit-Year (left) and Art-Unit-Year-USPC-Subclass (right) for Assignee (lower thresholds).



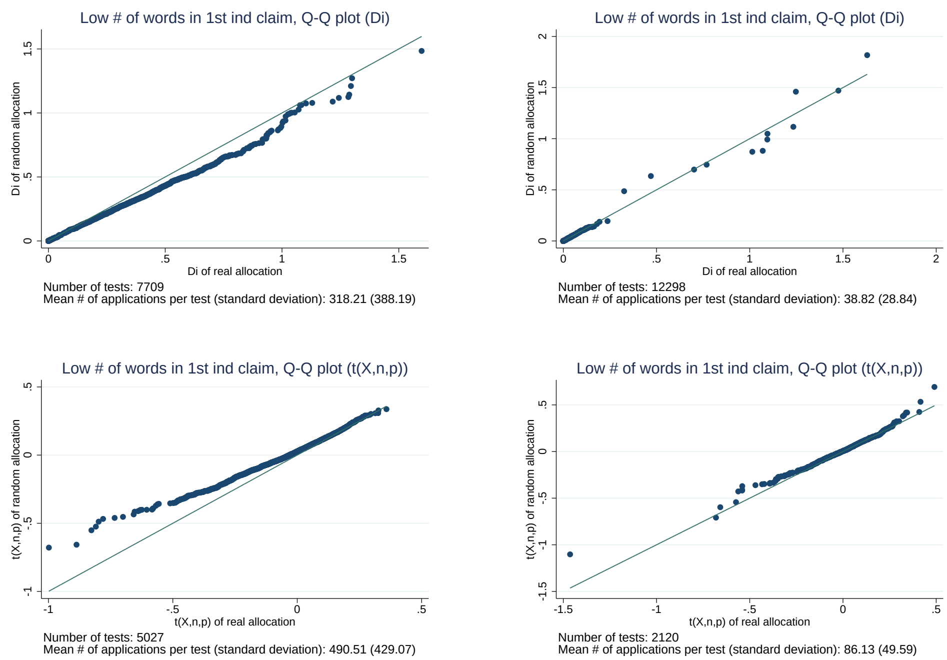
Each panel plots the quantiles of the D-index (top row) or MTAD statistic (bottom row) for the observed distribution (X-axis) against a simulated distribution under random assignment (Y-axis). Tests on subsamples with more than 10 applications for D-index and 25 applications for MTAD. If the observed distribution is random, the quantiles should be the same and the scatter points will fall along the 45-degree line. We observe large deviations from random assignment at the art-unit-year level, but much less evidence within art-unit-year-USPC-subclasses.

Figure A4: Quantile-Quantile Plots of D-index (top) and MTAD (bottom) by Art-Unit-Year (left) and Art-Unit-Year-USPC-Subclass (right) for DOCDB Family Size.



Each panel plots the quantiles of the D-index (top row) or MTAD statistic (bottom row) for the observed distribution (X-axis) against a simulated distribution under random assignment (Y-axis). Tests on subsamples with more than 20 applications for D-index and 50 applications for MTAD. If the observed distribution is random, the quantiles should be the same and the scatter points will fall along the 45-degree line. We observe some deviations from random assignment at the art-unit-year level, but almost no evidence within art-unit-year-USPC-subclasses.

Figure A5: Quantile-Quantile Plots of D-index (top) and MTAD (bottom) by Art-Unit-Year (left) and Art-Unit-Year-USPC-Subclass (right) for Words in $1st$ Claim.



Each panel plots the quantiles of the D-index (top row) or MTAD statistic (bottom row) for the observed distribution (X-axis) against a simulated distribution under random assignment (Y-axis). Tests on subsamples with more than 20 applications for D-index and 50 applications for MTAD. If the observed distribution is random, the quantiles should be the same and the scatter points will fall along the 45-degree line. We observe small deviations from random assignment at the art-unit-year level, and almost no evidence within art-unit-year-USPC-subclasses.

Table A1: Summary statistics for sample of applications.

Panel A: categorical variables

Variable	# of categories	Applications per category						
		Mean	Std dev	Min	5th percentile	Median	95th percentile	Max
Examiners	12,389	236.64	251.13	1	2	165	780	1,932
Art units	590	4,969.01	4,280.04	3	487	3,411	14,548	23,275
USPC subclasses	104,289	28.11	121.90	1	1	6	106	12,605
Assignees	173,242	13.06	299.50	1	1	1	19	60,177

Panel B: quantitative variables

Variable	N	Mean	Std dev	Min	5th percentile	Median	95th percentile	Max
DOCDB family size	2,904,291	2.83	5.55	1	1	2	8	378
Words in 1st claim	2,712,401	124.95	128.00	1	35	103	269	46,194

The number of applications characterized by a big DOCDB family and a low number of words in the first independent claim are respectively 115,068 and 116,667.

Table A2: Summary statistics for examiners' specialization and examination outcomes.

Variable	N	Mean	Std dev	Min	Median	Max
$Share_{ijt}$	1,985,470	0.06	0.13	0.00	0.01	1
Granted	1,936,297	0.67	0.47	0	1	1
Days	1,935,940	907.73	518.23	0	812	17,835
Words	1,077,041	49.21	87.73	-10,351	30	9,248

Appendix B

Appendix to Chapter 3

Table B1: Descriptive statistics for SEPs.

Variable	N	Mean	Sd	Min	Median	Max
Litigated [a]	4,346	0.09	0.28	0	0	1
Cites [a]	4,346	45.38	82.87	0	17	1,563
Granted	4,479	0.97	0.17	0	1	1
Abandoned	4,479	0.02	0.14	0	0	1
Words in patent's 1st ind. claim [a]	4,330	164.09	88.45	26	144	1,156
Claims in patent [a]	4,330	23.74	19.08	1	20	293
Reissue [a]	4,346	0.02	0.13	0	0	1
Reexamination [a]	4,346	0.02	0.13	0	0	1
CON children	4,479	0.92	2.09	0	0	32
CIP children	4,479	0.25	1.62	0	0	53
DIV children	4,479	0.18	0.99	0	0	16
Days earliest priority-issue (all) [a]	4,345	1,690.21	866.15	223	1,530	8,009
Days earliest priority-issue (continuity) [a]	4,345	1,611.21	858.38	223	1,434	8,009
Days filing-issue [a]	4,345	1,185.44	575.85	175	1,079	4,296
Days to respond to non-final rejection [b]	3,723	111.04	62.96	0	95	1,559
Requests for extension of time	4,479	0.73	1.09	0	0	9
RCE after notice of allowance [c]	4,363	0.05	0.22	0	0	1
Applicant delays (days) [d]	1,996	74.84	119.81	0	33	1,782
USPTO delays (days) [d]	1,997	538.01	483.43	0	478	3,528
CON	4,479	0.20	0.40	0	0	1
CIP	4,479	0.09	0.29	0	0	1
DIV	4,479	0.06	0.24	0	0	1
Provisional application [e]	4,479	0.21	0.40	0	0	1
PCT application [f]	4,479	0.06	0.24	0	0	1
Published	4,479	0.38	0.49	0	0	1
Words in application's 1st ind. claim [g]	1,700	117.88	69.83	12	103	1,164
Claims in application [g]	1,704	26.65	20.93	1	21	205
Small entity	4,479	0.04	0.19	0	0	1
Foreign priority	4,479	0.21	0.41	0	0	1
Priority year	4,478	1,997.52	5.60	1,971	1,998	2,010
Filing year	4,478	1,998.71	5.70	1,971	1,999	2,011
Issue year	4,345	2,001.74	6.39	1,974	2,002	2,015
Disclosure year	4,456	2,004.89	4.59	1,974	2,006	2,012

[a] Variable defined only for issued patents. [b] Variable defined only for applications with a reply to the first non-final rejection. [c] Variable defined only for applications that receive a notice of allowance. [d] Variable defined only for applications filed after May 29, 2000. [e] Provisional applications available since June 8, 1995. [f] USA joined the PCT in 1978. [g] Variable defined only for published applications.

Table B2: SEPs by Technology Center.

Technology Center	N	%
TC 2600 - Computers/Communications	2,252	50.30
TC 2700 - Computers/Communications [a]	670	14.97
TC 2400 - Computers/Communications	435	9.72
TC 2100 - Computers/Communications/Electrical [b]	423	9.45
TC 2200 - Electrical [a]	173	3.86
TC 2800 - Electrical	166	3.71
TC 2300 - Computers/Communications [a]	158	3.53
TC 3600 - Miscellaneous	115	2.57
Other TCs	85	1.88
Total	4,477	100.00

[a] Technology Center no longer active. [b] Prior to 1998 TC 2100 examined applications in the Electrical area. After 1998 it examines applications in the Computers and Communications area. See Graham et al. (2015) and Kesan et al. (2014) for mapping Technology Centers of the USPTO to broad technological areas.

Table B3: Business Models of Big SEP assignees.

SEP assignee	N	Business Model
NOKIA	428	Downstream
MOTOROLA	356	Downstream
ERICSSON	326	Downstream
QUALCOMM	263	Upstream
CISCO SYSTEMS	148	Downstream
INTERDIGITAL	97	Upstream
NORTEL NETWORKS	89	Downstream
AT&T	72	Downstream
MICROSOFT	70	Downstream
IBM	66	Downstream
APPLE	61	Downstream
ALCATEL	44	Downstream
KINETO WIRELESS	43	Downstream
TEXAS INSTRUMENTS	42	Upstream
SIEMENS	41	Downstream
HEWLETT PACKARD	40	Downstream
DIGITAL FOUNTAIN	38	Upstream
JUNIPER NETWORKS	38	Downstream
INTEL	36	Upstream
PHILIPS	36	Downstream
MATSUSHITA ELECTRIC INDUSTRIAL	32	Downstream

Assignees with more than 30 SEPs.

Table B4: Comparison of SEPs with population of regular utility patent applications.

Variable	Mean other applications	Mean SEPs	p-value t-test	Normalized difference
Litigated [a]	0.01	0.09	0.00	0.36
Cites [a]	10.40	45.38	0.00	0.57
Granted	0.67	0.97	0.00	0.84
Abandoned	0.20	0.02	0.00	-0.61
Words in patent's 1st ind. claim [a]	175.97	164.09	0.00	-0.13
Claims in patent [a]	15.61	23.74	0.00	0.51
Reissue [a]	0.00	0.02	0.00	0.14
Reexamination [a]	0.00	0.02	0.00	0.15
CON children	0.22	0.92	0.00	0.42
CIP children	0.11	0.25	0.00	0.11
DIV children	0.08	0.18	0.00	0.11
Days earliest priority-issue (all) [a]	1,419.84	1,690.21	0.00	0.32
Days earliest priority-issue (continuity) [a]	1,282.31	1,611.21	0.00	0.39
Days filing-issue [a]	944.1	1,185.44	0.00	0.44
Days to respond to non-final rejection [b]	107.75	111.04	0.00	0.05
Requests for extension of time	0.57	0.73	0.00	0.15
RCE after notice of allowance [c]	0.03	0.05	0.00	0.13
Applicant delays (days) [d]	30.78	74.84	0.00	0.43
USPTO delays (days) [d]	199.24	538.01	0.00	0.79
CON	0.12	0.20	0.00	0.21
CIP	0.07	0.09	0.00	0.09
DIV	0.07	0.06	0.04	-0.03
Provisional application [e]	0.15	0.21	0.00	0.15
PCT application [f]	0.13	0.06	0.00	-0.24
Published	0.57	0.38	0.00	-0.39
Words in application's 1st ind. claim [g]	126.31	117.88	0.00	-0.10
Claims in application [g]	20.02	26.65	0.00	0.27
Inventors	2.81	2.76	0.12	-0.02
Small entity	0.24	0.04	0.00	-0.61
Foreign priority	0.40	0.21	0.00	-0.40
Filing year	2,001.3	1,998.71	0.00	-0.34

Sample includes all regular utility patent applications filed at the USPTO and available in Public PAIR as of January 24, 2015 (4,479 SEP applications and 7,026,494 other applications). The samples used for individual tests may be different because of variable definitions and missing values. Details on the samples for each test are available upon request. [a] Variable defined only for issued patents. [b] Variable defined only for applications that receive a non-final rejection. [c] Variable defined only for applications that receive a notice of allowance. [d] Variable defined only for applications filed after May 29, 2000. [e] Provisional applications available since June 8, 1995. [f] USA joined the PCT in 1978. [g] Variable defined only for published applications.

Table B5: Comparison of SEP applications to upstream and downstream organizations (unclassified and upstream pooled together): OLS regressions.

Outcome	Coefficient upstream	Standard error upstream	p-value upstream
Litigated [a]	0.04	0.01	0.01
Cites (log) [a]	0.07	0.08	0.42
Granted	-0.01	0.02	0.47
Abandoned	0.01	0.01	0.44
Words in patent's 1st ind. claim [a]	-0.03	0.02	0.12
Claims in patent [a]	-0.04	0.04	0.26
Reissue [a]	0.01	0.01	0.29
Reexamination [a]	0.00	0.00	1.00
CON children (log)	0.16	0.03	0.00
CIP children (log)	0.02	0.02	0.44
DIV children (log)	0.03	0.02	0.05
Days earliest priority-issue (all) [a] (log)	0.26	0.03	0.00
Days earliest priority-issue (continuity) [a] (log)	0.26	0.03	0.00
Days filing-issue [a] (log)	-0.07	0.03	0.01
Days to respond to non-final rejection [b] (log)	0.14	0.03	0.00
Requests for extension of time (log)	0.06	0.03	0.08
RCE after notice of allowance [c]	0.05	0.02	0.02
Applicant delays (days) (log)	0.40	0.12	0.00
USPTO delays (days) (log)	-0.43	0.17	0.01
CON	0.24	0.03	0.00
CIP	0.07	0.02	0.00
DIV	0.07	0.02	0.00
Provisional application	0.12	0.03	0.00
PCT application	-0.01	0.02	0.33

The sample contains only regular utility SEP applications. "Upstream" organizations and those unclassified are pooled together. The sample contains 946 SEP applications assigned to downstream organizations and 754 SEP applications assigned to upstream or unclassified organizations. The samples used for individual regressions may be different because of variable definitions and missing values. Outcomes taken in logarithms may be either the natural logarithm of the variable or the natural logarithm of one plus the variable if the latter contain zeros. Details on samples, variables and regressions are available upon request. All regressions are estimated by OLS and include art unit effects, filing year effects, indicators for small entity status and foreign priority and control for the natural logarithms of the number of words in the first independent claim of the application, the number of claims in the application and the number of inventors. Applications with missing values for these variables are excluded from the sample. Robust standard errors are clustered at art unit level. [a] Variable defined only for issued patents. [b] Variable defined only for applications with a reply to the first non-final rejection. [c] Variable defined only for applications that receive a notice of allowance.

Table B6: Individual strategies and issuance lag of SEP applications: Poisson regressions.

	(1)	(2)	(3)	(4)	(5)
Outcome	Issuance lag (days)	Issuance lag (days)	Issuance lag (days)	Issuance lag (days)	Issuance lag (days)
Serialized Continuation	0.47***	0.47***	0.53***	0.53***	0.45***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Provisional application	0.21***	0.10***	0.06***	0.03	0.10***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
PCT application	0.43***	0.24***	0.26***	0.21***	0.31***
	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)
Requests for extension of time	0.13***	0.10***	0.07***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
RCE after NOA	0.29***	0.16***	0.12***	0.11***	0.10***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Upstream					0.06**
					(0.02)
Filing year effects	N	Y	Y	Y	Y
Art unit effects	N	Y	Y	Y	Y
Small entity	N	Y	Y	Y	Y
Foreign priority	N	Y	Y	Y	Y
Inventors	N	Y	Y	Y	Y
Applicant delay (days)	N	N	Y	Y	Y
Applicant delay (binary)	N	N	Y	Y	Y
USPTO delay (days)	N	N	Y	Y	Y
USPTO delay (binary)	N	N	Y	Y	Y
Words & Claims	N	N	N	Y	Y
Observations	4,344	4,340	1,868	1,535	1,167

All regressions estimated with Poisson models on the sample of granted SEPs. Outcome in all regressions is the number of days between the earliest priority and the issue date of the SEP. The variables "Applicant delay (days)", "USPTO delay (days)", "Inventors", "Words" and "Claims" are the natural logarithms of, respectively, one plus the number of days of applicant delay for the computation of the patent term adjustment, one plus the number of non-overlapping days due to delay of the USPTO for the computation of the patent term adjustment, the number of inventors, the number of words in the first independent claim of the application and the number of claims in the application. The variables "Applicant delay (binary)" and "USPTO delay (binary)" are binary indicators equal to one if the application has days of, respectively, applicant delay and USPTO delay (zero otherwise). Robust standard errors in parentheses, clustered at art unit level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B7: Prosecution time and claim changes, only SEPs disclosed before issuance.

	(1)	(2)	(3)	(4)
Outcome	Words added	Words added	Words added	Words added
Days filing-issue (log)	89.95***	108.58***	46.89**	29.45
	(12.29)	(13.79)	(21.77)	(20.56)
USPTO delay (binary & days)	Y	Y	Y	Y
Filing year effects	N	Y	Y	Y
Serialized continuation	N	Y	Y	Y
Provisional application	N	Y	Y	Y
PCT application	N	Y	Y	Y
Small entity	N	Y	Y	Y
Foreign priority	N	Y	Y	Y
Inventors	N	Y	Y	Y
Art unit effects	N	N	Y	Y
Words & Claims	N	N	N	Y
Observations	521	521	521	521

All regressions estimated by OLS on the sample of granted SEPs filed on or after November 29, 2000, published before issuance and disclosed as essential while still pending. Outcome in all regressions is the number of words added to the first independent claim between the published application and the issued patent. The variables "USPTO delay (days)", "Inventors", "Words" and "Claims" are the natural logarithms of, respectively, one plus the number of non-overlapping days due to delay of the USPTO for the computation of the patent term adjustment, the number of inventors, the number of words in the first independent claim of the application and the number of claims in the application. The variable "USPTO delay (binary)" is a binary indicator equal to one if the application has days of USPTO delay (zero otherwise). Robust standard errors in parentheses, clustered at art unit level. *** p<0.01, ** p<0.05, * p<0.10.

Table B8: SEP disclosure and patent issuance, excluding ETSI.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Issued	Issued	Issued	Issued	Issued	Issued
Sample	Matched sample	Matched sample	Matched sample	SEPs	SEPs	SEPs
Disclosed	0.007***	0.005***	0.005***	0.011***	0.011***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
SEP	-0.005***	-0.004***	-0.004***			
	(0.001)	(0.001)	(0.001)			
Month effects	Y	Y	Y	Y	Y	Y
Age effects	Y	Y	Y	Y	Y	Y
Art unit effects	N	Y	Y	N	Y	Y
Serialized Continuation	N	Y	Y	N	Y	Y
Provisional application	N	Y	Y	N	Y	Y
PCT application	N	Y	Y	N	Y	Y
Small entity	N	Y	Y	N	Y	Y
Foreign priority	N	Y	Y	N	Y	Y
Inventors	N	Y	Y	N	Y	Y
Words and Claims	N	N	Y	N	N	Y
Observations	60,260	60,260	60,260	40,957	40,794	38,080
Number of applications	1,030	1,030	1,030	675	673	623

All regressions estimated by OLS. ETSI SEPs and their controls are not included in the samples for this table. Unit of observation is application-month. The matched sample contains SEPs declared before their issue date and matched control applications filed on or after November 29, 2000, that were published as applications. SEP and control applications are matched on filing year, art unit, examiner, USPC class, foreign priority and small entity applicant. Applications with missing values for variables used to match or control variables are dropped from potential matches. The SEPs sample contains only SEPs declared before their issue date with filing date on or after November 29, 2000. Applications at risk of issuance from their filing month and observed until issuance month or abandonment month or end of calendar year 2014 if they do not issue as patents. Outcome in all regressions is an indicator variable equal to one in the month of issue. The variables "Inventors", "Words" and "Claims" are the natural logarithms of, respectively, the number of inventors, the number of words in the first independent claim of the application and the number of claims of the application. Robust standard errors are clustered at art unit level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B9: SEP disclosure and continuation application filings, excluding ETSI.

	(1)	(2)	(3)	(4)
Outcome	# of CON children	# of CON children	# of CON children	# of CON children
Sample	Matched sample	Matched sample	SEPs	SEPs
Disclosed	-0.15	0.15	0.16	0.82***
	(0.16)	(0.18)	(0.15)	(0.21)
SEP	0.32**			
	(0.14)			
Month effects	N	Y	N	Y
Age^2, Age^3 and Age^4	N	Y	N	Y
Application Fixed Effects	N	Y	N	Y
Observations	56,551	18,850	38,464	15,007
Number of applications	1,030	350	675	262

Unit of observation is application-month. ETSI SEPs and their controls are not included in the samples for this table. All models estimated with Poisson regressions. The matched sample contains SEPs declared before their issue date and matched control applications filed on or after November 29, 2000, that were published as applications. SEP and control applications are matched on filing year, art unit, examiner, USPC class, foreign priority and small entity applicant. Applications with missing values for variables used to match or control variables are dropped from potential matches. The SEPs sample contains only SEPs declared before their issue date with filing date on or after November 29, 2000. Applications at risk of filing continuation applications from their filing month and observed until their issue month or their abandon month or the end of calendar year 2012 if they are still pending. Robust standard errors in parentheses, clustered at application level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B10: SEP's issuance lag and litigation, excluding ETSI.

	(1)	(2)	(3)	(4)
Outcome	Litigated	Litigated	Litigated	Litigated
Issuance lag (log)	0.08***	0.08***	0.08***	0.07***
	(0.02)	(0.03)	(0.02)	(0.02)
Issue year effects	Y	Y	Y	Y
Small entity	N	Y	Y	Y
Foreign priority	N	Y	Y	Y
Inventors	N	Y	Y	Y
Words & Claims	N	N	Y	Y
USPC class effects	N	N	N	Y
Observations	646	644	640	640

All regressions estimated by OLS. Unit of observation is a SEP. Sample contains only SEPs declared before the issue date. SEPs declared to ETSI are excluded. Outcome in all regressions is an indicator variable equal to one if the patent is litigated before year 2015. The variables "Inventors", "Words" and "Claims" are the natural logarithms of, respectively, the number of inventors, the number of words in the first independent claim of the patent and the number of claims of the patent. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table B11: SEP's issuance lag and litigation, 5-year window.

	(1)	(2)	(3)	(4)
Outcome	Litigated	Litigated	Litigated	Litigated
Issuance lag (log)	0.06* (0.04)	0.05 (0.03)	0.06 (0.03)	0.07* (0.04)
Issue year effects	Y	Y	Y	Y
Small entity	N	Y	Y	Y
Foreign priority	N	Y	Y	Y
Inventors	N	Y	Y	Y
Words & Claims	N	N	Y	Y
USPC class effects	N	N	N	Y
Observations	328	328	328	328

All regressions estimated by OLS. Unit of observation is a SEP. Sample contains only SEPs declared before the issue date and granted before year 2010. Outcome in all regressions is an indicator variable equal to one if the patent is litigated within 5 years from issuance. The variables "Inventors", "Words" and "Claims" are the natural logarithms of, respectively, the number of inventors, the number of words in the first independent claim of the patent and the number of claims of the patent. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Appendix C

Appendix to Chapter 4

Table C1: Summary statistics for the citation sample.

Variable	N	Mean	Sd	Min	Median	Max
CWTS	167,707	21.71	20.15	0.00	11.09	84.08
Text-based similarity	164,930	24.25	19.22	0.00	23.67	100.00
SEP	168,589	0.53	0.50	0.00	1.00	1.00
SEP disclosed	168,589	0.35	0.48	0.00	0.00	1.00
SEP disclosure year	89,045	2,001.72	4.50	1,982.00	2,002.00	2,009.00
Citation year	168,589	2,003.00	4.43	1,980.00	2,004.00	2,009.00
Patent age at citation	168,589	4.99	4.02	0.00	4.00	19.00
Litigated	168,589	0.05	0.21	0.00	0.00	1.00
Backward NPL citations	168,269	4.42	13.54	0.00	1.00	422.00
Backward patent citations	168,269	13.61	22.91	0.00	8.00	782.00
Inventors	167,457	2.57	1.73	1.00	2.00	16.00
Words in 1st claim	167,455	175.58	92.83	20.00	155.00	1,054.00
Claims	167,468	23.50	21.57	1.00	19.00	374.00
CON	168,269	0.16	0.37	0.00	0.00	1.00
CIP	168,269	0.09	0.29	0.00	0.00	1.00
DIV	168,269	0.05	0.22	0.00	0.00	1.00
Provisional application	168,269	0.08	0.28	0.00	0.00	1.00
PCT application	168,269	0.03	0.18	0.00	0.00	1.00
Foreign priority	168,269	0.22	0.42	0.00	0.00	1.00
Small entity	168,269	0.07	0.25	0.00	0.00	1.00
Published application	168,269	0.08	0.27	0.00	0.00	1.00
Issue year	168,269	1,998.01	4.75	1,980.00	1,999.00	2,009.00
Filing year	168,589	1,995.43	4.29	1,978.00	1,996.00	2,007.00

The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age. NPL means "Non-Patent Literature".

Table C2: Citation sample by SSO.

SSO	# of citations	# of cited patents
ANSI	6,212	179
ATIS	2,515	33
BBF	37	2
CEN	1	1
CENELEC	437	6
ETSI	38,883	1,695
IEC	299	27
IEC - JTC1	1,918	27
IEEE	13,615	432
IETF	8,241	280
ISO	1,028	27
ISO - JTC1	1,486	58
ITU	8,372	301
OMA	4,355	111
TIA	1,646	46
CONTROLS	79,544	3,196
Total	168,589	6,495

The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age.

Table C3: Citation sample by licensing terms.

SSO	# of citations	# of cited patents
FRAND	80,617	3,042
FREE	2,568	114
NONE	3,040	88
SPECIFIC	2,820	55
CONTROLS	79,544	3,196
Total	168,589	6,495

The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age.

Table C4: Summary statistics for the patent sample.

Variable	N	Mean	Sd	Min	Median	Max
Citing classes	6,198	3.40	3.64	0.00	2.00	39.00
SEP	6,198	0.50	0.50	0.00	0.50	1.00
SEP disclosure year	3,099	2,003.03	3.98	1,982.00	2,004.00	2,008.00
Litigated	6,198	0.04	0.19	0.00	0.00	1.00
Backward NPL citations	6,198	3.98	12.56	0.00	1.00	282.00
Backward patent citations	6,198	13.15	21.51	0.00	8.00	479.00
Inventors	6,171	2.53	1.71	1.00	2.00	16.00
Words in 1st claim	6,171	167.58	86.26	6.00	149.00	1,054.00
Claims	6,172	21.01	17.65	1.00	18.00	374.00
CON	6,198	0.19	0.39	0.00	0.00	1.00
CIP	6,198	0.07	0.26	0.00	0.00	1.00
DIV	6,198	0.07	0.25	0.00	0.00	1.00
Provisional application	6,198	0.12	0.32	0.00	0.00	1.00
PCT application	6,198	0.06	0.24	0.00	0.00	1.00
Foreign priority	6,198	0.32	0.47	0.00	0.00	1.00
Small entity	6,198	0.06	0.24	0.00	0.00	1.00
Published application	6,198	0.25	0.44	0.00	0.00	1.00
Issue year	6,198	2,000.44	5.25	1,980.00	2,001.00	2,008.00
Filing year	6,198	1,997.47	4.71	1,978.00	1,998.00	2,007.00

The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009. NPL means "Non-Patent Literature".

Table C5: Patent sample by SSO.

SSO	# of patents
ANSI	178
ATIS	30
BBF	2
CENELEC	6
ETSI	1,578
IEC	22
IEC - JTC1	109
IEEE	408
IETF	232
ISO	27
ISO - JTC1	57
ITU	312
OMA	90
TIA	48
CONTROLS	3,099
Total	6,198

The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009.

Table C6: Patent sample by licensing terms.

SSO	# of patents
FRAND	2,879
FREE	79
NONE	85
SPECIFIC	56
CONTROLS	3,099
Total	6,198

The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents issued before year 2009.

Table C7: "Marginal/selection effects" of SSOs on Citation Weighted Technological Similarity between cited and citing patents: OLS models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	CWTS	CWTS	CWTS	CWTS	CWTS	CWTS
Disclosed	-0.23***	0.01	0.01	0.00	-0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
SEP	0.25***	0.12***	0.12***	0.10***	0.10***	
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	167,707	167,707	167,387	167,363	167,363	167,707
# of cited patents	6,468	6,468	6,463	6,461	6,461	6,468

The unit of observation is a citation. The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age. The outcome in all models is the natural logarithm of Citation Weighted Technological Similarity between cited and citing patent. All regressions estimated with OLS models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent and non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table C8: "Marginal/selection effects" of SSOs on text similarity between cited and citing patents: OLS models.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.
Disclosed	-0.34***	-0.12***	-0.12***	-0.13***	-0.12***	-0.05**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
SEP	0.31***	0.18***	0.18***	0.18***	0.18***	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	164,930	164,930	164,610	163,789	163,789	164,930
# of cited patents	6,478	6,478	6,473	6,445	6,445	6,478

The unit of observation is a citation. The sample contains all citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age. The outcome in all models is the natural logarithm of one plus the text similarity between cited and citing patent (Younge and Kuhn, 2016). All regressions estimated with OLS models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent and non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table C9: "Marginal/selection effects" of SSOs on Citation Weighted Technological Similarity between cited and citing patents: Poisson models, excluding citations made by the examiners.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	CWTS	CWTS	CWTS	CWTS	CWTS	CWTS
Disclosed	-0.01	0.05*	0.04	0.04	0.04	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
SEP	0.09**	0.06*	0.06*	0.05	0.05	
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	35,630	35,630	35,481	35,471	35,471	35,046
# of cited patents	3,159	3,159	3,156	3,155	3,155	2,575

The unit of observation is a citation. The sample contains the citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age, excluding those made by the examiners. We include only cited patents granted after year 2000 because information on examiners' citations is available starting from year 2001. The outcome in all models is the Citation Weighted Technological Similarity between cited and citing patent. All regressions estimated with Poisson models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent and non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table C10: "Marginal/selection effects" of SSOs on Citation Weighted Technological Similarity between cited and citing patents: Poisson models, excluding citations made by the examiners.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.	Text Sim.
Disclosed	-0.10***	-0.05	-0.05*	-0.06*	-0.05	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
SEP	0.15***	0.12***	0.10***	0.10***	0.10***	
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
Year effects	N	Y	Y	Y	Y	Y
Age effects	N	Y	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	Y	Y	Y	Y	N
USPC class effects	N	Y	Y	Y	Y	N
Application controls	N	N	Y	Y	Y	N
Patent controls	N	N	N	Y	Y	N
Litigated	N	N	N	N	Y	Y
Cited patent effects	N	N	N	N	N	Y
Observations	34,357	34,357	34,208	34,198	34,198	33,496
# of cited patents	3,134	3,134	3,131	3,130	3,130	2,488

The unit of observation is a citation. The sample contains the citations from regular utility patents to matched SEPs disclosed before year 2010 and control patents between the issue year and calendar year 2009 (inclusive) that occur before the 21st year of patent age, excluding those made by the examiners. We include only cited patents after year 2000 because information on examiners' citations is available starting from year 2001. The outcome in all models is the text similarity between cited and citing patent (Younge and Kuhn, 2016). All regressions estimated with Poisson models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent and non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** p<0.01, ** p<0.05, * p<0.10.

Table C11: "Marginal/selection effects" of SSOs on citations from old USPC classes (deepening): Poisson models, excluding citations made by the examiners.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes	Old Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
Disclosed	0.63***	0.18*	0.11	0.11	0.13	0.09
	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)	(0.09)
SEP	-0.35**	0.06	0.09	0.04	0.01	
	(0.14)	(0.11)	(0.12)	(0.12)	(0.12)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	7,070	7,070	7,070	7,070	7,070	1,788
# of cited patents	1,680	1,680	1,680	1,680	1,680	386

The unit of observation is a patent-year. The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents filed after year 2000 and issued before year 2009. Patent-year observations before the issue year and after the 20th year of age are excluded from estimation, as well as citations that occur after year 2009. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that cited the patent in previous years, excluding those originating from examiner citations. "Past classes" is the number of USPC classes that cited the focal patent before a given year, excluding those originating from examiner citations. All regressions estimated with Poisson models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent and non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table C12: "Marginal/selection effects" of SSOs on citations from new USPC classes (broadening): Poisson models, excluding citations made by the examiners.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	New Classes	New Classes	New Classes	New Classes	New Classes	New Classes
Sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched sample	Matched SEPs
Disclosed	-0.24**	-0.35***	-0.00	0.04	0.06	0.55***
	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.16)
SEP	0.22**	0.30***	0.02	-0.01	-0.06	
	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	
Past classes	N	Y	Y	Y	Y	Y
Year effects	N	N	Y	Y	Y	Y
Age effects	N	N	Y	Y	Y	N
Age^2, Age^3 & Age^4	N	N	N	N	N	Y
Filing year effects	N	N	Y	Y	Y	N
USPC class effects	N	N	Y	Y	Y	N
Application controls	N	N	N	Y	Y	N
Patent controls	N	N	N	N	Y	N
Litigated	N	N	N	N	Y	Y
Patent fixed effects	N	N	N	N	N	Y
Observations	7,070	7,070	7,070	7,070	7,070	2,312
# of cited patents	1,680	1,680	1,680	1,680	1,680	509

The unit of observation is a patent-year. The sample contains the regular utility SEPs disclosed before year 2010 and matched control patents filed after year 2000 and issued before year 2009. Patent-year observations before the issue year and after the 20th year of age are excluded from estimation, as well as citations that occur after year 2009. The outcome in all models is the number of USPC classes that cite the focal patent in a given year that never cited the patent before, excluding those originating from examiner citations. "Past classes" is the number of USPC classes that cited the focal patent before a given year, excluding those originating from examiner citations. All regressions estimated with Poisson models. "Application controls" include indicator variables for patents that issue from continuation applications, continuations-in-part, divisionals, claim the benefits of provisional applications, PCT applications and foreign applications, are filed by a small entity and are published before they issue. "Patent controls" include the natural logarithms of the number of backward patent and non-patent literature citations plus one, number of inventors, number of claims and number of words in 1st independent claim of the patent. Robust standard errors in parentheses, clustered at cited patent level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

References

Aghion, P., Dewatripont, M., and Stein, J. C. (2008). Academic freedom, private-sector focus, and the process of innovation. *RAND Journal of Economics*, 39(3):617–635.

Alcacer, J. and Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics*, 88(4):774–779.

Allison, J. R., Lemle, M. A., Moore, K. A., and Trunkey, R. D. (2004). Valuable patents. *Georgetown Law Journal*, 92(3):435–479.

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

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In Nelson, R. R., editor, *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, Princeton, NJ.

Baron, J. and Pohlmann, T. (2016). Mapping standards to patents using declarations of standard-essential patents. *Searle Center on Law, Regulation and Economic Growth, Working Paper*.

Baron, J., Pohlmann, T., and Blind, K. (2016). Essential patents and standard dynamics. *Research Policy*, 45(9):1762–1773.

Baron, J. and Spulber, D. F. (2016). The Searle Center database of technology standards and standard setting organizations. *Searle Center on Law, Regulation, and Economic Growth, Working Paper*.

Bekkers, R., Catalini, C., Martinelli, A., Righi, C., and Simcoe, T. (2016). Declared essential patents. *Hoover IP2 Working Paper*.

Bekkers, R., Duysters, G., and Verspagen, B. (2002). Intellectual property rights, strategic technology agreements and market structure: The case of GSM. *Research Policy*, 31(7):1141–1161.

Berger, F., Blind, K., and Thumm, N. (2012). Filing behaviour regarding essential patents in industry standards. *Research Policy*, 41(1):216–225.

Besen, S. M. and Farrell, J. (1994). Choosing how to compete: Strategies and tactics in standardization. *Journal of Economic Perspectives*, 8(2):117–131.

Bessen, J. and Maskin, E. (2009). Sequential innovation, patents, and imitation. *RAND Journal of Economics*, 40(1):611–635.

Boldrin, M. and Levine, D. K. (2008). *Against intellectual monopoly*. Cambridge University Press, Cambridge.

Boudreau, K. J. (2012). Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. *Organization Science*, 23(5):1409–1427.

Bresnahan, T. F. (2010). General purpose technologies. In Hall, B. H. and Rosenberg, N., editors, *Handbook of the Economics of Innovation*, volume 2, pages 761–791. Elsevier, Amsterdam.

Bresnahan, T. F. and Greenstein, S. (1999). Technological competition and the structure of the computer industry. *Journal of Industrial Economics*, 47(1):1–40.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics*, 65(1):83–108.

Carley, M., Hedge, D., and Marco, A. (2015). What is the probability of receiving a US patent? *Yale Journal of Law and Technology*, 17(1):203–223.

Chiao, B., Lerner, J., and Tirole, J. (2007). The rules of standard-setting organizations: an empirical analysis. *RAND Journal of Economics*, 38(4):905–930.

Cockburn, I. M., Kortum, S., and Stern, S. (2002). Are all patent examiners equal? The impact of examiner characteristics. *NBER Working Paper 8980*.

Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not). *NBER Working Paper 7552*.

Cotropia, C. A., Quillen, C. D. J., and Webster, O. H. (2013). Patent applications and the performance of the US Patent and Trademark Office. *Federal Circuit Bar Journal*, 23(1):179–191.

Farre-Mensa, J., Hegde, D., and Ljungqvist, A. (2015). The bright side of patents. *Working Paper*. Available at `https://www.uspto.gov/sites/default/files/documents/Patents%20030216%20USPTO%20Cover.pdf`.

Farrell, J., Hayes, J., Shapiro, C., and Sullivan, T. (2007). Standard setting, patents, and hold-up. *Antitrust Law Journal*, 74(3):603–670.

Farrell, J. and Saloner, G. (1985). Standardization, compatibility, and innovation. *RAND Journal of Economics*, 16(1):70–83.

Farrell, J. and Simcoe, T. (2012a). Choosing the rules for consensus standardization. *RAND Journal of Economics*, 43(2):235–252.

Farrell, J. and Simcoe, T. (2012b). Four paths to compatibility. In Peitz, M. and Waldfogel, J., editors, *Oxford Handbook of the Digital Economy*, pages 34–58. Oxford University Press, New York, NY.

Feng, J. and Jaravel, X. (2016). Who feeds the trolls? Patent trolls and the patent examination process. *Working Paper*. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2838017`.

Frakes, M. D. and Wasserman, M. F. (2014). Is the time allocated to review patent applications inducing examiners to grant invalid patents?: Evidence from micro-level application data. *NBER Working Paper 20337*.

Furman, J. L. and Stern, S. (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review*, 101(5):1933–1963.

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

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

Galasso, A., Schankerman, M., and Serrano, C. J. (2013). Trading and enforcing patent rights. *RAND Journal of Economics*, 44(2):275–312.

Gallini, N. and Scotchmer, S. (2002). Intellectual property: when is it the best incentive system? In Jaffe, A. B., Lerner, J., and Stern, S., editors, *Innovation Policy and the Economy*, volume 2, pages 51–77. MIT Press.

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

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

Gaulé, P. (2015). Patents and the success of venture-capital backed startups: Using examiner assignment to estimate causal effects. *Working Paper*. Available at `https://www.cerge-ei.cz/pdf/wp/Wp546.pdf`.

Gawer, A. and Henderson, R. (2007). Platform owner entry and innovation in complementary markets: Evidence from Intel. *Journal of Economics & Management Strategy*, 16(1):1–34.

Graham, S. J. H. and Hegde, D. (2014). Do inventors value secrecy in patenting? Evidence from the American Inventor's Protection Act of 1999. *Working Paper*. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2170555`.

Graham, S. J. H. and Hegde, D. (2015). Disclosing patents' secrets. *Science*, 347(6219):236–237.

Graham, S. J. H., Marco, A. C., and Miller, R. (2015). The USPTO patent examination research dataset: A window on the process of patent examination. *Working Paper*. Available at `https://www.uspto.gov/sites/default/files/documents/PatEx%20Working%20Paper.pdf`.

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

Griliches, Z. (1992). The search for R&D spillovers. *Scandinavian Journal of Economics*, 94(Suppl):S29–47.

Hall, B. H. (2005). A note on the bias in Herfindahl-type measures based on count data. *Revue d'Economie Industrielle*, 110(1):149–156.

Hall, B. H. (2007). Patents and patent policy. *Oxford Review of Economic Policy*, 23(4):568–587.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. *NBER Working Paper 8498*.

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

Hall, B. H. and Trajtenberg, M. (2004). Uncovering GPTs with patent data. *NBER Working Paper 10901*.

Hall, B. H. and Ziedonis, R. H. (2001). The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995. *RAND Journal of Economics*, 32(1):101–128.

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

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

Harhoff, D. and Wagner, S. (2009). The duration of patent examination at the European Patent Office. *Management Science*, 55(12):1969–1984.

Heller, M. A. and Eisenberg, R. S. (1998). Can patents deter innovation? The anticommons in biomedical research. *Science*, 280(5364):698–701.

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

Henderson, R., Jaffe, A. B., and Trajtenberg, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment: Comment. *American Economic Review*, 95(1):461–464.

Imbens, G. (2014). Matching methods in practice: Three examples. *NBER Working Paper 19959*.

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

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2000b). The meaning of patent citations: Report on the NBER/Case-Western Reserve survey of patentees. *NBER Working Paper 7631*.

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

Kang, B. and Bekkers, R. (2015). Just-in-time patents and the development of standards. *Research Policy*, 44(10):1948–1961.

Kang, B. and Motohashi, K. (2015). Essential intellectual property rights and inventors' involvement in standardization. *Research Policy*, 44(2):483–492.

Katz, M. L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75(3):424–440.

Kesan, J. P., Marco, A. C., and Miller, R. (2014). Patenting - with Chinese characteristics. *Working Paper*. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2469957`.

Kuhn, J. M. (2016). The right to exclude: Property rights and frictions in the sale of patents. *Working Paper*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2865293.

Kuhn, J. M., Roin, B. N., and Thompson, N. C. (2016). Causal inference on patent protection. *Working Paper*. Available at http://jeffreymkuhn.com/wp-content/uploads/2016/09/Patent_Metrics.2016.09.20.pdf.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Lanjouw, J. O. and Schankerman, M. (2001). Characteristics of patent litigation: a window on competition. *RAND Journal of Economics*, 32(1):129–151.

Lemley, M. A. (2002). Intellectual property rights and standard-setting organizations. *California Law Review*, 90(6):1889–1980.

Lemley, M. A. and Moore, K. A. (2004). Ending abuse of patent continuations. *Boston University Law Review*, 84(1):63–123.

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

Lemley, M. A. and Shapiro, C. (2007). Patent holdup and royalty stacking. *Texas Law Review*, 85(7):1991–2048.

Lerner, J. and Tirole, J. (2006). A model of forum shopping. *American Economic Review*, 96(4):1091–1113.

Lerner, J. and Tirole, J. (2015). Standard-essential patents. *Journal of Political Economy*, 123(3):547–586.

Levin, R. C., Klevorick, A. K., Nelson, R. R., Winter, S. G., Gilbert, R., and Griliches, Z. (1987). Appropriating the returns from industrial research and development. *Brookings Papers on Economic Activity*, 1987(3):783–831.

Liebowitz, S. J. and Margolis, S. E. (1994). Network externality: An uncommon tragedy. *Journal of Economic Perspectives*, 8(2):133–150.

Marco, A. C., Myers, A. F., Stuart Graham, a. P. D., and Kucab, J. (2015). The USPTO patent assignment dataset: Descriptions, lessons, and insights. *Working Paper*. Available at https://www.uspto.gov/sites/default/files/documents/USPTO_Patents_Assignment_Dataset_WP.pdf.

Marco, A. C., Sarnoff, J. D., and deGrazia, C. (2016). Patent claims and patent scope. *Working Paper*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2844964.

Matutes, C. and Regibeau, P. (1988). "Mix and match": product compatibility without network externalities. *RAND Journal of Economics*, 19(2):221–234.

Mehta, A., Rysman, M., and Simcoe, T. (2010). Identifying the age profile of patent citations: New estimates of knowledge diffusion. *Journal of Applied Econometrics*, 25(7):1179–1204.

Mori, T., Nishikimi, K., and Smith, T. E. (2005). A divergence statistic for industrial localization. *Review of Economics and Statistics*, 87(4):635–651.

Murray, F., Aghion, P., Dewatripont, M., Kolev, J., and Stern, S. (2009). Of mice and academics: Examining the effect of openness on innovation. *NBER Working Paper 14819*.

Nagaoka, S., Tsukada, N., and Shimbo, T. (2009). The structure and the emergence of essential patents for standards: Lessons from three IT standards. In Cantner, U., Gaffard, J.-L., and Nesta, L., editors, *Schumpeterian Perspectives on Innovation, Competition and Growth*, pages 435–450. Springer, Berlin Heidelberg.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of Political Economy*, 37(3):297–306.

Nordhaus, W. D. (1967). The optimal life of a patent. *Working Paper*. Available at `https://ideas.repec.org/p/cwl/cwldpp/241.html`.

Putnam, J. (1996). *The value of international patent protection*. Phd thesis, Yale University.

Rysman, M. and Greenstein, S. (2005). Testing for agglomeration and dispersion. *Economics Letters*, 86(3):405–411.

Rysman, M. and Simcoe, T. (2008). Patents and the performance of voluntary standard-setting organizations. *Management Science*, 54(11):1920–1934.

Sampat, B. and Williams, H. L. (2015). How do patents affect follow-on innovation? Evidence from the human genome. *NBER Working Paper 15851*.

Scherer, F. M. (1983). The propensity to patent. *International Journal of Industrial Organization*, 1(1):107–128.

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

Shapiro, C. (2001). Navigating the patent thicket: Cross licenses, patent pools, and standard setting. In Jaffe, A. B., Lerner, J., and Stern, S., editors, *Innovation Policy and the Economy*, volume 1, pages 119–150. MIT press, Cambridge, MA.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and statistics*, 88(4):641–658.

Simcoe, T. (2012). Standard setting committees: Consensus governance for shared technology platforms. *American Economic Review*, 102(1):305–336.

Simcoe, T. (2013). Governing the anti-commons: Institutional design for standard setting organizations. In Lerner, J. and Stern, S., editors, *Innovation Policy and the Economy*, volume 14. University of Chicago Press.

Simcoe, T. S., Graham, S. J., and Feldman, M. P. (2009). Competing on standards? Entrepreneurship, intellectual property, and platform technologies. *Journal of Economics & Management Strategy*, 18(3):775–816.

Sukhatme, N. U. and Cramer, J. N. (2014). Who cares about patent term? Cross-industry differences in term sensitivity. *Working Paper*. Available at `http://scholar.princeton.edu/sukhatme/publications/patent_term_sensitivity`.

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

Thoma, G., Torrisi, S., Gambardella, A., Guellec, D., Hall, B. H., and Harhoff, D. (2010). Harmonizing and combining large datasets–an application to firm-level patent and accounting data. *NBER Working Paper 21666.*

Thompson, P. and Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, 95(1):450–460.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *Rand Journal of Economics*, 21(1):172–187.

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

Waguespack, D. M. and Fleming, L. (2009). Scanning the commons? Evidence on the benefits to startups participating in open standards development. *Management Science*, 55(2):210–223.

Weiss, M. B. and Sirbu, M. (1990). Technological choice in voluntary standards committees: an empirical analysis. *Economics of Innovation and New Technology*, 1(1–2):111–133.

Wright, B. D. (1983). The economics of invention incentives: Patents, prizes, and research contracts. *American Economic Review*, 73(4):691–707.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Younge, K. A. and Kuhn, J. M. (2016). Patent-to-patent similarity: A vector space model. *Working Paper*. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2709238`.

Ziedonis, R. H. (2004). Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. *Management Science*, 50(6):804–820.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

- [redacted]

[redacted]

- [redacted]

[redacted]

- [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

- [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted] [redacted]

[redacted]

[redacted]