RIETI Discussion Paper Series 18-E-046

The Choice of Examiner Patent Citations for Refusals: Evidence from the trilateral offices

WADA Tetsuo
Gakushuin University

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 18-E-046

July 2018

The Choice of Examiner Patent Citations for Refusals: Evidence from the trilateral offices [1]

WADA Tetsuo

Gakushuin University, Faculty of Economics

Abstract

This paper compares X/Y patent citations (i.e., those cited as grounds for rejections) between major patent offices. It reveals discrepancies between the offices, despite the common patentability criteria of novelty and inventive steps to generate citations. This paper also examines how the discrepancies of X/Y patent citations at the European Patent Office (EPO) and the U.S. Patent and Trademark Office (USPTO) relate to the characteristics of applications and longitudinal aspects of office actions. X/Y patent citations of both the EPO and USPTO commonly show that the range of patent application classes is positively correlated with divergent reasons for refusal. One novel methodological feature of this paper is that examiner citations across jurisdictions are comparable if we employ family-to-family citations and common criteria for the X/Y citation category. Furthermore, unlike the normal citation-generating process where a citing document adds citations to prior art only once, this paper represents the first attempt to analyze a citation network with multiple citing opportunities for separate parties, thereby providing a new perspective on the notion of breadth in citation impact.

Keywords: Examiner citation, X/Y citation, Patent family, Blocking patent

JEL classification: K29, O33, O34

[1]This study has benefited from the project "Study on Industrial Innovation Capability and Innovation Infrastructure" undertaken at the Research Institute of Economy, Trade and Industry (RIETI), and also from a project for the Research Institute of Science and Technology for Society (RISTEX) and the Japan Science and Technology Agency (JST). The author is grateful for helpful comments and suggestions by the RIETI project leader Professor Sadao Nagaoka (Tokyo Keizai University), Setsuko Asami, Yoshimi Okada, Stuart Graham, Lee Fleming, Alan Marco, Alfons Palangkaraya, Andrew Toole, and Guan-Cheng Li, seminar participants at RIETI, and the participants at the 12th Annual Conference of the European Policy for Intellectual Property (EPIP 2017) and at the 16th International Conference on Scientometrics and Informetrics (ISSI 2017).

Introduction

Although international family-to-family patent citation data have been made readily accessible by the PATSTAT[2] database, which is distributed by the European Patent Office (EPO), there have been few international comparative analyses of family-level citation data[3]. International family-level patent citation data create a novel opportunity for patent citation analyses, in that the same invention can have separate bundles of citation records left by separate examinations, conducted in a parallel or sequential manner, at more than one patent office. We can thus observe a variety of prior art identified by examiners for the same invention, meaning that multiple technological linkages to prior art are obtained through concurrent evaluations by examiners, whereas the same basic principle of patentability standards (in particular, novelty and inventive step which generate citations because of prior art) is shared by major patent offices. As far as the author is aware, no previous study has directly compared examiner patent citations for the same set of international families of patent applications across jurisdictions.

The setting may also be fundamentally new ground, even for citation analyses in general. Citation has been perceived as a device to trace knowledge flows where scientific advancement follows prior work. While many patent citation analyses have been conducted with the assumption that patent citations have similar properties to citations in scientific papers (Narin 1994), international patent citations have unique network characteristics that other citation data lack, i.e., examiners in different jurisdictions can independently add different sets of citations to a single international family of patent applications. Unlike the normal citation-generating process, where a citing document confirms the addition of citations to prior art/literature only once at the time of publication, an international patent application has multiple chances to receive patent citations, even if we limit our attention to examiner citations. Whereas authors of scientific papers are usually the sole source of citations in scientific bibliometrics, authors may not necessarily

---

[2] The DOCDB family citation table (tls_228) has been added since its 2015 spring edition.

[3] According to the latest PATSTAT data catalogue (Spring 2018), the table of citations for DOCDB families "contains one entry for each pair of DOCDB simple families, where one member of a family cites at least one member of another family. That means if multiple publications of one family cite one or multiple publication(s)/application(s) of another family, this is counted as one citation between these two families." Its primary purpose seems to be to normalize the simple citation counts by duplications in international applications, and not to compare citations internationally in a citing or cited document.

be neutral evaluators of prior literature, and their reasons to cite can vary. Furthermore, analysts of citation data have virtually no chance of observing counterfactual cases to inquire what the consequences would be if a citing document were independently evaluated by a third party. On the other hand, examiners are trained for prior art searches, and they produce citations as reasons to reject applications by applying the same novelty and nonobviousness (interchangeably, inventive step) patentability criteria among patent offices. Therefore, international family-level patent citations are a rare, if not ideal, environment in which to study the mechanisms of the citation-generating process, with multiple trails of citations for an application, whereas most previous citation studies consider the citing mechanism only from the point of view of those cited.

To focus on patent-family data containing identical inventions at the trilateral patent offices of the EPO, the US Patent and Trademark Office (USPTO), and the Japan Patent Office (JPO), this paper takes a first look at the concordances and discrepancies of "X/Y" categories of examiner citations, which are equivalently called "blocking patents" in the US, among the trilateral offices. It also sheds light on aspects of cited inventions that cause diversions of citations. More specifically, it explores how the discrepancies of X/Y patent citations relate to the characteristics of patent applications and the longitudinal aspect of office actions, i.e., the breadth of patent claims and the dates on which refusals were presented to applicants.

The next section first describes the background of patent citation studies. In particular, it explains how obstacles to international comparisons of family-to-family citations have been resolved. Next, the data source for this study, especially X/Y citations and their US equivalents, is described. The following section illustrates the discrepancies among X/Y patent citations, regardless of the definition of an international family, and introduces a simple methodology of distance measurement between cited patent families. The final section analyzes how the discrepancies of X/Y patent citations relate to the characteristics of patent applications and the longitudinal aspect of office actions.

Background

Patent citations have been widely utilized for empirical studies of patent systems, particularly for issues such as economic value, technological impact, and knowledge flows. However, almost all existing studies rely on a set of data from one patent office, e.g., the USPTO or EPO. There are three difficulties in combining patent citation data across different offices, although in practice, valuable inventions are submitted as patent applications to many countries, thereby providing citations across many jurisdictions. The first obstacle is that there are a variety of patent citations, such as applicant and examiner citations, and too little is known about each of them to assume common ground to compare citations across jurisdictions. Second, although a specific category of examiner citations, i.e., X/Y, has a reasonable basis for comparison across patent offices, equivalent information regarding US citations is difficult to obtain. Third, international patent family data were not readily available until the PATSTAT database made it available for research. These obstacles for research have been gradually resolved, as discussed below.

Patent citations have been classified into several categories, and the most obvious difference is in the citing entities. With some technical exceptions[4], applicant citations are added by applicants and examiner citations are added by examiners. This distinction has received attention from researchers, followed by a series of empirical studies (e.g., Alacer and Gittleman 2006; Criscuolo and Verspagen 2008; Hegde and Sampat 2009; Cotropia et al. 2013). An implication from the studies is that examiner citations should be eliminated as proxies for knowledge spillovers between inventors, as examiners evaluate prior art at different times from inventions using different search tools. Whereas examiner citations are considered to be noisy as a means for measuring knowledge flow between inventors, examiner citations have been shown to have an advantage over inventor citations for measuring the value of patents (Hegde and Sampat 2009), possibly because examiners are third parties. Inventors may want to express gratitude in the form of citations, but examiners are free from such human interactions over an invention.

Indeed, examiners are required to follow exact standards to add special citation categories, such as the EPO Guidelines for Examination, and the USPTO Manual of Patent Examining Procedure (MPEP). Examiner citations for refusals, which are assigned special categories of “X” and “Y”[5] at the EPO and other offices such as the JPO, have essential roles during patent prosecution. Examiners must indicate specific reasons to

[4] Not all US examiner citations are made purely by examiners, in that some “examiner” citations may have been recognized and recorded through added information by the Information Disclosure Statement (IDS) supplied by applicants. On the other hand, some applicant citations are added by patent attorneys rather than inventors.

[5] X and Y indicate “Particularly relevant documents,” as the EPO Guidelines for Examination, Part B, Chapter X, 9.2.1 specify.

refuse patent claims, for example, lack of novelty or inventive step[6]. Because the standards of novelty and inventive step are very similar[7] between the offices, the categories of citation can be the common ground for comparisons of citations across offices, at least between the EPO and JPO[8].

On the other hand, the terminology for the X/Y citation categories is not used at the USPTO. However, the requirement is virtually the same between the EPO and USPTO, in the sense that the offices must clearly show the reasons for an applicant's rejection if it is attributable to prior art. At the USPTO, Section 706 "Rejection of Claims" of the MPEP states that "In rejecting claims for want of novelty or for obviousness, the examiner must cite the best references at his or her command (37CFR 1.104 (c))." In fact, as will be explained in the next section, rejection documents, or "blocking patents," are available from the Public Patent Application Information Retrieval (Public PAIR) database of the USPTO for most of results examined, and we can interpret each cited document specified as grounds for rejection in an equivalent manner to X/Y citations at the EPO and JPO.

Although blocking patents have been known to be available from raw texts on each rejection document in the US, it has been too difficult to create a large-scale database until recently. An initial analysis of small-scale data was performed (Cotropia et al. 2013) using manual collection of blocking patents to show that examiners rely on prior art that they had obtained themselves, rather than on applicant citations, to reject applications. A recent study based on PAIR data hosted by Google empirically defines "patent race" from novelty rejections (Thompson and Kuhn 2017). Finally, the USPTO recently released office action data in bulk form (Lu, Myers, and Beliveau 2017), although the initial release of the data covered only US-domestic citations and lacked international citations.

---

[6] See, for example, the EPO *Guidelines for Examination* Part C, Chapter III, 4. *First communication*, Part B, Chapter III, 1.1 *Opinions in relation to the search report*, and Part B, Chapter II, 2 *Objective of the search*. At the EPO, search reports are initially given to indicate patentability information to applicants, as Part B, Chapter II, 2 specifies that the "objective of the search is to discover the state of the art which is relevant for the purpose of determining whether, and if so to what extent, the claimed invention for which protection is sought is new and involves an inventive step." X/Y citations constitute clear reasons to reject, whereas an "A" citation does not, as "A" citations imply "relevant" prior art, which is not fatal for patentability.

[7] The term "novelty" is identical between the EPO and the USPTO, and "inventive step" at the EPO is usually interchangeable with "nonobviousness" at the USPTO. For example, the Patent Cooperation Treaty (PCT), Article 33, states that "The objective of the international preliminary examination is to formulate a preliminary and non-binding opinion on the questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), and to be industrially applicable."

[8] The procedure at the JPO is also virtually the same as those at the EPO, except that the JPO does not always create search reports preceding examinations.

This paper uses a set of large-scale PAIR rejection documents obtained directly from the Public PAIR portal ("file wrapper" pages), and independently from the data above, to include international citations, as will be further explained below.

Thus, the present study takes advantage of a novel large-scale data set of US blocking patents, obtained from refusal documents available as file wrappers on the Public PAIR database of the USPTO, to compare X/Y patent citations as reasons for refusals. An essential component of this comparison is PATSTAT patent family information. A citing family can consist of multiple applications and grant numbers over different jurisdictions, and the same is true for a cited family. To compare citations across jurisdictions, we need a standard by which to define an international patent family. Because PATSTAT has provided patent family data based on two definitions of patent families[9], i.e., DOCDB and INPADOC, we are now able to compare individual reasons for refusal by trilateral patent offices through comparisons of family-to-family citations at the two levels of patent families.

Data source

Our primary domain of study comprises of family-level patent citations recorded by triadic patent applications, obtained from the EPO PATSTAT database (Spring 2016). A triadic patent application means a family of applications relating to an identical technical content[10] that have been filed simultaneously at the EPO, USPTO, and the JPO. The citing unit is a DOCDB family[11] (43,207 triadic families) where only a single DOCDB family ID is observed for a family, implying that the same technical content is kept within the family in the sample for this article. Whereas X/Y citation indicator variable for the EPO is directly available from PATSTAT citation category table (tls 215), that for the JPO is not recorded on the table. However, X/Y citation information at the JPO is also

---

[9] DOCDB family, or DOCDB simple patent family, indicates a collection of related patent applications that is covering *the same* technical content. On the other hand, INPADOC extended patent family indicates a collection of related patent applications that is covering *similar* technical content. See "Patent Families at the EPO (July 2017)," linked from https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families/docdb.html

[10] Note that it means a narrower definition of triadic patent in Triadic Patent database (Dernis and Khan 2004), which relies on INPADOC family. Ours relies on "twin application" idea to study differential results of patent grants from the triadic offices of the U.S., the EU, and Japan (Jensen et al., 2005; Webster et al., 2007; Webster et al., 2014).

[11] DOCDB patent family is a simple patent family, and every document in a simple patent family is considered to be covering exactly the same technical content. See "Patent Families at the EPO (July 2017)," linked from the page above.

readily available on PATSTAT, as X/Y citations at the JPO are provided for applicants on "rejection reasons" documents, which are coded as "EXA" as the source of citation[12] on PATSTAT. Therefore, X/Y citations are easily identifiable for both of the EPO and the JPO.

US blocking patent database was developed by extracting patent numbers from the main text of "CTNF" (nonfinal rejections) and "CTFR" (final rejections) documents available on the Public PAIR database of the USPTO. These document types are different from "PTO-892" (Notice of references cited by examiners), from which US examiner citations were extracted for most of prior studies, in the sense that examiners state (in the main text of CTNF/CTFR) which prior art they actually rely on to reject individual claims[13]. Downloading from the Public PAIR portal, optical character recognition, and natural language processing were conducted by Dr. Guan-Cheng Li, who has also contributed to numerous patent data projects at the University of California, Berkeley. Attorneys at the International Association for the Protection of Intellectual Property (AIPPI), Japan, have also contributed to this blocking patent database through manual verification of the approximately 1000 cited families, and the processing algorithm has been verified through manual comparisons. Our data were sourced directly from Public PAIR data, whereas a similar database project relied on files hosted by Google (Thompson and Kuhn 2017). Both of the data sets from these methodologies have missing observations. Our data were also developed independently from the USPTO office action

[12] The examination phase at the JPO produces two types of examiner citations. One is X/Y for rejections, and the other is "A," representing generally relevant prior art. The former is coded with "EXA" as its source of citation, and the latter is coded with "SEA" as its source of citation. This was confirmed with a JPO examiner.

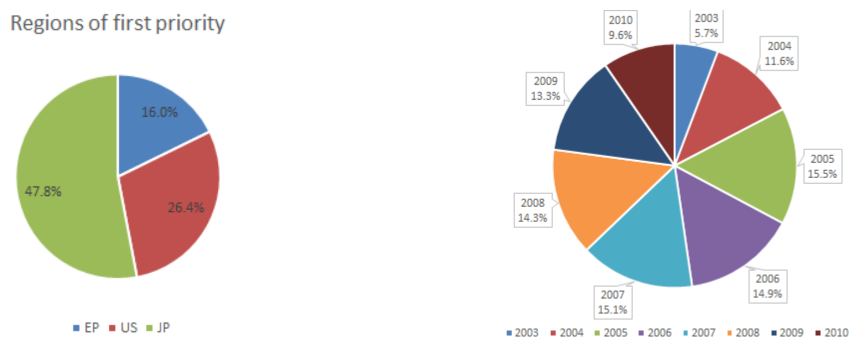
[13] Because the focal question concerns prior art utilized at the offices, we use rejection reasons based on Article 102 (novelty) or 103 (nonobviousness) of the US Code Title 35, and nonstatutory double patenting, which generate citations for prior art reasons. We extracted domestic and foreign patent numbers in the proximity of those key words. Not all cited patent numbers on CTNF/CTFRs are those of blocking patents, as examiners tend to cite "prior art made of record but not relied upon" at the end of each rejection documents. Those numbers after this key phrase were eliminated. The resulting list of blocking patents after this process closely matches the data provided in the USPTO office action data set (Lu, Myers, and Beliveau 2017), via the overlapping triadic sample with priority years between 2008 and 2010. However, because the USPTO office action data lack international citations, ours is the only data source for international comparisons. Although there are no complete blocking patent data as a perfect reference in the US, two-letter country codes of WO, JP, EP, GB, DE, FR, CN, CA, SU, KR, CH, AU, IT, DD, RU, NL, AT, BE, SE, TW, ES, and DK are covered in our data, and this set covers 4.0% of all patent citations cited in the US and recorded on tls211-212 tables of PATSTAT as ISR and PRS citations. The USPTO office action data set (Lu, Myers, and Beliveau 2017) seems to contain WO, JP, EP, DE, FR, KR and ES citations, albeit in a separate data field.

data (Lu, Myers, and Beliveau 2017), which utilized the office's internal data sources. Combined with the US blocking patent database, the EPO PATSTAT database (Spring 2016) has been used along with the Organisation for Economic Co-operation and Development (OECD) Triadic Patent database (Dernis and Khan 2004).

The domain of statistical analyses is the set of triadic applications from the Patent Cooperation Treat (PCT) and the Paris convention (non-PCT), consisting of 408,076 family citation pairs. Because of oversampling in triadic applications from Japan, approximately half of the domain consists of applications from Japan, as shown in Figure 1. The first priority years in a family were distributed between 2003 and 2010, also as shown in Figure 1. We use samples only when X/Y citations are added by all of the trilateral offices (specifically, first action allowances at one of the offices are not included). The cited unit is also a DOCDB family, so the unit of analysis is DOCDB family-to-family citation.

All citation data are patent citations, because of the limitation on the availability of DOCDB family-to-family citations. Lack of nonpatent citations is a weakness, although most observed examiner blocking citations are patents only. We combine citation date information from the EPO DOCDB back file and the USPTO PAIR Bulk data[14] to relate the citations to refusal timing in the prosecution processes for an international patent family.

Fig. 1 Sample composition (first priority by region and year)



[14] The date information for each office action in the US was obtained from the Patent Examination Research Dataset (Graham et al. 2015). https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-examination-research-dataset-public-pair

Simple Discrepancies at the DOCDB and INPADOC family level

First, as shown in Table 1 and Figure 2, the trilateral offices largely rely on their own reasons for refusals based on prior patents (X/Y citations or blocking patents). Based on 408,076 DOCDB family citations found from 43,207 citing families with single-DOCDB family IDs and priority years 2003–2010, we find a surprisingly small proportion of overlapping refusal reasons, even after the cited patents are consolidated by DOCDB international patent families. A larger ratio of "unique" X/Y citations is observed from both the USPTO and JPO than from the EPO. A contributing factor for the JPO may be oversampling.

Table 1 Dissimilarity between X/Y citations ("blocking patents," or examiner citations specified on CTNF/CTFR rejections by the USPTO) at the DOCDB family-to-family citation level

EP_XY	*JP_XY*	*US_XY_EQV*	*Counts of family-level citations*	
0	0	1	127,556	31.3%
0	1	0	147,868	36.2%
0	1	1	11,035	2.7%
1	0	0	75,441	18.5%
1	0	1	16,484	4.0%
1	1	0	19,764	4.8%
1	1	1	9,928	2.4%

("EP_XY" and "JP_XY" stands for X/Y citations at the EPO/JPO, respectively. "US_XY_EQV" stands for blocking patents obtained from CTNF/CTFRs at the USPTO.)

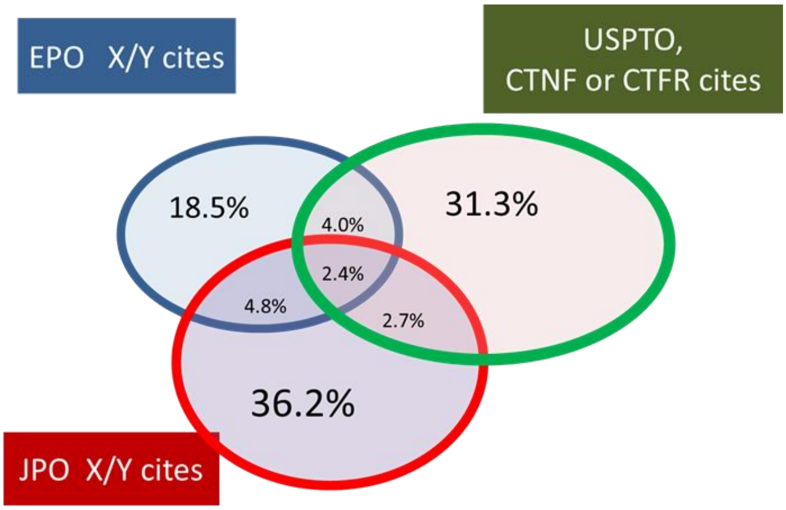


Fig. 2 Small overlaps between X/Y citations (blocking patents)

One might wonder that the small overlap is attributable to the prosecutions conducted over non-PCT applications because these do not carry over information from International Search Reports (ISRs), which could provide common information on prior art for the designated offices (DOs) conducting national prosecutions. However, as shown in Table 2, the distribution of the PCT sample is not much different from that of the combined PCT and non-PCT sample shown in Table 1. ISRs prepared by the receiving offices (ROs) for PCT applications may help searching activities at the DOs, but apparently, they do not affect the choice of X/Y citations during national prosecution processes.

Table 2 Dissimilarity between X/Y citations at the DOCDB family-to-family citation level, PCT only

EP_XY	*JP_XY*	*US_XY_EQV*	*Counts of family-level citations*	
0	0	1	67,877	32.0%
0	1	0	76,192	35.9%
0	1	1	7,179	3.4%
1	0	0	36,537	17.2%
1	0	1	7,732	3.6%
1	1	0	11,270	5.3%
1	1	1	5,447	2.6%

Furthermore, it is worth ascertaining whether the small overlap depends on the narrower definition of a DOCDB international patent family than that of an INPADOC extended family[15]. An INPADOC patent family indicates a *similar* technical contents being shared in a family, and has a wider definition than a DOCDB family, in that two patent publications belong to a single INPADOC family even if the two publications share a part of claimed inventions indirectly. Technically, applications that are members of an INPADOC extended patent family have at least one priority in common with at least one of the other members – directly or indirectly ("Patent families at the EPO," July 2017, p.23). For example, if application A is divided into two divisional applications, and one of the divisional applications is merged with a different application B, then all five documents—two divisional applications for A as well as A itself, and the original B and

[15] The definition of INPADOC patent family can be found at https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families/inpadoc.html

merged B—belong to a single INPADOC family, even though original A and B were separate initially. On the other hand, one DOCDB family has families only when the family members keep the identical technical content. Therefore, one might suspect that a patent cited in a country has a similar patent document as an INPADOC family member rather than as a DOCDB family member, and the discrepancies in documents cited by the same family may be attributable to the narrower definition of DOCDB family. Simply put, when an examiner cites a prior art of a DOCDB family while another examiner at another office cites a similar prior art of a different DOCDB family, the two prior art may belong to a single INPADOC family.

However, as shown below in Table 3, the ratio of overlaps between X/Y citations is very similar, even if cited patents are consolidated by the INPADOC extended family (the citing side is also consolidated by a DOCDB family, as above).

Table 3 Dissimilarity between X/Y citations by INPADOC family

EP_XY	*JP_XY*	*US_XY_EQV*	*Counts of family-level citations*	
0	0	1	115,618	30.3%
0	1	0	138,863	36.4%
0	1	1	10,810	2.8%
1	0	0	71,517	18.7%
1	0	1	16,095	4.2%
1	1	0	19,329	5.1%
1	1	1	9,694	2.5%

Thus, regardless of the definition of patent families and of application routes such as PCT or non-PCT, the overlaps of X/Y citations at the trilateral offices are small. We then utilize technological distance measurement to evaluate how the discrepancies of X/Y patent citations relate to the characteristics of patent applications and the longitudinal aspect of office actions.

Methodology for comparisons in technological discrepancies

A puzzle presented in the previous section is that examiners cite different prior art to reject a single invention with the same technological content, though the same basic principle of patentability standards (novelty and inventive step in particular) is shared by major patent offices. Cited prior art is determined here as “different” if it belongs to a different patent family. We do not have any information about how different cited patent families are, except binary information of the same or different patent family number.

Therefore, we then try to measure how distant a cited family is. Measurement of the difference of a cited patent family imposes a challenge, because we need to define what is measured when the objective is a bundle of international patents of a family, and also to define the reference point to be compared with the family.

There are several methods to measure diversity using patent information (Kuhn, Younge, and Marco 2017; Wang, Thijs, & Glänzel 2015). A distance measurement based on textual similarity (Kuhn et al., 2017) assumes common language, so it is not applicable in this paper. Although a variety of diversity index (Wang, Thijs, & Glänzel 2015) has been developed and is available, the objective here is a first look on how unique a particular cited patent is from other patents cited by other examiners, implying a focus on a single cited patent family, in relation to other cited families by other offices. Therefore, in the present paper, we employ distance measurement based on cosine vectors, following Jaffe (1986), to facilitate aggregations and comparisons across different judicial systems.

In order to obtain cosine similarity, two vectors are in question. A vector in question is a patent class vector for a family of cited patents, and the other vector is a "global citation vector" for all other cited patents, which is cited by the same citing family in the world, as a reference vector. Specifically, we use the cosine between bundles of citations using the Main Class level of the International Patent Classification (IPC), i.e.,

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|}$$

where $\vec{a}$ is a blocking patent family citation vector, i.e., a family of X/Y citation for an originating invention, specifically cited by an examiner at a patent office. An element of this vector is a count variable according to main class assignments of the X/Y citation, i.e., the number of main classes recorded in PATSTAT tls209 table (application IPC). No duplicate counts occur when several IPC subclasses, main groups, or subgroups in a main class are assigned to the X/Y citation. However, an X/Y citation may consists of two or more patents in the family, and it may cause two or more counts in a class of a vector. This is because each of the cited patent is assigned a classification according to national procedure, and the assignment of patent class may differ over jurisdiction, so we would like to take the classification variation in to account. Thus, more than one cited family member may result in more than one element in a main class.

By contrast, $\vec{b}$ is a global citation vector, intended as a reference vector based on all citations in the world. Namely, a bundle of all citations (including blocking and nonblocking examiner citations, as well as applicant citations) in the family are added and counted to compute the number of citations in main classes as elements of a technological vector. All DOCDB family members in the cited bundle are included to count the frequency of an IPC main class as each element of this global citation vector. Therefore, the number of counts of an IPC main class as an element of the vector may increase when many patents are cited by the original application in the bundle, and also when a cited patent has more patent family members. In summary, this cosine similarity measurement indicates how similar a blocking patent family of an office is to globally cited (blocking, nonblocking, and applicants' citations) bundles of patents.

If the cosine similarity is one, the technological vector of a particular blocking patent is identical to the technological vector of the bundle of all citations added to the same originating application worldwide. Figure 3 below shows an example. The X/Y citation in question is $\vec{a}$, a blocking patent cited in the US, which has a family member in other country. Including those patent families, all citations added to the same originating application constitute a global citation vector, $\vec{b}$.

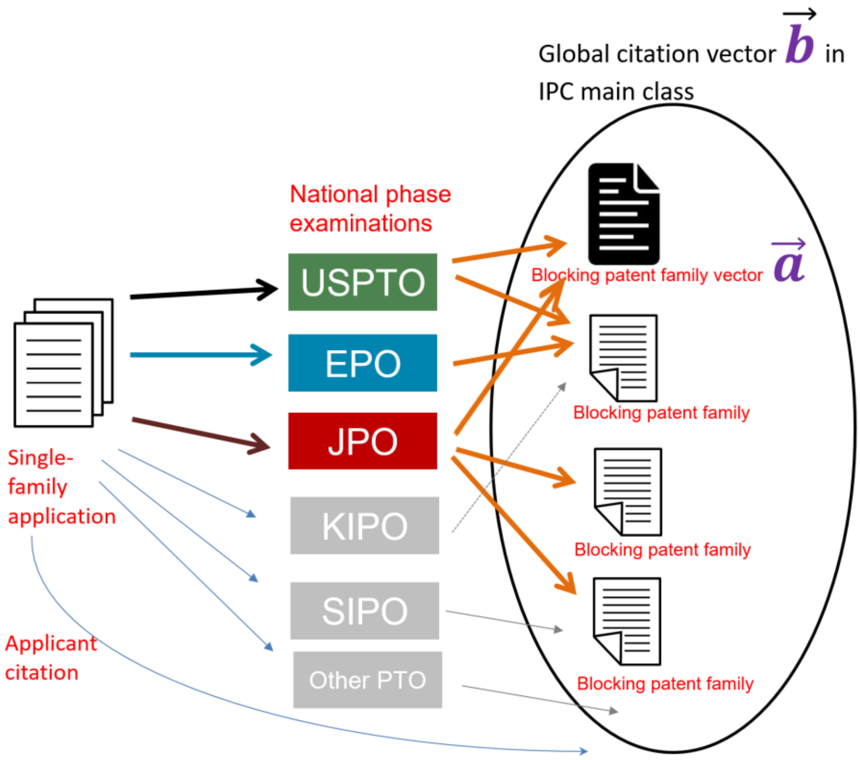


Fig. 3 Blocking citation vector $\vec{a}$ and global citation vector $\vec{b}$ for a US blocking patent

Dissimilarities by technology fields (World Intellectual Property Organization (WIPO) technology classification)

Utilizing the technology distance associated with a blocking patent, we describe the average cosine similarity distance in terms of the WIPO technology fields below (see Appendix for the names of each field). As is evident from Figure 4, the cosine similarity are systematically correlated across patent offices by field. This indicates the level of technology area at which there is a determinant of search (and decision) difficulty for examiners.

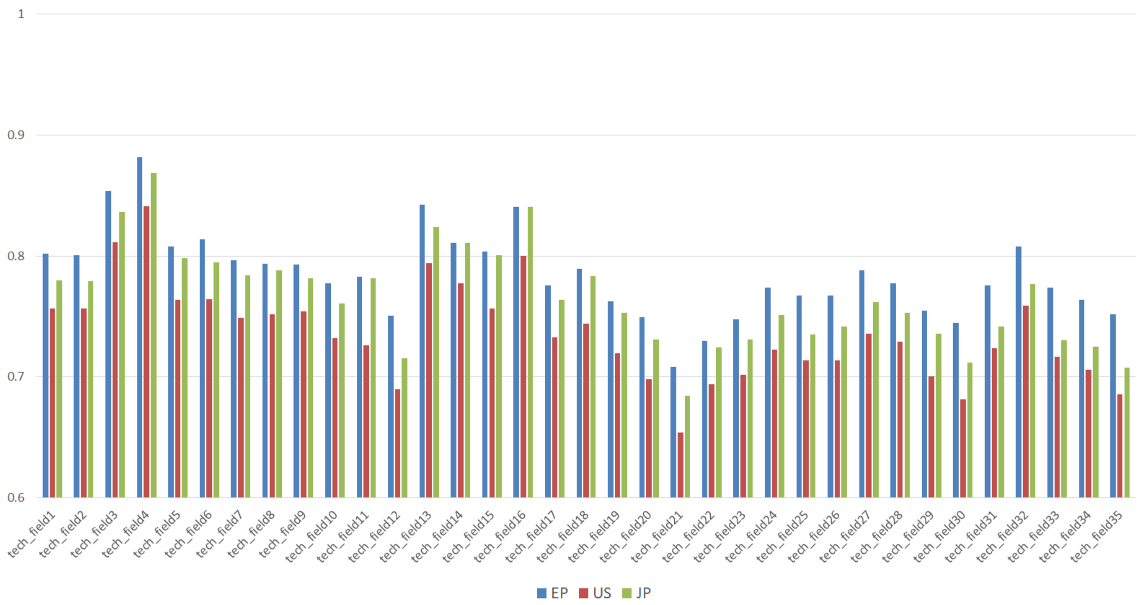


Fig. 4 Cosine similarity of X/Y citations by the trilateral offices from global citation vectors, by WIPO 35 technology fields

Complexity, longitudinal distribution, and technological distances of X/Y citations

It is virtually impossible to specify all reasons for the divergence of X/Y citations at the trilateral offices. However, we can test hypotheses for a few possible contributing factors that may change the likelihood of divergence because X/Y citations are added by human beings. Intuitively, we can hypothesize that the complexity of an application may drive the cost of examining it, which can result in different outcomes at different offices. A couple of predictions can then be made.

First, the discrepancies in X/Y patent citations may relate to the technological complexity of patent applications. This is because examiners in different offices may focus on diverse aspects of a patent application if it is technologically complex and has a broad scope. Previous studies have found that examiners have limited capabilities, depending on their experience and availability of processing time, in searching for prior art (Frakes and Wasserman, 2014). Thus, if an application is technologically complex, we expect that blocking citations added by examiners will diverge.

Moreover, the cost of evaluating an application can increase processing time. When an application is technologically complex, examiners need more time for prior art searches compared with when it is simple. Thus, we can predict that the lag between filing an application at an office and first office action may indicate the complexity of the application. The more time an examiner needs, the more divergent his/her rejection reasons will be.

To explore further the relationship between technological discrepancies in blocking citations and their longitudinal aspects, linear regression analyses were implemented, taking the cosine similarity measurement as the dependent variable. The unit of analysis is a pair of family-to-family citations. Unfortunately, office action date information at the JPO was not readily available at the time of analysis. Therefore, we can implement regression analyses only for the cosine similarity index of the EPO and USPTO data. Several explanatory variables are employed. The first is the breadth of an application, or the technological complexity of a triadic patent application, represented by the number of 35 WIPO technological fields assigned to the application (*techn_field_nr_counts*). We expect this variable to have a negative relationship with the dependent variable, the cosine similarity of an X/Y citation depicted by examiners. We also employ variables for the longitudinal aspects. The time lag in days between an entry of a triadic application into the US and the first rejection action by the USPTO is given by *us_action_lag_from_filing*. Similarly, the time lag in years between an entry of a triadic application into the EPO and the EP search report date is given by *EP_SEA_from_EP_app_year*. If an application is technologically complex, the size of this lag variable may increase, and we conjecture that this size may have a negative relationship with the similarity of refusal reasons. Unfortunately, office action timing is not available from the EPO for PATSTAT or DOCDB, except the timing of European Search Reports in years. Therefore, the lag variable for the EPO is in years, while the lag variable for the USPTO is in days.

In addition to the explanatory variables, we employ control variables for other characteristics. Specifically, *isr_cited_dummy* is a dummy variable indicating that the

X/Y citation was shown in the preceding international search reports. Whether an X/Y citation is used at other offices is coded as the binary variables of *ep_xy* and *jp_xy*, which mean the blocking patent is also an X/Y citation at the EPO and JPO, respectively. Patent grants ex post are controlled by the three dummy variables of *epo_granted*, *us_granted*, and *jpo_granted*. In addition, priority years and technological fields (35 WIPO technology field dummies) are included and controlled for citing families.

Table 4 shows the results of four estimation results. Model 1 and Model 2 employ cosine similarity between US blocking citation vector and global citation vector as the dependent variable, where Model 2 omits the lag variable at the EPO as an explanatory variable, because the focus is placed on the US blocking patents, and because the coefficient of US examination lag may correlate with the lag at the EPO. Model 3 and Model 4 employ cosine similarity concerning EP X/Y citation vector as the dependent variable, and Model 4 omits US lag variable, as in Model 2 compared with Model 1. Since the domain for Model 3 and 4 is EP X/Y citations, the explanatory variable of *ep_xy* is omitted.

As indicated by all of the linear estimation results in Table 4, both the US blocking patents and EP X/Y citations commonly show that the breadth of patent class of patent application (*techn_field_nr_counts*) is negatively and significantly correlated with the cosine similarity of refusal reasons to global citation bundle. This is consistent with an interpretation that costly examinations lead to diversified reasoning for grant/refusal decisions. On the other hand, office action lag variables show partly inconsistent results. Although the lag variable at the EPO has a negative and significant coefficient in all of the estimation results where the variable is employed, the lag at the USPTO does not have significance in the US (Model 1 and 2). Though the lag at the EPO seems to show that difficulty of publishing search report correlates with divergent X/Y citation, it is not clear from this evidence that longer evaluation times always lead to greater diversity of refusal reasons. A possible explanation is that because longer pendency in the US implies easier access to EP search reports available in the US, X/Y citations at the EPO and USPTO may become less divergent when pendency is longer in the US. This conjecture implies examination spillover between patent offices, which is beyond the scope of this paper.

Table 4 Linear regression, dependent variable: cosine similarity between X/Y citation vector (US/EP) and global citation vector. Unit of analysis: DOCDB family citation pairs

Dependent Var.	*Cosine similarity between US blocking patent vector and global citation vector*		*Cosine similarity between EP X/Y citation vector and global citation vector*	
	Model 1	Model 2	Model 3	Model 4
techn_field_nr_counts	-0.1006**** (0.00899)	-0.0984**** (0.00855)	-0.0611**** (0.0120)	-0.0699**** (0.00671)
us_action_lag_from_filing	-0.00000164 (0.00000202)	-0.00000145 (0.00000198)	-0.00000718** (0.00000319)	
EP_SEA_from_ EP_app_year	-0.00114** (0.00055)		-0.00249*** (0.000863)	-0.00168**** (0.000457)
isr_cited_dummy	0.00500 (0.00355)	0.000489 (0.00315)	-0.0128*** (0.00479)	-0.0132**** (0.00322)
ep_xy	0.0648**** (0.00182)	0.0654**** (0.00176)		
jp_xy	0.0606**** (0.00220)	0.0604**** (0.00209)	0.0228**** (0.00307)	0.0238**** (0.00169)
epo_granted	-0.00434 (0.00236)	-0.00400 (0.00218)	-0.00164 (0.00334)	-0.000324 (0.00193)
us_granted	0.0124**** (0.00239)	0.0126**** (0.00232)	0.0130**** (0.00359)	0.0123**** (0.00208)
jpo_granted	0.00614** (0.00242)	0.00702*** (0.00233)	0.00735** (0.00365)	0.00787**** (0.00201)
prio_year_2003	-0.00922 (0.00591)	-0.0116** (0.00556)	-0.00540 (0.00852)	-0.0119*** (0.00454)
prio_year_2004	-0.00155 (0.00457)	-0.00317 (0.00441)	-0.00241 (0.00706)	-0.00811** (0.00370)
prio_year_2005	-0.00975** (0.00447)	-0.0112*** (0.00434)	-0.00441 (0.00654)	-0.0103*** (0.00347)
prio_year_2006	-0.00656 (0.00440)	-0.00745 (0.00429)	0.00213 (0.00650)	-0.00575 (0.00343)
prio_year_2007	-0.00921** (0.00430)	-0.0100** (0.00420)	0.00422 (0.00644)	-0.00639 (0.00345)
prio_year_2008	-0.00958** (0.00437)	-0.00811 (0.00428)	-0.00573 (0.00642)	-0.00607 (0.00345)
prio_year_2009	-0.00251 (0.00439)	-0.00188 (0.00433)	0.0000549 (0.00642)	-0.00534 (0.00345)
tech_field1	0.0669**** (0.00945)	0.0632**** (0.00899)	0.0257** (0.0126)	0.0283**** (0.00707)
tech_field2	0.0626**** (0.00934)	0.0599**** (0.00891)	0.0199 (0.0129)	0.0252**** (0.00716)
tech_field3	0.0974**** (0.00962)	0.0945**** (0.00917)	0.0582**** (0.0131)	0.0545**** (0.00730)
tech_field4	0.1330**** (0.00958)	0.1321**** (0.00914)	0.0706**** (0.0131)	0.0878**** (0.00730)
tech_field5	0.0457**** (0.0116)	0.0421**** (0.0111)	0.0119 (0.0163)	0.0181** (0.00874)

tech_field6	0.0599**** (0.00948)	0.0582**** (0.00901)	0.0188 (0.0128)	0.0225*** (0.00712)
tech_field7	0.0816**** (0.0112)	0.0808**** (0.0108)	0.0141 (0.0184)	0.0339*** (0.00996)
tech_field8	0.0825**** (0.00967)	0.0782**** (0.00924)	0.0408*** (0.0130)	0.0458**** (0.00727)
tech_field9	0.0522**** (0.00957)	0.0510**** (0.00911)	0.0137 (0.0128)	0.0146** (0.00722)
tech_field10	0.0332*** (0.00977)	0.0307*** (0.00932)	-0.00410 (0.0133)	0.00308 (0.00734)
tech_field11	0.0710**** (0.0132)	0.0703**** (0.0123)	0.0585*** (0.0180)	0.0597**** (0.00982)
tech_field12	0.01582 (0.0105)	0.00993 (0.0102)	-0.0000895 (0.0146)	0.00313 (0.00818)
tech_field13	0.0922**** (0.00968)	0.0930**** (0.00922)	0.0504**** (0.0129)	0.0539**** (0.00720)
tech_field14	0.0733**** (0.0103)	0.0713**** (0.00981)	0.0185 (0.0142)	0.0328**** (0.00802)
tech_field15	0.0619**** (0.0115)	0.0603**** (0.0108)	0.0229 (0.0160)	0.0300*** (0.00889)
tech_field16	0.0946**** (0.0103)	0.0912**** (0.00978)	0.0563**** (0.0140)	0.0601**** (0.00801)
tech_field17	0.0832**** (0.00994)	0.0806**** (0.00949)	0.0530**** (0.0136)	0.0458**** (0.00782)
tech_field18	0.0646**** (0.0130)	0.0589**** (0.0124)	0.0376** (0.0174)	0.0228** (0.0101)
tech_field19	0.0571**** (0.0101)	0.0572**** (0.00966)	0.0208 (0.0138)	0.0258*** (0.00780)
tech_field20	0.0508**** (0.0105)	0.0480**** (0.0100)	0.0208 (0.0142)	0.0300**** (0.00808)
tech_field21	-0.00106 (0.0103)	-0.00320 (0.00984)	-0.0386*** (0.0143)	-0.0308**** (0.00815)
tech_field22	0.0704**** (0.0129)	0.0693**** (0.0120)	0.0289 (0.0197)	0.0268** (0.0113)
tech_field23	0.0409**** (0.0105)	0.0409**** (0.00999)	0.00500 (0.0140)	0.00511 (0.00790)
tech_field24	0.0666**** (0.0115)	0.0651**** (0.0108)	0.0274 (0.0162)	0.0350**** (0.0088)
tech_field25	0.0360*** (0.0105)	0.0309*** (0.0100)	0.0152 (0.0149)	0.0102 (0.0081)
tech_field26	0.0290*** (0.0109)	0.0267** (0.0104)	0.00698 (0.0145)	0.000614 (0.00820)
tech_field27	0.0159 (0.0104)	0.0156 (0.0100)	-0.0201 (0.0142)	0.00221 (0.00777)
tech_field28	0.0559**** (0.0101)	0.0499**** (0.0099)	0.0149 (0.0138)	0.0260*** (0.00781)
tech_field29	0.0509**** (0.0103)	0.0482**** (0.00985)	0.0135 (0.0135)	0.0276**** (0.0078)

tech_field30	0.00715 (0.0118)	0.00300 (0.0112)	0.00478 (0.0165)	-0.00417 (0.00894)
tech_field31	0.0409**** (0.0107)	0.0393**** (0.0102)	0.0145 (0.0148)	0.00844 (0.00819)
tech_field32	0.0655**** (0.0104)	0.0610**** (0.00990)	0.0412*** (0.0139)	0.0386**** (0.00764)
tech_field33	0.0390*** (0.0115)	0.0340*** (0.0111)	0.0123 (0.01636)	0.0223** (0.00919)
tech_field34	0.0499**** (0.0106)	0.0477**** (0.0101)	0.0259 (0.0149)	0.0205** (0.0083)
constant	0.837**** (0.00492)	0.8339**** (0.0047)	0.914**** (0.00730)	0.9114**** (0.00374)
n	271,210	292,505	46,234	135,313

Robust standard errors (clustered in citing family) are shown in the parentheses.

**** < 0.001, *** < 0.01, ** < 0.05

Conclusion

As a first look at the extent of technological discrepancies of X/Y citations, we found that the decision reasons used by examiners are very dissimilar between trilateral offices if similarity is defined in terms of coincidences between DOCDB or INPADOC patent families of X/Y citations. Second, we found evidence that is consistent with an understanding that X/Y citations at the EPO and USPTO become more divergent as the examination burden increases. This is consistent with prior findings (Frakes and Wasserman 2014; Lemley and Sampat 2012; Wada 2016) in that examiners are limited in searching capability for prior art; i.e., examiners are constrained in their information space. This result is probably the first finding on international comparison between family-level patent citation data over different jurisdiction, based on common standard of patentability to add X/Y citations.

Our understanding of the driving factors for the dissimilarity is still incomplete. For example, whereas this paper has utilized the number of technology fields only as the measurement of the breadth of a citing patent family, there are many other methods by which to measure technological diversity (Wang, Thijs, & Glänzel 2015). Also, technological distance between citing and cited families is measured only by cosine similarity, though it can potentially be measured in other ways, if we can normalize differences over languages (Kuhn, Younge, and Marco 2017). Moreover, patent claims can change during prosecution (Marco et al, 2016), and can affect the breadth of technology as well as citation, but those effects are ignored. Because of the limited data on office action timing, we have been unable to make efficient use of action lag data, especially at the JPO.

Although we have obtained insufficient detailed information on patent applications, changes of claims, the bias of examiners, and so on, this paper has added knowledge to the understanding of international family-level citations, as well as the ways in which examiner citations are generated. The methodology presented in this paper has two implications. One is that examiner citations across jurisdictions are comparable if family-to-family citations and common criteria for the X/Y citation category are employed. Furthermore, unlike normal citation-generating processes, where a citing document adds citations to prior art only once, this paper represents the first attempt to analyze a citation network with multiple citing opportunities by separate parties. We find that the variance of citation linkages depends negatively on how easily different citers evaluate prior art in the same way. Thus, the notion of breadth in citation impact has a new viewpoint. Although the analyses presented in this paper are in the early stage, and the attributes of citing entities could not be analyzed, this study has captured some factors causing citations to diverge. In other words, this paper adds a new perspective of the citation mechanism that affects the breadth of technological impact, which has usually been measured by citations without questioning the mechanism from the citers' point of view.

References

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

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

Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, 37(10), 1892-1908.

Dernis, H. & Khan, M. (2004). Triadic Patent Families Methodology. *OECD Science, Technology and Industry Working Papers*, OECD Publishing.

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. SSRN paper no. 2702637.

Frakes, M. and Wasserman, M. (2014) Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents?: Evidence from Micro-Level Application Data, *NBER Working Paper* No. 20337.

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, 105(3), 287-289.

Jaffe, A. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *American Economic Review*, 76(5), 984-1001.

Jensen, P.H., Palangkaraya, A. & Webster, E. (2005) Disharmony in international patent office decisions. *Federal Circuit Bar Journal*, 15, 679.

Kuhn, J.M., Younge, K.A., and Marco, A.C., (2017) Patent Citations Reexamined: New Data and Methods. SSRN paper https://ssrn.com/abstract=2714954

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

Lu, Q., Myers, A.F. and Beliveau, S. (2017) USPTO Patent Prosecution Research Data: Unlocking Office Action Traits. *USPTO Economic Working Paper* No. 10.

Marco, A. C., Sarnoff, J.D. and deGrazia, C. (2016) Patent Claims and Patent Scope. *USPTO Economic Working Paper* 2016-04.

Narin, F. (1994). Patent bibliometrics. Scientometrics, 30(1), 147-155.

Thompson, N., J. Kuhn. (2017). Does Winning a Patent Race lead to more follow-on Innovation? http://dx.doi.org/10.2139/ssrn.2899088

Wada, T. (2016). Obstacles to prior art searching by the trilateral patent offices: empirical evidence from International Search Reports, *Scientometrics*, 107: 701-722.

Wada, T. (2017) The choice of examiner citations for refusals: Evidence from the trilateral offices. *Proceedings from the 16th International Conference on Scientometrics and Informetrics*, Wuhan, P.R.China.

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: distinct effects of variety, balance, and disparity. *Plos One*, 10(5), e0127298.

Webster, E., Jensen, P. H. and Palangkaraya, A. (2014), Patent examination outcomes and the national treatment principle. *RAND Journal of Economics*, 45: 449–469.

Webster, E, Palangkaraya, A & Jensen, P.H. (2007) Characteristics of international patent application outcomes. *Economics Letters* 95, 362-368.

World Intellectual Property Organization. (2017). World Intellectual Property Indicators 2017. *WIPO Publication* No. 941E/17.

Table 5 Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
techn_field_nr_counts	544,403	2.328	1.207	1	10
us_action_lag_ from_filing (in days)	295,621	1112.1	474.8	38	4548
EP_SEA_from_ EP_app_year (in years)	507,642	2.513	2.074	0	11
Isr_cited_dummy	544,403	0.1029	0.3039	0	1
ep_xy	544,403	0.2684	0.4431	0	1
jp_xy	544,403	0.3823	0.4859	0	1
epo_granted	544,403	0.3924	0.4883	0	1
us_granted	544,403	0.6886	0.4630	0	1
jpo_granted	544,403	0.7050	0.4560	0	1

Appendix. WIPO technology fields

Field_number	Field_name
1	Electrical machinery, apparatus, energy
2	Audio-visual technology
3	Telecommunications
4	Digital communication
5	Basic communication processes
6	Computer technology
7	IT methods for management
8	Semiconductors
9	Optics
10	Measurement
11	Analysis of biological materials
12	Control
13	Medical technology
14	Organic fine chemistry
15	Biotechnology
16	Pharmaceuticals
17	Macromolecular chemistry, polymers
18	Food chemistry
19	Basic materials chemistry
20	Materials, metallurgy
21	Surface technology, coating
22	Micro-structural and nano-technology
23	Chemical engineering
24	Environmental technology
25	Handling
26	Machine tools
27	Engines, pumps, turbines
28	Textile and paper machines
29	Other special machines
30	Thermal processes and apparatus
31	Mechanical elements
32	Transport
33	Furniture, games
34	Other consumer goods
35	Civil engineering

(WIPO 2017, pp.220-221)