Available online at www.sciencedirect.com
ScienceDirect
Journal of Informetrics 1 (2007) 269–276
Journal of INFORMETRICS
An International Journal
www.elsevier.com/locate/joi

Measuring industrial knowledge stocks with patents and papers

Yoo-Jin Han *
Korea Institute of Intellectual Property, 3th FL, 647-9, Yeoksam-Dong, Kangnam-Gu, Seoul 135-980, Republic of Korea

Received 31 January 2007; received in revised form 30 March 2007; accepted 11 June 2007

Abstract

Under the National Innovation System (NIS) framework, knowledge stock has been recognized as a key factor for enhancing national innovative capabilities. However, despite the importance of patents and papers for measuring knowledge, previous research has not fully utilized patent and paper databases, and has instead relied on research and development (R&D) data. Therefore, in this research, I introduce a way to utilize both types of useful data when measuring industrial knowledge stocks. As primary data sources, the United States Patent and Trademark Office (USPTO) Web site for patents and the science citation index (SCI) for papers are used. In the case of Korea, the amount of knowledge stock proxied by patents and papers is different from that proxied by R&D, which indicates in turn that using a single indicator such as R&D may be misleading. Although the result may vary depending on the selected nation, the proposed method will be useful for gauging knowledge stocks in a more complementary way.

Keywords: Industrial knowledge stock; Patents; Papers; The USPTO Web site; The SCI database

1. Introduction

Under the framework of the National Innovation System (NIS), the concept of knowledge stock has been emphasized because in order to create more sophisticated knowledge, accumulated knowledge is considered more important than the one-time input of a large volume of knowledge (Metcalfe & Ramlogan, 2005; OECD, 1996). Therefore, many researchers have strived to measure knowledge stock by using R&D data (Griliches, 1979; Mansfield, 1980; Scherer, 1982). However, R&D data have its limitations in that it only represents the input-side of knowledge, thereby making us overvalue the stocks of knowledge (Kleinknecht, Van Montfort, & Brouwer, 2002; Lach, 1995; Park & Park, 2006). Thus, in order to cope with the shortcomings of R&D data, patents and papers, which indicate the output-side of knowledge, have been adopted recently in shaping science and technology (S&T) policies (Archibugi & Pianta, 1996; Moon & Lee, 2005; Okubo, 1997). However, their usage has been limited to a certain technology or scientific discipline, not incorporating the issues relating to industrial policies within a nation.

An industry, as a unit of analysis, is important since it shares a common knowledge ground in terms of technological development and production, and acts as a critical element in the NIS. Therefore, Pavitt (1984), Breschi and Malerba (1997), and Marlerba (2002) focused on elaborating the peculiar patterns of industries and emphasized an industry as the primary unit for analyzing important innovations.

However, despite the aforementioned potential usefulness of patents and papers for measuring knowledge stocks at the industry level, there has not been any sufficient effort to utilize patents and papers so far, for the purpose mentioned.

* Tel.: +82 2 2189 2634; fax: +82 2 2189 2699.
E-mail address: yjhan@kiip.re.kr.

1751-1577/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2007.06.001

Therefore, in this research, I introduce a way to use patents and papers for measuring industrial knowledge stocks. In addition, I will present the case of Korea for the period of 1994–2000 when the knowledge stocks significantly increased, as an illustration.

2. Two knowledge measures

2.1. Usefulness of patents

Due to the importance of knowledge, international economic organizations such as the OECD have made great efforts to suggest measures for knowledge through large-scale workshops.[1] Among others, patents nowadays have drawn great attention and have become the focus of discussion due to their representativeness and comprehensiveness as a knowledge indicator. Earlier, Archibugi and Pianta (1996) elaborated the advantages and disadvantages of patents when they are used in policy making. The main advantages of patents are as follows: (i) they are the direct outcome of inventions with the aim of being used commercially; (ii) they not only harbor information on the rate of inventive activities but also that of their development directions; (iii) they contain large volumes of information related to inventors, descriptions, patent classes, claims, and diagrams of technological outputs; (iv) they are easily accessible. On the other hand, there exist disadvantages such as the following: (i) not all inventions are patented; (ii) patent propensity is different across firms and sectors; (iii) the classification scheme of patents does not correspond to that of economic fields; (iv) each country has its own different patent system.

2.2. Usefulness of papers

The importance of papers is more directly emphasized because compared with patents, they seem to pursue the accumulation of fundamental knowledge which, in turn, may not be directly applied to industrial use. However, Rosenberg and Nelson (1994) and Nelson and Rosenberg (1998) explicitly pointed out the importance of scientific knowledge for industrial development. Therefore, in order to examine the knowledge stocks within the NIS, we should incorporate scientific knowledge. Like patents, scientific papers have their merits and demerits. The merits include the following: (i) scientific papers are the only media that show scientific achievements (Wouters, 1998); (ii) citation patterns can be a useful tool for examining the knowledge exchange among scientists as well as the interdependencies of disciplines (Small & Garfield, 1985). In contrast, the demerits are summarized as follows: (i) the propensity to publish is different depending on the field (Hicks & Katz, 1996; Meyer, 2002); (ii) papers not written in English are often disregarded (Meyer, 2002).

3. Measuring knowledge stocks

The stock in "knowledge stock" is a concept which is equivalent to that in "capital stock" in economics. That is, knowledge stock at a certain point is composed of four distinctive parts—the knowledge input at a certain point, the knowledge stock from the previous period, the depreciation rate, and the time-lag caused by the gestation period. Therefore, we can obtain the knowledge stock of industry i at time t as follows:

$$\mathrm{KS}_{it} = \mathrm{KI}_{i(t-m)} + (1 - \delta_{it})\mathrm{KS}_{i(t-1)} \tag{1}$$

where KS is the knowledge stock, KI the knowledge input, δ the depreciation rate, m the time lag, i the industry, and t is the time.

Then since the knowledge stock of time point $t-1$ can be expressed as the sum of the knowledge inputs from the base point to $t-1$, we can get the following equation:

$$\begin{aligned}\mathrm{KS}_{it} = {} & \mathrm{KI}_{i(t-m)} + (1 - \delta_{it})\mathrm{KI}_{i(t-m-1)} + (1 - \delta_{it})(1 - \delta_{i(t-1)})\mathrm{KI}_{i(t-m-2)} \\ & + \cdots + (1 - \delta_{it})\cdots(1 - \delta_{i(t-b)})\mathrm{KI}_{i(b-m)}\end{aligned} \tag{2}$$

[1] 2006 Workshop on Patent Statistics for Policy Decision Making (Vienna on 23–24 October 2006); conference on research use of patented inventions (Madrid on 18–19 May 2006); and so on.

where KS is the knowledge stock, KI the knowledge input, δ the depreciation rate, m the time lag, i the industry, t the time, and b is the base point.

If we assume that the depreciation rates of each point are the same, then we can transform Eq. (2) into (3) as follows:

$$\mathrm{KS}_{it} = \mathrm{KI}_{i(t-m)} + (1-\delta_i)\mathrm{KI}_{i(t-m-1)} + (1-\delta_i)^2\mathrm{KI}_{i(t-m-2)} + \cdots + (1-\delta_i)^b\mathrm{KI}_{i(b-m)} \tag{3}$$

where KS is the knowledge stock, KI the knowledge input, δ the depreciation rate, m the time lag, i the industry, t the time, and b is the base point.

Subsequently, if we consider that the knowledge stock at the base point can be obtained from the cumulative forms of the knowledge inputs from the zero time point, and the knowledge inputs are increased by the constant growth rate across all periods, then we can get the knowledge input at the base point as follows:

$$\mathrm{KI}_{i(b-m)} = \mathrm{KI}_{i0}(1+g_i)^{(b-m-1)} \tag{4}$$

where KI is the knowledge input, g the average growth rate, m the time lag, i the industry, and b is the base point.

Therefore, as noted in Eq. (2), the knowledge stock at the base point can also be earned as follows:

$$\mathrm{KS}_{ib} = \mathrm{KI}_{i(b-m)} + (1-\delta_i)\mathrm{KI}_{i(b-m-1)} + \cdots + (1-\delta_i)^{(b-m-1)}\mathrm{KI}_{i0} \tag{5}$$

where KS is the knowledge stock, KI the knowledge input, δ the depreciation rate, m the time lag, i the industry, t the time, and b is the base point.

As noted in Eq. (4), since the knowledge inputs across all time points can be expressed by using the knowledge input at the zero time point and the growth rate, Eq. (5) can be transformed as follows:

$$\mathrm{KS}_{ib} = \mathrm{KI}_{i0}(1+g_i)^{(b-m-1)} + (1-\delta_i)\mathrm{KI}_{i0}(1+g_i)^{(b-m-2)} + \cdots + (1-\delta_i)^{(b-m-1)}\mathrm{KI}_{i0} \tag{6}$$

where KS is the knowledge stock, KI the knowledge input, δ the depreciation rate, m the time lag, i the industry, t the time, and b is the base point.

Finally, if we organize the right side of Eq. (6), then we can get the knowledge stock at the base point as follows:

$$\mathrm{KS}_{ib} = \frac{1+g_i}{g_i+\delta_i}\mathrm{KI}_{i(b-m)} \tag{7}$$

where KS is the knowledge stock, KI the knowledge input, δ the depreciation rate, m the time lag, i the industry, t the time, and b is the base point.

Using Eqs. (1) and (7), we can calculate the amount of knowledge stock in each industry. Since I proposed the use of patents and papers for the proxy measures for knowledge, we can measure knowledge stocks at the industrial level by replacing KI with patent counts and paper counts.

4. The Korean case

4.1. Data collection and manipulation

4.1.1. Patents

As mentioned above, since every country has its own different patent system, we need to select one country's patent data in order to avoid home country bias (Criscuolo, 2006). In this study, I selected US patents since their values are the highest (Criscuolo, 2006; Watabane, Tsuji, & Griefy-Brown, 2001). Therefore, I collected data from the USPTO Web site. The fields needed were the Assignee Country, Issue Date, Patent Number, and Current US Classification. Since Korea is an exemplary case here, I only gathered the patents issued by Korean assignees for the period of 1994–2000. In particular, I searched for the patents corresponding to the reference country and period by using the query statement ACN/KR AND ISD/19940101 → 20001231. Consequently, 15,541 patents were retrieved as shown in Table 1. Then by referring to Hall, Jaffe, and Trajtenberg's (2001) sub-categories,[2] I re-matched the classes into six important industries for the Korean economy, namely, the chemical, machinery, and transportation industries – traditional sectors – and

[2] Hall et al. (2001) matched technological classification to industrial categorization.

Table 1
The fields extracted in the case of patents

Country	Year	Patent number	Current U.S. classification	Hall et al.'s (2001) sub-category
KR	1994	5275008	62	69
KR	1995	5377781	184	59
KR	1996	5480824	438	46
KR	1997	5590483	37	64
KR	1998	5704136	34	19
KR	1999	5855077	34	19
KR	2000	6009599	16	59

Note: The total number of patents issued by Korean assignees from 1994 till 2000 was 15,541—923 for 1994; 1159 for 1995; 1471 for 1996; 1896 for 1997; 3249 for 1998; 3559 for 1999; 3284 for 2000.

Table 2
The fields extracted in the case of papers

Country	Year	Article number	Source title	ISI subject category
Korea	1994	1	Journal of the Agricultural Association of China	Agriculture, Dairy and Animal Science
Korea	1995	1	Library and Information Science	Information Science and Library Science
Korea	1996	1	Algorithms and Computation	Computer Science, Theory and Methods
Korea	1997	1	Combinatorial Pattern Matching, Proceedings	Computer Science, Theory and Methods
Korea	1998	1	Algorithms and Computations	Computer Science, Theory and Methods
Korea	1999	1	Science of Engineering Ceramics II	Materials Science, Ceramics
Korea	2000	1	Cerebellar Modules: Molecules, Morphology, and Function	Neurosciences

Note: The total number of papers published by Korean authors from 1994 till 2000 was 69,396—4328 for 1994; 6453 for 1995; 7968 for 1996; 10,054 for 1997; 11,730 for 1998; 13,744 for 1999; 15,119 for 2000.

the IT information technology (IT), biotechnology (BT), and nanotechnology (NT) industries—emerging sectors. The matching results are shown in Appendix 1.

4.1.2. Papers

As for the proxy for scientific knowledge, I used the SCI database provided by the ISI. It has limitations in that most of the achievements were written in English and may therefore not accommodate those shown in other languages. However, since the SCI holds the most renowned journals that scientists wish to publish their works in, it can be used as a source to represent scientific knowledge. In addition, the SCI recently provided bibliographic data through their Web site, which is called the Web of science (WoS), thereby allowing us to have access to the most recent information. The fields needed were the address, publication year, article number, and source title. As mentioned above, I only gathered the papers published by Korean authors for the period of 1994–2000. That is, I searched for the papers corresponding to the reference country and period by using search words like "Korea" in the address field and "1994, 1995, 1996, 1997, 1998, 1999, and 2000" in the publication year field. Consequently, 69,396 papers were retrieved as shown in Table 2. Afterward, since the SCI covers over 6000 journals, we had to re-categorize them into the industries of our interests. Therefore, by referring to the subject categories[3] by which the ISI classified the numerous journals depending on their academic disciplines, I converted each paper into the six industries listed above. The matching results are shown in Appendix 2.

4.2. Measuring

First, as mentioned above, knowledge stock is calculated by substituting KI of Eq. (1) with patents and papers. Then we delete the time-lag term as shown by Lach (1995). With regard to the depreciation rate, Lach (1995) used 0.15,

[3] The ISI provides the scope notes for each subject category on the Web (http://scientific.thomson.com/mjl/scope/scope_sci.html). By referring to the explanations on the Web page, we matched the scientific fields to the corresponding industries.

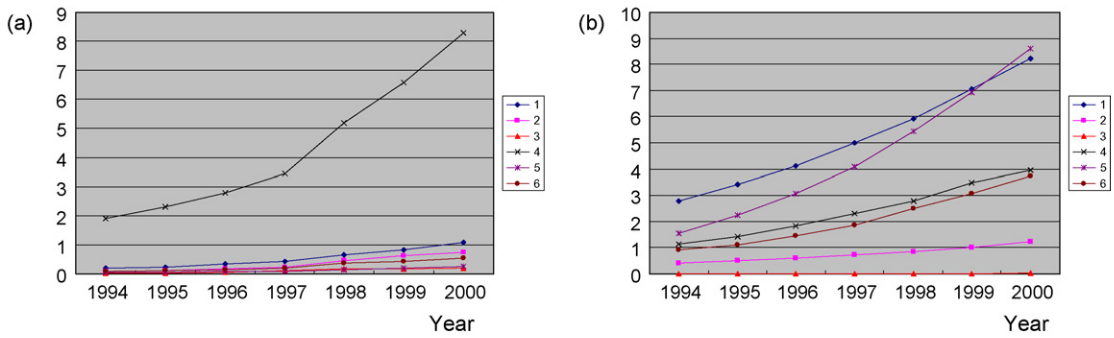

Fig. 1. Industrial knowledge stocks: (a) patent-based (thousand); (b) paper-based (thousand).

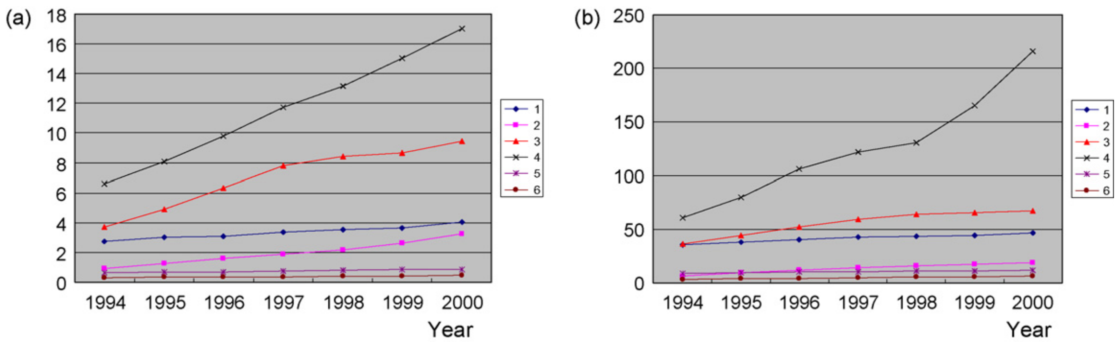

Fig. 2. Industrial knowledge stocks: (a) R&D expenditure-based; (b) R&D personnel-based.

while Park and Park (2006) employed different ratios across sectors. Since the depreciation rate varies depending on the measurement method, we use 0.13 which is the depreciation rate used for most industries in the study of Park and Park (2006).

As a result, Fig. 1 shows industrial knowledge stock measured with patents and papers, respectively. It is very clear that when using patents, the knowledge stock in the IT industry is strikingly immense, while those in other industries are not very visible. On the other hand, when using papers, the distribution of the knowledge stocks is more dispersed across industries. In addition, the knowledge stocks in the chemical and BT industries are more than the knowledge stock in the IT industry. Moreover, since the industrial structure has become more high technology-oriented since 1999, the BT industry has taken the lead in the chemical industry.

Fig. 2 shows industrial knowledge stock measured with R&D, the left of which is measured with R&D expenditure, and the right of which is gauged based on R&D personnel.[4] The patterns of the two graphs are analogous, except that we can see that there is more knowledge stock for the Transportation industry when using R&D expenditure as a proxy. However, the trend markedly varies as compared to the cases of patents and papers. Therefore, these results indicate that providing an explanation of knowledge stocks by simply considering R&D data may be misleading.

5. Discussion

Above, I compared the knowledge stocks proxied by simple patent and paper counts with those measured with R&D expenditure and personnel. However, although simple patent and paper counts represent the knowledge stocks

[4] R&D expenditure and R&D personnel are the most widely used R&D data when measuring knowledge (OECD, 1996).

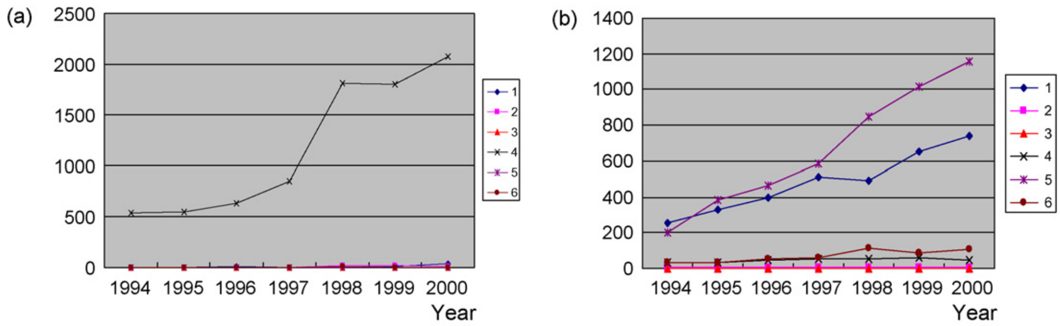

Fig. 3. Industrial knowledge stocks: (a) weighted patent counts-based; (b) weighted paper counts-based.

within a country in a broad sense, citation-weighted patent and paper counts have been recently known to reflect knowledge stocks better (Abdih & Joutz, 2006). Therefore, after collecting citation data from the USPTO Web site and SCI database, I re-calculated the knowledge stocks as follows:

$$\mathrm{CKS}_{it} = \frac{\mathrm{NC}_{it}}{\sum_{i=1}^{6}\mathrm{NC}_{it}}\mathrm{KS}_{it} \tag{8}$$

where CKS is the citation-weighted knowledge stock, NC the number of citations, KS the knowledge stock, i the industry, and t the time.

Fig. 3 shows the citation-weighted knowledge stocks at an industrial level for the same reference period. As a result, it was found that Korea had accumulated the most knowledge stock in the IT industry when knowledge stock was measured with patents, and in the BT and chemical industries when such was gauged with papers. Also, the amount of knowledge stock in other industries was evaluated less as compared to the case in which simple patent and paper counts were used.

6. Concluding remarks

In this research, I proposed the use of patents and papers for measuring industrial-level knowledge stocks in order to overcome the shortcomings of R&D data. Moreover, this research is meaningful in that it introduced a way to collect a large set of patent and paper data. First, in terms of patent data collection, while previous research made use of secondary database such as the National Bureau of Economic Research (NBER),[5] I directly collected necessary data from the USPTO Web site in order to enhance timeliness. Second, with respect to paper data collection, previous research utilized the SCI database for a limited technology or discipline, whereas I comprehensively gathered data corresponding to one country from the WoS. Finally, despite the importance of scientific achievements in industrial development, there has not been any sufficient effort to link science to industry. However, in this research, I included the ISI subject categories as the criteria through which we can group scientific disciplines into industries.

In the case of Korea, it has been found that the amounts of knowledge stock at the industrial level differ depending on the proxy measures for knowledge. Therefore, this indicates that we should not consider only one indicator such as R&D. Furthermore, it should be noted that the amounts of knowledge stock across industries may vary from nation to nation due to idiosyncratic national S&T policies and industry specialization. However, the proposed method will enable S&T policy makers to measure industrial knowledge stocks in a more complementary way.

[5] www.nber.org offers a collective set of US patent data from 1963 to 1999.

Appendix A

Sector	Industry	Hall et al.'s (2001) sub-category
Traditional	1. Chemical	12: coating; 13: gas; 14: organic compounds; 15: resins; 19: miscellaneous-chemical
	2. Machinery	53: motors, engines and parts; 59: miscellaneous-mechanical
	3. Transportation	55: transportation
Emerging	4. IT	21: communications; 22: computer hardware and software; 23: computer peripherals; 24: information storage; 41: electrical devices; 42: electrical lighting; 45: power systems; 46: semiconductor devices; 49: miscellaneous-Elec.; 116, 123, 181, 279 from 69: miscellaneous-others
	5. BT	31: drugs; 33: biotechnology; 39: miscellaneous-drug and med
	6. NT	32: surgery and medical instruments; 43: measuring and testing, optics except 399

Appendix B

Category	Sub-category	ISI subject category
Traditional sectors	1. Chemical	Chemistry, analytical; chemistry, applied; chemistry, inorganic and nuclear; chemistry, multidisciplinary; chemistry, physical; electrochemistry; engineering, chemistry; engineering, petroleum; materials science, coatings and films; materials science, composites; polymers science
	2. Machinery	Engineering, manufacturing; engineering, mechanical; mechanics
	3. Transportation	Transportation science and technology
Emerging sectors	4. IT	Automation and control systems; computer science, artificial intelligence; computer science, cybernetics; computer science, hardware and architecture; computer science, information systems; computer science, interdisciplinary applications; computer science, software engineering; computer science, theory and methods; engineering, electrical and electronic; imaging science and photographic technology; telecommunications
	5. BT	Biochemical research methods; biochemistry and molecular biology; biology; biophysics; biotechnology and applied microbiology; chemistry, medicinal; chemistry, organic; cell biology; critical care medicine; developmental biology; emergency medicine; engineering, biomedical; evolutionary biology; genetics and heredity; integrative and complementary medicine; marine and freshwater biology; materials science, biomaterials; medical informatics; medicine, general and internal; medicine, research and experimental; microbiology; pharmacology and pharmacy; reproductive biology
	6. NT	Instruments and instrumentation; materials science, characterization and testing; microscopy; mycology; nanoscience and nanotechnology; neuroimaging; optics; radiology, nuclear medicine and medical imaging; spectroscopy; surgery

References

Abdih, Y., & Joutz, F. (2006). Relating the knowledge production function to total factor productivity: An endogenous growth puzzle. *IMF Staff Papers*, *53*(2), 242–271.

Archibugi, D., & Pianta, M. (1996). Measuring technological change through patents and innovation surveys. *Technovation*, *16*(9), 451–468.

Breschi, S., & Malerba, F. (1997). Sectoral innovation system: Technological regimes, Schumpeterian dynamics, and spatial boundaries. In C. Edquist (Ed.), *Systems of innovation: Technologies, institutions and organizations* (pp. 130–156). London/Washington: Pinter.

Criscuolo, P. (2006). The 'home advantage' effect and patent families: A comparison of OECD triadic patents, the USPTO and the EPO. *Scientometrics*, *66*(1), 23–41.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, *10*(1), 92–116.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citations data file: lessons, insights and methodological tools. *NBER working paper*, 8498.

Hicks, D., & Katz, S. (1996). Systemic bibilometric indicators for the knowledge-based economy. In *Conference on new S&T indicators for the knowledge-based economy*

Kleinknecht, A., Van Montfort, K., & Brouwer, E. (2002). The non-trivial choice between innovation indicators. *Economics of Innovation and New Technology*, *11*(2), 109–121.

Lach, S. (1995). Patents and productivity growth at the industry level: A first look. *Economic Letter*, *49*, 101–108.

Mansfield, E. (1980). Basic research and productivity increase in manufacturing. *American Economic Review*, *70*(5), 863–873.

Marlerba, F. (2002). Sectoral systems of innovation production. *Research Policy*, *31*(2), 247–264.

Metcalfe, J. S., & Ramlogan, R. (2005). Limits to economy of knowledge and knowledge of the economy. *Futures*, *37*, 655–674.

Meyer, M. (2002). Tracing knowledge flows in innovation systems. *Scientometrics*, *54*, 193–212.

Moon, H. S., & Lee, J. D. (2005). A fuzzy set theory approach to national composite S&T indices. *Scientometrics*, *64*(1), 67–83.

Nelson, R., & Rosenberg, N. (1998). Science, technological advance and economic growth. In A. D. Chandler, P. Hagstrom, & O. Solvell (Eds.), *The dynamic firm* (pp. 45–59). Oxford: Oxford University Press.

OECD. (1996). *The Knowledge-based Economy*. Paris.

Okubo, Y. (1997). Bibliometric indicators and analysis of research systems. *OECD STI working paper* 1997/1.

Park, G., & Park, Y. (2006). On the measurement of patent stock as knowledge indicators. *Technological Forecasting and Social Change*, *73*(7), 793–812.

Pavitt, K. (1984). Sectoral patterns of technical change: Toward a taxonomy and a theory. *Research Policy*, *13*(6), 343–373.

Rosenberg, N., & Nelson, R. R. (1994). American universities and technological advance in industry. *Research Policy*, *23*, 323–348.

Scherer, F. (1982). Interindustry technology flows and productivity growth. *Review of Economics and Statistics*, *64*, 627–634.

Small, H., & Garfield, E. (1985). The geography of science: Disciplinary and national mappings. *Journal of Information Science*, *11*, 147–159.

Watabane, C., Tsuji, Y. S., & Griefy-Brown, C. (2001). Patent statistics: Deciphering a 'real' versus a 'pseudo' proxy of innovation. *Technovation*, *21*(12), 783–790.

Wouters, P. (1998). The signs of science. *Scientometrics*, *41*, 225–241.