

# **Construction of Japanese Patent Database for Research on Japanese patenting activities \***

Akira Goto<sup>1</sup> and Kazuyuki Motohashi<sup>1</sup>

## **Summary**

In this paper, we present the methodology of developing patent database using Japanese patent office data files. Then preliminary attempt was made to investigate the innovation activities in Japan by using this dataset. Patent based databases have been developed and used for various innovation researches in US and Europe. IIP Patent Database based on JPO patent information fills a big hole in patent base database development in the triad.

One of focus of this paper is innovation analysis using citation information in IIP Patent Database. We compared the citation pattern of this dataset to those of the datasets in US and Europe, by using OECD patent family data, and found that citation information in this dataset is useful as a tool for analyzing the value of patent and cumulative nature of innovation. In this paper, the trend of backward and forward citation is examined. In addition, the citation propensity and citation related indicators (originality and generality indices) are compared across type of applications and technology fields. The results confirm the usefulness of patent citation data for analyzing the nature of innovation process.

Key words: Patent database, Japan, citation data, innovation indicator

---

\* The development of the patent database discussed herein has benefited from the cooperation of numerous individuals. We would like to express gratitude to: Koji Manaka, Jun Suzuki, Yusuke Naito, Setsuko Asami, Junko Sato, Sadao Nagaoka, Yosuke Okada, Tetsuo Wada, and Schumpeter Tamada. Financial Support from the Japan Patent Office for this project is also acknowledged. The Institute of Intellectual Property also helped us in many ways. Usual disclaimer applies.

<sup>1</sup> Research Center for Advanced Science and Technology, University of Tokyo.

## **1. Introduction**

Creating and learning technological knowledge, along with acquiring, maintaining, and strengthening such capabilities, is crucially important to advancing the quality of people's lives continuously. However, the sources of data that can serve as a basis for studying technological knowledge are extremely limited. This stems from the difficulties inherent to the measurement of knowledge. Patent data is one of the few precious sources of standardized information related to technological knowledge. The patent system grants time-limited exclusive rights in order to secure incentives for innovation which is novel, non-obvious, and useful for the industry, but requires that applications be filed soon after invention and makes those applications public 18 months after the filing date. As a result, patent data is an enormously valuable collection of technological information.

Since the National Bureau of Economic Research of the United States made public its NBER Patent Citation Data File, research on innovation made a major progress. Many papers have been written using this valuable database. Following this example, we developed the patent database based on the patents filed with the Japan Patent Office. We call this database the IIP Patent Database. IIP stands for the Institute of Intellectual Property, and the database is available through its website. In this paper, we present the methodology of developing IIP Patent Database, as well as insights for innovation activities using this database.

When Schmoockler's pioneering research (Schmoockler,1966) was published in 1966, it made a significant impact on the study of innovation processes and, at the same time, served to widely disseminate the notion that patents are a valuable source of primary data for innovation research. However, patent data was not much utilized since then for several reasons, including the way patent documents were filed, the sheer number of patents and the many kinds of specialized knowledge required to understand the technical content of the patents as well as the patent system itself.

In recent years, however, economic research utilizing patent data has been making rapid progress. Three factors are in the background behind the recent flowering of patent data based innovation research. The first factor is the increased interest in technological innovation. Corporate managers, policy makers, and others are increasingly interested in technological innovation, driven by the realization that technology is an crucial determinant of corporate and national "competitiveness". Economists are also increasingly concerned about the inadequate treatment of technological innovation, and

started to work to remedy this, leading to the endogenous growth theory in macroeconomics and analysis of technological innovation in applied microeconomics. The second factor is the greater ease of using patent data for analyzing technological innovation thanks to the creation of electronic patent databases. Across various countries and regions, patent data has been computerized and made easier to use. As discussed below, because the computerization was done with the purpose of facilitating searches for individual patents, using the databases for statistical purposes requires reformatting the data, which requires enormous amount of time and costs. Nevertheless, statistical analysis has become possible. The third factor is intimately related to the preceding factor: progress in computers, software, and the Internet has facilitated the use of large amounts of data and made possible statistical processing of qualitative data.

NBER patent database for US patent data is a path breaking achievement in this context, and it opens up great possibilities for researchers to work on innovation studies (Hall et. al, 2001; Jaffe and Trajtenberg, 2002). Recently, EPO has made significant effort to prepare for patent database based on DOC-DB. Moreover, OECD's creation of patent family data covering international patent applications in Japan, Europe, and the United States, indicates that the preparation of patent databases is quickly advancing on an international front as well (Dernis and Khan, 2004). In this sense, the construction of a Japanese patent database carries an extremely important significance for advancing the study of technological innovation, and, moreover, for enabling evidence-based and meaningful deliberation of the patent system and technology policy.

The rest of this paper is organized as follows. The next section, section 2, describes the methodology of constructing Japanese patent database and its basic features, followed by section 3 that presents basic features of innovation activities using this database. Section 4 compare the citation data in the IIP Patent Database with that in the NBER and EPO databases. Section 5 concludes the paper.

## **2. Methodology and Basic Features of Japanese Patent Database**

The Japan Patent Office (JPO) has been disseminating patent database called "Seiri Hyojunka Data" (JPO Patent Database, hereafter).<sup>2</sup> This data contains published information for individual patents as well as internal records by JPO in a process of examinations and post grant infringement settlements. The original datasets are

---

<sup>2</sup> Currently, NCIPI (National Center for Industrial Property Information and Training), an independent administrative institute attached to JPO is in charge of data dissemination.

organized in SGML format, and all information is tabulated by chronological order of events, such as patent application, request for examination and registration. The JPO Patent Database has been published regularly, such as bi-weekly for patent application and registration information.

The first step of data construction is to put together all historical data from various versions of JPO Patent Database. Then, original datasets by SGML format was converted to SQL database, by deleting redundant information and normalizing the data structure.<sup>3</sup> The data coverage extends to patent applications filed from early 20<sup>th</sup> century through January 2004.<sup>4</sup> It contains more than 56,000,000 records and the size of data is over 100 Giga bite.

Since the size of this dataset is too large for data processing under personal computer environments, we selected the most commonly used variables by using NBER patent database as a reference. In addition, we limited the data to those in 1964 and afterwards, since the IPC code information (for JPO examiners' searching purpose) is available for only for these years.

The whole database is organized as the following five files, and now the file size goes down to some hundred Mega bite level, which is manageable by PC environment. These files are open to public, and can be downloaded from the web-site of Institute of Intellectual Property<sup>5</sup>.

1. Patent application file (9,027,486 records)
  - Application number
  - Application date
  - Examination request date
  - Applicant number
  - Number of claims at patent application
  - Lead IPC code at patent application
  - Aggregated technology category
2. Patent registration file (2,618,699 records)
  - Application number

---

<sup>3</sup> See Naito (2005) for a detailed description of how the standardized data supplied by JPO has been processed.

<sup>4</sup> We used JPO standardized data through its 22<sup>nd</sup> release for the year 2004. The database includes patent applications published through January 28, 2004 and data entered by the JPO through January 23, 2004.

<sup>5</sup> <http://www.iip.or.jp>

- Registration number
  - Registration date
  - Rights expiration date
  - Rights holder number
  - Number of claims at patent registration
  - Lead IPC code at patent registration
  - Aggregated technology category
3. Applicant (holder of right) file (626,708 records)
    - Applicant number
    - Applicant name
    - Applicant type
    - Country and state code
    - JPO applicant code
  4. Rights holder file (204,622 records)
    - Rights holder number
    - Rights holder name
  5. Citation information file (5,318,225 records)
    - Citing patent application number
    - Cited patent application number
    - Citation type

In order to construct the database, it was necessary to clean data and add some new variables which are useful for researchers to use this database. The following describes this process.

#### (1) Technology classification

The JPO Patent Database uses IPC code for technology classification, but it should be noted that IPC code system have been revised regularly. Therefore, users should check with the corresponding version of IPC coding system, depending on the timing of application or registration of each patent.

In addition, as the IPC code can be too detailed to examine the overall patenting activities, it is useful to provide more aggregated classification. Thus, we also included the technology classification shown in Table 1, which is used in the WIPO statistical reports. This technology classification can be matched with the top level categories at NBER patent database with six types of technology, namely (1) chemical, (2) computer and communications, (3) drugs and medical, (4) electrical and electronics, (5)

mechanical, (6) others, as is shown also in Table 1.

(Table 1)

#### (2) Request for examination information

The system of request for examination system started in 1971. Before the introduction of this system, all applications were examined. After 1971, applications are examined only when applicant request examination, which can be done within seven years after application. This period was shortened to three years in 2001. In the JPO Patent Database, information pertaining to examination requests is kept as internal codes of JPO corresponding to the various application processing steps such as patent revisions, request for proof of priority rights, etc. Request for examination information is created by using this internal code.

#### (3) Number of claims

In JPO Patent Database, both the numbers of claims at application and that at registration are available. Both type of data are provided in the IIP database, because these two can be different. It should be noted that this data is missing for most patents prior to 1970. However, in practice, this is not a major problem since the number of claims for almost all patents used to be one, before improved multiple claim system was introduced in 1988.

#### (4) Applicant file

Detailed information including names and locations of applicants is provided in a separate applicant file. For patents that have multiple applicants or right holders, the database lists all applicant (or right holder) codes by using “;” (semi-colon) as a separator.

In JPO’s original data, there is significant number of missing information on applicant information, and applicant information had not been systematically managed before JPO introduced applicant coding system in 1992. Therefore, it was necessary to conduct very time-consuming applicant name cleaning and matching works. Names of major applicants (with more than 500 patent applications throughout the period covered by the dataset) were checked manually.

In addition, the following information on applicant type and location is added to applicant file by using information from JPO.

- Applicant type: 1: individual, 2: private firms or 3: government and non profit organization
- Location: Country code, and prefecture code in case of Japanese applicants

Details of data construction method and cleaning process are found in Motohashi (2005).

#### (5) Citation Information

The JPO Patent Database includes two types of citation information data. First, for some patent applications for which examination was requested, prior art references were given by a patent examiner to the applicant as grounds for rejection. This information on prior art cited by examiners can be used as a basis for citation. Second, there is also citation data in Patent Gazette for registered patents. Although this data is limited to registered patents, there are cases in which a successful patent application, while not rejected, nevertheless includes prior art references when the patent examiner decided to be important. The JPO Patent Database includes this information as well, but only data from 1985 as a year of publication.

Figure 1 shows, by year of application filing, the number of registered patents that include citation information. As for citation information drawn from patent rejections (type 1), the share of citing patent to total patent application varies greatly year-to-year. This could be due to missing data or changes in examination policies by the JPO. There is also a problem with citation information drawn from Patent Gazette, because that data only goes back to late 1980's. To provide the flexibility to meet the needs of different kinds of analysis, the IIP Patent Database includes citation information drawn from both sources but also provides a code to identify the source of citation information.<sup>6</sup>

(Figure 1)

Citation information is one of the most important part of patent data because it can be used for various purposes including tracing the process of technology development and evaluating the importance of a patent. However, it is important to understand the quality and the limitations of the citation data in the IIP database explained above when researchers try to use it. JPO patent citation information is compared with those of US and EU patent data in section 5 of this paper.

---

<sup>6</sup> Citation type is 1 (by only first type citation), 2 (by only second type citation) and 3 (by both types of citations).

### 3. Overview of innovation activities by IIP patent Database

The number of patents covered by the IIP Patent Database is summarized in Table 2. There are 9,027,486 patent applications, of which 4,427,840 have received a request for examination and 2,594,044 have been successfully registered. As mentioned above, this database covers data from 1964 through January 2004. This means that, for example, most of patent data applied before 2002 or later are not published by January 2004, and are not included in the database. There is also significant data truncation in patent registration data due to the fact that it takes a long time for a patent application to pass through the request for examination stage and finally get to registration stage.

(Table 2)

Figure 2 is a breakdown of the approximately 9 million patent applications by technology. The share of patent applications in the ICT, electrical & electronics, and pharmaceuticals & medical sectors has been rising while the share of applications in chemical & mechanical has been falling.

Figure 3 is a breakdown of the same data by type of applicant. Over 90 percent of patent applications to JPO come from domestic applicants, and the overwhelming fraction of these Japanese applicants is firms. The share of foreign applicants is only about 7 percent.

(Figure 2) and (Figure 3)

Figure 4 shows the annual volume of registered patents broken down by technology. It is clear that the share of information technology and electrical & electronics patents is rising. The decrease in registered patents in recent years is due to the data truncation problem mentioned above. The sudden jump in patents in 1996 is likely due to the introduction of the post-grant opposition system in that same year.

(Figure 4)

Although the number of patent applications is often used as an indicator of innovation output, it should be noted that number of claims in one application increased rapidly after the change of the patent system in 1989. With the introduction of the Revised System of Multiple Claims in 1989, inventions that previously had to be filed as separate patent applications could be filed as a single application with multiple claims. Figure 5 shows the trend in the number of claims (at application time) broken down by technology. All technology sectors exhibit an increase in the average number of claims

upon the introduction of the Revised System of Multiple Claims. The number of claims in pharmaceutical & medical and chemical patents has increased rapidly, but in electrical & electronic, ICT, and mechanical sectors the increase in claims was more gradual. In the most recent period, sectors with the largest average number of claims are electrical & electronics and ICT. A drop in the number of applications in the early 1990's may be, at least to certain extent, due to this change of the system.

(Figure 5)

Lastly, Figures 6 and 7 show the period for patent life. Although patent rights can last up to 20 years from the date of application (priority claim date), many patents are abandoned prior to the end of this period. These figures show the actual length of patent life. Figure 6 shows that foreign applicants and Japanese individual applicants are abandon patents earlier than Japanese firms. Japanese government-related patents were maintained longest. In fact, it can be seen that almost all of the government-related patents are maintained until close to the end of the 20-year period. This may be due to the fact that patents owned by the government are subject to patent fee exemptions.

Viewing patent life by the time of application (Figure 7), we observe that compared to patents applied during the 1960's and 1970's, the life length for patent applied during the 1980's has been somewhat shorter. This implies that the speed of technological obsolete is becoming faster. It should be noted that the date during the 1990's and afterwards are subject to data truncation problem.

(Figure 6) and (Figure 7)

#### **4. JPO Citation Data: Discussion and Analysis**

There are two types of citation information in original JPO Patent Database. Both of citations are made by patent examiner.<sup>7</sup> In this sense, this citation data is similar in nature to the citation information derived from search reports produced by examiners at the European Patent Office. Furthermore, it is different from U.S. citation information, derived from patent application process<sup>8</sup>. As discussed above, U.S. citation is mostly

---

<sup>7</sup> Independent from this, Tamada et al. (200x) developed a software to pick up citations made by applicant from the text of the description of technology in application file, and examined citation patterns in Japan.

<sup>8</sup> On September 1, 2002, Japan adopted the prior art reference disclosure practice. According to this practice, applicants must file a detailed list of prior art references known at the time of the patent application. However, there are no associated penalties for failing to do so. It seems most applicants nevertheless list references extensively. This can be a useful source of citation in the future.

given by applicants and thus differs in character from Japanese and European citation data, which are supplied by patent examiners. However, it must be noted that many of the citations contained in U.S. patents are given by patent attorneys working for applicants, and the USPTO examiners as part of the examination process. Beginning in 2001, U.S. patent citation data is classified into citations provided by the applicant and citations provided by the examiner. According to this data, over 40% of citations are provided by examiners (Alacer and Gittelman, 2004). Moreover, according to a survey of inventors on U.S. patents, less than 20% of inventors are reported to be aware of the cited patents prior to the invention and about 30% reported that they did not know about the cited prior art prior to their invention. (Jaffe et. al, 2000). Therefore, U.S. patent data seems to include significant amount of information beyond the pure technology spillover effect, i.e., beyond inventors actually innovating cumulatively on top of a base of prior art.

In this paper, citation information from IIP Patent Database was compared with that of USPTO and EPO. All of three patterns of citation information are converted to a group of comparable inventions by using OECD patent family database (Dernis and Khan, 2004), and citation lag, backward and forward citation counts are compared. It is found that IIP data and EPO citation show a similar pattern of citation lag, because substantial numbers of citations are made to application patents. In case of US, the system of publication of applied patents was not introduced until recently, so that only granted patents can be cited. In terms of citation counts, we can find a significant positive correlation, particularly for that of forward citation. Details of these results can be found in Appendix of this paper.

In this section, IIP citation data are used for comparing innovation activities across technology fields as well as types of innovator (individual, firms or public organizations). There are 5,318,225 pairs of citing and cited patents in IIP Patent Database. However, substantial numbers of pairs are those for not granted patents. In this section, citing and cited pairs for granted patents, which represents cumulative nature of innovation both for citing and cited patents, are used. The number of citation pairs is reduced to 1,602,130 for 2,618,699 granted patents from 1964 to 2003.

Figure 8 shows the share of patents with citing (or cited) patents by application year. It is found that the share of patents with citing patents (backward citations) is increasing, but even recently, only 60% of patents make citations. In case of USPTO data, it is difficult to find a patent without citing patents. The share of patents received citations (forward citations), increases until the middle 1980's, but decreases afterwards.

Increasing citing propensity pushes up the cited propensity in general, while the truncation effect is found in recent period. The jump in 1971 can be explained by introduction of application patent publication system.

(Figure 8)

In Figure 9, the trend of indicators for patents with citation made (backward citation) is presented. The average number of citation made has increased from around 1 in 1964 to 2 in 2000. The originality index is defined as  $1 - \sum_i s_i^2$ , where  $s$  is the share of citing patents by technology fields.<sup>9</sup> Here, we use the technology classification of patents with 33 sectors presented in Table 1. The originality index represents the diversity of technologies in citing patents. Since this index is increasing, an invention process becomes more complex, based on different kinds of existing technologies.

(Figure 9)

In Figure 10, the trend of indicators for patents with any citation received (forward citation) is presented. The average number of citation received increases until the middle of 1980's, but decreases afterwards. The generality index is based on a similar idea as the originality index. This index represents the diversity of technologies in cited patents. A high value of this index implies this technology is used by various kinds of future inventions. The generality index also increases until the middle 1980's, but decreases afterwards. Since a recent drop can be explained by data truncation, this index may be in an increasing trend, which implies again that innovation process becomes complex in a sense of increasing technology spillover across technology fields.

(Figure 10)

It is also interesting to see differences of these citation related indicators across technology fields and applicant types. Therefore, we conduct descriptive regressions of 6 types of indicators presented in Figure 8, 9 and 10. Right hand variables in this models are the following dummy variables.

- Timings of application by decade (1960's as a base)
- Types of patent applicant ('Individuals' as a base)
- 33 types of technology fields ('agriculture' as a base)

The results are presented in Table 3.

---

<sup>9</sup> In order to control for discrete variable biases, this index is adjusted by  $N/N-1$ , where  $N$  is the number of citation made (Hall et. al, 2001).

(Table 3)

First, the time trends in Figure 8, 9 and 10 are confirmed after controlling for applicant type and technology field. Second, it is found that the firm's patent shows higher citing and cited propensity. Patents by public organizations, mainly by public research institutes, do not show any distinctive pattern, and look similar to those by individuals. Finally, a significant heterogeneity across technology fields can be found. In general, a technology field with higher backward citation propensity has higher forward citation propensity. This suggests innovation process evolves within each technology field. But in some technologies, we can find relatively high generality indices, implying positive technology spillovers to other technology fields. It is interesting to see that higher values in generality index can be found in chemicals (drugs, organic chemicals, organic molecules and biotechnologies). In contrast, in electronics and semiconductor fields, generality indices are in the same level as that of agriculture. In terms of number of citations, both technologies show cumulative nature of innovations, but spillover to other technology can be seen more in chemicals.

## 5. Conclusion

In this paper, we present the methodology of developing patent database using Japanese patent office data files. Then preliminary attempt was made to investigate the innovation activities in Japan by using this dataset. Patent based databases have been developed and used for various innovation researches in US and Europe. IIP Patent Database based on JPO patent information fills a big hole in patent base database development in the triad.

One of focus of this paper is innovation analysis using citation information in IIP Patent Database. We compared the citation pattern of this dataset to those of the datasets in US and Europe, by using OECD patent family data, and found that citation information in this dataset is useful as a tool for analyzing the value of patent and cumulative nature of innovation. In this paper, the trend of backward and forward citation is examined. In addition, the citation propensity and citation related indicators (originality and generality indices) are compared across type of applications and technology fields. The results confirm the usefulness of patent citation data for analyzing the nature of innovation process.

Famous physicist Niels Bohr said that we cannot say that something exists until it can be measured. These words are often quoted to emphasize the importance of measurement. Although the importance of knowledge is emphasized, with the coming

of knowledge society, the measurement of knowledge is not a simple matter. In this context, patent data is a valuable source information about technological knowledge. However, in their original form, it is difficult to use patents for quantitative analysis. The already large and yet still growing contributions from the NBER database show the importance of converting this potentially valuable source of technological knowledge into an easy-to-use database and then using the database to analyze the processes by which technological knowledge is created and utilized. Patent statistics have opened new horizons for our analysis. The database that we described in this article is the first database for comprehensive quantitative analysis of Japanese patents. As compilers of this database, it is our hope that it will be used widely and will make a contribution to research in innovation in Japan.

## References

- Alcacer, J. and M. Gittelman, (2004), How do I now what you know? Patent Examiners and the Generation of Patent Citations, mimeo
- Dernis, H. and M. Khan (2004), Triadic Patent Families Methodology, STI WORKING PAPER 2004/2, OECD Paris
- Hall, B., Jaffe, A. and M. Trajtenberg (2000), Market Value and Patent Citations: A First Look, NBER WP 7441
- Hall, B., Jaffe, A. and M. Trajtenberg (2001), The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools, NBER WP 8498
- Jaffe, A. and M. Trajtenberg (2002), *Patents, Citations and Innovations, A Window on the Knowledge Economy*, MIT Press
- Jaffe, A., Trajtenberg M. and M. Fogarty (2000), The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees, NBER WP 7631
- Lanjouw, J. O and M. Schankerman (1999), The Quality of Ideas: Measuring Innovation with Multiple Indicators, NBER Working Paper #7345
- Michel, J. and B. Bettels (2001), Patent Citation Analysis: A Closer Look at the Basic Input Data from Patent Search Report, *Scientometrics*, vol. 51, No. 1, pp. 185-201
- Motohashi, K. (2005), Development of Patent Database for Researchers by JPO's Seiri Hyojunka Data, in Research Report on Economic Analysis of Innovation by Patent Data to JPO, 2005 March (in Japanese)
- Naito, Y. (2005), Development of Patent Database, in Research Report on Economic Analysis of Innovation by Patent Data to JPO, 2005 March (in Japanese)
- Schmookler, J. (1996) *Invention and Economic Growth*, Harvard University Press, 1966.
- Webb, C., Dernis, H., Harhoff, D. and K. Hoisl (2005), Analyzing European and International Patent Citations: A Set of EPO Patent Database Building Blocks, STI WORKING PAPER 2005/9, OECD Paris

Table 1: Aggregated Technology Classification

Tech No	Title	Corresponding IPC	NBER
1	Agriculture	A01 (Except A01N)	6
2	Food Stuffs	A21 ~ A24	6
3	Personal and Domestic Articles	A41 ~ A47	6
4	Health and Amusement	A61 ~ A63 (Except A61K)	3
5	Drugs	A61K	3
6	Separating, Mixing	B01 ~ B09	1
7	Machine tools, Metal working	B21 ~ B23	5
8	Casting, Grinding, Layered Product	B24 ~ B32 (Except B31)	5
9	Printing	B41 ~ B44	6
10	Transporting	B60 ~ B64	5
11	Packing, Lifting	B65 ~ B68	5
12	Non organic chemistry, Fertilizer	C01 ~ C05	1
13	Organic chemistry, Pesticides	C07 , A01N	1
14	Organic molecule compounds	C08	1
15	Dyes, Petroleum	C09 ~ C11	1
16	Biotechnology, Beer, Fermentation	C12 ~ C14	3
17	Genetic Engineering	C12N15/	3
18	Metallurgy, Coating metals	C21 ~ C30	5
19	Textile	D01 ~ D07	6
20	Paper	D21 , B31	6
21	Construction	E01 ~ E06	6
22	Mining, Drilling	E21	6
23	Engine, Pump	F01 ~ F04 , F15	5
24	Engineering elements	F16 ~ F17	5
25	Lighting, Steam generation, Heating	F21 ~ F28	6
26	Weapons, Blasting	F41 ~ F42 , C06	6
27	Measurement, Optics, Photography	G01 ~ G03	4
28	Clock, Controlling, Computer	G04 ~ G08	2
29	Display, Information Storage, Instruments	G09 ~ G12	2
30	Nuclear physics	G21	4
31	Electronics components, semiconductor	H01 ~ H02 , H05	4
32	Electronics circuit, communication tech.	H03 ~ H04	2
33	Others	B81 , B82	6

Table 2: Patent counts in Japanese Patent Database

	Application	Request Exam.		Registration		
	A	B	B/A	C	C/B	C/A
1964	36850	0		27,963		76%
1965	41115	0		34,630		84%
1966	45358	0		40,375		89%
1967	47317	0		43,090		91%
1968	55446	0		51,100		92%
1969	61411	0		57,023		93%
1970	77010	0		71,823		93%
1971	103,227	74263	72%	44,128	59%	43%
1972	127,426	91509	72%	54,288	59%	43%
1973	141,477	99706	70%	58,256	58%	41%
1974	146,240	102979	70%	59,069	57%	40%
1975	156,654	109398	70%	61,070	56%	39%
1976	157,911	106779	68%	60,547	57%	38%
1977	157,997	106861	68%	60,404	57%	38%
1978	162,979	111377	68%	62,571	56%	38%
1979	172,567	115016	67%	64,859	56%	38%
1980	188,124	121933	65%	68,762	56%	37%
1981	215,450	136220	63%	75,640	56%	35%
1982	234,061	142584	61%	78,424	55%	34%
1983	251,374	150704	60%	82,736	55%	33%
1984	281,199	164186	58%	88,555	54%	31%
1985	296,934	163440	55%	92,179	56%	31%
1986	311,280	163441	53%	96,735	59%	31%
1987	329,263	165665	50%	102,345	62%	31%
1988	328,055	161975	49%	105,030	65%	32%
1989	337,059	167907	50%	111,176	66%	33%
1990	353,765	178096	50%	117,123	66%	33%
1991	354,565	180780	51%	117,099	65%	33%
1992	350,286	183551	52%	114,554	62%	33%
1993	347,385	194872	56%	112,879	58%	32%
1994	335,186	191757	57%	92,287	48%	28%
1995	349,484	200610	57%	77,163	38%	22%
1996	356,748	203658	57%	64,006	31%	18%
1997	369,726	159190	43%	53,323	33%	14%
1998	379,601	136483	36%	43,461	32%	11%
1999	357,485	117032	33%	28,966	25%	8%
2000	430,355	109564	25%	14,809	14%	3%
2001	407,285	81616	20%	4,159	5%	1%
2002	163,362	30996	19%	1,373	4%	1%
2003	8,469	3692	44%	64	2%	1%
Total	9027486	4427840		2594044		

Figure 1: Patent Counts by Backward Citation Type

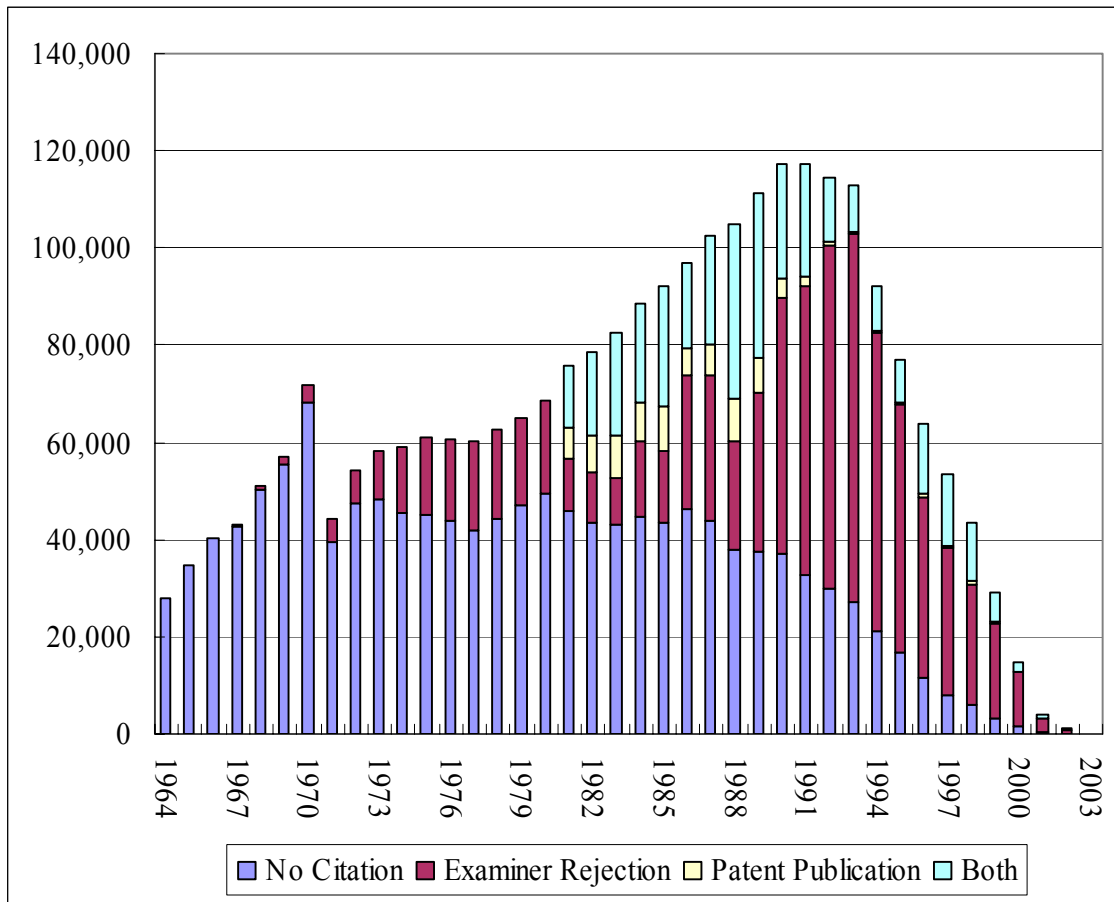


Figure 2: Patent Application by Technology

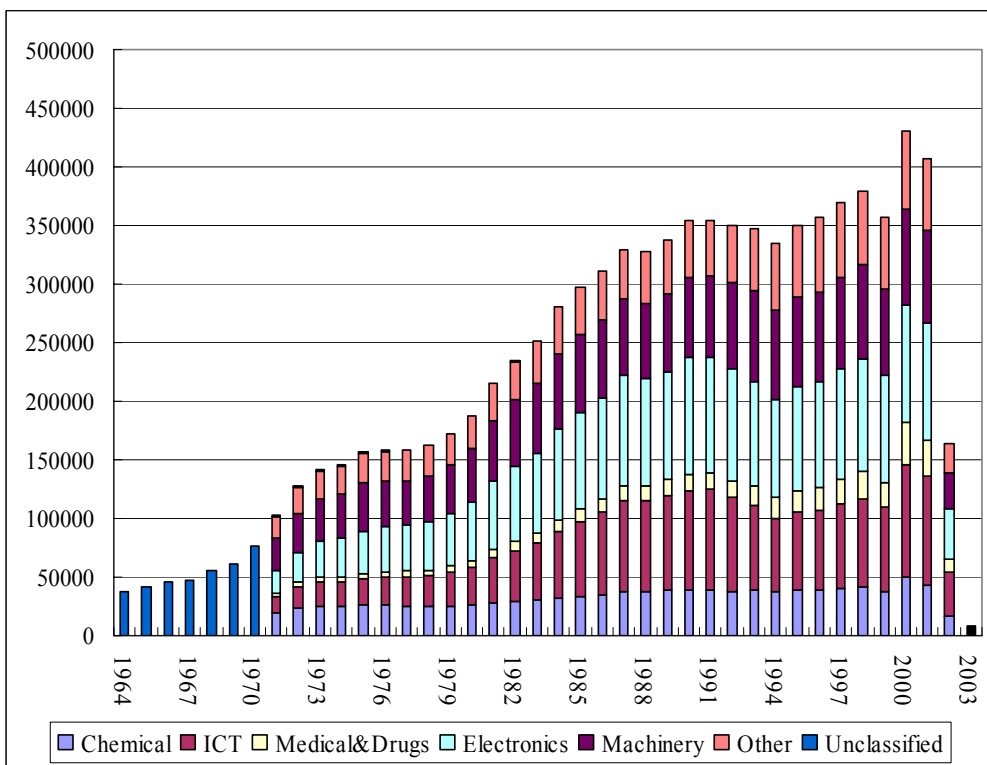


Figure 3: Patent Application by Applicant Type

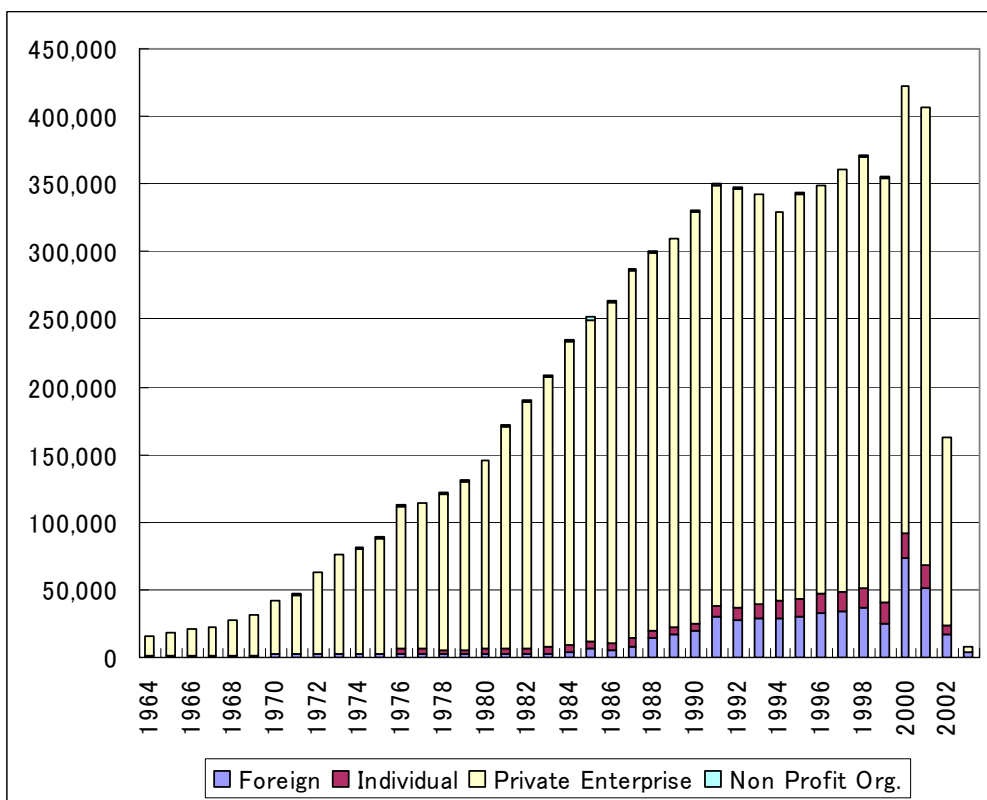


Figure 4: Patent Registration by Technology

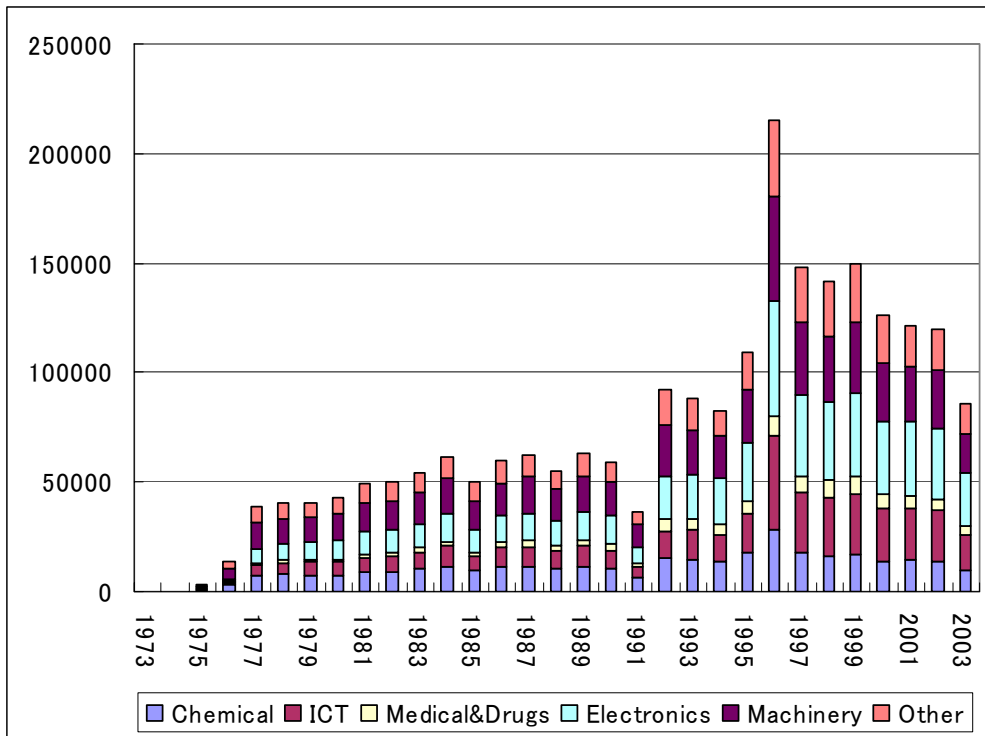


Figure 5: Number of Claims by Technology

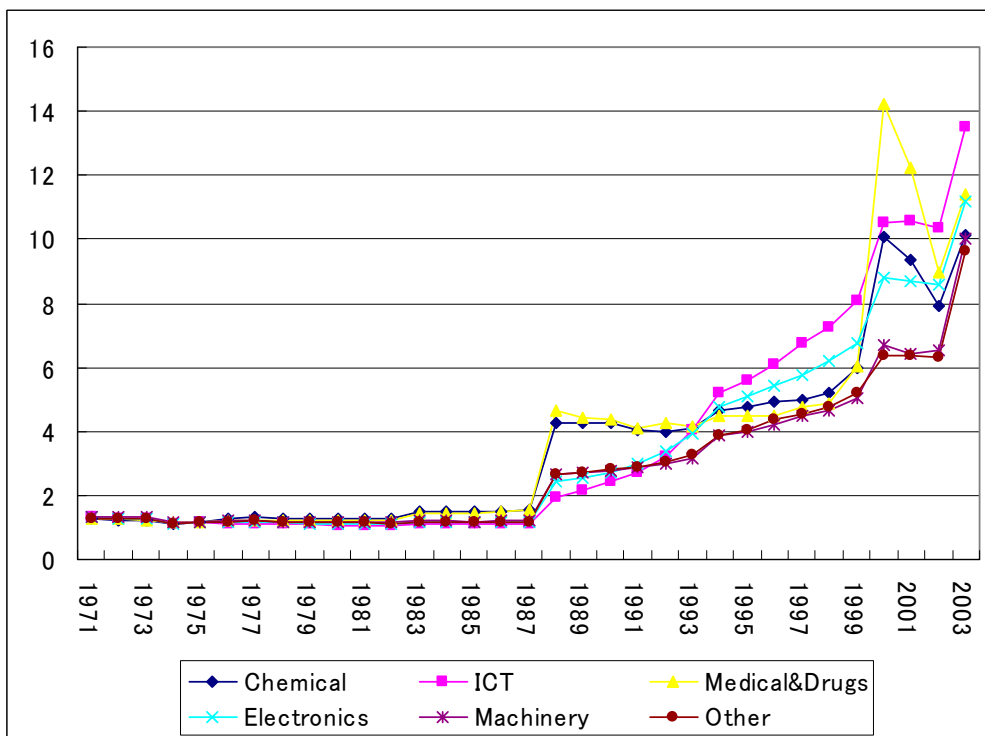


Figure 6: Patent Life Length by Applicant Type

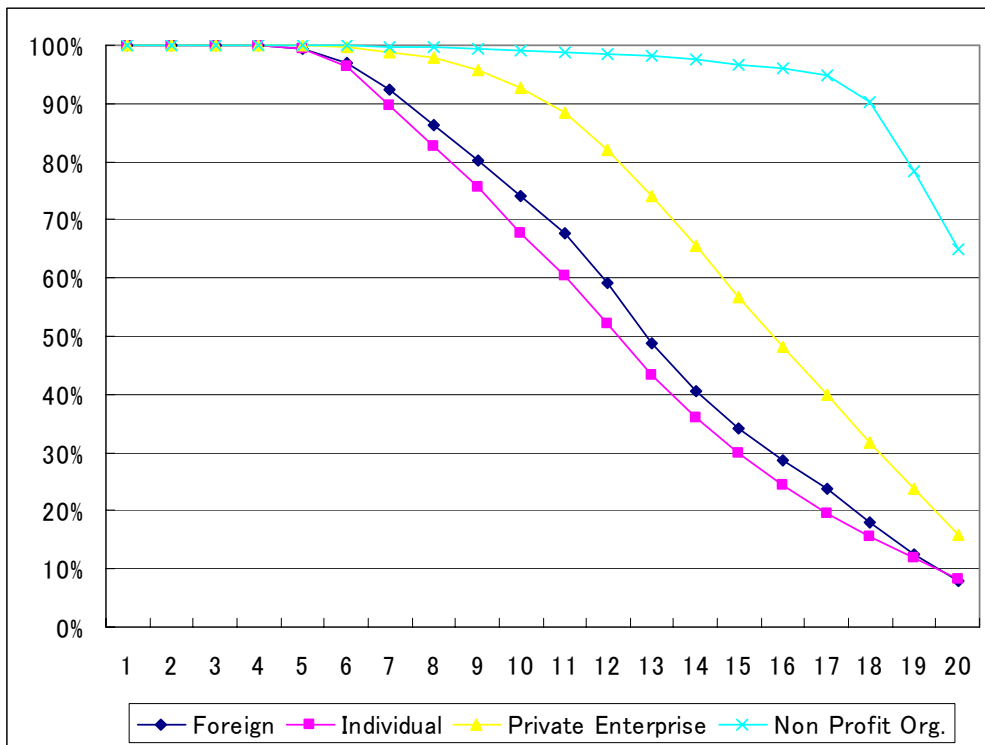


Figure 7: Patent Life Length by Application Date

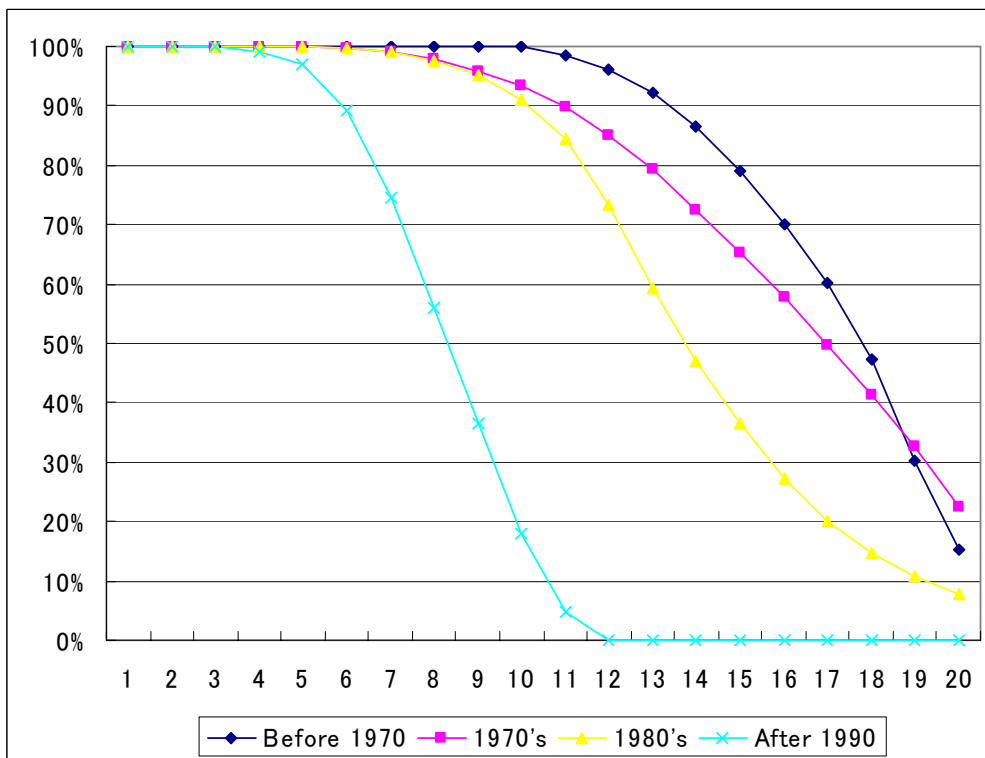


Figure 8: Share of patents with citing (or cited) patents

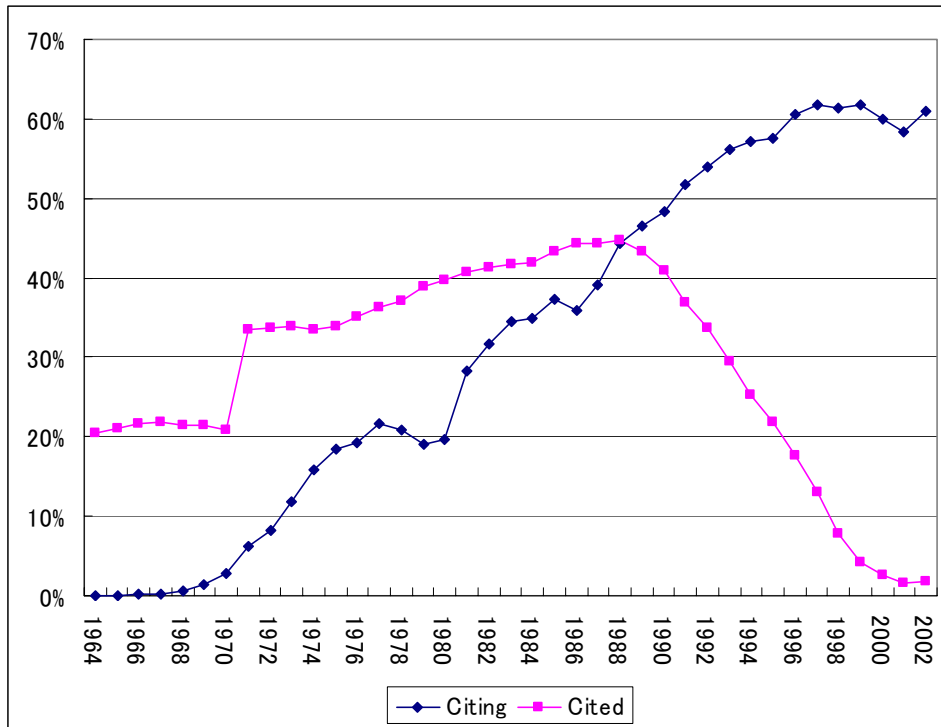


Figure 9: Indicators for Citation Made (Backward Citation)

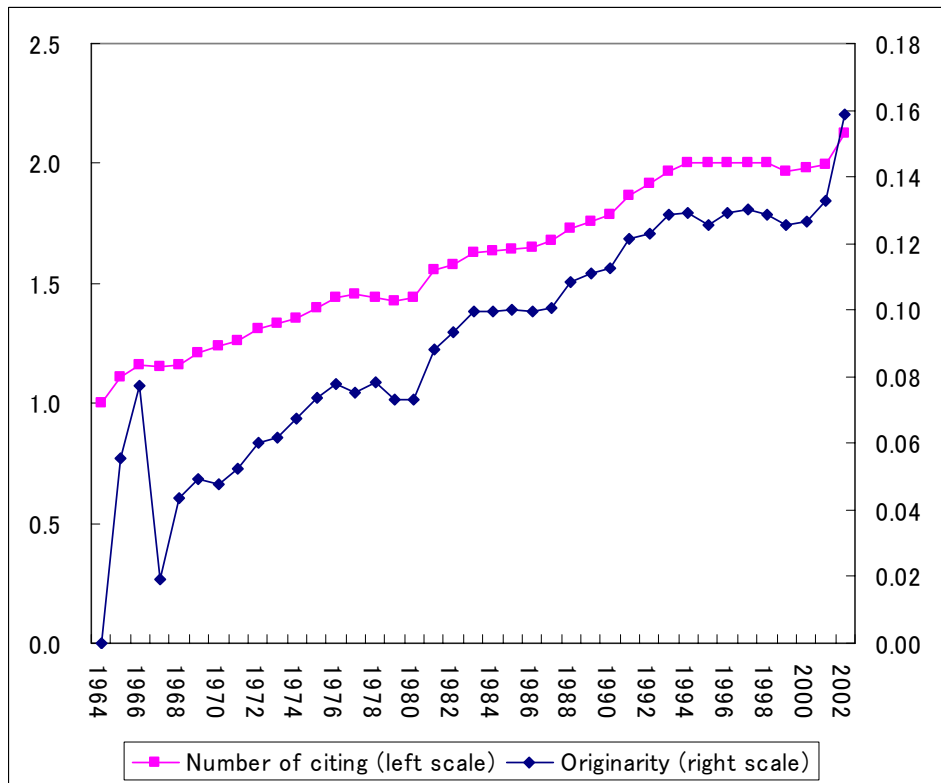


Figure 10: Indicators for Citation Received (Forward Citation)

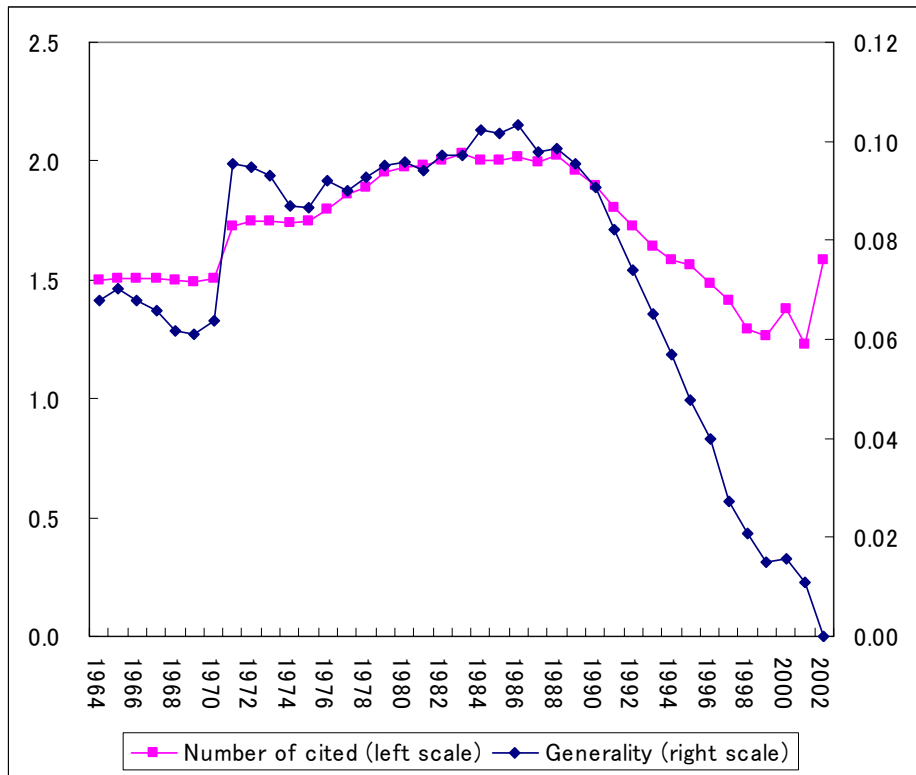


Table 3: Regression result of citation indicators

	Share Citing (%)	Mean # Citing	Mean Originality	Share Cited (%)	Mean # Cited	Mean Generality
1970's	0.000 (.)	0.000 (.)	0.000 (.)	0.000 (.)	0.614 (11.53)**	0.086 (9.85)**
1980's	0.144 (8.07)**	0.197 (4.63)**	0.021 (1.28)	0.037 (2.95)**	0.626 (11.80)**	0.096 (11.11)**
1990's	0.299 (16.82)**	0.416 (9.81)**	0.048 (2.89)**	-0.128 (10.27)**	0.353 (6.66)**	0.060 (6.89)**
2000's	0.376 (21.00)**	0.453 (10.43)**	0.059 (3.50)**	-0.319 (25.34)**	0.000 (.)	0.000 (.)
Firms	0.057 (3.75)**	0.118 (3.26)**	-0.012 (0.82)	0.036 (3.33)**	0.152 (3.83)**	-0.013 (1.95)
Public Organization	0.005 (0.33)	-0.021 (0.56)	-0.001 (0.06)	0.026 (2.38)*	-0.007 (0.17)	-0.001 (0.18)

(continuing to the next page)

Table 3: contunued

Food Stuffs	0.103 (2.05)*	0.379 (3.06)**	0.028 (0.58)	0.056 (1.60)	0.008 (0.06)	0.024 (1.11)
Personal and Domestic Articles	0.031 (0.63)	0.111 (0.89)	0.063 (1.32)	-0.024 (0.69)	-0.122 (0.90)	0.029 (1.30)
Health and Amusement	0.125 (2.51)*	0.309 (2.55)*	0.026 (0.56)	0.077 (2.18)*	0.310 (2.30)*	0.047 (2.12)*
Drugs	0.076 (1.52)	0.608 (4.91)**	0.109 (2.26)*	0.075 (2.14)*	0.254 (1.83)	0.079 (3.49)**
Separating, Mixing	0.152 (3.03)**	0.355 (2.93)**	0.102 (2.16)*	0.045 (1.29)	0.164 (1.24)	0.104 (4.83)**
Machine tools, Metal working	0.215 (4.28)**	0.452 (3.73)**	0.063 (1.34)	0.047 (1.34)	0.106 (0.80)	0.052 (2.43)*
Casting, Grinding, Layered Product	0.106 (2.12)*	0.340 (2.74)**	0.090 (1.88)	0.040 (1.13)	0.145 (1.10)	0.090 (4.19)**
Printing	0.194 (3.78)**	0.603 (4.87)**	0.112 (2.34)*	0.083 (2.31)*	0.416 (3.08)**	0.063 (2.88)**
Transporting	0.107 (2.13)*	0.188 (1.55)	0.023 (0.49)	0.013 (0.37)	0.142 (1.05)	0.039 (1.76)
Packing, Lifting	0.125 (2.43)*	0.128 (1.03)	0.022 (0.46)	-0.003 (0.07)	-0.043 (0.33)	0.050 (2.31)*
Non organic chemistry, Fertilizer	0.210 (4.20)**	0.486 (4.01)**	0.083 (1.77)	0.073 (2.08)*	0.213 (1.62)	0.078 (3.65)**
Organic chemistry, Pesticides	0.042 (0.84)	0.429 (3.54)**	0.031 (0.67)	-0.006 (0.18)	0.042 (0.32)	0.043 (1.99)*
Organic molecule compounds	0.164 (3.27)**	0.519 (4.28)**	0.067 (1.42)	0.109 (3.10)**	0.408 (3.02)**	0.132 (5.99)**
Dyes, Petroleum	0.191 (3.82)**	0.674 (5.56)**	0.141 (3.01)**	0.079 (2.23)*	0.114 (0.87)	0.103 (4.81)**
Biotechnology, Beer, Fermentation	0.018 (0.36)	0.256 (2.12)*	0.060 (1.28)	-0.021 (0.60)	-0.080 (0.59)	0.059 (2.68)**
Genetic Engineering	-0.059 (1.13)	0.120 (0.95)	0.080 (1.63)	-0.035 (0.94)	-0.146 (1.02)	0.074 (3.15)**
Metallurgy, Coating metals	0.197 (3.93)**	0.602 (4.97)**	0.086 (1.82)	0.094 (2.66)**	0.263 (2.03)*	0.103 (4.88)**
Textile	0.219 (4.37)**	0.608 (5.01)**	0.133 (2.83)**	0.116 (3.31)**	0.193 (1.49)	0.051 (2.43)*
Paper	0.156 (3.04)**	0.335 (2.64)**	0.052 (1.05)	0.117 (3.24)**	0.308 (2.28)*	0.084 (3.80)**
Construction	0.159 (3.18)**	0.167 (1.38)	0.008 (0.16)	0.018 (0.52)	0.070 (0.53)	0.033 (1.54)
Mining, Drilling	0.218 (4.35)**	0.439 (3.54)**	0.111 (2.32)*	0.038 (1.08)	0.362 (2.53)*	0.066 (2.84)**
Engine, Pump	0.136 (2.71)**	0.313 (2.58)*	0.068 (1.44)	-0.003 (0.09)	0.069 (0.51)	0.028 (1.27)
Engineering elements	0.160 (3.18)**	0.209 (1.72)	0.052 (1.10)	0.002 (0.07)	0.271 (2.05)*	0.043 (2.01)*
Lighting, Steam generation, Heating	0.125 (2.50)*	0.230 (1.90)	0.054 (1.14)	-0.009 (0.27)	0.089 (0.66)	0.038 (1.75)
Weapons, Blasting	0.170 (3.39)**	0.173 (1.43)	0.109 (2.32)*	-0.035 (1.01)	-0.042 (0.30)	0.054 (2.40)*
Measurement, Optics, Photography	0.178 (3.55)**	0.462 (3.81)**	0.051 (1.08)	0.085 (2.41)*	0.303 (2.35)*	0.043 (2.04)*
Clock, Controlling, Computer	0.165 (3.29)**	0.266 (2.19)*	0.078 (1.67)	0.073 (2.09)*	0.281 (2.17)*	0.060 (2.83)**
Display, Information Storage, Instruments	0.181 (3.61)**	0.293 (2.42)*	0.035 (0.75)	0.019 (0.55)	0.334 (2.53)*	0.042 (1.93)
Nuclear physics	0.163 (3.25)**	0.345 (2.79)**	0.152 (3.16)**	0.063 (1.79)	-0.079 (0.58)	0.094 (4.27)**
Electronics components, semiconductor	0.189 (3.78)**	0.411 (3.39)**	0.050 (1.07)	0.061 (1.74)	0.282 (2.18)*	0.038 (1.79)
Electronics circuit, communication tech.	0.178 (3.55)**	0.454 (3.75)**	0.066 (1.40)	0.052 (1.48)	0.222 (1.69)	0.058 (2.72)**
Others	0.425 (6.47)**	0.303 (1.93)	0.282 (4.63)**	-0.063 (1.37)	-0.124 (0.39)	0.050 (0.95)
Constant	0.016 (0.41)	0.992 (10.51)**	0.027 (0.73)	0.293 (10.93)**	0.972 (9.16)**	-0.037 (2.13)*
Observations	384	376	376	384	337	337
R-squared	0.67	0.45	0.19	0.77	0.48	0.48

Absolute value of t statistics in parentheses

\* significant at 5%; \*\* significant at 1%

Appendix: Examining JPO's citation information by comparing those of EPO and USPTO

In this appendix, we compare Japanese patent citation data with those of Europe and the U.S. The OECD patent family database allows us to compare citations given to the same patent in Japan, Europe and United States. The OECD patent family database is comprised of patents on inventions which have been patented, with the same priority date, in at least two of the following three jurisdictions: Japan, U.S., and Europe (Hernis and Khan, 2004). Assuming that all patents belonging to a particular patent family cover the same underlying technology, this database allows us to directly compare citing and cited reference pairs across Japan, U.S., and Europe.

First, we took citing-cited patent pairs in Japanese, U.S., and European database, and assigned them to their corresponding patent families. Specifically, even if only a single patent in a family included a particular citation, we assigned that as a citation to the family as a whole. For U.S. citation information we employed the NBER patent database and the update data compiled by Hall (Hall et. al, 2000). For European citation information we employed the test-user version of citation data provided by OECD and EPO (Webb et. al, 2005). Because the period covered by the OECD patent family database is 1978 – 2002, both the cited references and the citing references in the data after our conversion to family-level analysis are restricted to this period. Table A1 compares the citation information among three at patent family base.

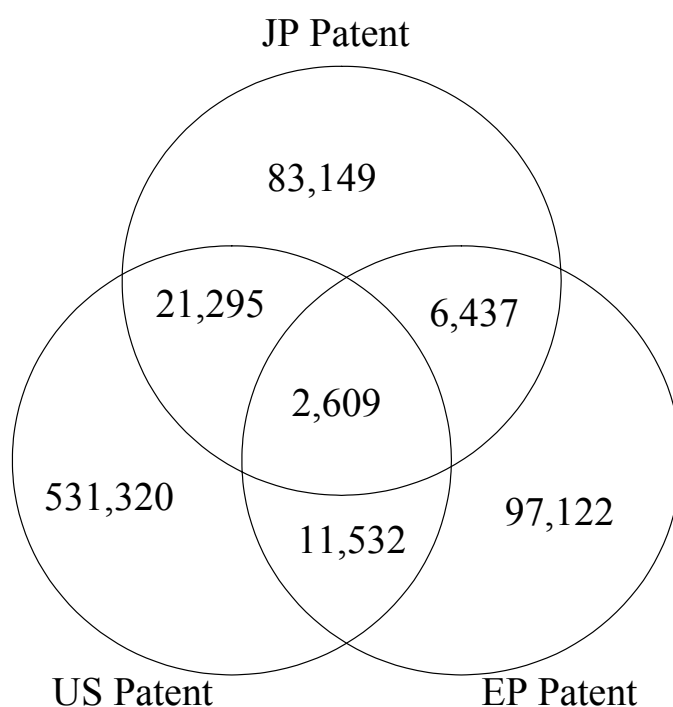
Table A1. Comparison of the number of citing and cited references in patent families

	Total Family #	Citation Number	Average Family #
JP	117700	77045	1.53
EP	113490	80991	1.40
US	566756	205974	2.75

The number of patent families involved in citation is about 80,000 in both Japanese and European data, but is over 200,000 – more than twice as large – in U.S. data. Also when viewed in terms of the average number of citations per patent family, the 2.75 average citations in U.S. appears large relative to 1.53 in Japan and 1.40 in Europe. Both Japanese and European citation data are limited to the smallest necessary number of citations judged important by patent examiners, whereas U.S. data is supplied by patent applicants who tend to be very broad in their selection of references to cite. (Michel and Bettels, 2001) This tendency is confirmed at the level of the patent family.

To what extent do the family-level citing-cited reference pairs match among the three jurisdictions? Figure A1 portrays this relationship. Only 2,609 pairs – a mere fraction of the total – appear in all three jurisdictions. Of the 117,700 Japanese pairs, 9,046 have a match in European data and 23,904 have a match in U.S. data. This reveals a considerable difference in the content of citations across the three data sets.

Figure A1: Overlapping numbers of citing and cited pairs (patent family base)

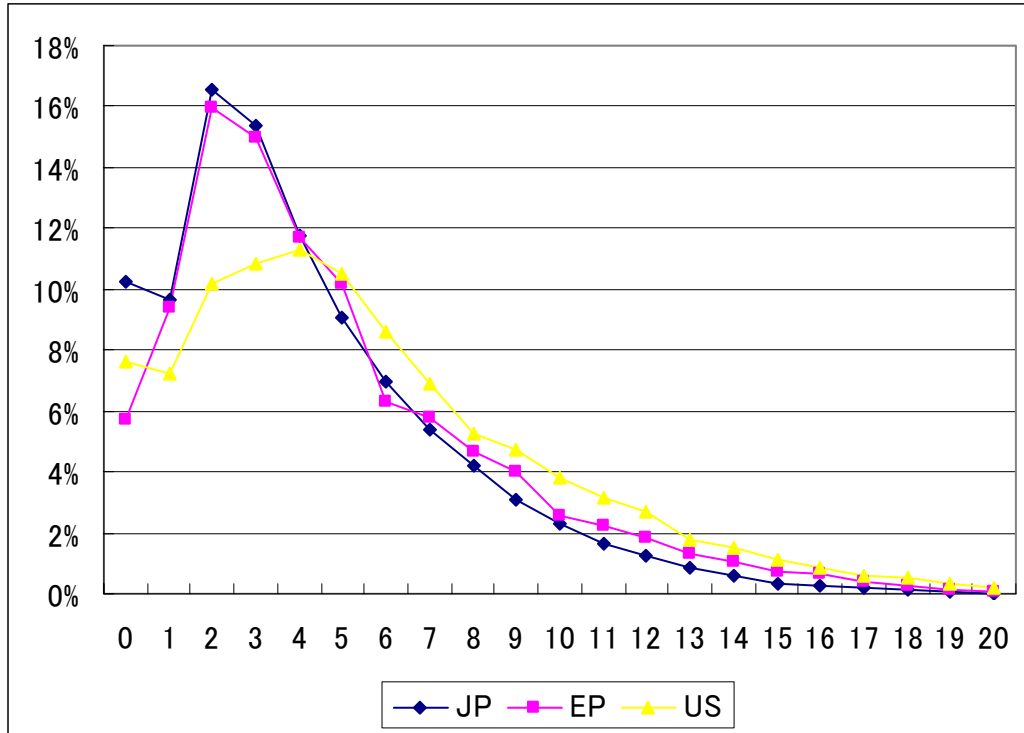


This difference can be considered to be a consequence of the aforementioned difference in how and by whom citations are made in Japan and Europe, on the one hand, and in the United States, on the other. Moreover, under the Japanese and European systems, all patent applications are published, and a large portion of citations refer to these published applications. In contrast, under the U.S. system, patent applications are not published until recently and therefore cited patents are available only for granted patents. Thus, the time lags between the citing patent and the cited patent in Japanese and European citation data can be considered to be shorter than such lags in U.S. data.

Figure A2 compares citation time lags (in terms of year of priority) in the family-level data in Japan, Europe and the U.S. It confirms that the citation time lag is shorter in Japan and Europe: the lag peaks at 5 years in the U.S. data and at 2 years in Japanese and European data. Although the extent of correspondence between citing-cited

reference pairs in Japanese and European citation data is rather low, the pattern of citation lags appears to be similar.

Figure A2: Citation Lag (OECD Patent Family base)



In light of the above, citing-cited reference pairs in Japanese, European, and U.S. patents do not always indicate the same correspondence relationship between inventions (patents). In addition to being caused by differences in patent systems such as whether applications are published, this could also be a consequence of a “home country bias,” a bias towards citing own-country patents. Patent citation data is can be a valuable source of information to understand the process of cumulative innovation. To examine this process in Japan, Europe, and the U.S., we compared backward and forward citation patterns using family-level citation patterns among three regions. This comparison yielded the following results.

Table A2 shows the pair-wise correlations between backward citation counts found in Japanese, U.S., and European citation data. All three pairs, namely Japan-Europe, Japan-U.S., and U.S.-Europe, have correlation coefficients in the range of 0.22 – 0.24, which indicates that some degree of correlation is present.

Table A2. Correlation of backward citation frequencies

	EP	US	JP
EP	-		
US	0.22	-	
JP	0.23	0.24	-

In Table A3, we present the pair-wise correlation coefficients of forward citation counts. The degree of correlation becomes higher than that of backward citation counts. It is found that a forward citation count represents the value of patent very well, based on the USPTO citation information (Lanjouw and Schankerman, 1999). In this respect, the citation information from IIP Patent Database can be used also for similar type of analysis.

Table A3. Correlation of forward citation frequencies

	EP	US	JP
EP	-		
US	0.42	-	
JP	0.42	0.36	-