We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Exploring Characteristics of Patent-Paper Citations and Development of New Indicators

Yasuhiro Yamashita

Additional information is available at the end of the chapter

Abstract

In this study, the characteristics of "papers cited in patents" are examined and impact indicators of them based on existing bibliometric indicators are developed. First, the nature of patent-paper citations is examined for Japanese scientific papers as the basic knowledge for developing indicators. Second, the patent-paper citation index (PPCI) indicator, which was proposed in the previous study, is revised. Third, a set of indicators, named High Feature Valued Patent-Paper Citation Index, which is based on three feature values of citing patents, is proposed. Evidence using our new indicators is presented and the tendency of patent-paper citations of Japanese three sectors such as university, public institute, and corporation is discussed. Finally, issues to be addressed are discussed.

Keywords: patent-paper citations, impact indicators of papers, bibliometrics, institutional sectors, normalized citation impact, patent-paper citation index, technological impact, high feature valued patent-paper citation index

1. Introduction

Today, scientific research is expected not only to create knowledge but also to contribute to the development of industrial technology and the solution of social problems. Citations of scientific papers from patents (hereafter patent-paper citations) are rare data representing knowledge flows between coded scientific knowledge (scientific papers) and coded technological knowledge. Although there have been controversies over what is meant by patent-paper citations, it is deemed as data representing knowledge flows and used in the public statistics at present (e.g., see [1–3]).

As an indicator representing the relationship between science and technology, the number of cited scientific documents per patent (it is known as "science linkage") has been widely used. It is relatively straightforward to introduce science linkage, since it does not require identification of each scientific paper cited in patents and match to a specific record in databases of academic papers, such as Web of Science (WoS) and Scopus. However, science linkage only provides information on vicinity of science from technology, not vicinity of technology from science.

Along with the research utilizing science linkage as an index as described above, the nature of patent-paper citations itself has been studied. Such studies needed identification of bibliography of papers which appeared in patent documents. For example, Branstetter and Ogura [4] used data of patent-paper citations provided by CHI Research and analyzed the relationship between probabilities of occurring patent-paper citations and some variables obtained from both patents and papers for California. Such research had been relatively scarce, since they required a large-scale data set with identified paper data. However, in recent years, Ahmadpoor and Jones analyzed a large citation network, which consisted of patent-patent, paper-paper, and patent-paper citations, based on a large data set of US patents and scientific papers indexed in the Web of Science database provided by Clarivate Analytics and comprehensive patent-paper citation data [5]. They dealt with both patent-patent and paper-paper citations symmetrically and handled patent-paper citations like it bordered between these two networks and then uncovered differences in various aspects of them. Fukuzawa and Ida [6] analyzed the features of patent-paper citations from the paper side for 100 top researchers who were awarded the twenty-first-century COE. They found some important characteristics of patent-paper citations, such as the time lag of the former was longer than the latter, and the more the papers were cited from other papers, the more they tended to be cited from patents.

While these findings are important for practical use of patent-paper citations, there are almost no existing studies on the development of impact indicators of papers cited in patents.

On the other hand, the demand for methods of analysis and empirical indicator data of "papers cited in patents" in practical context has been expanded recently. For example, the Fifth Science and Technology Basic Plan which is the current Japanese five-year national plan for the promotion of science and technology between FY 2016 and 2020 requires monitoring of the performance. "Scientific papers cited in patents" is one of the key performance indicators of the plan. However, an effective method for showing performance using patent-paper citations is still unclear; therefore, it is indispensable to develop valid indicators of patent-paper citations.

My motivation is to develop impact indicators for scientific papers to show technological impact at meso (institutional sector in a country, research funding, and so on) to macro levels (country), based on the statistical nature of patent-paper citations. In the field of bibliometrics, many indicators have been developed and verified by many researchers (see [7]) and practical uses such as Leiden Ranking and Scimago Journal & Country Rank. Therefore, by developing robust impact indicators based on patent-paper citations symmetrical to existing bibliometric impact indicators, it should be possible to overview both the scientific and technological impacts of researches at the same time.

Moreover, from the view of patents, there have been many indicators for measuring patent quality (major indicators were written in [8]). For evaluating scientific papers from the aspect

of contributions to technological development, citations of scientific papers from "high-impact" patents seem to be good indicators of scientific papers. As far as my survey, I could not find any empirical study of indicators from the view mentioned above.

According to the aforementioned problem consciousness, I develop the new impact indicators of papers in the aspect of patent-paper citations. To secure the validity of new indicators, we investigate the nature of patent-paper citations in the dataset prior to the development of the indicators.

This article consists of the following sections. In Section 2, I explain data and time scheme of the study. I analyze relationships between probabilities of occurrence of paper citations from the patents and feature values of the scientific papers, using logistic regression analysis in Section 3. Based on the result of the analysis in Section 3, I improve the patent-paper citation index which we developed recently [9] (Section 4) and develop a set of new indicators from the aspect of patents' feature values (Section 5). Then, issues to be tackled are discussed in Section 6.

2. Data and their process

I utilized data sources and decided time scope in the study in the following process.

2.1. Patent data

I used worlds' patent data contained in the 2016 spring edition of the Patstat database produced by European Patent Office (EPO). The database contains patent applications filed until January 2016 and publications published until February 2016.

To avoid overrating the same inventions, patent data were counted by the DOCDB patent family. Only patent families which contain published patents, neither utility models nor design patents, were included in the dataset for securing consistencies of their statistic natures. Patent families are counted by their application year. The application year of the patent family was defined as the earliest filing year of the applications that constituted the family. Patent families which no application belonged to any of technology field defined in [10] were excluded, since percentiles of patent-patent citations were calculated by technology field.

2.2. Data of scientific papers

The Science Citation Index Expanded collection of the WoS database was used for this study. The WoS database contained bibliographic records of scientific papers which were published between 1981 and 2015. Each scientific paper in the WoS was classified to 1 of 22 scientific disciplines of the Essential Science Indicators. As for journals classified in "Multidisciplinary" by Clarivate Analytics, each of their papers was classified into 1 of the other 21 disciplines using their information on both forward and backward citations. Papers which were not classified into any of the 21 disciplines by the process were classified into "Multidisciplinary." They were excluded from the study because most of them obtained no or only a few citations and tended to be overestimated in the calculation of percentiles in the "Multidisciplinary" discipline. Disciplinary classification used in the study is shown in **Table 1**. Hereafter, I designated the codes for disciplines in the figures in this article.

Code	Discipline	Code	Discipline
AGS	Agricultural sciences	MTS	Materials science
BBI	Biology and biochemistry	MIC	Microbiology
CHE	Chemistry	MOL	Molecular biology and genetics
CLM	Clinical medicine	NEB	Neuroscience and behavior
CPS	Computer science	PHT	Pharmacology and toxicology
ECB	Economics and business	PHY	Physics
ENE	Environment/ecology	PLA	Plant and animal science
ENG	Engineering	PSS	Psychiatry/psychology
GSC	Geosciences	SPA	Space science
IMU	Immunology	SSS	Social sciences, general
MAT	Mathematics		

Table 1. Disciplinary classification of the study.

2.3. Linking non-patent literatures in the Patstat to specific papers in the WoS

All non-patent literatures appeared in the TLS214_NPL_PUBLN table of the Patstat and were matched to each bibliographic record of the WoS, so that citation links between them were identified. As a result of this process, 11,753,856 patent-paper citation links from Patstat to the WoS were identified. Number of WoS papers cited in the Patstat were 2,669,386, excluding duplications.

2.4. Attribution of institutional sectors to authors' organizations

Institutional sectors of authors' organizations were needed to be attributed to analyze tendencies of patent-paper citations by institutional sector in the following sections. The Connection Table between "Web of Science Core Collection" (WoSCC) and "NISTEP Dictionary of Names of Universities and Public Organizations" publicly provided by National Institute of Science and Technology Policy, Japan, was used for the purpose. The table consists of IDs of scientific papers in the WoS (UT), organization names, and sector and some other information extracted from the NISTEP Dictionary of Names of Universities and Public Organizations. The table contains UTs of Japanese papers published between 1998 and 2015 of which document types were "Article" or "Review." Therefore, the scope of data used in the study was limited to these document types and publication years.

The sectoral classification of the research was derived by combining the categories of the NISTEP table as shown in **Table 2**.

2.5. Time scheme of the study

As a result of the preprocess mentioned above, a scheme of time periods for analysis was set as **Figure 1**. A 6-year citation window (7 years including publication year of the scientific papers)

| Sector classification in the study	Sector classification in NISTEP table
University	National university, public university, private university, interuniversity research institute
Public Institute	National institute, government-affiliated public corporation/independent administrative institution, institute of local government
Corporation	Corporation

Table 2. Institutional sector classification in the study.
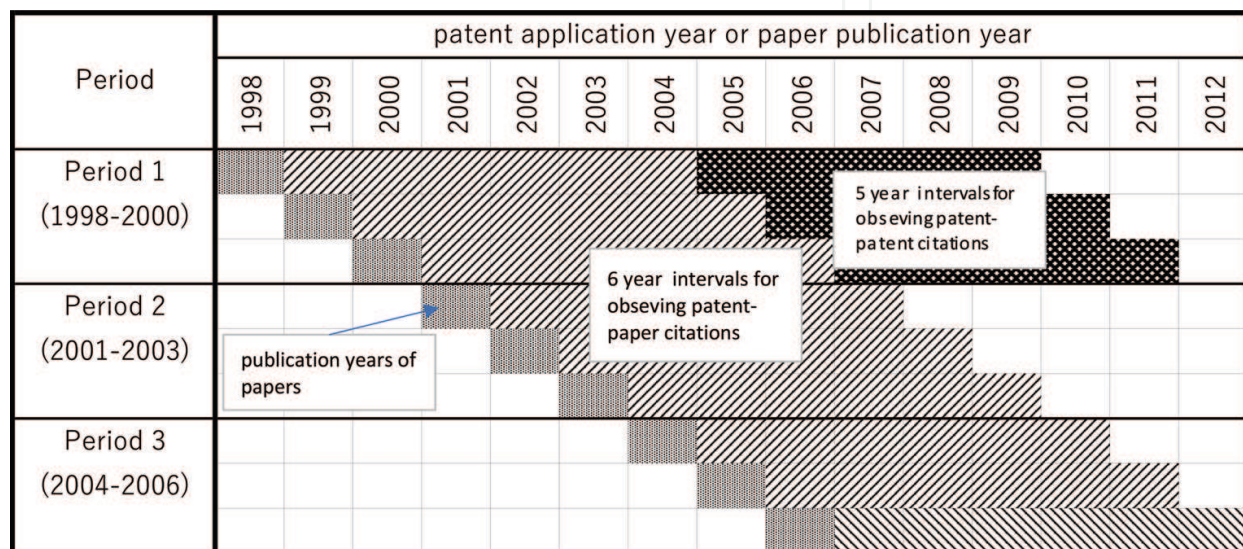
Figure 1. Time scheme of the study.

was secured for both patent-paper and paper-paper citations. The 6-year citation windows were defined in our previous study based on the criterion that at least a half of observable patent-paper citations could be grasped [9]. As for the earliest period (Period 1), 5-year citation windows were set according to [8] for observing citations from patents to patents citing target papers.

2.6. Basic statistics of the dataset

As a result of the abovementioned process, a dataset for the study, which consisted of 6,962,541 records of the worlds' scientific papers published between 1998 and 2006, was obtained. The number of Japanese papers by institutional sector counted fractionally by the number of addresses appearing in each paper in the dataset was shown in **Figure 2**. Japanese universities published 72.4% of Japanese papers; public institutes and corporations published 13.3 and 8.6%, respectively. When rate of papers cited in patents in papers of each sector was calculated, the above orders were reversed; the rate of papers cited in patent of corporation, public institutes, and universities was 21.6, 11.2, and 10.2%, respectively.

Number of the worlds' papers published between 1998 and 2006 by discipline was shown in **Figure 3**. Both clinical medicine and chemistry showed large numbers of papers, and that
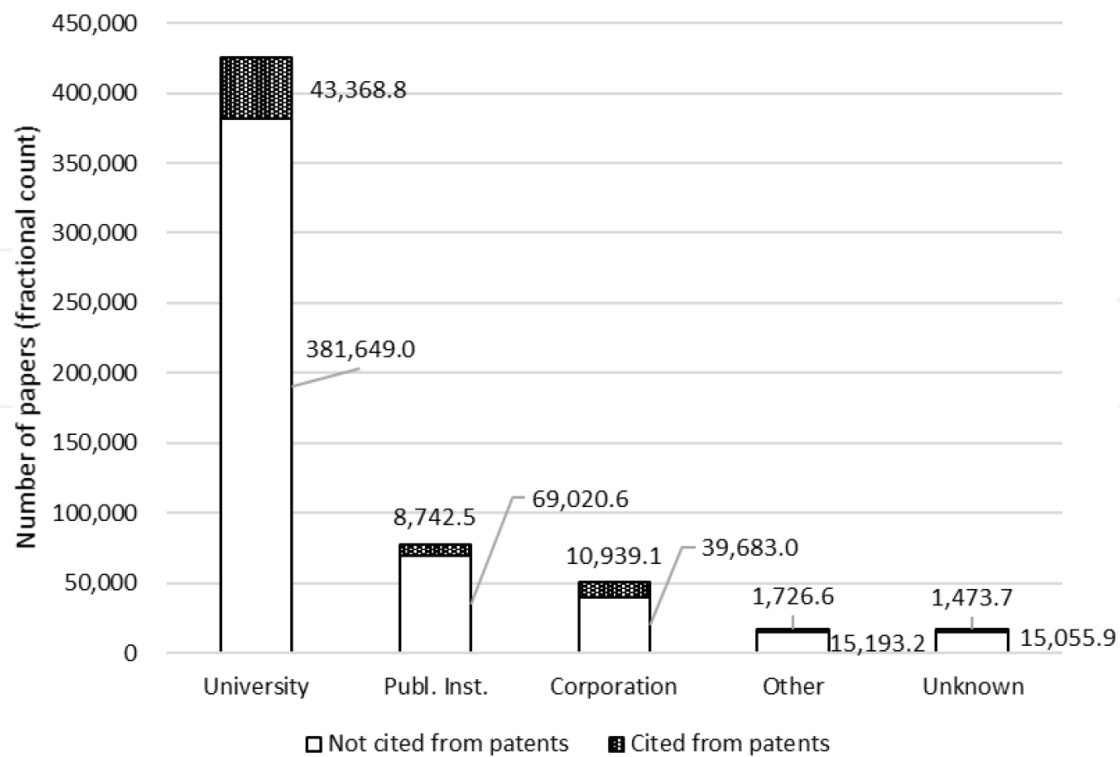
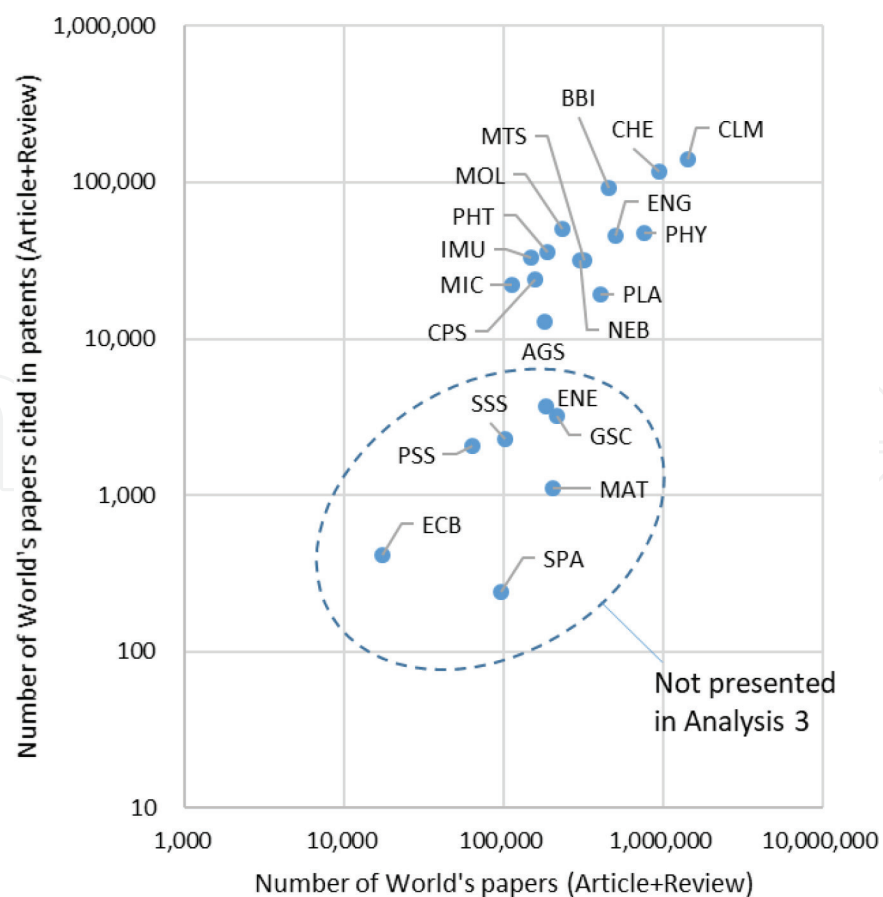**Figure 2.** Number of Japanese papers by sector in 1998–2006.

Figure 3. Number of publications and papers cited in patents between 1998 and 2006.

was cited in the patents. Biology and biochemistry showed relatively smaller numbers of papers but showed comparatively close number of that cited in patents to clinical medicine and chemistry. Therefore, it showed a relatively higher rate of papers cited from patents per their papers. Seven disciplines surrounded by the dotted circle in **Figure 3** showed both small number of papers and that was cited in the patents. These disciplines were excluded from presentation in analysis 3 (Section 5), in which analysis was executed and presented by discipline. However, these seven disciplines were included in the calculation as in other analyses, i.e., analysis 1 (Section 3) and analysis 2 (Section 4).

3. Relationships between feature of papers and patent-paper citations (analysis 1)

3.1. Research question

Patent-paper citations are different from paper-paper citations in their statistic nature, such as their small amount compared to that of the latter. Therefore, some indicators developed in bibliometrics cannot be applied to patent-paper citations. To develop valid indicators, many aspects of their tendencies, especially which kind of papers were preferred to cite in patent, should be grasped. Although some studies tackled this question partially [4–6, 11], their analyses were restricted to the US patents [4, 5, 11] or limited numbers of "top" researchers [6].

Moreover, it is still unknown how papers were cited from patents of which feature values were relatively high (hereafter, they are called as high-feature-valued patents). Branstetter [12] addressed the question whether patents citing papers tended to be high feature valued. However, his approach was done from the patent side, not the paper side. Patent-paper citations from high-quality patents seemed to be more valuable from the view of possibility of occurrence of innovation in many cases.

Here, I tried to grasp statistical tendencies of relationship between patent-paper citations from both all patents and those with high feature values. I intended to show the difference between them and to obtain basic knowledge of paper citations from high-feature-valued patents to develop valid indicators and show tendencies of (Japanese) scientific research from multi-aspects of patent-paper citations in the following sections.

3.2. Relationship between feature values of patents and their patentability

Although many "quality indicators" have been proposed, it might be questionable whether all of them exactly reflect patent quality. Since they each focused on different aspects of patents, they might represent different features of patents, not all of which represent "quality." To facilitate a precise understanding of the results of analysis of patent-paper citations from patents with high-"quality indicators" (hereafter they are called as "feature values" since they were not necessarily representative of quality), and the meaning of the new indicators proposed in Section 5, here I tried to show differences in meaning of the various major patent feature values.

In this subsection, I focused on the relationship between the three major feature values of patents: patent family size, forward citations (hereafter it is called as patent-patent forward

Independent variable	Coefficient	Std. err	Z value	Pr(>|z|)	Signif. codes
Intercept	0.474038	0.004213	112.51	<2e-16	***
Patent family size	0.257991	0.001038	248.62	<2e-16	***
Patent-patent forward citation (Top 1%)	0.029541	0.000276	107.02	<2e-16	***
Patent Generality Index	−0.188278	0.006765	−27.83	<2e-16	***

Signif. codes: "***" 0.001, "**" 0.01, "*" 0.05, "." 0.1, ""1.

Table 3. Result of logistic regression analysis of patent feature values.

citations to distinguish it from other kinds of citations), and patent generality index. They are three of the four components of "composite index 4" presented in [8]. "Claims," which was the rest of the four, was not included in the study because it was not included in the Patstat comprehensively (only the US patents and European patents comprehensively included it exceptionally). As for "patent-patent forward citations," a dummy variable which distinguished whether patents obtained the top 1% of citations from other patents or not (it was presented as a "breakthrough" indicator in [8]) was used. The percentile of patent-patent citations was calculated by each of the 35 technology fields defined in [10].

Here, logistic regression analysis, of which independent variables were three patent feature values mentioned above, was executed. "Granted" flag in TLS201_APPLN table in the Patstat was selected as dependent variable, since it should represent an aspect of patent quality. Please note that this analysis was executed in the initial stage of the study before the specification of dataset was decided; therefore, all types of patents (such as utility models) were included.

The results are shown in **Table 3**. All coefficients of the three independent variables were significant at 0.1 percent level. Two of them (patent family size and patent-patent forward citations) were positive, and the rest was negative. As far as grant of patents was regarded as representative of patent quality, the former represents some aspects of patent quality. Patent family size could be thought of as quality assessed by applicants themselves (self-assessed quality), since "applicants might be willing to accept additional costs and delays of extending protection to other countries only if they deem it worthwhile" (p. 14) [8], while patent-patent forward citations could be deemed as quality assessed mainly by other applicants or examiners. On the other hand, the patent generality index seemed not to represent patent quality in the aspect of patentability.

3.3. Relationships between features of scientific papers and their citedness from all/high-feature-valued patents

In this subsection, I explored which features of papers affect their citedness from patents to grasp basic nature of patent-paper citations which might influence the nature of indicators presented in the following sections. Since we utilized information on patent-patent citations in which patents citing papers obtained, the analysis in this section was executed for Period 1 (PY1998–2000) in **Figure 1**.

Independent variable	(a) Cited/not		(b) Large patent family (> = 15)		(c) High patent-patent forward citation (top 1%)		(d) High patent generality index (> = 0.85)	
	Coefficient		Coefficient		Coefficient		Coefficient	
(Intercept)	−2.504476	***	−4.30065	***	−4.91949	***	−5.27141	***
Review	0.125596	*	0.29487	*	0.30569	*	0.20671	.
Int Coauthored					−0.09241		0.08866	.
IF	0.269865	***	0.15193	***	0.14490	***	0.13507	***
Top 10%	1.417856	***	1.42854	***	1.65927	***	1.59834	***
University	−0.281680	***	−0.42518	***	−0.35581	***		
Publ Inst	−0.038220	.	−0.36932	***	−0.11208	.	0.17765	***
Corporation	0.837952	***	0.83858	***	0.81681	***	0.62083	***
Other								
AGS	−0.268111	***	−0.39318	*	−0.91885	**		
BBI	0.895510	***	0.33564	***	0.57985	***	0.85431	***
CHE	0.044250	.	−0.35768	***	0.13846	*	0.76895	***
CPS	0.296150	***	−2.09236	***	0.80914	***		
ECB	−0.806569							
ENE	−1.403637	***	−2.40567	***	−1.92041	**	−1.21992	*
ENG	−0.144508	***	−3.77031	***	0.29438	***	0.27416	**
GSC	−3.268167	***	−15.37014		−2.16536	***	−3.29013	**
IMU	1.074738	***	1.19463	***	0.89635	***	0.49992	***
MAT	−4.296640	***	−15.18047		−13.65028		−13.66243	
MTS	−0.426886	***	−2.19212	***	0.25507	**	0.76189	***
MIC	0.829376	***	0.31394	*	−0.71977	*	0.29761	
MOL	1.063727	***			0.53478	***	0.94839	
NEB					−0.22540	.	−0.20025	
PHT	0.402472	***	0.71171	***	0.19559	.		
PHY	−0.559729	***	−3.77438	***				
PLA	−0.475982	***	−1.35393	***	−0.49401	***	−1.22287	***
PSS	−1.228774	***	−1.60205		−13.72792		−13.74086	
SPA	−4.640363	***	−15.22141		−13.74022		−13.82943	
SSS	−1.694540	***	−2.23959	*	−1.75911	.	−13.78140	

Signif. codes: "***" 0.001, "**" 0.01, "*" 0.05, "." 0.1, "" 1.

Table 4. Result of logistic regression of rate of patent-paper citations.

I tried to include broad feature values of papers which might affect their citedness from patents as widely as possible to grasp characteristics of patent-paper citations comprehensively. Six feature values (document type, international co-authorship, impact factor (hereafter IF), paper-paper citations, institutional sectors and disciplines) shown in **Table 4** were selected from [13]. In **Table 4**, the variable "Review" and "Int-Coauthored" represents the feature value "document type" and "international co-authorship," respectively, and the variables "University" to "Other" and "AGS" to "SSS" represent "institutional sectors" and "disciplines," respectively.

I executed logistic regression analyses of which independent variables were six feature values of papers mentioned above and dependent variables were distinct from whether papers were cited from (all or high-feature-valued) patents (1) or not (0). To ignore the shape of distributions of patent-paper citations, I discarded information on the number of citations but used distinction of cited or not.

IFs were obtained from the Journal Citation Reports produced by Clarivate Analytics. Since IFs changed every year, years of IFs were defined as publication years of papers. This was because I intended to use them as the journals' quality indicators independent of the target papers. IFs in a year Y were calculated using papers published in years Y-1 and Y-2; therefore, they did not contain the target papers in the calculation. As it was well known, values of IFs differed largely by discipline; therefore, they were normalized by the following process: (1) IFs were attributed to each paper in the WoS (but IFs could not be given to some papers exceptionally); (2) mean values of IFs attributed to papers by ESI discipline were calculated for each year; (3) IF attributed to each paper was normalized by mean IF of its ESI discipline.

The threshold values of feature values of patents were decided according to the criteria: number of papers cited in high-feature-valued patents should be almost the same. As the number of papers cited from the top 1% patent-patent forward citation patents was predetermined, it was used as the reference value of number of papers cited from high-feature-valued patents. Threshold values were set to 15 for patent family size, 0.85 for patent generality index. Therefore, patents of which patent-patent forward citations were within top 1% or patent family sizes or patent generality indexes were equal to or more than the abovementioned thresholds were defined as high-feature-valued patents in this study.

Document types "Article" and discipline "Clinical Medicine (CLM)" were set to reference, since they were classified exclusively.

The results of the logistic analyses were shown in **Table 4**. Since patent-paper citations from high-feature-valued patents ((b), (c), (d)) were subsets of the whole patent-paper citations, they showed somewhat similar tendencies.

As for document type, reviews showed positive relationships to probabilities of being cited from both patent ((a)) and all three types of high-featured-valued patents ((b)-(d)). The result on patent ((a)) reinforced the result by Hicks et al. [11]. This result showed that indicators should be weighted by document type as far as possible.

International co-authorship showed no statistically significant relationship to any kinds of paper citedness. While Japan's co-authorships with any country were combined into the same

flag, it might show a statistically significant difference if difference of countries was taken into account. However, the number of international co-authored papers was limited, so we did not divide them into specific countries.

IF showed positive relationships with all kinds of patent-paper citations. This result reinforced analysis of Guan and He [14]. They showed nine of ten journals most frequently appeared as non-patent literatures in Chinese inventors' US patent were ranked within the top ten in their categories in the Journal Citation Report. Therefore, papers published in prestigious journals tended to be more cited than those published in lesser known journals.

The top 10% of paper-paper citations also showed positive relationships with all kinds of patent-paper citations, as many previous studies [5, 6, 11].

Institutional sectors showed some interesting tendencies; corporations showed relatively strong tendencies to be cited from all four kinds of patents ((a)-(d)). Although university and public institutes tended not to be cited from patents generally, they were not so from patents with high patent generality indexes. Latter tendencies might be explained that universities and public institutes produce generic knowledge, not focus on specific industrial applications, so patents citing them tended to also have a generic nature.

As for disciplines, some of the life sciences (biology and biochemistry, immunology, microbiology, molecular biology and genetics, pharmacology and toxicology) showed tendencies to be more cited (than clinical medicine, which was a reference discipline), while most physical sciences (engineering, materials science, physics) showed opposite tendencies. Similar results were reported in previous studies, such as [11]. However, it also showed some interesting tendencies when citations from high-feature-valued patents were focused on. For example, computer science tended to be more cited relatively, while they tended to be less cited from large patent families; engineering and materials science tended not to be cited from patents, while they tended to be cited from patents of top 1% patent-patent forward citations; microbiology showed an opposite tendency in that they tended to be cited from patents, while they tended not to be cited from patents of top 1% patent-patent forward citations. What caused such differences? To answer this question, further investigation from the patent side is needed.

4. Improvement of the patent-paper citation index (PPCI) (analysis 2)

4.1. Definition of improved PPCI

In the previous study, we proposed an impact indicator of patent-paper citations, named patent-paper citation index (PPCI) [9]. PPCI is based on rates of the papers cited from patents in the targets' publications. We proposed a method to overview targets' research activities from both scientific and technological impacts compared to the world average by using normalized citation impact (NCI) [13] in combination. Differences in both document types and disciplines were ignored in the previous study [9]. However, the analysis in Section 3.3 revealed their effects on papers' tendencies to be cited from patents. Therefore, I propose an improved version of PPCI in this section.

NCI, which was the basis of PPCI, is the ratio of the number of paper-paper citations which the target paper got to the expected value of that of the same cohort papers in the world. NCI is calculated for paper by paper, so when it is applied to an aggregate, such as institutions or countries, the average per their publications' NCI is applied. On the other hand, PPCI is based on the rate of papers cited in patents in targets' publications. Indeed, it is preferable to apply the same definition as NCI to secure symmetry; we applied the abovementioned definition to avoid influence of limited highly cited papers, since the rate of papers cited from patents was relatively smaller than that from papers.

Improved PPCI was defined as Eq. (1):

$$p_{ijd} = \frac{\left(n'_{ijd}/n_{ijd}\right)}{\left(N'_{id}/N_{id}\right)} \tag{1}$$

where.

n_{ijd}: number of target j's papers with document type d published in discipline i;

n'_{ijd}: number of target j's papers cited in patents with document type d published in discipline i;

N_{id}: number of total papers with document type d published in discipline i; and

N'_{id}: number of total papers cited in patents with document type d published in discipline i.

Target j's field weighted PPCI was calculated as follow:

$$P_j = \frac{\sum_i \sum_d p_{ijd} \times n_{ijd}}{\sum_i \sum_d n_{ijd}} = \frac{\sum_i \sum_d \left(N_{id} \times n'_{ijd}/N'_{id}\right)}{\sum_i \sum_d n_{ijd}} \tag{2}$$

To increase visibility, we normalized PPCI by Eq. (3):

$$\text{Normalized } P_j = \frac{\left(P_j - 1\right)}{\left(P_j + 1\right)} \tag{3}$$

Hereafter, improved Normalized PPCI (Eq. (3)) is merely called as PPCI.

While the whole counting method was used to count Japanese sectors' publications in the previous study [9], the fractional counting method by number of addresses which appeared in each paper was used. The whole counting method always attributed one count to each target appeared in a paper, so they are easy to understand intuitionally; however, it often causes overrating to multiauthored papers.

4.2. Chronological changes of NCI and PPCI of Japanese sectors

Next, I tried to apply PPCI to three Japanese sectors (university, public institute, corporation) to show how PPCI could describe the scientific and technological impact of aggregate of meso (sector) level. This was mainly aimed to figure out on which level of aggregates PPCI could be used. The chronological change of both NCI and PPCI was shown in **Figure 4**.
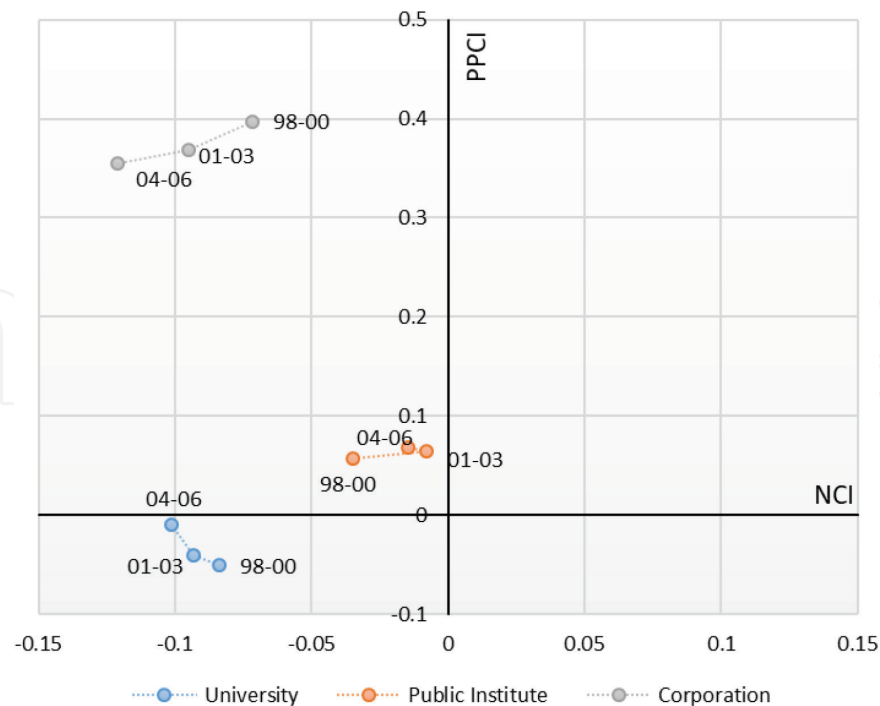
Figure 4. Chronological change of NCI and PPCI of three Japanese sectors.

All three sectors were located on the left half of the plane, which meant average scientific impacts of them were below world average during three periods. Two sectors, public institute and corporation, were located on the second quadrant; therefore, their average technological impacts were above the world average. In particular, corporation showed a remarkably high PPCI values and seemed to have been specializing in technological impact only period by period. University, which published most of the Japanese papers, was located on the third quadrant, which meant both scientific and technological impacts were below world average. However, their PPCI had been increasing period by period.

5. Development of high-feature-valued patent-paper citation index (analysis 3)

5.1. Definition

I showed that tendencies of paper citations from high-feature-valued patents differed from whole patents in some cases. It is suggested that indicators based on high-featured-valued patents might reveal hidden structure of the targets' research performance.

I tried to develop another indicator symmetrical to the PPCI to use them in combination. Here, we introduced the indicators based on paper citations from high-feature-valued patents, named high-feature-valued patent-paper citation index (HFPPCI). HFPPCI is a generic name of set of indicators, since there were many kinds of patent feature values. Of the many kinds of patent feature values, I will show the analysis of three patent feature values (patent

family size, patent-patent forward citations, and patent generality index) of Japanese sectors to examine the nature of HFPPCI as well as to show the tendencies of the Japanese sectors.

HFPPCI of target j in discipline i was defined as Eq. (4):

$$p_{ij}^h = \frac{\left(m_{ij}'/n_{ij}\right)}{\left(M_i'/N_i\right)} \tag{4}$$

where.

n_{ij}: number of target j's papers published in discipline i;

m_{ij}': number of target j's papers cited in high-feature-valued patents published in discipline i;

N_i: number of total papers published in discipline i; and

M_i': number of total papers cited in high-feature-valued patents published in discipline i.

To increase visibility, we normalize HFPPCI by Eq. (5):

$$\text{Normalized } P_{ij}^h = \frac{\left(p_{ij}^h - 1\right)}{\left(p_{ij}^h + 1\right)} \tag{5}$$

Here, the difference in document types was ignored, since the number of review papers cited from high-feature-valued patents was very few. Eq. (2) could be applied to aggregate p_{ij}^h into the whole target level; however, the selection of disciplines was inevitable because paper citations from high-feature-valued patents occurred rarely and M_i' might be zero in some cases.

5.2. Japanese sectors' PPCI and HFPPCI by discipline

In this subsection, I tried to analyze the Japanese three sectors' technological impacts by discipline in Period 1 (1998–2000). HFPPCIs of three patent feature values were called as large patent family paper citation index (LPFPCI) for large patent family, high forward citation patent-paper citation index (HFCPCI) for the patents of high patent-patent forward citations, and high generality patent-paper citation index (HGPCI) for patents with a high patent generality index. Definition of high-feature-valued patents was same as Section 3.3: equal or more than 15 for patent family size, top 1% for patent-patent forward citations, and equal or more than 0.85 for patent generality index. In the following subsections, document types were ignored in the calculation of both PPCI and HFPPCI. Both PPCI (X-axis) and HFPPCI (Y-axis) were plotted in bubble charts, and the number of papers cited from high-feature-valued patents was presented as size of the circles in **Figures 5–13**.

5.2.1. University

For LPFPCI, each discipline in **Figure 5** was positioned in line to some extent. This roughly means that large patent families of most of the disciplines in the sector appeared in proportion to papers cited in patents. In this case, there were not very much special information that could be obtained from the LPFPCIs, because PPCI contained almost the same information as LPFPCI. However, it was suggested that the LPFPCI functioned robustly, since there were only few deviating cases.
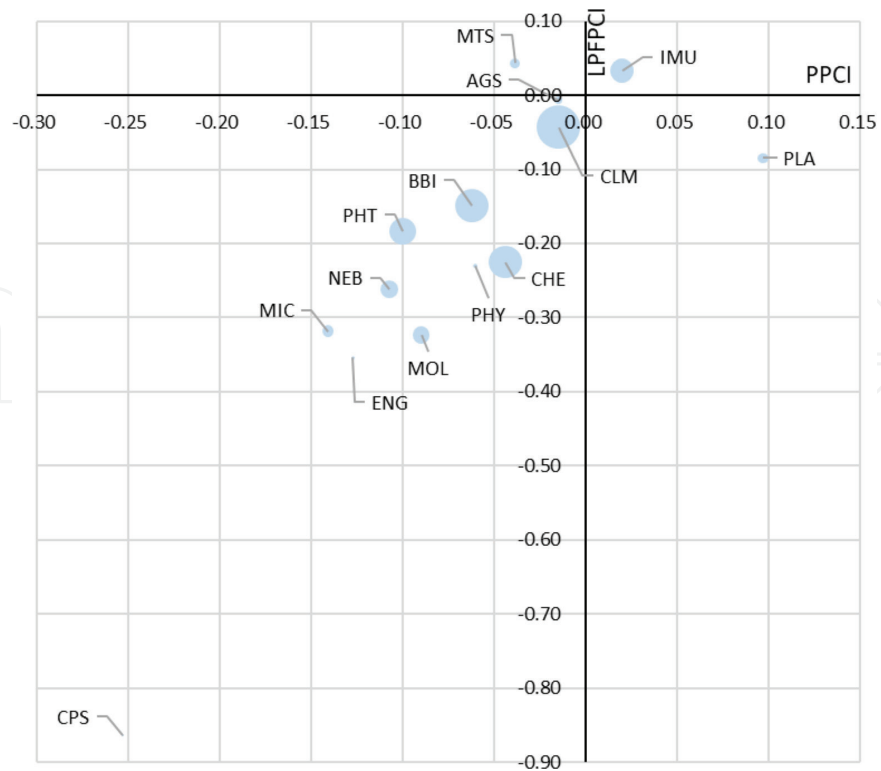
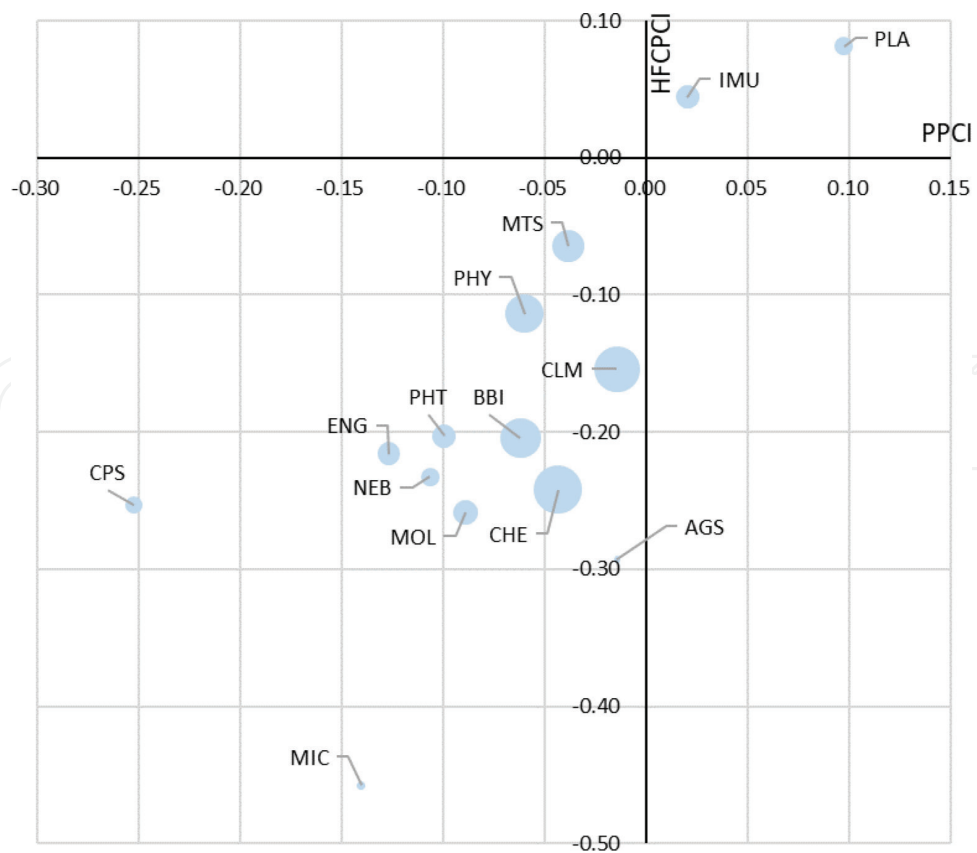**Figure 5.** PPCI and LPFPCI of Japanese university sector by discipline (1998–2000).

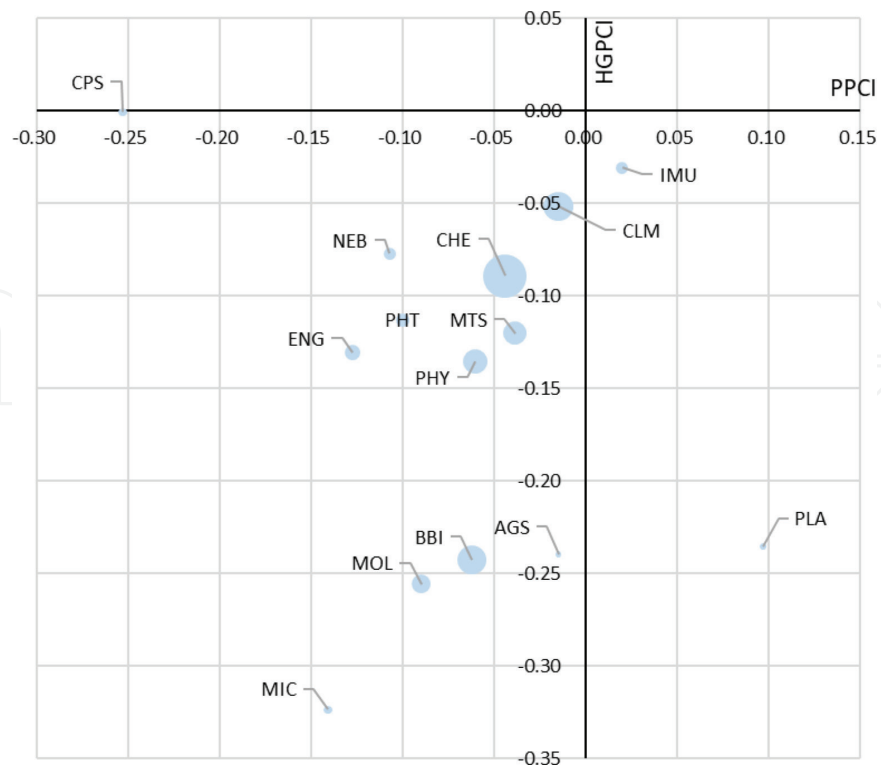**Figure 6.** PPCI and HFCPCI of Japanese university sector by discipline (1998–2000).

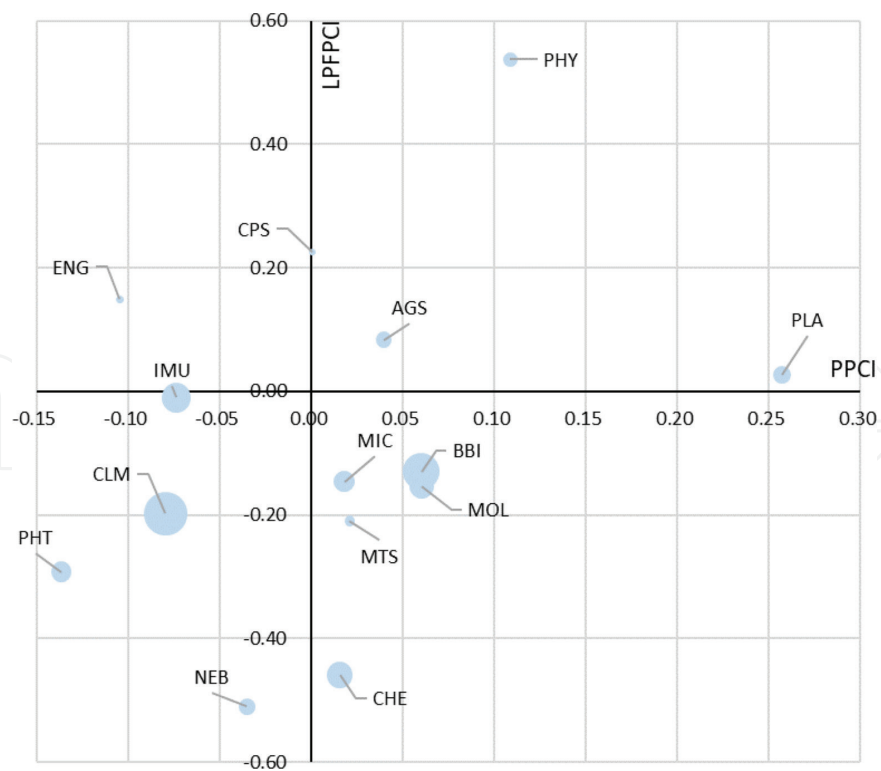**Figure 7.** PPCI and HGPCI of Japanese university sector by discipline (1998–2000).

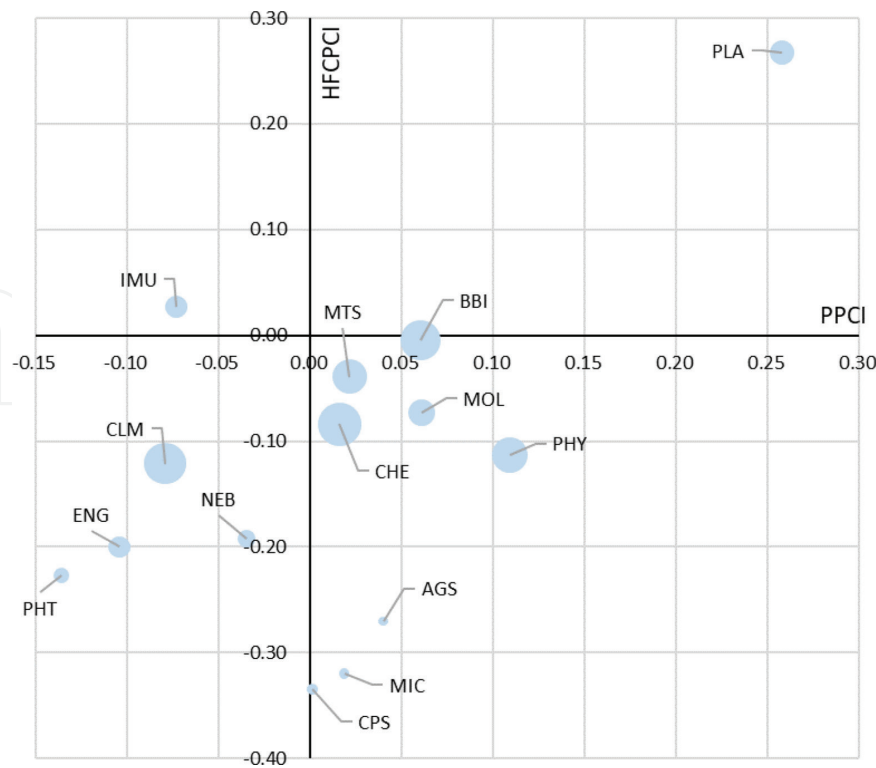**Figure 8.** PPCI and LPFPCI of Japanese public sector by discipline (1998–2000).

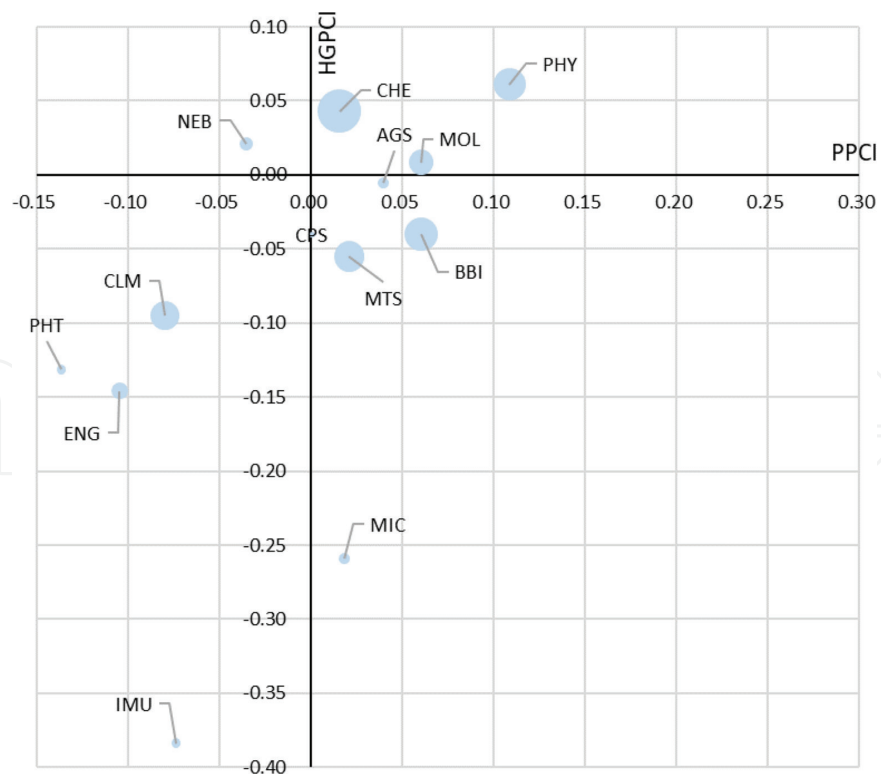**Figure 9.** PPCI and HFCPCI of Japanese public sector by discipline (1998–2000).

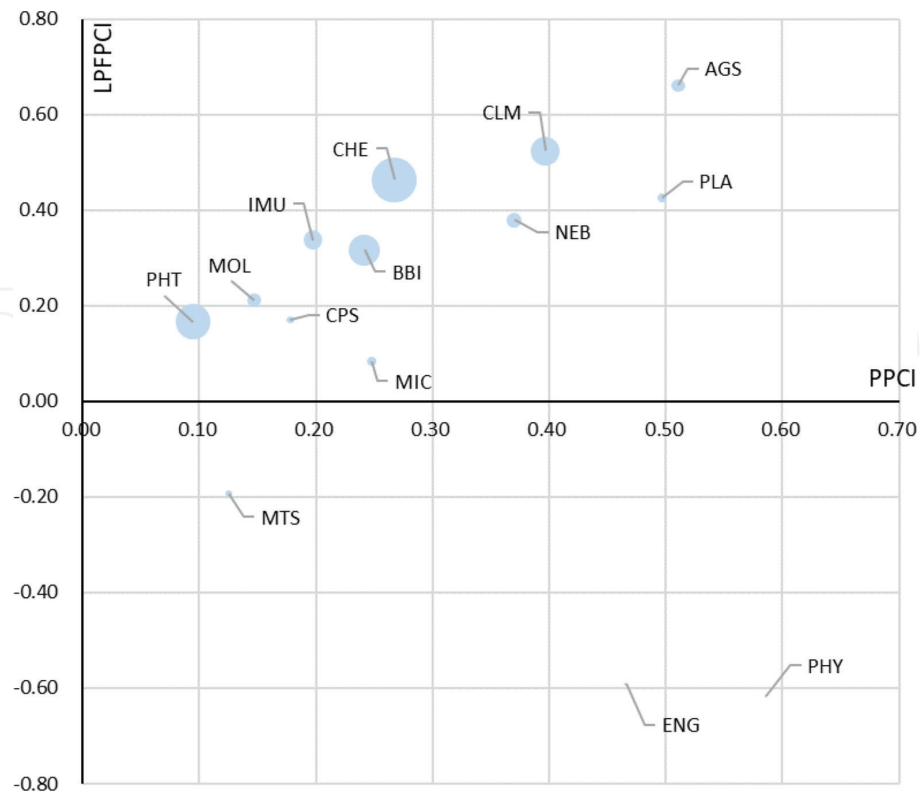**Figure 10.** PPCI and HGPCI of Japanese public sector by discipline (1998–2000).

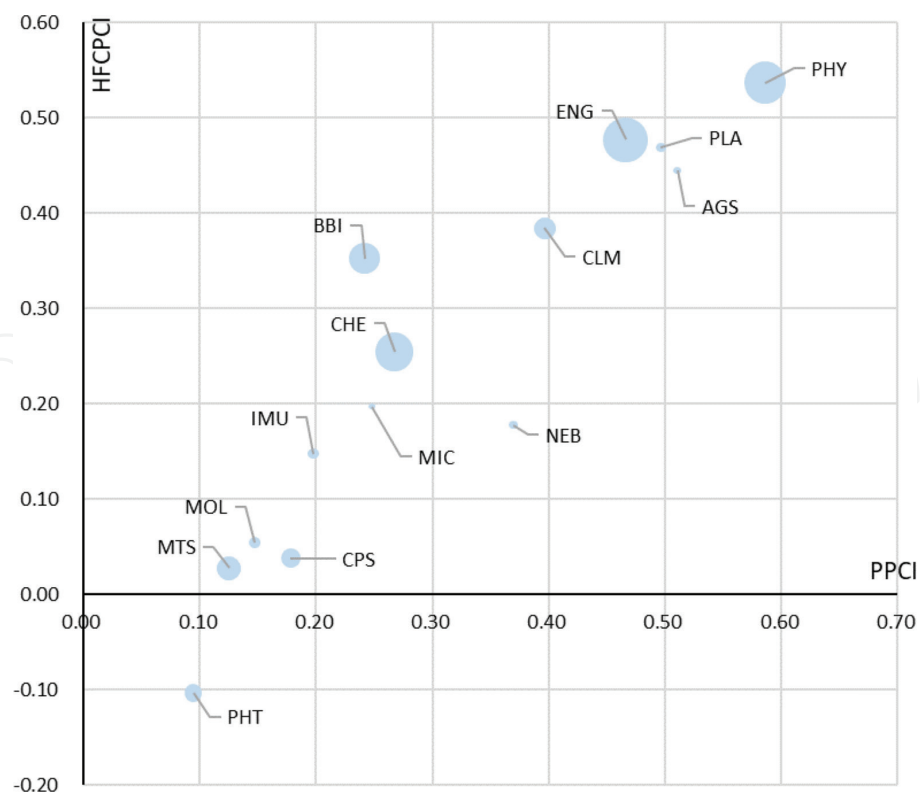**Figure 11.** PPCI and LPFPCI of Japanese corporation sector by discipline (1998–2000).

Figure 12. PPCI and HFCPCI of Japanese corporation sector by discipline (1998–2000).
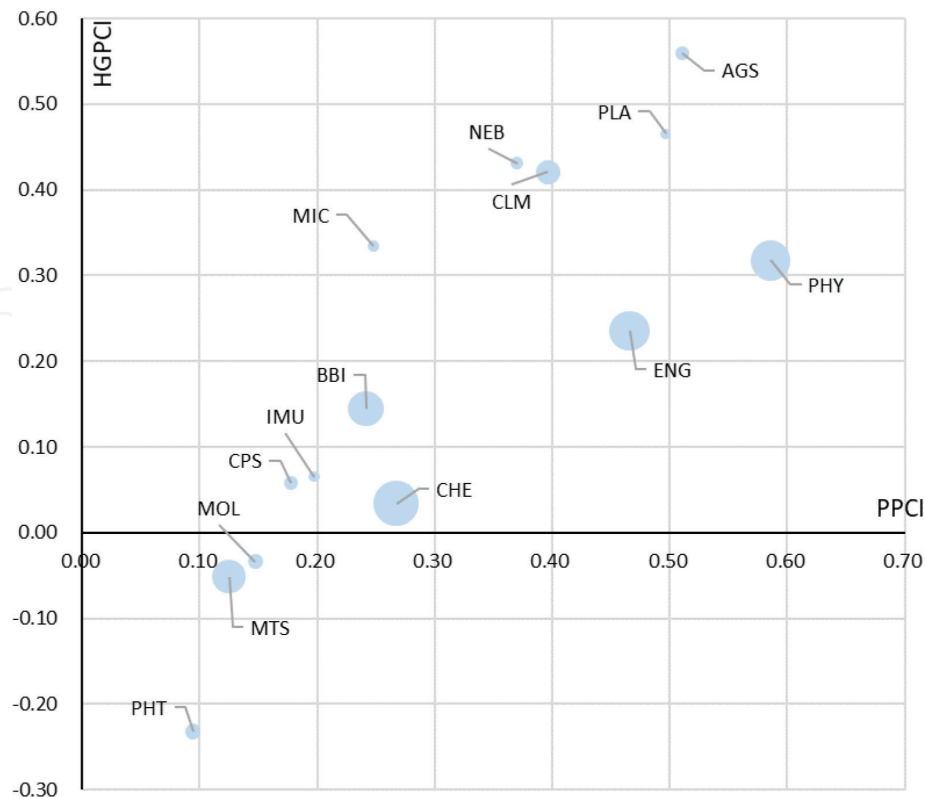
Figure 13. PPCI and HGPCI of Japanese corporation sector by discipline (1998–2000).

For HFCPCI, most disciplines seemed to distribute vertically, suggesting their relatively inconsistent natures in terms of HFCPCI within the sector (**Figure 6**). Two disciplines, immunology and plant and animal science, showed relatively high impact in both PPCI and HFCPCI.

For HGPCI, the university sector seemed to consist of two clusters divided vertically, except for two small disciplines, plant and animal science and computer science (**Figure 7**). The upper cluster consisted of both physical and life sciences, while the lower consisted of life sciences concerning biotechnology.

5.2.2. Public institute

There seemed to be almost no correlation between PPCI and LPFPCI shown in **Figure 8**, and it seemed interesting that relatively smaller circles were located above X-axis while larger circles were opposite. This arrangement was caused by the fact that the disciplines located above the X-axis tended not be cited from the large-sized patent families as a whole, so the Japanese public institute was positioned above average regardless of their small number of papers cited from large-sized patent families.

Most of the disciplines, of which number of papers cited in patents ranked within the top 1% patent-patent forward citations were relatively large, were located on the fourth quadrant (**Figure 9**). Therefore, papers' impact on highly cited patents seemed to be below X-axis totally. This agrees with the coefficient of public institute's patent-patent forward citations, which was below zero as shown in **Table 4**.

For HGPCI shown in **Figure 10**, two relatively large disciplines—Chemistry and Physics—which were located above the X-axis, seemed to make a trend of public institute, because the coefficient of the sector in the column of high patent generality index in **Table 4** was positive.

5.2.3. Corporation

Corporation's prominent performance in both PPCI and HFPPCIs could be seen in **Figures 11–13** in which most disciplines were located on the first quadrant. It was also interesting that all three figures showed a correlation between the two indicators, except for two disciplines (engineering and physics) in **Figure 11**. Therefore, three indicators functioned robustly, regardless of the limited number of papers cited in high-feature-valued patents and corporation's relatively small share of publications in Japan.

Engineering and physics showed opposite impacts in LPFPCI (**Figure 11**) compared to HFCPCI (**Figure 12**) and HGPCI (**Figure 13**). They showed very low values of LPFPCI and limited number of papers cited in large-sized patent families. However, they showed high values of both HFCPCI and HGPCI and relatively large numbers of papers which were cited in patents with the top 1% patent-patent forward citations and with high patent generality index. Although further analysis was needed to show the correct factors of the phenomenon, this might be caused by characteristics of the industries which cited these disciplines.

6. Discussion and conclusion

In this study, three issues were tackled: investigation of the statistical nature of patent-paper citations, development of indicators, and tendencies of Japanese sectors' characteristics concerning patent-paper citations. Here, I discuss the findings and issues needed to be addressed:

1. Investigation in the study revealed the statistical nature of patent-paper citations, i.e., review papers, papers published in high IF journals, and papers highly cited from papers tended to be more cited than papers not so. These characteristics had been reported by previous studies which utilized different datasets and methodologies. Therefore, these results should reveal precise characteristics of patent-paper citations and suggest that fostering excellent scientific research might serve not only science itself but also technological development to some extent.

2. Results of both the logistic regression analysis and analysis by new indicators showed corporation sector's prominence from the view of patent-paper citations. Why were their papers cited more frequently than that of other sectors? To know the reason, identification of patent applicants might be needed, since information on who cited their paper is important to guess the motivation of citations.

3. I showed that (improved) PPCI and HFPPCI could be used to obtain an overview of technological performance of target, whereas there were some problems intrinsic to the rare

and long-tailed nature of citations. If these indicators were used as monitoring tools, a long citation window would be a bottleneck for practical use. Exploring the possibilities of development of methods for shorter-time measurement and to show their availability and limitations should be an important theme.

4. HFPPCI might be inevitably sensitive to small changes in time sequence. Paper citation from high-feature-valued patents is a rarer phenomenon than that from all patents—even the latter is rare. Therefore, only a few citations might yield large changes to values of indicators. Chronological changes of HFPPCIs should be traced to grasp to what extent they are sensitive, and also possibilities for relaxing the threshold to increase samples should be addressed.

Acknowledgements

Author details

Yasuhiro Yamashita

Address all correspondence to: yasuhiro.yamashita@jst.go.jp

Japan Science and Technology Agency, Tokyo, Japan

References

[1] OECD. OECD Science, Technology and Industry Scoreboard 2013. Innovation for Growth. Paris: OECD Publishing; 2013. DOI: 10.1787/sti_scoreboard-2013-en

[2] OECD. OECD Science, Technology and Industry Scoreboard 2015. Innovation for Growth and Society. Paris: OECD Publishing; 2015. DOI: 10.1787/sti_scoreboard-2015-en

[3] NISTEP. Japanese Science and Technology Indicators 2017. NISTEP Research Material No. 261. Tokyo: National Institute of Science and Technology Policy; 2017. DOI: 10.15108/rm261

[4] Branstetter L, Ogura Y. Is academic science driving a surge in industrial innovation? Evidence from patent citations. NBER Working Paper No. 11561; Issued in August 2005

[5] Ahmadpoor M, Jones BF. The dual frontier: Patented inventions and scientific advance. Science. 2017;**357**:583-587. DOI: 10.1126/science.aam9527

[6] Fukuzawa N, Ida T. Science linkages between scientific articles and patents for leading scientists in the life and medical sciences field: The case of Japan. Scientometrics. 2016;**106**:629-644. DOI: 10.1007/s11192-015-1795-z

[7] Waltman L. A review of the literature on citation impact indicators. Journal of Informatics. 2016;**10**:365-391. DOI: 10.1016/j.joi.2016.02.007

[8] Squicciarini M, Dernis H, Criscuolo C. Measuring Patent Quality: Indicators of Technological and Economic Value. OECD Science, Technology and Industry Working Papers, No. 2013/03. Paris: OECD Publishing; 2013. DOI: http://dx.doi.org/10.1787/5k4522wkw1r8-en

[9] Yamashita Y, Jibu M. Exploration of new performance indicator of academic paper citations from patents. JAPIO Yearbook. 2017;**2017**:144-155

[10] Schmoch U. Concept of a technology classification for country comparisons. Final Report to the World Intellectual Organisation (WIPO); June 2008

[11] Hicks D, Breitzman A Sr, Hamilton K, Narin F. Research excellence and patented innovation. Science and Public Policy. 2000;**27**:310-320. DOI: 10.3152/147154300781781805

[12] Branstetter L. Exploring the link between academic science and industrial innovation. Annales d'Économie et de Statistique. 2005;**79**(80):119-142

[13] Thomson Reuters (present Clarivate Analytics) [Internet]. InCites Indicator Handbook. Available from: http://ipscience-help.thomsonreuters.com/inCites2Live/8980-TRS/version/default/part/AttachmentData/data/InCites-Indicators-Handbook-6%2019.pdf [Accessed: September 20, 2017]

[14] Guan J, He Y. Patent-bibliometric analysis on the Chinese science - technology linkages. Scientometrics. 2007;**72**:403-425. DOI: 10.1007/s11192-007-1741-1