A New Proxy for Strategic Patents and Its Application to Japanese and US Patent Data (*)

Invited Researcher: Henri de BELSUNCE (**)

This paper develops a new proxy for the strategic patents of firms and applies it to the Japanese IIP patent dataset and the US NBER patent dataset. Specifically, we identify patent clustering à la Rivette and Kline (2000) in the data ex-post. Initial results - in line with previous findings in the literature – indicate that strategic patenting is an important problem in Japan, even more so than in the US. We acknowledge caveats to our analysis, in particular with respect to the controlling for patent portfolio sizes. These drawbacks are focal points for further research, but this paper already presents the potential of a fully-fledged proxy for strategic patenting that could be of great use for regulators and policy makers at the patent office.

I Introduction

Many studies rely on patent data to investigate the effects of patent scope on innovation (Qian[1] 2007, Motohashi[2] 2004, Lerner[3] 2009). In these studies innovation is often proxied for by observable patent statistics. However, these statistics as measures of innovation may be flawed when firms file patents for strategic reasons.

There exist numerous theoretical papers on how firms may strategically want to optimise their patent portfolio. The consensus of these papers lies on the point that firms may have an incentive to shade their patent applications in order to achieve other goals than just the protection of an idea. Possible reasons may include the blocking of the research of competitors, enhancing royalty income or, more generally, create barriers to entry for competitors.

An important strategy discussed in the literature is the creation of patent thickets (Shapiro[4], 2001), where firms demand many, overlapping patent rights. When entering a market in the presence of intellectual property (IP) protection requires either the licensing or invalidating of those patent rights, these IP rights create barriers to market entry. The larger and more obscure a patent portfolio of an incumbent is, the more costly it is for entrants to tackle the issue of dealing with these existing IP rights. A theoretical foundation that rationalises excessive patenting as a means to reduce market entry is given in Belsunce[5] (2014). In practice, a significant strategy of firms, which leads to the creation of patent thickets, is that of *patent clustering* (Rivette and Kline[6], 2000), where firms file for additional related patents of minor quality around a core patent that protects a valuable innovation. These additional related patents are often of low quality and represent a burden on the patent system (Moore[7] 2005), let alone exacerbate the pressure on the patent offices. It is a common argument that low quality patents are inefficient and should be avoided. However, devising policies reducing the use of weak patents is very complicated since they are difficult to identify ex-ante.

In order to study the problem of weak patents, identifying them ex-post is a vital step. We contribute to this task by developing a statistic to identify the additional related patents (called *excess patents*) in the vicinity of valuable patents in the data. Such a proxy is mandated by the fact that no clear theoretical prediction exists for the strategic reaction of firms to patent reforms.

Our approach is purely based on metadata, i.e. citation patterns, and thus easily implementable for use in practice. The analysis is composed of (i) devising a measure of proximity between patents within a firm's patent portfolio. The separation of a firm's patents into core or related patents is conditional on the two being related. We use backwards citation overlap as our measure of the proximity. (ii) Among patents that are technologically close to each other, we distinguish the core from related patents by using forward citation numbers as a proxy for their impact on future innovations. (iii) We combine the two measures into an overall proxy, which can be read similar to a probability, that a patent is a core or a related, excess patent. We then look at averages of these proxies for patents of yearly cohorts to better understand the evolution of strategic patenting over time and across countries.

Our results indicate that excessive patenting has been constant over time in both the US and Japan. Furthermore, our proxy indicates that excessive patenting

(*) This is a summary of the report published under the 2014FY Industrial Property Research Promotion Project (2014FY-2016FY) Entrusted by the Japan Patent Office.

(**) Doctoral Research Fellow, Max Planck Institute for Innovation and Competition, at our institute over a period of approximately 5 months from May 25, 2015 through October 17, 2015, as an Invited Researcher for the Fiscal Year 2015.

has been on average higher in Japan than in the US – a result that coincides with earlier findings by IIP[8] (1998).

While these findings are interesting, we emphasise some important caveats - that have not been thoroughly tested yet - in the discussion in section VI of the full paper.

Still, this work and the developed proxy may be of particular importance for policy makers and the patent office concerned with the large number of patent applications. The indicator of excess patenting developed here could help to identify strategic patent applicants and help devise policies to prevent such behaviour. In particular, the proxies developed seem to be adapted for the analysis of individual firm's patent portfolios. Avoiding excess patents may bring a social welfare improvement as it could foster market entry and thus competition. Such a rationale may be even more valid in a context with many start-ups and small and medium enterprises. These companies are particularly sensitive to barriers to entry created by intellectual property due to the expertise required in dealing with protection rights. We look forward to more research in this domain.

This analysis is strongly related to the literature on devising statistics to measure patent thickets. Graevenitz et al.[9] (2009) are the first to devise a proxy to measure the density of patent thickets directly.

Two further important indices used to identify patent thickets have been proposed in the literature. The fragmentation index developed by Ziedonis[10] (2004) and the blocking index proposed by Siebert and von Graevenitz[11] (2008). In contrast to their approach, our proxy for strategic patenting does not rely on the relationship of patent portfolios between different companies, but solely focuses on individual firm's patenting activity. The angle of analysis of patent thickets is thus not the same.

We also note that Clarkson[12] (2005) derives an identifier of patent thickets, by analysing the citation density of patent networks.

With respect to Japanese data, two papers that study the patent thicket problem are particularly notable. Doi and Zhang[13] (2014) use the blocking index of Clarkson[14] (2005) and Zhang et al.[15] (2013) to show that the threat of patent blocking has a positive, significant effect on patent licensing. Nagaoka and Nishimura[16] (2014) use both complementarity and fragmentation of IP rights as proxies for patent thickets in order to investigate their impact on the first-mover-advantage that is understood to be a key driver of R&D activities. We contribute an analysis of a newly devised proxy for patent thickets on Japanese patent data. Furthermore, we compare the results for Japan with those from the US.

II Data

For Japan, we use the IIP patent database, the NISTEP company dataset[17] and the inventor citation dataset from Artificial Life Laboratory for the analysis. The first dataset consists of Japanese patent applications since the late 1960s until 2012. Its specificities are detailed in Goto et al.[18] (2007).

For the United States, we use the NBER patent database. This dataset contains detailed information on patents granted in the US between 1976 and 2006. The specificities of this data are detailed in Hall et al.[19] (2001). Furthermore, company characteristics are given by the upfront match to firms listed in the COMPUSTAT database.

We make a few data adjustments in our analysis, mainly in order to ensure comparability between the US and Japanese datasets. All adjustments with notes on their impacts can be found in the full report. We highlight that the resulting subsamples are representative of the full samples.

III Empirical strategy

1 Overview of Identification Strategy

We aim to identify patents in the data, which are likely to be of a strategic nature. Specifically, we focus on the type of excessive patenting described in detail by Rivette and Kline[20] (2000), called patent clustering. This refers to strategically patenting weaker, related ideas around a core invention. We see this as a driver for patent thickets, when firms file for patents on patentable side-aspects of a core invention since we expect the risk of an IP overlap between the core patent and the related patent to be high.

We devise a proxy, which identifies these patents. We take care to clean this proxy of noise in the data, e.g. truncation and inflation of patent citations – for details see section IV-3. We investigate the evolution of this statistic over time and, by applying the same methodology to patent data from the US, compare the results for both countries and aim to trace back differences to fundamentals of the patent system.

The identification of these excessive patents is based on three main components, which are detailed in the following paragraphs. First and given that our focus is on excessive patenting in vicinity of a core invention, we assume that the core and related patent are both filed by the same company[21]. Companies have the index j in what follows.

In the second step, we devise alpha, our measure of technological similarity of a patent and the other patents in that firm's patent portfolio. This measure is based on the citation overlap of patent i of firm j with all other patents in firm j's patent portfolio. The simplest example is a patent portfolio consisting of two patents A and B. If patent A has

3 backwards citations and patent B has 4 backwards citations, 2 of which are identical to the backwards citations of A, then the citation overlap percentages are 66% and 50% respectively, indicating a measure of similarity between the patents.

When there are more than two patents in a firm's portfolio, alpha does not measure the similarity between pairs of patents, but measures the similarity of each patent with all the other patents together that exist in that firm's portfolio at the time of application of the patent in question. In the previous example, add a patent *C*, which has a single citation that *C* only shares with patent *A*. Then overlap percentages are: 100% for patent *A*, 50% for patent *B* and 100% for patent *C*.

This percentage gives an abstract measure of the technical proximity of a patent to the other patents in the portfolio of firm *j* at the time of application of the patent in question. The higher this figure is, the closer we expect the patents to be and thus a possible situation where excessive patenting may occur is given. However, while alpha indicates which patents may be part of a group of core and related patents, it does not give us information on which patent is of which type, i.e. which one is a core or an excessive patent. We do this in the next step.

In the third step, we devise beta, our measure of quality (or impact) of the patents and use this to separate the core from excessive patents. This measure ultimately relies on Trajtenberg[22] (1990), who demonstrated a strong correlation between forward citations of a patent and a measure of the (social) value of the patents. This measure of patent value is also confirmed by Lanjouw and Schankerman[23] (1999) as well as by Harhoff et al.[24] (1997) and has been used for example by Hall and Ziedonis[25] (2001). In our simple previous example of patent *A* and *B*: If *A* has 8 forward citations and *B* has 1 forward citation, then we consider that patent *A* is significantly more valuable than patent *B*.

Finally, we combine alpha and beta into a single proxy, called gamma, as follows:

$$\gamma_{i,j} = \frac{\alpha_{i,j}}{(1 + \beta_{i,-j}{}^{2})}$$

where index *i* refers to the patent, *j* to the firm and –*j* refers to patents of other firms. Thus, we ignore forward self-citations when computing beta. Furthermore, Trajtenberg[26] (1990) finds that the utility of forward citations as an indicator of patent value is increasing in the number of citations, i.e. the informational content of a citation increases at the margin. We incorporate this insight by squaring the impact of beta. This squared term enables us to get a clearer separation of high value patents compared to low value patents.

We encourage the reader to refer to the full report for further details on the implementation of this algorithm and the precise definition of the relative measures of alpha, beta and gamma.

Gammadash, which will be presented only in this summary, is defined as the ratio of alphadash and (1+betadash^2). Alphadash compares the degree of citation overlap of the patents of firm *j* with the average citation overlap of all comparable patent portfolios. Betadash measures the relative quality in comparison to all patents of the same cohort.

We will use this proxy for analysis – however, we recommend the reader to take note of the caveats of this proxy described in the full report.

IV Results

In the following we present parts of the results. We show the computed measures for gammadash. The left figures present results for Japan, the right ones refer to the US. Furthermore, the graphs represents the average over all technology classes.

1 Gammadash – Japan/US Comparison

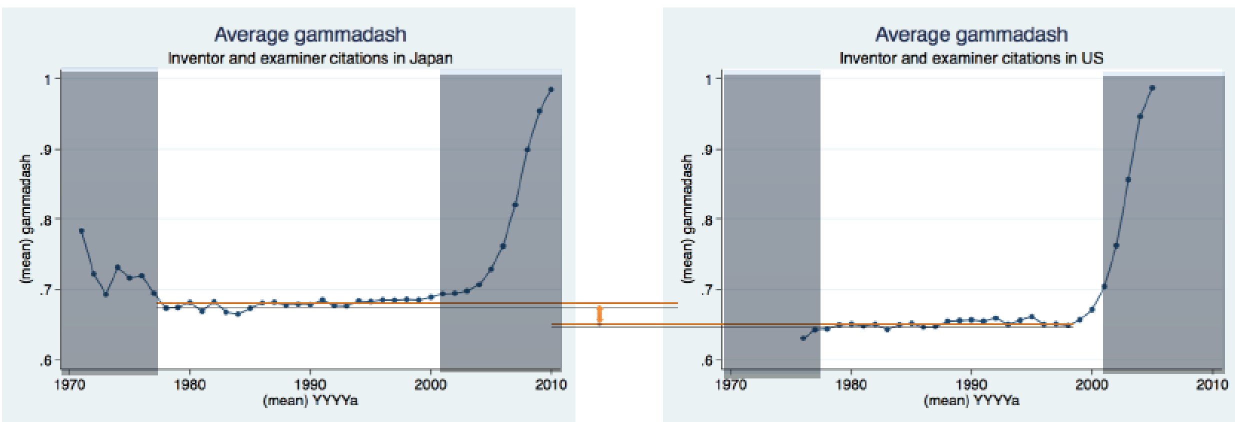

Figure 8: Evolution of average gammadash over time in Japan and the US.

Figure 8 summarises much of this work. For reasons outlined in the full paper, we focus on the timeframe from approx. 1980 to 2000, indicated by non-shaded parts of the top row of figure 8. Gammadash is a combination of the relative measures alphadash and betadash, which both measure patents' similarity and value in comparison to their peers, i.e. other patents applied for in the same year and of the same technology class. Combined, we take these measures as an indicator whether granted patents may be classified as excessive patents that have been filed for strategic reasons, e.g. for a patent clustering strategy.

First, we observe that gammadash is constant in the timeframe of our analysis. This is an interesting result, since it states that, on average and controlling for time trends, excessive patenting has remained roughly constant over time. This insight lends support to the understanding that the use of strategic patents has prevailed to a similar extend over the two decades from 1980 to 2000.

Second, we observe that gammadash is on average higher in Japan than in the US. This result is particularly interesting as it suggests that Japanese firms engage in strategic patenting activity to a higher extend than US firms. Furthermore, this result coincides with previous findings by IIP (1998)[27]. IIP compares the magnitude of patent costs, which are composed of application and maintenance costs of patents, in relation to operating figures (e.g. total R&D costs or Sales) in Japan and the US. The report finds that the ratio of patent costs as percentage of sales is equal to 0.20% in Japan and 0.10% in the US. Similarly, the ratio of patent costs as a percentage of R&D costs is significantly higher in Japan than in the US (4.71% versus 2.69%, respectively). While his results suggest that Japanese firms spend additional resources on their IP strategies, our results is congruent with his and indicates that the high fund allocation of Japanese firms to their patenting activities could be related to excessive patenting.

While graphs of the average values of alphadash and betadash would not yield any insights[28], gammadash shows variation because it is essentially a matching of the two measures of alphadash and betadash. Gammadash is an average taken over all patents, not within firm portfolios. Hence gammadash is not constant, even under firm symmetry, because it depends on which value of alphadash "meets" a value of betadash. And since, alpha and beta are not perfectly correlated, we observe some meaningful variation.

However, we acknowledge that so far, we do not control for firm size in our analysis – a factor that might be an important driver of our results. Especially, if we assume that the US is composed of a high number of firms with patent portfolios of size one[29]. This is a valid concern since we use the COMPUSTAT firm identifier for the US database, which is significantly wider than the NISTEP firm identifier in Japan[30]. By definition, alpha (and thus gamma) for patents of size-one-portfolios will be zero, since there is no patent owned by the same firm to measure a citation overlap with. Adding large numbers of single-patent-owners to the averages, will introduce a strong downwards bias of the indicators. This is an important caveat and we emphasise its relevance for the results. Controlling for this factor is a natural extension of this piece of work and vital to take the proxy further towards usefulness in practice.

V Discussion

While the newly developed proxy is not yet in its final version, we believe that this work has many qualities. First of all, it is a transparent measure of the strategic patenting behaviour of firms, which is computed at the level of the individual patents and may be aggregated to portfolio or even country level. Thereby, the perspective of analysis is flexible and adapted for policy makers.

Second, our proxy for excessive patenting is computed on metadata only. This means that it relies solely on patterns in the citation data and does not require any specific analysis of the individual patent's contents in the first place. Computationally, this is considerably less demanding and it therefore becomes a feasible tool for the screening of large patent landscapes and allows to flag patent of specific interest for further analysis.

We discuss the potential impact of firm portfolios sizes in the full and thus are cautious to rely overly on the present results from the cross-country comparison. Our results, however, motivate further work on the development of this proxy.

The potential of this proxy is significant. Being able to capture the strategic filing nature of individual patents in a single index would allow to condition broader studies on the effect of patent rights on innovation on the strategic behaviour of players in the patent system. From a patent office perspective, a broader study on the determinants of strategic patenting could be developed using our gamma proxy as the outcome variable. Such a regression would allow the regulator to anticipate strategic patenting behaviour by firms and fine-tune the patent system in accordance.

VI Conclusion

This paper is a first step in devising a new proxy for strategic patenting, specifically excessive patenting, in the vicinity of a core innovation. Initial results - in line with previous findings in the literature – indicate that strategic patenting is an important problem in Japan, even more so than in the US. We acknowledge caveats to our analysis, in particular with respect to the controlling for patent portfolio sizes. These drawbacks are focal points for further research, but this paper already presents the potential of a fully-fledged proxy for strategic patenting.

[1] Yi Qian, "Do National Patent Laws Stimulate Domestic Innovation in a Global Patenting Environment? A Cross-Country Analysis of Pharmaceutical Patent Protection 1978-2002", The Review of Economics and Statistics, Vol. 89, No. 3 (Aug., 2007), pp. 436-453

[2] Motohashi, Kazuyuki. "Japan's patent system and business innovation: reassessing pro-patent policies." Patents, innovation and economic performance: OECD conference proceedings. OECD Publishing, 2004

[3] Josh Lerner, "The Empirical Impact of Intellectual Property Rights on Innovation: Puzzles and Clues", The American Economic Review, Vol. 99, No. 2, Papers and Proceedings of the One Hundred Twenty-First Meeting of the American Economic Association (May, 2009), pp. 343-348

[4] Shapiro, Carl (2001), "Navigating the Patent Thicket: Cross Licenses, Patent Pools and Standard Setting." In Innovation Policy and the Economy, Volume 1, 119–150, MIT press.

[5] de Belsunce, Henri „Do More Patents Mean Less Entry?" Working paper, 2014. Please ask for the latest version. An interesting quote from the discussion with the CEO of a start-up company in an industry with important IP rights reveals that artificially inflated patent portfolios are indeed a problem when competitors and investors „count patents by the kilo" (referring to the fact that the length and number of patents often trumps the intellectual quality of the patents).

[6] Rivera, Kevin G. and David Kline (2000). Discovering new value in intellectual property. *Harvard Business Review*, *55*.

[7] Moore, Kimberly A. (2005), "Worthless Patents." Technical report, George Mason University School of Law Working Papers Series : 27.

[8] Institute of Intellectual Property. (1998), Research and Study on the Economic Analysis of the Intellectual Property System, 130-131.

[9] Graevenitz, Georg v. and Stefan Wagner and Dietmar Harhoff, "How to measure patent thickets – a novel approach", Economic Letters, Vol. 111, pages 6-9 (2011)

[10] 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.

[11] Siebert, R., & von Graevenitz, G. (2008). *Does Licensing Resolve Hold Up in the Patent Thicket?* (No. 2008-01). Discussion Papers in Business Administration.

[12] Clarkson, G. (2004). Objective identification of patent thickets: A network analytic approach. *Harvard Business School Doctoral Thesis*.

[13] Doi, Noriyuki & Xingyuan Zhang (2014): Patent Thickets and Licensing:Empirical Findings from Japanese Listed Companies. OECD.

[14] Footnote 20.

[15] From Zhang, Xingyuan, Michael R. Ward and Yoshifumi Nakata. 2013. "The market value of patenting: New findings from Japanese firm level data." Unpublished manuscript., based on source in footnote 21.

[16] Nagaoka, Sadao and Yoichiro Nishimura, "Complementarity, Fragmentation and the Effects of Patent Thickets", RIETI Discussion Paper Series 14-E-001 (2014)

[17] Japanese Name: „Nistep Kigyomei Jisho"

[18] Goto, Akira and Kazuyuki Motohashi, "Construction of a Japanese Patent Database and a first look at Japanese Patenting Activities", Research Policy, Vol. 36, Issue 9 (2007)

[19] Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498.

[20] See footnote 9.

[21] For Japanese data, we use the NISTEP company identifier since it is more stable over time (i.e. across mergers of applicants, name changes, etc.) than the applicant number from the IIP patent database. See the discussion in section VI for the limitations due to that choice. For the US data, we use the provided company identifier of the NBER patent database (which originally is drawn from the Compustat database).

[22] See footnote 4.

[23] Lanjouw, Jean O., and Mark Schankerman. The quality of ideas: measuring innovation with multiple indicators. No. w7345. National Bureau of Economic Research, 1999.

[24] Harhoff, Dietmar, et al. "Citation frequency and the value of patented inventions." *Review of Economics and statistics* 81.3 (1999): 511-515.

[25] Hall, Bronwyn H., and Rosemarie Ham Ziedonis. "The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995." RAND Journal of Economics (2001): 101-128.

[26] See footnote 4.

[27] Institute of Intellectual Property. (1998), Research and Study on the Economic Analysis of the Intellectual Property System, 130-131.

[28] By definition of these measures, the average over all patents is defined equal to 1, i.e. a straight horizontal line.

[29] In other words, companies that own only a single patent.

[30] See section III for details on the datasets.