Cognitive Distances in Prior Art Search by the Triadic Patent Offices: Empirical Evidence from International Search Reports

Tetsuo Wada

tetsuo.wada@gakushuin.ac.jp
Gakushuin University, Faculty of Economics, Mejiro, Toshima-ku, Tokyo 171-8588 (Japan)

Abstract
Despite large numbers of empirical studies are conducted on examiner patent citations, few have scrutinized the cognitive limitations of officials at patent offices in searching for prior art to add citations during patent prosecution. This research takes advantage of the longitudinal gap between International Search Reports (ISRs) required by the Patent Cooperation Treaty (PCT) and subsequent examination procedure in national phase. It inspects whether several kinds of distances actually affect the probability that a piece of prior art is caught at the time of ISRs, which is much earlier than national phase examinations. Based on triadic PCT applications for all of the triadic patent offices (EPO, USPTO, and JPO) between 2002 and 2005 and their citations made by the triadic offices, evidence shows that geographical and organizational distances negatively affect the probability of prior patents being caught in ISRs, while lag of prior art positively affects the probability. Also, technological complexity of an application negatively affects the probability, whereas the size of forward citations of prior art affects positively.

Conference Topic
Patent Analysis (foundation of examiner patent citations, in particular)

Introduction
Patent citations have been widely utilized for empirical studies of patent systems, particularly for such issues as economic value and knowledge flows. Several empirical studies have examined whether examiner citations are different from inventor citations. One of the studies on the subject was conducted by Alacer and Gittleman (2006), who showed the similarity between examiner citations and inventor citations with respect to geographical distance in particular. While previous studies have compared examiner citations and inventor citations in other aspects such as the relationship with renewal rates, there have not been enough analyses concerning how patent offices are influenced by several kinds of "distances" that can limit cognitive boundary during prior art search. This study focuses on ISRs as a basis for measuring the search obstacles of the triadic patent offices, and tests how officials are bounded by "distances," including similar kinds of cognitive obstacles against prior art search, without relying on comparison with inventor citations. In conducting the analyses, we consider applicants' self-selection, since applicants from the U.S. and Japan can choose the European Patent Office as their search agency, where the EPO has reputation for its complete search (applicants who seek stringent search may choose the EPO ex ante).

The methodology: PCT and ISR as the basis of empirical measurement
This project proposes and implements a method of measuring the search obstacles, namely binding conditions on search capability, of the triadic patent offices by focusing on ISRs issued by different ISAs, specifically the patent offices in Europe, the U.S. and Japan, according to the PCT. In particular, binary choice models are employed for each of cited patents (which are added in the national phase in all of three jurisdictions) about whether or not they were already caught at the earlier time of ISR issued by the triadic offices. We limit our samples to those PCT applications made to and examined at all of the three offices. There are advantages to employ this methodology.

First, ISRs are issued under the common search criterion imposed by the WIPO under the PCT system. Under the PCT, "an applicant must file an application with a receiving office and choose an international searching authority to provide an international search report and a written opinion on the potential patentability of the invention." "The applicant generally has at least 30 months from the filing (priority) date to decide whether to enter the national phase in the countries or regions in which protection is sought" (WIPO, 2014). The guideline at the WIPO applies to every ISA when issuing ISRs, whereas applicants in some countries are allowed to choose ISAs. The same criterion for prior art search is applied over different patent offices, while national phase examinations do not have such standardized rules.

Second, the lag mentioned above between ISRs and national phase examinations allows a "level" testing ground for search completeness. While ISRs are issued at an early stage, more searches are conducted in national offices later. Since knowledge is geographically localized (Jaffe et al. 1993; 1999), and knowledge diffusion takes time, additional time between ISRs and national phase search facilitates more complete search in the later stage. We limit our samples to those PCT applications that are examined at all of the three triadic offices, meaning that localized knowledge in any of these areas at the time of ISRs is more likely to be caught by the offices at the national phase in a less localized way. See Figure 1 below for the lag and collective searches made at later stages in national phase.

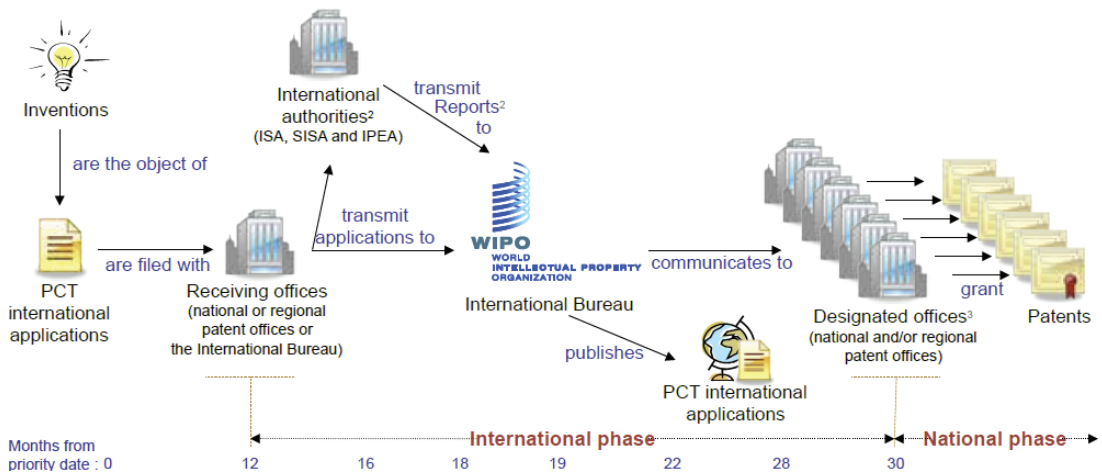

Figure 1. PCT procedure (replicated from WIPO, 2014, p.13).

Following the logic above, we retrospectively define the probability of every cited patent depicted in national phase, identified at the INPADOC family level, to have been already caught in the ISR of the originating PCT application. Taking this probability (a binary variable *found_in_ISR*, empirically) as the dependent variable, we implement PROBIT analyses at INPADOC family level with explanatory variables representing the various "distances" between citing and cited patents, including technological complexity of originating applications, and other related indicators.

Applicants' (inventors') citations are excluded from the analysis, since the objective is to evaluate the determinant of search completeness by the ISAs. However, self-selection of the U.S. and Japanese applicants to choose the EPO as their ISA is considered in the analyses, since the EPO has high reputation of examination standard and therefore applications with higher quality from the U.S. and Japan may choose the EPO as the ISA.

Although actual ISR search is sometimes outsourced to non-PTO agencies, we consider ISRs as a basis of evaluating PTOs, since they are issued under the name of the patent offices, not private search agencies. Only citations made by the triadic offices are considered in the current analyses. Since PATSAT, our primary data source, records non-patent literature in

non-standardized formats, we could not consolidate the same non-patent literature across different records. For this reason, we employ patent citations only at this time.

Hypotheses

Since ISR searchers (examiners/searchers for patent offices) are affected by cognitive obstacles from various "distances," we hypothesize that a prior patent (that was found in ISR or national phase) is more likely to be found in ISR when "distances" are less problematic, i.e.,
H1) a relevant prior patent is closer in geography (physical distance),
H2) prior patent is older (knowledge diffusion time),
H3) prior patent is from the same applicants (organizational distance),
H4) prior patent has more number of forward citations (knowledge diffusion probability), and
H5) application for which an ISR is issued has less scope, less number of claims, less number of inventors, and less number of international family (complexity against diffusion).
In addition, we consider if applicants' self-selection of ISAs affects the outcome variable.

Data source

The empirical domain of analysis is the triadic patent applications through PCT, with their earliest priority date within its international family between 2002 and 2005. Triadic PCT patent applications are defined here as INPADOC families that contain all of EPO, USPTO and JPO applications recorded on EPO's PATSTAT database, with only one "WO (PCT)" application in a family, meaning that a single PCT application initiates international phase for all applications in a family. The number of international families for the analysis is 97,828. Although international applications to and from China and Korea has increased dramatically in the last ten years, the triadic patent offices of the EPO, the USPTO and the JPO represented the vast majority before 2005, which is our observation period.
EPO PATSTAT (2013 OCT version) is used, and INPADOC family is the unit of analysis. Citation data also comes from PATSTAT (2013 OCT), although JPO citation data is augmented by Seiri-Hyojunka data (JPO's standardized patent prosecution data). US citations are not complete as well on PATSTAT, since citations for rejected applications are not registered on PATSTAT. The lack of the US citations for rejected applications may affect the result of the analysis, but this has not been verified yet. Applicant identifiers are consolidated by the EEE-PPAT database developed by ECOOM (Du Plessis et al., 2009; Magerman et al., 2009; Peeter et al., 2009).

Variables

We employ several categories of explanatory variables, representing each of hypotheses above, in PROBIT analyses taking the probability of a cited patent being caught in the previous ISR as the binary dependent variable ("*found_in_ISR")*. The unit of analysis is a pair of citing and cited international families, both consolidated at INPADOC family level.
For H1, three variables of *euro_cited* (cited family has its 1st priority, i.e., the earliest date, in EPC countries within a family, derived from tls201 and tls219 tables of PATSTAT), *us_cited* (cited family has its 1st priority in the U.S.), and *jp_cited* (cited family has its 1st priority in Japan) are defined. When a cited family has its origin in the same region where ISR is issued, the ISA of the region is expected to have geographical advantage over the relevant technology. Expected sign is positive for each region, e.g., positive *jp_cited* coefficients for applications originating from Japan.
For H2, citation lag between the 1st priority of a citing family and that of a cited family is defined as *fam_cite_lag* (derived from tls201 and tls219 tables of PATSTAT). The longer the lag is, the easier the prior art will be to be found at the time of ISR.

For H3, *self* is defined as a binary variable, taking the value of one if one of patents in a cited family and one of patents in a citing family belongs to the same applicant, based on PATSTAT (tls207) combined with EEE-PPAT, using "L2" id. Patent office will find it easier to locate prior relevant art within the same applicant.
For H4, *fwd_cite_of_the_cited* is defined and obtained from PATSTAT (tls217) as the number of forward examiner citations, counted at publication level (but consolidated at family level), and made out to the cited patent family.
For H5, we first use scope indicators. *IPC4_count* is the total net count of IPC subclasses (4-digit IPC, derived from tls209) assigned in a citing INPADOC family. Since patent classification of an application may change during prosecution process both in international phase and in national phase, we include all IPC subclasses to capture the breadth of a family. The number of claims of a patent is correlated with the complexity of the technological content. As an indicator of the number of claims, we obtain *publn_claims_max_tls211*, which is the maximum number of claims registered on PATSTAT (tls211 table) in a citing INPADOC family. We do not simply rely on claims data from a single office such as from the EPO, since an application can be modified during its prosecution internationally. We also employ *invt_nr*, the maximum number of inventors in an application included in a citing INPADOC family, from PATSTAT (tls207). The size of international family, *family_size*, is a count variable of applications in different countries in a citing INPADOC family (tls211/219).
In addition to the variables above, which are used to test hypotheses directly, we define three variables to address self-selection of ISAs by applicants. The first two represent the potential of the applicant. The first of the two is *total_count*, which is the number of total applications that an applicant has made, taken from EEE-PPAT. The second one is *applicant_avg_cited*, which is the number of average forward citations that an applicant has received, calculated by PATSTAT (tls212) and EEE-PPAT. Both are supposed to represent the experience level of the applicant, and are used as instrument variables for instrumented PROBIT on the variable *ISA_CHANGED*. This binary variable *ISA_CHANGED* indicates that the U.S. and Japanese applicants choose the EPO as their ISA (the EPO can be chosen from the U.S. and Japanese applicants, but not vice versa). This information can be obtained for PCT applications on PATSTAT, since the citation table tls212 has a field on "citation origin" where "ISR" is shown for PCT applications. Since first application country (RO) in a family is available from tls201, switching from RO to a different ISA can be coded. The correlation coefficient between *ISA_CHANGED* and the dependent variable *found_in_ISR* is low at 0.0348.
Control variables for originating areas, which are *JP_app* and *US_app* (applications from Japan and the U.S., respectively), are used. Technology class is controlled by thirty-five WIPO technology classification dummies (results not shown for space reason).

Estimation results

The result shown in the Model 1 of Table 1 employs all samples from the triadic regions. As is evident from the negative sign for *JP_app* and *US_app*, the baseline ISA (EPO) is found to be advantaged in finding prior art at the time of ISR. The positive sign of *ISA_CHANGED* also indicates that prior art is easier to be identified at the time of ISR if applicants from the U.S. or Japan choose the EPO as their ISA (for which robustness is checked in Model 4 and 5). These are consistent with the EPO's good reputation from international applicants. H1 is supported from the positive sign of *euro_cited*. Likewise, H2, H3, H4 and H5 are all supported o this model, except that the number of inventors has an insignificant coefficient.
Model 2 uses applications from Japan only in order to examine the locality of knowledge in Japan. As is expected in H1, *jp_cited* has a positive and significant sign, whereas *us_cited* has negative and significant sign. Other variables show similar results with the Model 1 and are consistent with hypotheses, except *self* indicates the negative sign. Model 3 uses U.S.

applications only, and the results are just consistent with the hypotheses. Model 4 and 5 limit the citation data to non-self citations only for robustness checks, while employing two instrument variables on the variable *ISA_CHANGED*. For Japanese applications, the coefficient for *ISA_changed* lost the significance in the Model 4, suggesting that the advantage provided by the ISA change from JPO to EPO is due to the applicants' self-selection. However, this effect is not observed for the U.S. applications in the Model 5.

Table 1. PROBIT analyses on the probability of ISR coverage; dep. var.=*found_in_ISR.*

Model 4 and 5 use "*total_count*" and "*applicant_avg_cited*" as instruments for "ISA_CHANGED."

****<0.001 ***<0.01 **<0.05 Robust standard errors are in the parentheses (clustering on citing family).

Model & sample	Model 1 (all of triadic samples/ baseline=EP_app)	Model 2 (JP app only)	Model 3 (US app only)	Model 4 (JP app & non-self only)	Model 5(US app & non-self only)
method	Probit	Probit	Probit	IV Probit	IV Probit
euro_cited	0.1419984**** (0.0080393)	-0.031025 (0.0160179)	0.1776262**** (0.0120059)	0.0203394 (0.0174625)	0.148418**** (0.0253879)
us_cited	-0.0620007**** (0.0078305)	-0.3377195**** (0.0155267)	0.050351**** (0.0114757)	-0.2974986**** (0.0169034)	0.0777813**** (0.0159886)
jp_cited	0.0393056**** (0.0082601)	0.8054234**** (0.0151802)	-0.4295359**** (0.0121628)	0.8367819**** (0.0175193)	-0.3751166**** (0.0427623)
fam_cite_lag	0.0030127**** (0.000212)	0.0023379**** (0.0004175)	0.0046464**** (0.000329)	0.0005303 (0.0004425)	0.0026492**** (0.0005495)
self	0.2091817**** (0.0047187)	-0.1759722**** (0.0082345)	0.1123806**** (0.0076398)		
fwd_cite_of_the_cited	0.0000359**** (0.00000321)	-0.00000566 (0.00000781)	0.0000573**** (0.00000437)	-0.00000566 (0.00000799)	0.0000551**** (0.00000526)
IPC4_count	-0.0165033**** (0.0013614)	-0.0176023**** (0.002381)	-0.0215867**** (0.0022476)	-0.0170435**** (0.0026306)	0.0099131 (0.011092)
publn_claims_max_tls211	-0.0080901**** (0.0001942)	-0.0029271**** (0.0003468)	-0.0094453**** (0.0002733)	-0.0033284**** (0.0004149)	-0.0081833**** (0.0010323)
invt_nr	0.0000932 (0.0011831)	-0.0007108 (0.002112)	-0.0058672*** (0.0018111)	0.0008906 (0.0023144)	-0.0089979*** (0.0026535)
family_size	-0.006626**** (0.0007439)	-0.0142835**** (0.0021553)	-0.0053694**** (0.0011327)	-0.0091501*** (0.0032126)	-0.0138593**** (0.002496)
JP_app	-0.0667862**** (0.0069462)				
US_app	-0.2808785**** (0.0072769)				
ISA_CHANGED	0.3096426**** (0.0066579)	0.2758815**** (0.0169662)	0.380766**** (0.0074961)	0.0109491 (0.1314658)	1.35421**** (0.3121653)
Technology class dummies	included	included	included	included	included
n	1031127	325990	455830	264805	363328

Discussion and further development

Overall results are consistent with the hypotheses, suggesting that examiners (and searchers working for the PTOs) are bound by various kinds of "distances," including technological complexity of applications. These are intuitive, and are supported by the novel methodology for the first time. An interesting interpretation is that examiners (unlike inventors) are

required to find prior art by law, but that they are naturally bound by informational horizons they have. This has policy implications, since Patent Prosecution Highways (PPH) rely on outcomes from previous patent offices. Most prior studies using examiner citations do not incorporate these informational obstacles born by examiners, but they cannot be ignored. For example, prior studies on the difference of examination outcomes between patent offices (Jensen et al., 2005; Webster et al., 2007, 2014) do not explicitly consider them, but the cost of prior art search may affect the results. The results with instrument variables suggest the self-selection is working, but is evident for the Japanese samples only. Further scrutiny is needed.

Acknowledgments

This interim output is drawn from the collaborative project with Professor Setsuko Asami (Tokyo University of Science) and Professor Yoshimi Okada (Hitotsubashi University). The entire project is supported by RISTEX/JST. The comparison of search quality between PTOs at aggregated level was previously presented at the International Workshop on Patent System Design for Innovation at Hitotsubashi University (Wada and Asami, 2014) and at the 2014 Annual Conference of the Asia-Pacific Innovation Network. The idea of the probability of ISR coverage of this paper evolved out of the idea of aggregate ISR coverage ratio, which Professor Asami first thought of. The author also acknowledges the support from the MEXT/JSPS (Grant #22330122) for the analyses of citations and firm boundaries.

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.

Du Plessis, M., Van Looy, B., Song, X & Magerman, T. (2009). Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation. EUROSTAT Working Paper and Studies, Luxembourg.

Jaffe, A., & Trajtenberg, M. (1999). International knowledge flows: evidence from patent citations. *Economics of Innovation and New Technology 8*, 105-136.

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

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

Magerman T, Grouwels J., Song X. & Van Looy B. (2009). Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization." EUROSTAT Working Paper and Studies.

Peeters B., Song X., Callaert J., Grouwels J., & Van Looy B. (2009). Harmonizing harmonized patentee names: an exploratory assessment of top patentees. EUROSTAT working paper and Studies.

Wada, T., & Asami, S. (2014). Quality comparison of International Search Reports (ISRs) by selectable International Search Authorities (ISAs) under the Patent Cooperation Treaty (PCT) system, *a paper presentation at the 2014 International Workshop on Patent System Design for Innovation,* Hitotsubashi University, Tokyo, Japan.

Webster, E., Jensen, P. H. & Palangkaraya, A. (2014). Patent examination outcomes and the national treatment principle. *The 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. (2014). Patent Cooperation Treaty Yearly Review: The International Patent System. *WIPO Economics and Statistics Series*.