Kyoto University,
Graduate School of Economics
Research Project Center Discussion Paper Series

Science linkages focused on star scientists in the life and medical sciences:

The case of Japan

Naomi Fukuzawa
Takanori Ida

Discussion Paper No. E-14-006

Research Project Center
Graduate School of Economics
Kyoto University
Yoshida-Hommachi, Sakyo-ku
Kyoto City, 606-8501, Japan

August, 2014

Science linkages focused on star scientists in the life and medical sciences: The case of Japan

Naomi Fukuzawa

National Institute of Science and Technology Policy, Ministry of Education, Culture, Sports, Science, and Technology
naomi.fukuzawa27@gmail.com

Takanori Ida

Graduate School of Economics, Kyoto University
ida@econ.kyoto-u.ac.jp

Abstract

We analyze the distributions of paper-paper and paper-patent citations and estimate the relationship between them, based on a sample of 4,763 published papers for which the corresponding authors were among the top 100 researchers in the life and medical sciences in Japan. We find that paper-paper citations peak at an average of 4 years after the publication of a paper, while the corresponding lag for paper-patent citations is 6 years. Although there is a time lag before papers can be put to practical use, this lag has shortened in recent years. Moreover, the quality of a paper is important for being cited by a patent, and a paper's quality increases the number of paper-patent citations. In addition, we show that an inverse U-shaped relationship exists between the amount of research grant funding and research quality, and we can derive the efficient amount of research grant funding that maximizes research quality. We find that the relationship between research quality and the total number of papers written by the researcher(s) is U-shaped, and we derive the number of papers that minimizes research quality.

Keywords

Science linkages, Non-patent references, Technology transfer

1. Introduction

Academic knowledge is important for promoting R&D and technological development in industry. Mansfield (1991) surveys the percentage of firms' new products and processes commercialized during 1975–1985 that could not have been developed in the absence of recent academic research. He indicates that about 11 percent of major American firms' new products and about 9 percent of their new processes could not have been developed in the absence of recent academic research. Zucker and Durby (2001) focus on technological transfer, and find that collaborations between particular university star scientists and firms have a large positive impact on firms' research productivity. They use the number of U.S. patents as indicators of firms' innovative output, while their explanatory variable is the number of star scientists' papers in the Japanese biotechnology field. Their results reveal that collaborations increase the average firm's biotech patents by 34 percent, products in development by 27 percent, and products on the market by 8 percent. The above studies suggest that universities' academic knowledge is useful for R&D in industry.

Moreover, according to Sanberg et al. (2014), research universities are expected to play a central role in the knowledge-centered economy by the technology transfer. They also point out that universities should expand their criteria to treat patents, licensing, and commercialization activity by faculty as an important consideration for merit, tenure, and career advancement, along with publishing, teaching, and service. However, measurement of science-technology transfer is one of the most difficult challenges (Tijssen et al. 2000). Although there are many previous studies about the relationship between academic research and industrial technology, the features characterizing this relationship have not been clarified completely.

Therefore, non-patent references (NPRs), that is, the documents other than the patent in the patent application, are used as indices to analyze the relationship between academic research and industrial technology, which is called *science linkage*. NPRs measure the strength of the relationship between science and technology, or *science intensity* (Meyer 2000; Tijssen et al. 2000). According to Tijssen et al. (2000), NPRs represent explicit connections between scientific research and technological innovations and as a consequence can reasonably describe the features of science-technology linkages. However, it is the important background knowledge playing an important indirect link rather than a direct link (Meyer 2000). Anderson et al. (1996) find strong linkages between human genetic technology and basic science research, and Narin et al. (1997)

show that 73% of the papers cited by patents were authored at public science institutions. They use the NPRs cited on the front page of U.S. patents. Moreover, MacMillian et al. (2000) indicate that the biotechnology industry depends on public science much more heavily than other industries.

These previous studies show that knowledge from the academia played a key role in patent development, which proxies for industrial technology. How about the academic standing of the papers cited by the patents? In other words, do papers that are heavily cited by industry also have a strong impact in academia? The pioneering studies that examine this question include Tijssen et al. (2000). They analyze the correlation between the number of citations to papers (paper-paper citations) and the number of citations to patents (paper-patent citations) by using 2,241 Dutch research papers that were cited by at least one USPTO patent during 1993–1996. They find a correlation of 0.16 (p=0.01). While this is low, papers that are highly cited by patents also tend to be cited by research papers. Hicks et al. (2000) examine 1993–1995 U.S. papers cited in 1997 U.S. patents. They categorize 6,595 papers in terms of their paper-paper citations, grouping them into the top 1%, top 2–10%, and the top 11–50%. They show that papers in the top 1 % are also most cited by patents. Although these are pioneering studies, they are restricted to correlation analysis. In addition, they target only papers that have at least one paper-patent citation, and examine a short time period. Moreover, all the above studies use the backward citation approach that extracts NPRs from patents.

Furthermore, the time lag in citations has been treated as a separate issue in the previous literature. Recent studies of the time lag in paper-paper citations have employed two approaches, the synchronous approach and the diachronous approach. The former uses the time lag between selected papers and the publication dates of the references that they cite (backward citation), while the latter uses the time lag between paper publication dates and the dates when these papers are cited (forward citation) (Bouabid and Lariviére 2013). The diachronous approach is the appropriate method to characterize citation processes, and measures of citation impact should be based on this approach (Glänzel 2004). According to Bouabid and Lariviére (2013), the total citation counts of papers published in Japan in 1995 and 2000 peaked 3 years after publication. Finardi (2014) finds that citations to papers in chemistry peaked 2 years after publication.

As for the time lag of paper-patent citations, Mansfield (1991) finds that the mean time lag between academic research findings and the commercial introduction of the product or process is about 7 years. Van Vianen et al. (1990) find that a 4-year lag was most common for chemistry patents. Verbeek et al. (2002) indicate that the time lag was 3 years on average, based on USPTO patents during 1992–1996. Lo (2010) finds an average citation age of 9.8 years in genetic engineering technology, while Finardi (2011) finds that a time lag of around 3–4 years is most common in the nanosciences and nanotechnology fields.

In this paper, we extract the number of paper-paper and paper-patent citations, focusing on the top 100 researchers[1] in the Japanese life and medical sciences fields. This is the first such attempt using the forward citation approach. We construct data on how many times papers published during 1996–2009 were cited by papers and patents during 1996–2012. Since we cover papers that are not cited by patents, we can analyze the relationship between journal papers and patents more accurately. In addition, the effect of science on technological development is clarified by using the number of papers written by each researcher, and the relevant grant amounts. Although previous studies have shown that papers with many citations tend to be more heavily cited by patents, the relationship between papers and patents was not analyzed in sufficient depth. Estimating the relationship between papers and patents, using both the number of papers and the amounts of grant funding, is useful to reveal the relationship between science and technology. Furthermore, our approach is distinctive in that we use the forward citation approach. None of the papers cited above have examined paper-paper citations and paper-patent citations at the same time, though they have analyzed them separately.

This paper is organized as follows. Section 2 describes the hypotheses, while Section 3 provides details regarding data. Section 4 contains descriptive statistics. Section 5 presents methods and estimation results regarding the citation distribution, while Section 6 presents methods and estimation results regarding paper-patent citations. Section 7 concludes.

2. Hypotheses

This section provides a description of our hypotheses. We explore the following three hypotheses.

H1: The lag of paper-paper citations is shorter than the lag of paper-patent citations.

We clarify the peak lag of paper-paper citations (peak in academia) and paper-patent

[1] Although Zucker and Darby (2001) define star scientists as authors who discover nucleotide sequences, we defined top scientists as authors with many total citation counts (see section 3).

citations (peak in industry), showing citation distributions that indicate the time necessary for published papers to impact papers versus patents. We then analyze the differences in the shapes of the two distributions. According to the literature (Bouabid and Lariviére 2013, Finardi 2014), citations to papers published in 1995 and 2000 peak after 3 years, and the lag was 2 years in the chemistry field. On the other hand, the peak lag for paper-patent citations is 3 to 10 years. Therefore, we expected that the lag for paper-patent citations to exceed that for paper-paper citations.

H2: The higher the quality of a paper, the more it is cited by patents.

We analyze the relationship between paper-paper citations and paper-patent citations. As previously shown by Tijssen et al. (2000) and Hicks et al. (2000), these two types of citations are correlated. Therefore, considering papers with many paper-paper citations to be of high academic quality, we verify whether high-quality and high-impact papers are more heavily cited by patents.

H3: There is a tradeoff between the quality and quantity of papers, and there exists an optimal amount of research grant funding that maximizes the quality of papers.

Rassenfosse (2013) shows the tradeoff between quality[2] (measured by family sizes) and quantity of patents in firms. Similarly, we verify whether there is a tradeoff between papers' quality and the number of papers written by their authors. We also verify whether some particular level of research funding maximizes paper quality.

Although we use patents to analyze the relationship between scientific papers and technology, patents do not cover all of the technological effects. We assume that "patents can be used as a proxy of the codification of technological knowledge" and "also assume that citations in patents are always pertaining to the content of the patent" (Finardi 2011).

3. Data

In this section, we describe our data. We extract paper-paper citations and paper-patent citations from journal articles. The extraction strategy is as follows. Our publication data are for researchers in the life sciences and medical sciences who had been selected for the

[2] Citation is the one of the most commonly used indices of patent quality. Rassenfosse (2013) uses family size because of a selection bias in his survey, as it includes many countries.

21st Century Centers of Excellence (COE) program established in 2002 by the Japan Society for the Promotion of Science. The aim of this program was to cultivate a competitive academic environment among Japanese universities by providing targeted support for the creation of world-class research and education bases.

First, we used the Scopus database to document each researcher's number of papers and citations. As Scopus assigns author IDs, we could perform full name searches and searches based on the researchers' institutions. Since this process could lead to accidental omissions, we contacted each researcher via e-mail and confirmed the list of their publications to the extent possible. We targeted 1,232 researchers who have at least one paper. Next, we extracted the 100 researchers with the most citations to papers published between 1996 and 2009. Therefore, among the excellent Japanese researchers, we concentrate on the ones with the most citations.

These top 100 researchers published a total of 20,661 papers as journal articles during the past 13 years. However, we only target 4,763 papers, for which these researchers were listed as corresponding authors. Corresponding authors have the responsibility of understanding and explaining the content of the paper. Wren et al. (2007), who conduct a survey about authorship credit for medical school, categorize credit as "initial conception", "work performed," and "supervision." They show that the last author as the corresponding author deserved most credit for supervision and initial conception. Therefore, this analysis essentially targets papers for which the chosen researchers were the principal investigators and supervisors.

The data on these 4,763 papers' paper-paper citations and paper-patent citations have been extracted using Scopus's "citedby" function (with a focus on citing papers and patents during a 16-year period (1996–2012). For our data on patents, we used the Espacenet Patent search offered by the European Patent Office. Espacenet offers free access to more than 80 million patent documents worldwide, covering the time period from 1,836 to the present. Data on paper-patent citations were extracted from all parts of the patent text (full text), not only from the references. We refer to the publication date of the patent as the citing year.

In addition, we use the KAKEN database for data on research grants (http://kaken.nii.ac.jp/). KAKEN is the database of Grants-in-Aid for Scientific Research. These grants are awarded through a peer review process, in order to promote creative and pioneering research across a wide spectrum of scientific fields. In this study, we use only research studies headed by our chosen researchers, and we include only direct expenses during 1996–2009.

4. Descriptive statistics

This section is about the descriptive statistics of our data.

4.1. Overview of the number of citations

Of the 4,763 papers that we target, 134 papers (2.8%) do not have any paper-paper citations, while 2,922 papers (61.4%) do not have any paper-patent citations. The threshold number of paper-paper citations for the 25^{th} percentile in the distribution of these citations is 10, while the corresponding number is 0 for paper-patent citations. The median number of paper-paper citations is 23, while that for paper-patent citations is 0. The number of paper-paper citations marking the 75^{th} percentile of the distribution is 55, while the corresponding number for paper-patent citations is 2. In addition, the numbers of paper-paper and paper-patent citations marking the 90^{th} percentile are 118 and 7, respectively. Therefore, the distribution of citation is biased to the left, and only a small portion of papers has a very strong impact.

Table 1 contains descriptive statistics. The average number of paper-paper citations is 55, while the average number of paper-patent citations is 4. The average total amount of research grant funding is $5.5 million (the median is $3.1 million), given $1=¥100. The average total number of papers authored by the selected researchers during the last 13 years is 236 (the median is 213). The average numbers of paper-paper and paper-patent citations by researcher are 2,594 and 168, respectively. The average number of papers for which the selected researchers were listed as corresponding authors is 48.

<Table 1 here>

As highlighted earlier, our analysis uses the forward citation approach based on citing papers and patents. This approach has never been used in previous science linkage studies. As long as all papers and all patents are used, the backward citation approach and the forward citation approach differ in extracting NPRs. Only 38.7% of the papers in this analysis are cited by patents. Therefore, if only these papers are considered while analyzing the relationship between paper-paper citations and paper-patent citations, the selection bias of extraction becomes important. If the relationship is analyzed based on our method, this selection bias problem can be avoided.

4.2. The top 10 most cited papers

We show the top 10 most cited papers (Table 2). Shinya Yamanaka's paper published in 2006 in Cell has the highest number of paper-paper citations (2,670). The same author's paper published in 2007 in Cell has the maximum number of paper-patent citations (414). Many of the journals are top journals such as Cell, Nature, and Science. Moreover, the same researcher appears several times in the ranking list of the leading researchers in this field. Most researchers list stem cell biology, immunology, and diabetes as their disciplines. All 10 of the papers with the highest paper-paper citations are also cited by patents. This suggests that papers of high academic quality and impact have an effect on patents.

<Table 2 here>

5. Distribution of citations

We describe the methods of analysis of citation distributions, and our estimation results.

5.1 Methods of analysis

First, we calculate the distribution of the time lag between papers published during 1996–2009 and the papers citing them that were published during 1996–2012. The lag is 0 when a paper is cited in the same year in which it is published. We use average citation counts, calculated as follows. Let a and b represent the year of publication and the year of citation respectively, with $a \le b$. The lag is $l = b - a$, and l takes values from 0 to 16. Then, let $CITATION_{a,b}$ be the number of citations of a paper that was published in year a and cited by a paper published in year b, and let $ARTICLES_a$ be the number of papers published in year a. In addition, let $MEAN_{b-a=l}$ be the average citation counts per paper published in year a and cited in year b with lag l. $MEAN_{b-a=l}$ can be written as follows:

$$MEAN_{b-a=l} = \frac{\sum_{b-a=l} CITATION_{a,b}}{\sum_{b-a=l} ARTICLES_a} \tag{1}$$

The graph of citation distribution is plotted as $MEAN_{b-a=l}$ versus l.

Next, we test for differences in the expected value of peak lag between paper-paper citations and paper-patent citations. If there is more than one peak, it is counted separately. Since the lag l is a discrete random variable, let $P(X = l)$ denote its probability distribution. The expected value of the lag at which citations peak is calculated as follows:

$$E(X) = \sum_l \{l \times P(X = l)\} \tag{2}$$

In addition, we test for the equality of the peak lag between paper-paper citations and paper-patent citations. We tested whether the citation distribution of peak lags is normally distributed, using the Kolmogorov-Smirnov test. This test rejects the null hypothesis of a normal distribution at the 1% level. Therefore, we test the difference between paper-paper citations and paper-patent citations at the median using the nonparametric Wilcoxon rank-sum test. Furthermore, we estimate the probability that the lag for paper-patent citations is longer than that for paper-paper citations.

5.2 Estimation results

First, we show the citation distribution profile that shows the lag from publication to citation (Figure 1). As for paper-paper citations, we find that the average citation count is high from two to five years after publication. The number of average citations peaks three years after publication, at 6.57 counts, and gradually declines thereafter. In particular, since there is an immediate sharp increase at the two-year mark, it is understood that papers impact academia at once. On the other hand, the average paper-patent citation lag is larger—five to ten years—with citations peaking five years after publication, at 0.47 counts. Patents cite papers for a longer time than other papers do. It is suggested that the lapse of time is necessary to put papers to practical use.

<Figure 1 here>

Turning now to the expected value of lags, we show the results on expected value and the Wilcoxon rank-sum test based on the 4,629 papers that have at least one paper-paper citation and the 1,841 papers with at least one paper-patent citation (Table 3). The expected peak lag for paper-paper citations and paper-patent citations is 4.31 and 6 years, respectively, while the medians are 4 and 5 years, respectively. The Wilcoxon rank-sum test rejects the null hypothesis of equality of the medians at the 1% level. Thus, the peak lag for paper-patent citations is longer than that for paper-paper citations. Furthermore, the probability that paper-patent citations have a longer lag than paper-paper citations is 0.664. This suggests that being cited by a patent requires more time than being cited by a paper.

<Table 3 here>

In addition, to verify the difference of lags within paper publication years, we analyze four delimited periods (96–98, 99–01, 02–04, and 05–07). We find that the median lag for paper-paper citations lag is consistent at four years excluding the period 05–07; however, the median lag for paper-patent citations lag has shortened in recent years (from seven years to six, then five, and then three years). Moreover, since the Wilcoxon rank-sum test rejects the null hypothesis of equality between the lags of paper-paper and paper-patent citations at the 1% level, excluding the period 05–07, the lag for paper-patent citations is longer than that for paper-paper citations. The probability that paper-patent citations have a longer lag is 0.714, 0.637, and 0.595, for 96–98, 99–01, and 02–04, respectively. Therefore, although the speed of impact on academia does not differ by year of publication, the time taken to produce an impact on technology has fallen in recent years. An approach to open innovation and industry-university cooperation are thought to underlie this result[3].

6. Paper-patent citations

In this section, we describe the method of the Tobit model with instrumental variables, and present the estimation results.

[3] Since recently published papers have only had the opportunity to be cited for a few years, it is necessary to note that citations to these papers may not have peaked yet. However, because the median lag for paper-patent citations is five years, there is a high possibility that papers published during 05–07 have experienced their peak.

6.1 Methods of analysis

To analyze the impact of the paper-paper citations on the paper-patent citations, we use the paper-patent citations as the dependent variable and the paper-paper citations as explanatory variables in our estimation. The Tobit model is used because paper-patent citations have a non-negative value. Since there is a possibility of endogeneity in that paper-paper citations might be correlated with the error term, we use instrumental variables that are correlated with paper-paper citations, but not with the error term. Therefore, we used the total amount of research grant and the total number of papers of each researcher as instrumental variables. As for the relation between instrumental variables and paper-paper citations, there is a possibility that researchers who secure a lot of grant funding may have high research ability, and can therefore publish high quality papers. In addition, because of focusing only on volume, the papers might be low quality. Then, we use these instrumental variables to analyze the optimal amount of grant funding for high quality publications, while also looking at the relationship between the number of papers and research quality. The estimation is performed using Newey's (1987) two-step efficient estimator. The estimation model is below.

Let i and p represent researcher and paper, respectively.

$$\begin{aligned} y_{ip} &= \beta x_{ip} + u_{ip} \\ x_{ip} &= \pi z_i + v_{ip} \end{aligned} \tag{3}$$

$$y_{ip} = \begin{cases} y_{ip}^{*} & \text{if} \quad y_{ip}^{*} > 0 \\ 0 & \text{if} \quad y_{ip}^{*} \leq 0 \end{cases}$$

In equation (3), y_{ip} denotes paper-patent citations for paper *p* by researcher *i* (y_{ip}^{*} is the corresponding latent variable), while x_{ip} denotes paper-paper citations for paper *p* by researcher *i*. z_i denotes instrumental variables, and we use the total amount of grant funding and the total number of papers authored by the selected researchers. In view of the non-linear relationship between instrumental variables and paper-paper citations, we use a model with a quadratic term in instrumental variables (Model 1). Moreover, considering the non-linear relationship between paper-paper citations and paper-patent citations, we use a quadratic term in paper-paper citations as an endogenous variable (Model 2).

6.2 Estimation results

Our estimation results are shown in Table 4. The null hypothesis of exogeneity of the instrumented variable is rejected at a statistically significant level (p=0.005), while the overidentification test that checks the correlation between the error term and instrumental variables does not reject the null hypothesis of no correlation (p=0.980). Therefore, the adequacy of the estimation model is verified.

<Table 4 here>

Let us now show the two-step estimation results. Model 1 uses only a linear term in paper-paper citations as an explanatory variable. The estimated coefficient is 0.154, which is statistically significant at the 1% level. Model 2 uses both a linear term and a quadratic term in paper-paper citations as explanatory variables. Although the linear term has an estimated coefficient of 0.168, which is statistically significant at the 1% level, the estimated coefficient for the quadratic term is -0.0000162, and is not statistically significant.

We thus focus on Model 1 below. The point of the results is that paper-paper citations have a positive and significant relationship with paper-patent citations, and it is found that papers of high academic quality are more heavily cited by patents. The marginal effect evaluated at the median is 0.036[4]. Therefore, the number of paper-patent citations increases by 0.036 when the number of paper-paper citations increases by one.

Turning now to the relationship between instrumental variables and paper-paper citations, since the coefficient on the linear term in the total amount of grants for each researcher is significant and positive, while that on the corresponding quadratic term is significant and negative, there is an inverse U-shaped relationship. On the other hand, since the coefficient on the linear term in the total number of papers is significant and negative, while that on the corresponding quadratic term is significant and positive, we have a U-shaped relationship. We calculate the amount of research grant that maximizes paper quality, and the number of papers that minimize paper quality, using the following function based on the results of the first-step estimation:

---

[4] Since only the linear prediction of marginal effects is available for two-step estimation, the marginal effect is estimated through the maximum likelihood estimator. Since the coefficient on the quadratic term of paper-paper citations is not statistically significant, we calculate the marginal effect using Model 1.

$$x = 111.9714 + 0.000392 z_1 - 0.000000000794 z_1^2 - 0.5647761 z_2 + 0.0008635 z_2^2 \quad (4)$$

In equation (4), x denotes paper-paper citations, z_1 the total amount of grant funding, and z_2 the total number of papers.

We find that the quality of papers is highest when the research grant is $1.9 million per year ($24.7 million for 13 years). The quality declines gradually above this level of funding. On the other hand, when the number of papers is less than 25.2 papers per year (327.0 papers for 13 years), an increase in the number of papers lowers quality, while quality becomes higher when the number of papers exceeds this threshold. The saddle point that is calculated using the amount of research grant and the number of papers is (z_1, z_2) = (246,851.4, 327.0), and the citation count is 68.0 at this point.

Thus, we find that there exists an efficient amount of research grant funding that maximizes paper quality. There is a tradeoff between paper quality and quantity up to 25 papers per year. As for grants, it is found that it is difficult to manage grants that exceed about $2 million per year. Our results suggest that the papers that are tapped to the maximum in industry also have a very strong impact on academia.

7. Conclusion

In this paper, we investigate the shapes of the distributions of paper-paper citations and paper-patent citations, using 4,763 papers published between 1996 and 2009, for which the corresponding authors were among the top 100 Japanese researchers in the life and medical sciences fields. Furthermore, we analyze the relationship between paper-paper citations and paper-patent citations, incorporating the amount of research grant and the number of papers of each researcher into the framework. As a result, we confirm the following hypotheses.

H1: The expected value of the paper-paper citations lag (four years) is shorter than the corresponding value for paper-patent citations (six years), and the lag of paper-patent citations has fallen in recent years.

H2: Papers of high academic quality are more heavily cited by patents, and the number of paper-patent citations increases by 0.036 when the number of paper-paper citations increases by one.

H3: Since we find a U-shaped relationship between the quality and quantity of papers, there is a tradeoff up to a certain number of papers. Moreover, since there is an inverse U-shaped relationship between the quality of papers and the amount of grant funding, an efficient amount of grant funding maximizes research quality.

However, the above-mentioned conclusions regarding the amount of grant funding and the number of papers are obtained from analyzing papers for which the top Japanese researchers in the medical and life sciences fields are the principal investigators or supervisors. Therefore, it is necessary to note that these results do not necessarily apply to other fields, or to all researchers in Japan. In particular, the efficient grant size may vary across fields and laboratory environments, and therefore the level that we find does not necessarily maximize research quality for all researchers in all fields of the life sciences or the medical sciences.

References

1. Anderson J, Williams N, Seemungal D, Narin F, Olivastro D (1996) Human genetic technology: Exploring the links between science and innovation. *Technology Analysis and Strategic Management* 8(2): 135-156.
2. Bouabid H. and Lariviére V. (2013) The lengthening of paper's life expectancy: A diachronous analysis. *Scientometrics* 97(3): 695-717.
3. Finardi U (2011) Time relations between scientific production and patenting knowledge: The case of nanotechnologies. *Scientometrics* 89(1): 37-50.
4. Finardi U (2014) On the Time evolution of received citations, in different scientific fields: An empirical study. *Journal of Informetrics* 8(1): 13-24.
5. Glänzel W (2004) Towards a model for diachronous and synchronous citation analysis. *Scientometrics* 60(3): 511-522.
6. Hicks D, Breitzman A, Kimberly S, Narin F (2000) Research excellence and patented innovation. *Science and Public Policy* 27(5): 310-320.
7. Lo Azu-chia (2010) Scientific linkage of science research and technology development: a case of genetic engineering research. *Scientometrics* 82(1): 109-120.
8. Mansfiels E (1991) Academic research and industrial innovation. *Research Policy* 20(1): 1-12.
9. McMillian G, Narin F, Deeds D (2000) An analysis of the critical role of public science in innovation: the case of biotechnology. *Research Policy* 29(1): 1-8.
10. Meyer M (2000) Does science push technology? Patents citing scientific literature. *Research Policy* 29(3): 409-434.
11. Narin F, Hamilton K, Olivastro D (1997) The increasing linkage between U.S. technology and public science. *Research Policy* 26(3): 317-330.
12. Newey W (1987) Efficient estimation of limited dependent variable models with enfogenous explanatory variables. *Journal of Econometrics* 36(3): 231-250.
13. Rassenfosse G (2013) Do firms face a trade-off between the quantity and the quality of their inventions? *Research Policy* 42(5): 1072-1079.
14. Sanberg P, Gharib M, Harker P, Kaler E, Marchase R, Sands T, Arshadi N, Sarkar S (2014) Changing the academic culture: Valuing patents and commercialization toward tenure and career advancement. *Proceedings of the National Academy of Sciences of the United States of America* 111(18): 6542-6547.
15. Tijssen R, Buter R, Van Leeuwen Th (2000) Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics*

47(2): 389-412.

16. Van Vianen B, Moed H Van Raan A (1990) An exploration of the science base of recent technology. *Research Policy* 19(1): 61-81.
17. Verbeek A, Debackere K, Marc L, Andries P, Zimmermann E, Deleus F (2002) Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* 54(2): 399-420.
18. Wren J, Kozak K, Johnson K, Deakyne S, Schilling L, Dellavalle R (2007) A survey of perceived contributions to papers based on byline position and number of authors. *EMBO reports* 8(11): 988-991.
19. Zucker L, Durby M (2001) Capturing technological opportunity via Japan's star scientists: Evidence from Japanese firm's biotech patents and products. *Journal of Technology Transfer* 26(1-2): 37-58.

Figure 1: Distribution of citations

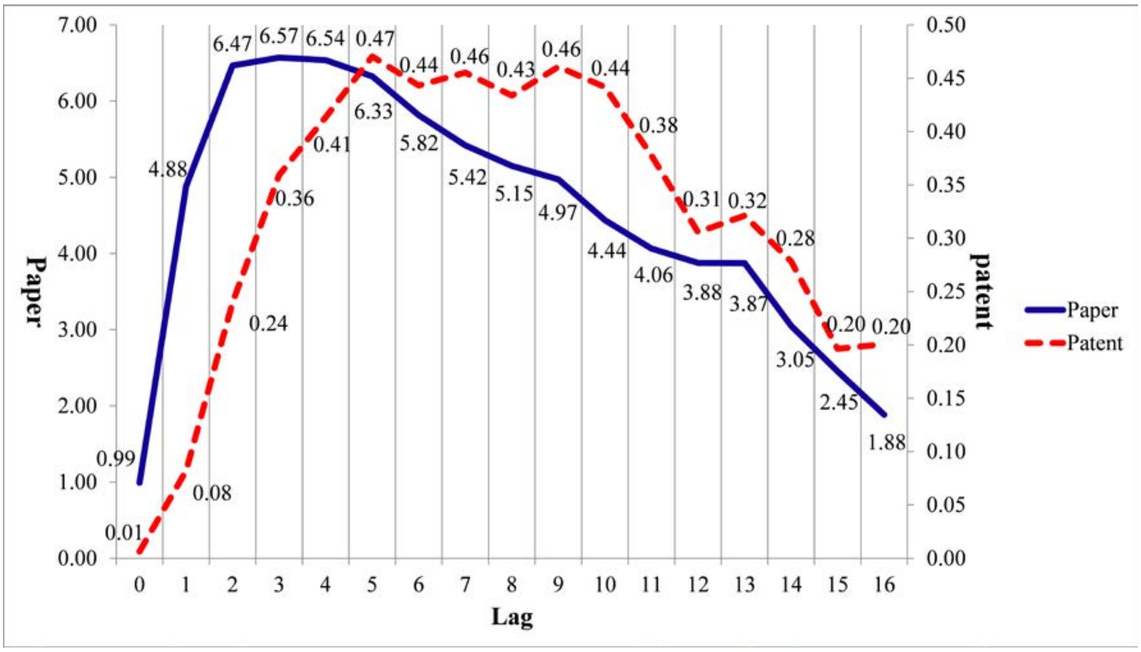

Table 1: Descriptive statistics

	Number of observations	Mean	Standard deviation	Median	Min	Max
Number of paper-paper citations by paper	4,763	55	121	23	0	2,670
Number of paper-patent citations by paper	4,763	4	16	0	0	414
Total amount of research grant of each researcher ($100)	4,763	55,052	74,646	31,440	0	512,270
Total number of papers authored by each researcher	4,763	236	123	213	31	733
Number of paper-paper citations by researcher	100	2,594	2,896	2,064	4	20,536
Number of paper-patent citations by researcher	100	168	279	76	0	1,594
Number of corresponding author papers by researcher	100	48	41	37	1	210

Table 2: The top 10 papers ranked by paper-paper and paper-patent citations

Rank	Name, Journal, Number of paper-paper citations (paper-patent citations), Major	Name, Journal, Number of paper-patent citations (paper-paper citations), Major
1	Shinya Yamanaka, Cell (2006), 2670 (377), Stem cell biology	Shinya Yamanaka, Cell (2007), 414 (2217), Stem cell biology
2	Shimon Sakaguchi, Science (2003), 2371 (86), Immunology	Shinya Yamanaka, Cell (2006), 377 (2670), Stem cell biology
3	Shizuo Akira, Nature (2000), 2218 (322), Immunology	Takashi Kadowaki, Diabetes (1998), 350 (130), Diabetes
4	Shinya Yamanaka, Cell (2007), 2217 (414), Stem cell biology	Keiichi Hiramatsu, Lancet (2001), 347 (722), Microbiology
5	Shigekazu Nagata, Nature (1998), 1618 (115), Integrated Biology	Shizuo Akira, Nature (2000), 322 (2218), Immunology
6	Shizuo Akira, Journal of Immunology (1999), 1511 (104), Immunology	Hajime Nawada, Nature (1996), 193 (966), Immunology
7	Shizuo Akira, Immunity (1999), 1488 (93), Immunology	Shinya Yamanaka, Nature (2007), 169 (916), Stem cell biology
8	Shimizu Nobuyoshi, Nature (1998), 1353 (100), Molecular biology	Shizuo Akira, Nature Immunology (2002), 169 (690), Immunology
9	Hidenori Ichijo, Science (1997), 1012 (61), Cell signaling	Hajime Nawada, Nature (1998), 144 (611), Immunology
10	Yoshihide Tsujimoto, Nature (1999), 998 (54), Molecular biology	Tasuku Honjo, Journal of Experimental Medicine (2000), 142 (568), Diabetes

Table 3: Expected value of lag and the results of the Wilcoxon rank-sum test

Papers cited at least once	Paper (all)	Patent (all)	Paper (96-98)	Patent (96-98)	Paper (99-01)	Patent (99-01)	Paper (02-04)	Patent (02-04)	Paper (05-07)	Patent (05-07)
Number of papers	4629	1841	1002	502	1131	580	1071	451	951	247
Expected value of lag	4.31	6	4.98	7.55	4.90	6.26	4.40	5.22	3.56	3.62
Standard deviation	2.92	3.22	3.65	3.59	3.25	2.98	2.51	2.48	1.71	1.71
Median	4	5	4	7	4	6	4	5	3.5	3
Wilcoxon rank-sum test (p value)	0.000		0.000		0.000		0.000		0.656	
Probability that the lag for patents is longer than the lag for papers	0.664		0.714		0.637		0.595		0.508	

Table 4: Estimation results

Dependent variable: Number of paper-patent citations	Model 1	Model 2
Number of paper-paper citations	0.154*** (0.0141)	0.168*** (0.0601)
Quadratic term of the above		-0.0000162 (0.0000717)
Intercept	-17.89*** (0.897)	-18.37*** (2.061)
Instrumental variables for the linear term		
Total amount of research grants for each researcher	0.000392***(0.0000609)	0.000392*** (0.0000609)
Quadratic term of the above	-7.94e-10***(1.42e-10)	-7.94e-10*** (1.42e-10)
Total number of papers of each researcher	-0.565***(0.0475)	-0.565*** (0.0475)
Quadratic term of the above	0.000864***(0.0000729)	0.000864*** (0.0000729)
Intercept	112.0***(7.518)	112.0*** (7.518)
Instrumental variables for the quadratic term		
Total amount of research grants for each researcher		0.217** (0.0959)
Quadratic term of the above		-5.68e-07** (0.000000223)
Total number of papers of each researcher		-509.8*** (74.88)
Quadratic term of the above		0.814*** (0.115)
Intercept		73217.9*** (11844.7)
Overidentification test. H0: E (zu) =0.	0.980	0.975
Wald test of exogeneity of instrumented variable. H0: No endogeneity.	0.005	0.020

Standard errors in parentheses, * p<0.1, ** p<0.05, *** p<0.01