Hitotsubashi University
Institute of Innovation Research

Impact of Patent Scope on subsequent Inventions: Findings from a new Measure

Yoshimi Okada
Yusuke Naito
Sadao Nagaoka

IIR Working Paper WP#17-03

Feb. 2017

Institute of Innovation Research
Hitotsubashi University

Tokyo, Japan
http://www.iir.hit-u.ac.jp

Impact of Patent Scope on subsequent Inventions: Findings from a new Measure[†]

Yoshimi Okada[a*], Yusuke Naito[b], Sadao Nagaoka[c]

a Institute of Innovation Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan
b Institute of Innovation Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan
c Faculty of Economics, Tokyo Keizai University, 1-7-34 Minamicho, Kokubunji, Tokyo 185-0021, Japan

February 2016

Abstract

While patent scope defined by patent claims provides crucial information on the contribution of underlying inventions to the state of the art, its existing measures do not seem to appropriately capture it, especially with respect to the generality of the inventive concept. This study investigates how significantly the breadth of the first claim can predict the patent's knowledge impact on subsequent inventions in complex and discrete technologies using the inverse of the first claim length as the indicator. There are two major findings. First, this indicator has very significant predictive power for the knowledge impact of the underlying invention as measured by applicant forward citations, controlling for two existing indicators of patent scope (the number of patent claims and the number of different patent classification codes assigned) in both technology areas. Second, its predictive power for the incidence of top-ranked patents increases in higher quantiles in the complex but not the discrete technology area, unlike the other indicators. This is consistent with an economic model predicting that the knowledge impact of an invention with broad scope has a high variance, depending on the emergence of complementary inventions that enhance the impact of the initial invention.

Keywords: patent scope; claim breadth; first claim length; knowledge; complex technology
JEL classification: O34

[†] This work was supported by JSPS KAKENHI Grant Number 26885031 and 16K13380. We would like to thank Research Assistant Junko Morikawa for checking the samples.

1. Introduction

A patent system promotes innovation not only by enhancing the appropriability of inventions but also by promoting their disclosure and thereby the knowledge available for subsequent inventions. Surveys among R&D managers and inventors confirm that patent literature is, in fact, one of the most important sources of information suggesting new research projects and contributing to their implementation (see Giuri et al. [2007] for the inventor survey in Europe and Walsh and Nagaoka [2009] for surveys in the US and Japan).

In a patent, the very crucial technological information, the contribution of the underlying invention to the state of the art, is indicated in the independent claim(s), especially in the first independent claim[1] because the scope of the patent is defined by claims and the first claim generally conveys the broadest inventive concept that must satisfy the novelty and the inventive step requirement. As the generality of the inventions will increase with the breadth of the scope of patents, we can expect that a patent with a broad scope has a large potential knowledge impact: many follow-on inventions for cumulative improvements, various applications in different technological areas, and a wide range of combination possibilities with other inventions, in particular, in complex technology areas. This paper examines how significantly a new measure of the breadth of claim

---

[1] Claims that define all the essential components of the invention by itself are called 'independent claims.' Claims that refer to one or more previous claims in order to avoid redundant expressions are called 'dependent claims.' A dependent claim does not by itself define all the components of the invention.

scope, the inverse of the first claim length of a patent, can help predict its knowledge impact on subsequent inventions in complex and discrete technologies. Of the information that has been directly obtained from patent claims, the existing empirical studies predominantly have used the number of claims of a patent to assess the breadth of the scope of the patent right (see, for example, Lanjouw and Schankerman [2004]). The breadth of independent claims, however, is a distinct dimension of the patent's scope, that is, the generality of the inventive concept, which cannot be captured adequately by simply counting the number of claims. While a patent with many claims often covers a large number of variations or application areas known at the stage of the patent filing, a patent application with few but broad claims may be filed for an embryonic invention that is followed by a large number of subsequent complementary inventions. It is thus highly likely that we miss substantial information on the patent's scope if we do not measure the breadth of claim scope directly. Partially based on this understanding, other studies, since Lerner 's (1994) study, have used the number of different International Patent Classification (IPC) subclasses assigned to each patent as a proxy for the patent's scope. It is again highly likely, however, that this measure is very imprecise because the indicator is not directly derived from the claims and depends on the specific structure of the patent classification system. The first objective of this study is to demonstrate that the inverse of the description length of the first claim of a patent has significant predictive power with respect to the knowledge impact of the invention on subsequent inventions, even controlling for major existing patent scope indicators, including the number of claims and the number of different

IPC subclasses, examining a large scale of Japanese patents.

More specifically, we test the inverse of the number of characters involved in the first independent claim in a patent as a new proxy for the breadth of the claim scope of the patent in predicting the knowledge flow to subsequent inventions, based on the following intuition. The description of an independent claim grows longer as the number of elements limiting the scope of the exclusive right increases; thus, one can expect that the inverse of the description length of an independent claim is positively correlated with the breadth of the scope of this claim. Because, in general, the first independent claim conveys the broadest inventive concept in a patent, we focus on this claim and use the inverse of the description length of the first independent claim as the measure of the breadth of claim scope, examining whether this indicator significantly predicts applicant forward citations of the underlying invention.

Moreover, as a combination of complementary inventions is important for innovations in a complex technology area, it is likely that the knowledge impact of an invention depends heavily on how extensively the complementary inventions will emerge in this technology area. We thus expect that the knowledge impact of a patent with a broad claim has a more skewed distribution in the complex technology area than in the discrete technology area.[2] Given that a patent with a broad claim is likely to have more technological opportunities for combination with future complementary inventions, such a patent will appear significantly more frequently among top-ranked patents in

[2] Cohen, Nelson, and Walsh (2000) discuss how these two areas differ, focusing on appropriability conditions.

terms of the scale of its knowledge impact on subsequent inventions. On the other hand, such a relationship between the breadth of claims and the skewed distribution of the impacts on subsequent inventions will be less apparent in the discrete technology area, as an invention in such a technology area is more likely to be exploited as a stand-alone invention and will depend less on combinations with other inventions. The second objective of this study is to examine whether such a difference in the dependence of knowledge impact on the breadth of claim scope exists between complex and discrete technology areas.

Despite the potential importance of the first claim length as a predictor of the knowledge impact of underlying inventions to subsequent inventions, to the best of our knowledge, there are no systematic studies examining this subject. Furthermore, no studies have examined how the effect of the breadth of claim scope differs in terms of its impact on subsequent inventions, depending on the nature of the innovation process, namely, the difference between complex and discrete technology areas. This study aims to uncover how well the first claim length predicts the knowledge flow to subsequent inventions and how such predictive power differs in the complex and discrete technology areas, depending on the rank of the impact to subsequent inventions.

The rest of the paper is organized as follows. The next section briefly reviews prior literature. Section 3 offers a discussion of our data construction. Section 4 presents a model of an invention’s knowledge impact that depends on subsequent complementary inventions in order to clarify why the quantile regressions are essential for our analysis as well as the potential underlying mechanism of the observed skewed outcome. Section 5 presents the results and discussion. Section 6 concludes.

2. Prior literature

Although a significant number of empirical studies exist on patent value and its skewness (Scherer and Harhoff 2000; Harhoff, Scherer, and Vopel 2003; Nagaoka, Motohashi, and Goto 2010; Squicciarini, Dernis, and Criscuolo 2013), empirically, the scope of a patent is difficult to operationalize and measure (Harhoff, Scherer, and Vopel 2003). Most of existing empirical studies use the number of claims and the number of different IPC subclasses to assess the breadth of the scope of a patent right. Lanjouw and Schankerman (2004) use the number of claims in addition to forward and backward citations and family size as the value indicators of a patent. Lerner (1994) proposed the number of different subclasses (based on the four-digit code) that a patent examiner assigned to the patent as a proxy for patent scope, focusing on the technological scope based on the International Patent Classification (IPC) scheme, and demonstrated its explanatory power of the forward citations of a patent and the value of start-up firms in the biotechnology industry. Since then, some have followed Lerner’s approach (Shane 2001; Harhoff Scherer, and Vopel 2003; Nerkar and Shane 2003) and others furthered this approach, identifying rare combinations of different IPCs (Fleming 2001; Verhoeven, Bakker, and Veugelers 2016). Novelli (2015) discusses the difference in dimensions that are measured by the number of claims and by the number of different IPC subclasses. However, neither the number of claims nor the number of different IPC subclasses directly measure the generality of the inventive concept, thus, these indicators, even if combined, measure patent scope imperfectly.

Only a very limited number of prior studies use description length of independent claims as a measure of the breadth of patent scope. Malackowski and Barney (2008) used the average number of words per independent claim in an issued US patent as a proxy of the quality of the patent examination by the United States Patent and trademark Office (USPTO). Their assumption was that the constancy of invention quality applied for the patent over the period. They found that the number of words in independent claims increased through the examination process, from 111.1 on filing to 153.2 on issuance (that is, the patent scope narrowed on average). They also found that independent claims of US patents issued in 2007 had 4.4% more words compared to that of patents issued in 2003, which was considered as evidence that the quality of the patent examination had not decreased during that period. Although they examined the length of independent claims as a proxy for the quality of the patent examination procedure, they did not examine its explanatory power as a predictor of the knowledge contribution of the underlying invention to subsequent inventions.

Harhoff (2016) states that the inverse of the length of the first independent claim is used as a proxy measure of patent breadth by analysts at patent authorities. He points out that the average number of words of European Patent intendant claims became significantly longer over time: 165

words in 1990, 175 words in 2008, and 183 words in 2014; and attributed such change to the narrower scope in the reformed fee structure and the restriction of divisional applications. However, the relation between the claim length and the social impact of a patent is unknown.

Jansen (2009) examined the relationship between the private patent value and the claims by investigating about 2700 European patents in technological fields relevant to the company Phillips. Contrary to practitioners' views, he concluded that the description length of independent claims was not a significant predictor of the patent value, unlike the number of claims, controlling for the family size. He may have got this result because his estimation was based on the assessment of the mean effect of the claim length on the value ranking of a relatively small number of patents (2700), mostly in the complex technology area.

Regarding text data other than claims, Reitzig (2004) examined how the number of words describing the state of the art and those describing the technical problem predict the likelihood of opposition, together with procedural indicators (such as accelerated examination request). However, he did not investigate the first claim length or the difference between complex and discrete technology areas.

Webster, Jensen, and Palangkaraya (2014) investigated the incidence of changes in independent claims in 236 European patents (EP) and 82 Japanese patents (JP). They found that foreign inventors filing patent applications at the European Patent Office (EPO) and the Japanese Patent Office (JPO) were more likely to receive a grant with claim changes than domestic inventors.

The purpose of their investigation was to investigate a potential violation by national patent offices of the obligation of respecting the national treatment principle stipulated in the Trade Related Aspects on Intellectuals Properties in the World Trade Organization (WTO) agreement. Osenga (2011) focused on the number of words in claims. However, he used the claim length as a proxy of readability or comprehension of claims and not as a proxy of breadth of patent scope.

3. Data construction

Generally, the first independent claim conveys the broadest inventive concept and we focus on this claim. Because the Japanese language does not use spaces between words and thus it is hard to count the number of words automatically, we use the number of characters instead of the number of words for the metric of the description length of the first claim in a patent. Hereafter, 'first claim length' denotes the number of characters in the first claim. We measure the breadth of the claim scope of a patent by the inverse of the first claim length and referred to it as 'IFCL.'

We prepared our data sets from Japanese patent databases purchased from Artificial Life Laboratory, Inc., which covers all text data in patent publications as well as applicant citation data. In order to identify self- and non-self-citations, we utilized a dictionary of all major Japanese company names and the connection table for patent application provided by the National Institute of Science and Technology Policy (NISTEP) and considered only those patents filed by applicants identified by the NISTEP database. Further, we used PATSTAT (2014 autumn, the European Patent Office) to obtain US patent family data.

Considering that the relationship between the IFCL and the impact of the underlying invention on subsequent inventions may differ significantly between product and process patents, we developed a program that divided patents into categories of ‘product’ and ‘process.’[3] We focus on product patents in the main part of our analysis here, which account for more than 80% of total patents.[4] We find very similar results for process patents, which are reported in Appendix B.

We restrict our assessment to the patents filed between January 1991 and August 2002. The ending limit (August 2002) was set to eliminate the influence of the patent law change related to disclosure of prior arts on our measures of backward citations as well as to secure a wide window for applicant forward citation flow. The beginning limit was set because of the availability of the patent claims text data. For simplicity, we eliminated the divisional applications from our data sets.

We divided the technological fields into six large categories: ‘Chemical (Chem.),’ ‘Computers & Communications (C & C),’ ‘Drugs and Medical (D & M),’ ‘Electrical & Electronic (E & E),’ ‘Mechanical (Mech.),’ and ‘Others’ (see Appendix Table A.1 for the IPC correspondence

---

[3] The randomly chosen 150 samples showed no errors.

[4] The rates of product invention by technology are as follows: Total, 81.4%; Complex, 82.8%; Discrete, 78.3%; Computers & Communications, 85.8%; E & E, 79.1%; Mechanical, 84.5%; Chemical, 65.5%; Drugs and Medical, 90.9%; and Others, 83.4%.

table). ‘C & C,’ ‘E & E,’ and ‘Mech.’ belong to the complex technology area, and ‘Chem.,’ ‘D & M’ and ‘Others,’ the discrete technology area.

We eliminated the patents involving chemical structure formulae, mathematical formulae, and/or tables in the first claim utilizing ‘tag information’ and its ilk, as the crucial part of the patent right is provided by image data rather than text data in these patents. The amount of eliminated data was less than 5% in each field except for Chem.[5]

Figure 1 shows the distribution of the natural logarithm of the IFCL of product patents aggregated by all fields. Its shape is similar to that of a normal distribution.

[Figure 1 near here]

We use the number of forward applicant citations as an indicator of knowledge impact. Forward citations have been extensively used in prior literature as indicators of knowledge flow, although it is well recognized that it is a very noisy measure (Jaffe, Trajtenberg, and Fogarty 2000). We do not use examiner citations because they are added by examiners ex post when examining the patent application and they contain many patent applications that the applicants did not know when they had been engaging in the research of the focal invention. Our data contain applicant citations

[5] The rates of the eliminated data are: all, 3.9%; Complex, 1.8%; Discrete, 9.0%; C & C, 0.4%; E & E, 3.6%; Mech., 1.0%; D & M, 4.8%; and Chem., 25.9%; Others, 1.1%.

that are constructed by searching all patents or patent applications cited by the applicants in entire patent application documents, which include not only the citations of the early patent literature (granted as well as applications) cited as prior art but also citations cited as references to research tools or methods used to invent or implement the focal inventions. We thus believe that our measure is less noisy and more comprehensive in capturing important knowledge contributions to subsequent inventions.

4. Analytical framework and estimation model

4.1 Knowledge impact of an invention on subsequent inventions

We consider a statistical model explaining the variation of the knowledge impact of an invention where the emergence of complementary inventions stochastically enhances the knowledge utility of the focal initial inventions. This process seems to be much more important in the complex technology area than in the discrete technology area. We assume that the social value of the invention as a knowledge source for subsequent inventions (v) increases multiplicatively with the occurrence of N complementary inventions ($z_i > 1$)[6]:

[6] The private value of invention itself is highly skewed and follows a log normal distribution (Scherer and Harhoff 2000). This suggests that both the value and the knowledge impact follow the multiplicative process.

$$v = qz_1z_2 \dots . z_N \quad (1)$$

Here q is the stand-alone knowledge impact of the focal invention for subsequent inventions with the expected value of its logarithmic value $\overline{lnq}$ and its variance σ^2_{lnq}. Variable z_i represents the multiplicative contribution of a complementary inventions, which will emerge after the focal invention. If we regard v as the number of forward citations, it will include citations by N complementary inventions, but this number is dominated by other citations when N is large, as v increases exponentially with N. We assume that the contributions of the complementary inventions are independently distributed among themselves and the logarithmic value has a common mean $\overline{lnz}$>0 and a common variance σ^2. In this case, we have

$$E(ln\ v) = \overline{lnq} + N\overline{lnz} \quad (2)$$

$$Variance(ln\ v) = \sigma^2_{lnq} + N\sigma^2 \quad (3)$$

We assume that the breadth of claim scope (θ) affects its impact on subsequent inventions not only through $\bar{q}$ but also through N when the opportunity for the emergence of complementary inventions exists: N increases with the breadth of claim scope (θ), considering that an embryonic invention invites many complementary inventions, which thereby enhances the knowledge value of the original invention. This means that an invention with broader claims has a high variance in the social value of the invention as knowledge for subsequent inventions (v). For simplicity, let us further assume the following:

$$\overline{lnq} = \alpha_0 + \alpha_1\theta,\ with\ \alpha_1 \geq 0 \quad (4)$$

$$N = \beta_c\theta,\ with\ \beta_c \geq 0 \quad (5)$$

Given this, we have the following:

$$(ln\ v|s) = \alpha_0 + (\alpha_1 + \beta_c\overline{lnz})s \quad (6)$$

Denoting the inverse of the standard normal cumulative density function by $\varphi^{-1}(\tau)$ for quantile τ, the conditional quantile function (which shows the value for quantile τ for a given patent scope breadth) is given by

$$Q_\tau(ln\ v|s) = \alpha_0 + \left(\alpha_1 + \beta_c\overline{ln\theta}\right)s + \sqrt{\sigma_{lnq}^2 + \beta_c\theta\sigma^2}\varphi^{-1}(\tau) \quad (7)$$

It follows from equations (6) and (7) that the quantile regression coefficient of the breadth of claim scope θ is larger than that of ordinary least squares (OLS) for upper quantiles. Importantly, if the breadth of claim scope contributes to the knowledge impact of inventions entirely by increasing the number of complementary inventions and has only negligible effect on their stand-alone knowledge impact on average ($\overline{lnz} \cong 0$) and ($\alpha_1 = 0$), the above equation (6) cannot detect the importance of the breadth of claim scope, even if patents with broad claims enhance significantly the subsequent inventions as knowledge sources because they disproportionately exist among top impact patents[7]. On the other hand, quantile regression can detect this, as shown in equation (7).

4.2 Estimation model

Given the above analytical results, we use both the OLS method as well as the quantile regression method, which can accommodate the possibility that the impact of a patent with a broad claim scope on subsequent inventions has a high variance. We assessed how the IFCL can be utilized to predict the number of forward applicant citations, adding to the predictive power of the combination of all conventionally used major indicator variables, including the number of claims, the number of different IPC subclasses, the number of inventors, the existence of corresponding US patents,[8] the existence of corresponding patent applications to the EPO, backward applicant citations (self and non-self), backward examiner citations, and the existence of divisional applications based on the focal patent as well as technology by priority year dummies.

Specifically, we used the following model where Q_x is the x quantile or mean. We choose the following three quantiles (.70, .90 and .99) for the presentations of our estimations, given that close

---

[7] The patents in the top 10% quantile account for more than 80% of the aggregate value (Scherer and Harhoff 2000).

[8] The US adopted a pre-grant publication system for patent applications filed on and after November 29, 2000. Before this legal reform, a patent application document was published only when a patent was granted. Hence, we used the existence of corresponding US patents instead of US patent applications.

to a half of the patents do not have forward citations[9] and we focus especially on the predictive power of the IFCL on top impact patents.

$$
\begin{aligned}
&Q_x(ln\_N\_F\_Citation|independent\ variable)\\
&= \beta_0\ ln\_IFCL + \beta_1\ ln\_N\_claims\ +\ \beta_2\ ln\_N\_IPC\_subclass\\
&+ \beta_3\ ln\_N\_inventors\\
&+\beta_4\ US1EP1\_dmy + \beta_5\ US1EP0\_dmy + \beta_6\ US0EP1\_dmy\\
&+ \beta_7\ ln\_N\_Bnonself\_Citn + \beta_8\ ln\_N\_Bself\_Citn\\
&+ \beta_9\ ln\_N\_Bttl\_ExCitn + \beta_{10}\ ln\_N\_Bttl\_CoCitn\\
&+ \beta_{11}\ subAppln\_dmy\ \ + \beta_{12}\ joint\_dmy\\
&+ \beta_{year}\ filing\_year\_dmys + \beta_{tech}\ tehnology\_dmys\\
&+ \beta_{year,tech}\ filing\_year\_dmys\ \times\ tehnology\_dmys\\
&+ constant +\ \varepsilon \qquad (8)
\end{aligned}
$$

The definitions of the variables and descriptive statistics are shown in Table 1 for product patents; descriptive statistics for process patents are presented in the Appendix Table B1.

[Table 1 near here]

---

[9] The values of 50, 70, .90 and .99 quantile of *ln_N_F_Citation* for product patents are as follows, respectively: complex, 0.00, 0.69, 1.61, 2.89; discrete, 0.00, 0.69, 1.61, 3.04; for process patents: complex, 0.00, 0.69, 1.61, 3.00; discrete, 0.69, 1.10, 1.79, 3.14.

5. Results and discussion

Table 2 shows the summary results for complex and discrete technology areas in the case of product patents. According to the results of the OLS estimation, the signs of the coefficients for the IFCL (indicated as ln_IFCL) are positive and statistically highly significant in the two technology areas. The IFCL, therefore, has a significant explanatory power for predicting the applicant forward citations, even controlling for the existing major indicators. Its explanatory power in the complex technology area, however, is only a quarter of that in the discrete technology area (0.014 vs. 0.051). Conversely, the estimated coefficients for the number of claims and the number of IPC subclasses are similar (0.073 (complex) vs. 0.071 (discrete) and 0.067 (complex) vs. 0.064 (discrete)) and statistically significant in both areas. Although the explanatory power of the IFCL is smaller than that of the number of claims and the number of different IPC subclasses, it is still comparable even if we consider the size of the standard deviations of the IFCL, the number of claims, and the number of different IPC subclasses (0.51 (0.61), 0.82 (0.76) and 0.43 (0.45), in the complex (discrete) technology areas, respectively).

[Table 2 near here]

The results of the quantile regressions reveal that the predictive power of the IFCL increases significantly in higher quantiles in the complex technology area (the highest (0.16) in the 0.99 quantile; and negligible (0.005) in the 0.70 quantile), whereas it is stable for all quantiles in the discrete technology area (0.060 in the 0.99 quantile; and 0.062 in the 0.70 quantile). As a result, the IFCL predicts the 0.99 quantile far more in the complex technology area than in the discrete technology area, but much less in the 0.70 quantile. Figure 2 compares the coefficient value for the IFCL relative to that of the number of different IPC subclasses and Figure 3 compares the coefficient value for the IFCL relative to that of the number of claims. Given that the number of different IPC subclasses and the predictive power of the number of claims are relatively stable across all quantiles in both technology areas, the IFCL predicts the 0.99 quantile more in complex technology than in discrete technology, but much less in the 0.70 quantile in complex technology than in discrete technology.

Table B.2, Figure B.1, and Figure B.2 in the Appendix show that very similar findings (that is, the rise of the relative predictive power of the IFCL in higher quantiles in the complex technology area and its relative stability in the discrete technology area) hold for process patents too.

[Figure 2 near here]

[Figure 3 near here]

Figure 4 shows the relative values of the coefficients estimated by 99% quantile regression against those estimated by the OLS method for each explanatory variable for product patents. The value for the IFCL in the complex technology area is outstanding, more than 12, whereas others are less than four; the values are less than two for the number of IPC subclasses and the number of claims. Thus, the IFCL very significantly predicts top ranked patents in terms of applicant forward citations in the complex technology area. Figure B.3 in the Appendix shows the corresponding values for process patents with similar results.

The different effect of the breadth of claim scope between the two technology areas is consistent with the prediction from the analytical framework introduced in Section 4.1 and supports the following interpretation. In the complex technology area, the impact of a pioneering patent with a broad claim on subsequent inventions depends on the emergence of many complementary inventions due to the necessity of combining many technologies to produce a product; hence, the impact of a patent with a broad claim in the complex technology area is more uncertain and tends to have a high variance. In contrast, the value of a patent in the discrete technology area depends much more on standalone value, and thus, the impact of a patent with a broad claim on subsequent inventions is more certain.

The IFCL therefore detects a very important and hitherto unrecognized characteristic of patent scope in the complex technology area, as articulated by our analytical framework explained in Section 4.1: patents in the complex technology area can display a higher variance of impact on subsequent inventions due to the varied enhanced impact depending on complementary inventions. The IFCL is therefore a unique proxy for the breadth of claim scope in that it effectively explains the difference between complex and discrete technology areas in terms of the mechanism of the contribution of a high impact patent on subsequent inventions.

[Figure 4 near here]

6. Conclusions

While patent scope defined by patent claims provides crucial information on the contribution of underlying inventions to the state of the art, its existing measures (the number of claims and the number of different IPC subclasses) do not seem to appropriately capture it, especially with respect to the generality of the inventive concept. This study proposed a new measure of the patent scope, the IFCL (inverse of the first claim length), and assessed the significance of this new measure in predicting the knowledge impact of the invention. The results validate that this new measure significantly predicts the number of applicant forward citations, controlling for the conventional proxies for patent scope and other major bibliographic indicators. The explanatory power of the IFCL is indeed comparable with that of the number of claims and the number of different IPC subclasses. Moreover, whereas its predictive power is stable across the 70% to 99% quantiles in the discrete technology area, it becomes increasingly more significant for top-ranked patents in the complex technology area, unlike the other measures of patent scope. These findings are consistent with an economic model predicting that the impact of a pioneering patent with few limitations on its claim on subsequent inventions is more uncertain in the complex technology area where such impact depends significantly on the emergence of complementary inventions that enhance the impact of the initial focal invention. Unlike the conventional two proxies of patent scope, this new IFCL proxy therefore provides important hitherto unexploited information on a patent scope's crucial characteristic, which is a source of the variation of the skewness of its impact on subsequent inventions (especially high-impact patents).

Our findings open up a number of follow up research possibilities. One immediate question is how significantly the IFCL can contribute to accounting for the highly skewed distribution of private patent values, and another question is how the IFCL could be used as a measure of the contributions of patent examinations. We are currently addressing these questions. One important policy implication will be that, whereas ex-post value indicators, such as forward citations or the renewal of patent maintenance, need a relatively long time to measure, the IFCL provides one useful ex-ante information for a policy maker in identifying potentially high-spillover inventions both in the complex and discrete technology areas, which may be useful in monitoring the performance of R&D projects funded by the government in the future.

References

Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2000. *Protecting Their Intellectual Assets: Appropriability Conditions and Why the US Manufacturing Firms Patent (or Not).* No. w7552. National Bureau of Economic Research.

Fleming, L. 2001. "Recombinant Uncertainty in Technological Search." *Management science* 47 (1): 117-132.

Giuri, P., M. Mariani, S. Brusoni, G. Crespi, D. Francoz, A. Gambardella, W. Garcia-Fontes et al., 2007. "Inventors and Invention Processes in Europe." *Research Policy* 36: 1107–1127.

Harhoff, D., F. M. Scherer, and K. Vopel. 2003. "Citations, Family Size, Opposition and the Value of Patent Rights." *Research Policy* 32 (8): 1343-1363.

Harhoff, D. 2016. "Patent Quality and Examination in Europe." *The American Economic Review: Papers & Proceedings 2016* 106 (5): 193-197.

Jaffe, A., M. Trajtenberg, and M. Fogarty. 2000. "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." *American Economic Review* 90: 215-218.

Jansen, W., 2009. *Examining the Relation Between Patent Value and Patent Claims*. Eindhoven: Eindhoven University of Technology.

Lanjouw, J. O., and M.A Schankerman. 2004. "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators." *The Economic Journal* 114: 441–465.

Lerner, J. 1994. “The Importance of Patent Scope: An Empirical Analysis.” *The RAND Journal of Economics*: 319-333.

Malackowski, J., and J. Barney. 2008. “What is Patent Quality? A Merchant Banc's Perspective.” *Nouvelles-Journal of the Licensing Executives Society* 43 (2): 123-134.

Nagaoka, S., K. Motohashi, and A. Goto. 2010. “Patent Statistics as an Innovation Indicator.” *Handbook of the Economics of Innovation* 2: 1083-1127.

Nerkar, A., and S. Shane. 2003. “When do start-ups that exploit patented academic knowledge survive?” *International Journal of Industrial Organization* 21 (9): 1391-1410.

NISTEP. 2015. Company name dictionary.

URL: http://www.nistep.go.jp/research/scisip/data-and-information-infrastructure (accessed 04,17.,15).

NISTEP. 2015. Connection table to other databases. URL:http://www.nistep.go.jp/research/scisip/data-and-information-infrastructure (accessed 04,17, 15).

Novelli, E. 2015. “An Examination of the Antecedents and Implications of Patent Scope.” *Research Policy* 44 (2): 493-507.

Osenga, K. 2011. “Shape of Things to Come: What We Can Learn from Patent Claim Length.” *Santa Clara Computer & High Tech.* LJ 28: 617.

Reitzig, M. 2004. “Improving Patent Valuations for Management Purposes—Validating New Indicators by Analyzing Application Rationales.” *Research Policy* 33 (6): 939-957.

Scherer, F. M., and D. Harhoff. 2000. “Technology Policy for a World of Skew-Distributed Outcomes.” *Research Policy* 29 (4): 559-566.

Shane, S. 2001. “Technological Opportunities and New Firm Creation.” *Management science* 47 (2): 205-220.

Squicciarini, M., H. Dernis, and C. Criscuolo. 2013. *Measuring Patent Quality: Indicators of Technological and Economic Value*. No. 2013/3. OECD Publishing.

Verhoeven, D., J., Bakker, and R. Veugelers. 2016. Measuring Technological Novelty with Patent-Based Indicators. *Research Policy* 45 (3): 707-723.

Walsh, J. P., and S. Nagaoka. 2009. “How ‘Open’ Is Innovation in the US and Japan? Evidence from the RIETI-Georgia Tech Inventor Survey.” *Research Institute of Economy, Trade and Industry Discussion Paper* 09-E-022.

Webster E., P. H. Jensen, and A. Palangkaraya. 2014. “Patent examination outcomes and the national treatment principle.” *The RAND Journal of Economics* 45 (2): 449–469.

Table 1. Descriptive statistics for the data (product patents)

Variables	Definition	Complex N=601, 435		Discrete N=231,841	
		Mean	Std. Dev.	Mean	Std. Dev
ln_N_F_Citation	logarithm of “the number of forward citations + 1”	0.553	0.733	0.605	0.775
ln_IFCL	logarithm of the inverse of *first-claim- length*	-5.78	0.506	-5.51	0.614
ln_N_claims	logarithm of the number of claims	1.18	0.818	1.05	0.763
ln_N_IPC_subclass	logarithm of the number of IPC subclass	0.308	0.434	0.335	0.450
ln_N_inventors	logarithm of the number of inventors	0.586	0.603	0.745	0.601
US1EP1_dmy	1 if both the corresponding US patent and the corresponding EP application exist; otherwise, 0	0.0933	0.291	0.0681	0.252
US1EP0_dmy	1 if there is a corresponding US patent but no corresponding EP application; otherwise, 0	0.143	0.350	0.0457	0.209
US0EP1_dmy	1 if there is no corresponding US patent and there is an EP patent application; otherwise; 0	0.00816	0.0899	0.00100	0.0993
ln_N_Bnonself_Citn	logarithm of “the number of backward non-self-citations + 1”	0.272	0.489	0.345	0.580
ln_N_Bself_Citn	logarithm of “the number of backward self-citations + 1”	0.143	0.354	0.180	0.391
ln_N_Bttl_ExCitn	logarithm of “the number of “examiner citations + 1”	1.28	0.638	1.15	0.675
ln_N_Bttl_CoCitn	logarithm of “the number of backward citations cited both by the applicant and the patent examiner + 1”	0.101	0.268	0.124	0.300

subAppln_dmy	1 when subsequent divisional applications exist; otherwise; 0	0.0406	0.197	0.0414	0.199
joint_dmy	1 if the patent is jointly held; 0 if the patent holder is single	0.0613	0.240	0.0705	0.256
filing_year_dmy	earliest priority year of application				
technology_dmy	6 Categories				

Table 2. Summary results of the regression (product patents)

	Dependent variable: *ln_N_F_Citation*							
	Complex (N = 601,435)				Discrete (N = 231,841)			
Independent variables	OLS	.99 quantile	.90 quantile	.70 quantile	OLS	.99 quantile	.90 quantile	.70 quantile
ln_IFCL	.0135***	.158***	.0385***	.00511**	.0511***	.0599***	.0768***	.0615***
	(.00192)	(.0113)	(.00444)	(.00214)	(.00283)	(.0143)	(.00624)	(.00345)
ln_N_claims	.0725***	.156***	.125***	.0674***	.0712***	.146***	.117***	.0724***
	(.00133)	(.00757)	(.003)	(.00152)	(.00229)	(.0119)	(.00516)	(.00276)
ln_N_IPC_subclass	.0666***	.141***	.110***	.0683***	.0644***	.135***	.105***	.0689***
	(.00225)	(.0131)	(.00508)	(.00281)	(.0038)	(.0200)	(.00869)	(.00445)
ln_N_inventors	.0711***	.173***	.133***	.0711***	.0597***	.117***	.100***	.0568***
	(.00166)	(.00951)	(.00383)	(.00205)	(.0027)	(.0148)	(.00624)	(.00325)
US1EP1_dmy	.201***	.486***	.363***	.298***	.295***	.536***	.478***	.414***
	(.00388)	(.0205)	(.00849)	(.00509)	(.00814)	(.0372)	(.0172)	(.0107)
US1EP0_dmy	.115***	.270***	.209***	.161***	.155***	.293***	.291***	.200***
	(.00293)	(.017)	(.0067)	(.00552)	(.00859)	(.0338)	(.0213)	(.0143)
US0EP1_dmy	.0853***	.138***	.143***	.131***	.119***	.188***	.197***	.211***
	(.0111)	(.0532)	(.0228)	(.0235)	(.0182)	(.0516)	(.0382)	(.0353)
ln_N_Bnonself_Citn	.144***	.412***	.257***	.174***	.133***	.292***	.222***	.170***
	(.00257)	(.0128)	(.00535)	(.00395)	(.00358)	(.022)	(.00808)	(.00474)
ln_N_Bself_Citn	.0896***	.354***	.172***	.102***	.0764***	.151***	.118***	.0908***
	(.0037)	(.0201)	(.00713)	(.00480)	(.00499)	(.0294)	(.0105)	(.00571)
ln_N_Bttl_ExCitn	.0687***	.196***	.129***	.0608***	.0584***	.161***	.110***	.0538***
	(.00156)	(.00915)	(.00363)	(.00169)	(.00263)	(.0143)	(.00599)	(.00322)
ln_N_Bttl_CoCitn	-.0538***	-.244***	-.126***	-.0713***	-.0187***	-.160***	-.0459***	-.0162*
	(.00441)	(.0267)	(.00976)	(.00648)	(.00669)	(.0350)	(.0147)	(.00843)
subAppln_dmy	.338***	.752***	.576***	.466***	.347***	.840***	.610***	.452***
	(.00623)	(.0248)	(.0126)	(.00886)	(.0100)	(.0477)	(.0219)	(.0150)
joint_dmy	-.0272***	-.0746***	-.0599***	-.0302***	-.00422	-.0563*	-.000349	-.0106
	(.00393)	(.0245)	(.0091)	(.00434)	(.00625)	(.0307)	(.0149)	(.00651)
filing_year	yes	yes	yes	yes	yes	yes	yes	yes
tech	yes	yes	yes	yes	yes	yes	yes	yes
filing_year × tech	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	.0701				.0985			
Adjusted R-Squared	.0700				.0983			

Note. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

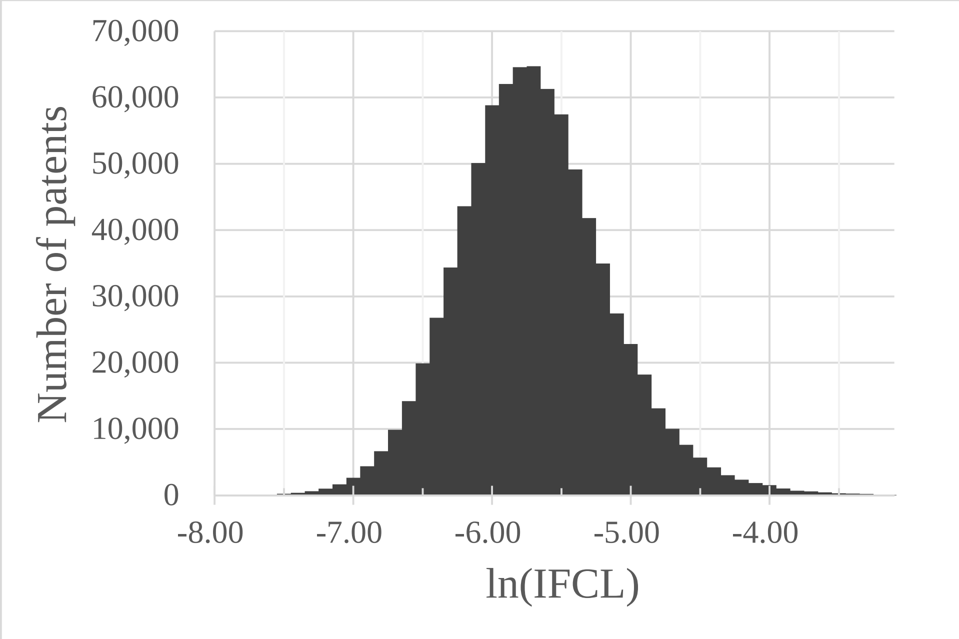

Figure 1. Distribution of the natural logarithm of IFCL

Note: The figure shows the distribution of the natural logarithm of the inverse of the first claim length (IFCL) of product patents aggregated by all technological fields.

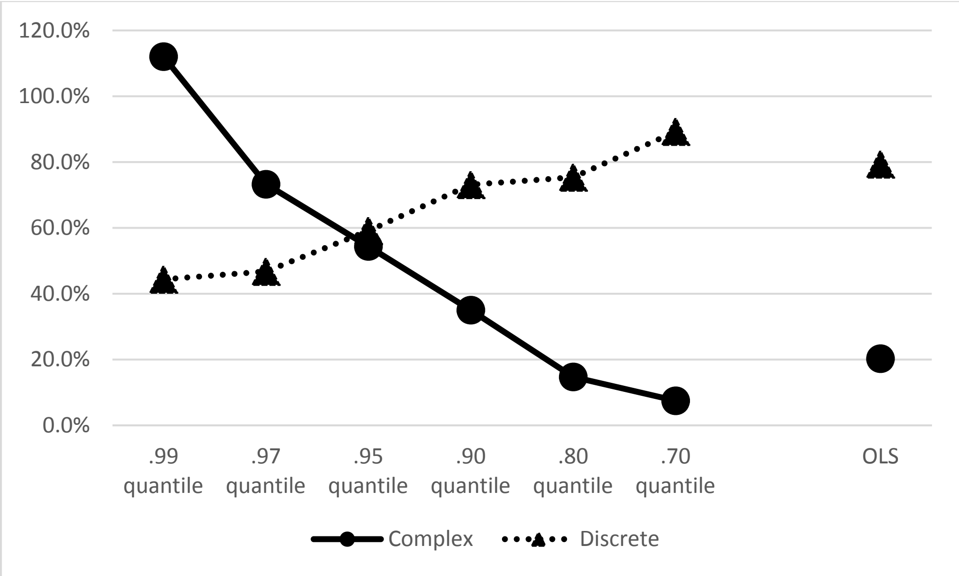

Figure 2. Relative value of the coefficient for the IFCL against that of the number of IPC subclasses (product patents)

Note: The figure shows the values of the following ratio, coefficient for the IFCL/coefficient for the number of different IPC subclasses, of product patents estimated by quantile regression where the quantiles are 0.99, 0.97, 0.95, 0.90, 0.80, and 0.70 as well as those estimated by the OLS method.

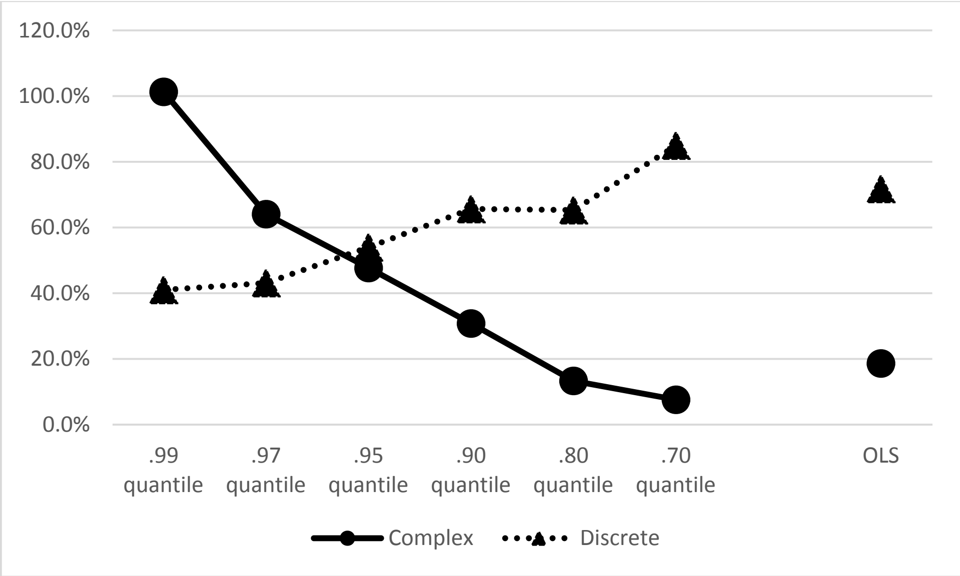

Figure 3. Relative value of the coefficient for the IFCL against that of the number of claims (product patents)

Note: The figure shows the values of the following ratio, coefficient for the IFCL/coefficient for the number of claims, of product patents estimated by quantile regression where the quantiles are 0.99, 0.97, 0.95, 0.90, 0.80, and 0.70 as well as those estimated by the OLS method.

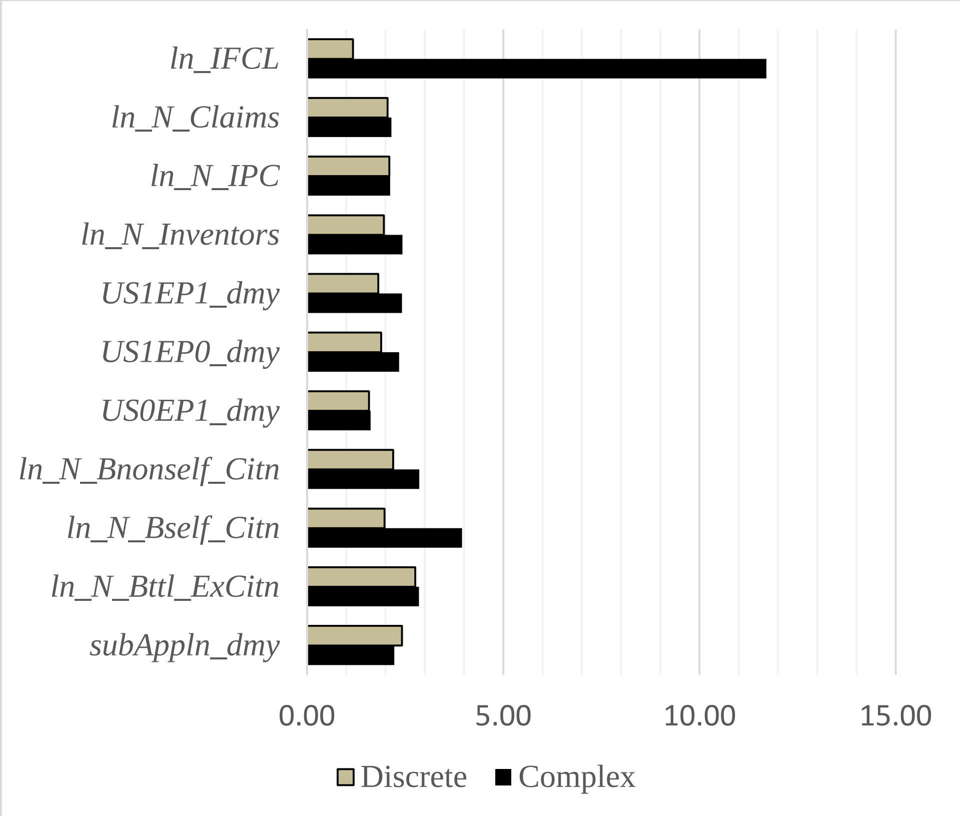

Figure 4. The relative values of the coefficients estimated by 99% quantile regression against those estimated by the OLS method (product patents)

Appendix A

Table A.1 Technology sector classification

	Technology sector	IPC
Complex technology	Computers & Communications	G04-G12, H03-H04
	Electrical & Electronic	G01-G03, H01-H02,H05
	Mechanical	B21-B32(excludes B31)-B44, B60-B68, F01-F04,F15-F17
Discrete technology	Chemical	A01N, C01-C11(excludes C06), C21-C30
	Drugs & Medical	A61-A63, C12-C14
	Others	A01(excludes A01N), A21- A47, B01-B09, B31, B81,B82, C06, D01-D21, E01-E21, F21-F42, G21

Appendix B

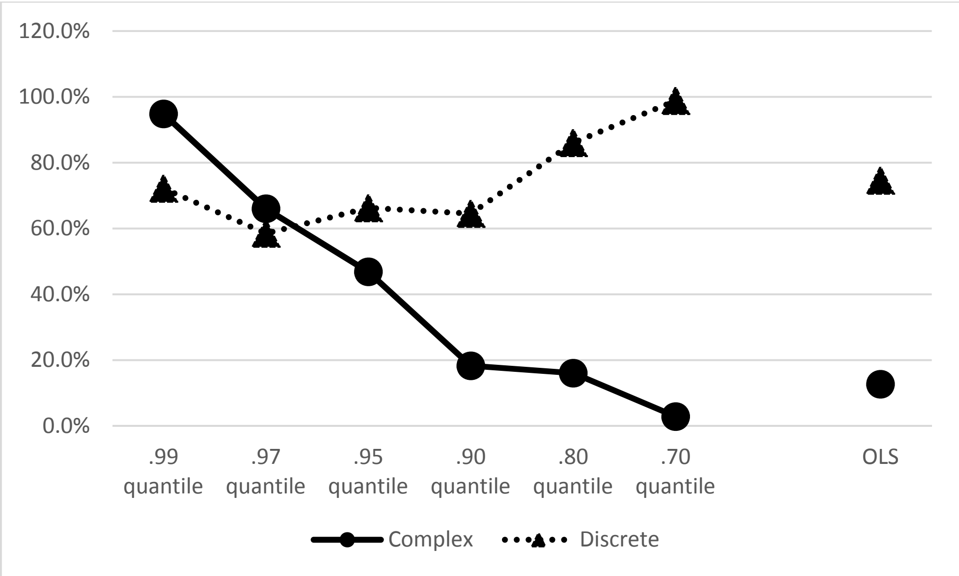

Figure B.1 Relative value of the coefficient for IFCL against that of the number of IPC subclasses (process patents)
Note: The figure shows the values of the following ratio, coefficient for the IFCL/coefficient for the number of IPC subclasses, of the process patents estimated by quantile regression where the quantiles are 0.99, 0.97, 0.95, 0.90, 0.80, and 0.70 as well as those estimated by the OLS method.

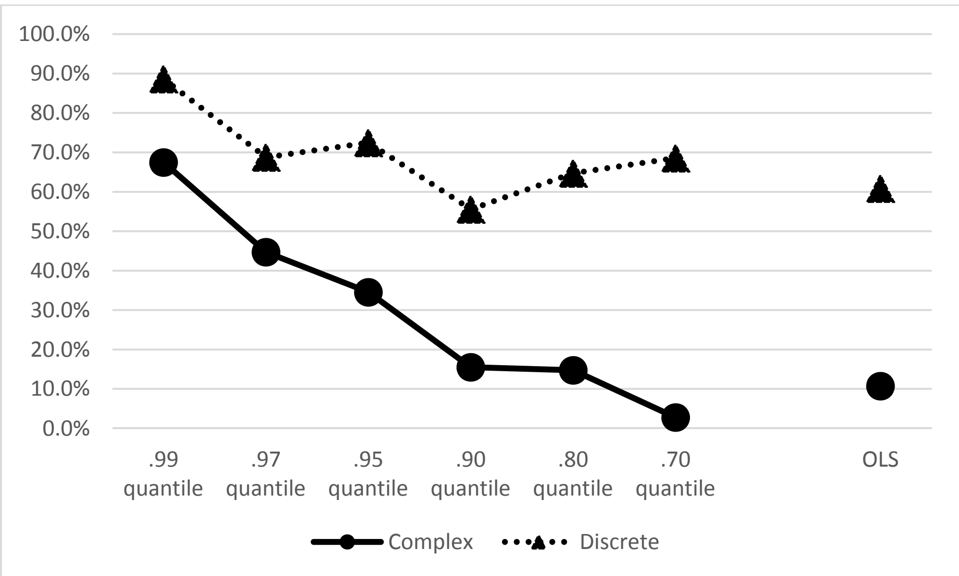

Figure B.2 Relative value of the coefficient for the IFCL against that of the number of claims (process patents)
Note: The figure shows the values of the following ratio, coefficient for the IFCL/coefficient for the number of claims, of the process patents estimated by quantile regression where the quantiles are 0.99, 0.97, 0.95, 0.90, 0.80, and 0.70 as well as those estimated by the OLS method.

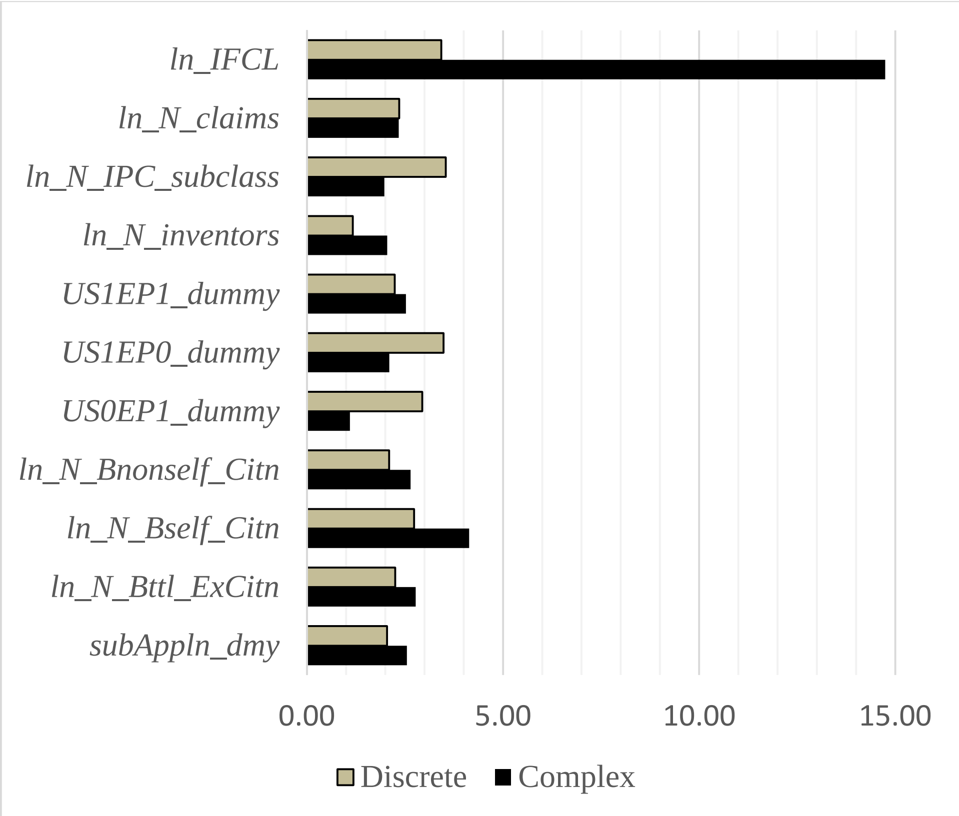

Figure B.3. The relative value of coefficients estimated by 0.99 quantile regression against that estimated by the OLS method (process patents)

Table B.1. Descriptive statistics for the data of process patents that are used in regression analysis reported in Table B.2

	Complex (N= 125,721)		Discrete (N= 64,333)	
Variables	Mean	Std. Dev.	Mean	Std. Dev.
ln_N_F_Citation	0.570	0.751	0.660	0.805
ln_IFCL	-5.67	0.511	-5.35	0.540
ln_N_claims	1.24	0.847	1.05	0.768
ln_N_IPC_subclass	0.307	0.430	0.369	0.459
ln_N_inventors	0.662	0.603	0.904	0.563
US1EP1_dmy	0.0856	0.280	0.0913	0.288
US1EP0_dmy	0.133	0.340	0.0427	0.202
US0EP1_dmy	0.00798	0.0890	0.0127	0.112
ln_N_Bnonself_Citn	0.310	0.535	0.497	0.654
ln_N_Bself_Citn	0.176	0.410	0.255	0.468
ln_N_Bttl_ExCitn	1.35	0.616	1.25	0.648
ln_N_Bttl_CoCitn	0.111	0.285	0.178	0.356
subAppln_dmy	0.0443	0.206	0.0308	0.173
joint_dmy	0.0565	0.231	0.0724	0.259

Table B.2. Summary results of the regression (process patents)

	Dependent variable: *ln_N_F_Citation*							
	Complex (N = 125,721)				Discrete (N = 64,333)			
Independent variables	OLS	.99 quantile	.90 quantile	.70 quantile	OLS	.99 quantile	.90 quantile	.70 quantile
ln_IFCL	.00746*	.111***	.0183**	.00154	.0432***	.148***	.0639***	.0515***
	(.00419)	(.0242)	(.00927)	(.00331)	(.00589)	(.0206)	(.0124)	(.00701)
ln_N_claims	.0696***	.163***	.118***	.0557***	.0710***	.167***	.115***	.0751***
	(.00291)	(.0156)	(.00613)	(.00271)	(.00451)	(.0158)	(.00940)	(.00587)
ln_N_IPC_subclass	.0588***	.116***	.100***	.0538***	.0579***	.205***	.0990***	.0521***
	(.00503)	(.0258)	(.0112)	(.00511)	(.00711)	(.0256)	(.0154)	(.00922)
ln_N_inventor	.0805***	.165***	.149***	.0643***	.0672***	.0787***	.134***	.0762***
	(.00362)	(.0205)	(.00805)	(.00363)	(.00565)	(.0213)	(.0125)	(.00703)
US1EP1_dmy	.211***	.533***	.386***	.326***	.233***	.522***	.382***	.347***
	(.00914)	(.0339)	(.0185)	(.0115)	(.0132)	(.0449)	(.0240)	(.0159)
US1EP0_dmy	.128***	.269***	.219***	.208***	.153***	.533***	.298***	.220***
	(.00684)	(.0371)	(.0149)	(.0121)	(.0179)	(.0491)	(.0425)	(.0188)
US0EP1_dmy	.122***	.134**	.298***	.127**	.138***	.406***	.322***	.196***
	(.0268)	(.0576)	(.0691)	(.0509)	(.0331)	(.0574)	(.0497)	(.0416)
ln_N_Bnonself_Citn	.137***	.362***	.253***	.169***	.125***	.262***	.210***	.193***
	(.00513)	(.0271)	(.0107)	(.00886)	(.0059)	(.0207)	(.0115)	(.00796)
ln_N_Bself_Citn	.0962***	.398***	.218***	.102***	.0637***	.174***	.131***	.0855***
	(.00717)	(.0355)	(.0154)	(.0104)	(.00801)	(.0224)	(.0147)	(.0122)
ln_N_Bttl_ExCitn	.0577***	.160***	.109***	.0448***	.0533***	.120***	.0884***	.0623***
	(.00358)	(.0207)	(.00781)	(.00289)	(.00537)	(.0193)	(.0115)	(.00637)
ln_N_Bttl_CoCitn	-.0284***	-.151***	-.111***	-.0469***	-.0179	-.0851***	-.0392*	-.0100
	(.00918)	(.0425)	(.0200)	(.0146)	(.0109)	(.0330)	(.0233)	(.0161)
subAppln_dmy	.348***	.888***	.608***	.448***	.411***	.840***	.651***	.504***
	(.0135)	(.0412)	(.0295)	(.0145)	(.0236)	(.0511)	(.0571)	(.0289)
joint_dmy	-.0384***	-.227***	-.0953***	-.0332***	.00923	-.103***	-.0239	-.00320
	(.00884)	(.0426)	(.0213)	(.00679)	(.0121)	(.0328)	(.0317)	(.0192)
filing_year	yes	yes	yes	yes	Yes	yes	yes	yes
tech	yes	yes	yes	yes	Yes	yes	yes	yes
filing_year × tech	yes	yes	yes	yes	Yes	yes	yes	yes
R-squared	.067				.062			
Adjusted R-Squared	.0667				.0613			

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10