ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Master's Thesis

Technology portfolio analysis using patent data: printer manufacturing firm's performance

Lee Do Hyun

Department of Management Engineering

Graduate School of UNIST

2019

Technology portfolio analysis using patent data: printer manufacturing firm's performance

A thesis/dissertation
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of

Master of Science

Lee Do Hyun

12. 28. 2018

Approved by

____

Advisor
Woo Hangyun

Technology portfolio analysis using patent data: printer manufacturing firm's performance

Lee Do Hyun

This certifies that the thesis/dissertation of Lee Do Hyun is approved.

12. 28. 2018

signature

____

Advisor: Woo Hangyun

signature

____

Kim Young Choon: Thesis Committee Member #1

signature

____

Choi Young Rok: Thesis Committee Member #2

Abstract

Firms participating in printer industries have invested their constrained resources into technology development in order to sustain their competitiveness in the industry. Considering the fast-changing market circumstances, each firm's own investment decisions on technology portfolio may directly affect their performance.

In this study, we analyzed patent data, namely number of forward citations and technological classification data (CPC). Using this data, the technological portfolio of a specific firm can be identified, which can further help our understanding on firms' R&D investment strategies. Number of studies mainly focused on patent class combinations of individual technology level, but portfolios of patent class at a firm level have been understudied.

In this study, we tracked the change of class composition within each firms' technological patents' portfolio and attempted to identify practical and theoretical implications to portfolio management. We utilized Entropy Index, Co-occurrence and cosine similarities measurements for each indicating diversification, patent scope and portfolio similarities within each patents' classification subclasses. Additionally, performance evaluation of each portfolio is conducted using forward citation data.

This paper shows that in-depth patent data analysis can allow us to explore deeper insights at various levels, individual technology, products and product lines, and firms sufficing different stories.

Contents

1. Introduction

1.1 Research Background

Studies using patent data have long been conducted in various contexts from the past. Additionally, in product development, innovativeness has a positive effect on the performance of the company (Kleinschmidt, 1991), so the need for R & D management for innovativeness of products has emerged. Therefore, Research on the R & D capability of firms using open patent data has been conducted in the past to improve understanding of investment in technological development of companies. Previous researches showed that higher R&D expenditures positively affects the value of patent issue (Pakes, 1980). They also analyzed the network between the patent classes using classification data such as patent IPC. Research on the proximity of patents, distance between patents, distance measurement and mapping between patent classes have been actively studied (Cho, 2011). However, the discussion of such classifications basically focuses on the mapping of relations between classes itself. This study categorizes the internal patent portfolios of companies participating in the printer manufacturing industry on the base of subclasses, investigating whether a company is actively conducting research and showing performance in certain subclasses. As the technology invented by the company becomes more attractive as the direction of the market becomes more consistent (H. Ernst, 2003), the direction of technology development will be influenced by the market and the situation of the industry.

Previous studies used the co-occurrence between subclasses matrix data and cosine distance and similarities data to measure the distance and similarity between corresponding technology classes to study the correlation between patent subclasses within industry.

The ability of companies to develop and patent their technology development capabilities is seen as a decisive factor in corporate performance and innovation in product development. The importance of investing in technology development and patent issuance is becoming increasingly important for each company and developing a technology that is relatively superior to other companies has a positive impact on the company. However, in highly competitive environments, the benefits of technologies and patents developed by each company cannot be temporary and permanent. Over time, however, a relatively steady flow of technological advances allows companies to continue to generate high levels of profitability. Therefore, effective project or R&D process management has been a major concern for many business and business researchers. This study investigated key issues related to effective technology development or patent development for a company.

First, we investigate the relationship between the performance of the patent issued by the company and the patent issued by the competitor of the industry belonging to the company. Previous research and literature have used patent data and found that corporate spending on R & D investments has a positive impact on the value of issued patents (Pakes, 1980). Holger Ernst (1998) identifies patent strategies in an enterprise-wide patent portfolio and identifies the relative technical position of competitors. Existing research and development management consists mainly of a portfolio of projects rather than a patent portfolio, and research has been conducted to select valuable R & D projects (Christian Stummer and Krt Hidenberger, 2003)

A portfolio of patent data provides different traits and insights from the portfolio of technology development projects, and research is expected to provide relatively meaningful insights into additional technology strategies. In particular, we tested the effect of the relationship between a company's patent portfolio and its patent portfolio on its portfolio performance.

This study examines how a company's patent portfolio using patent data affects the performance of patents issued by patents. In addition, because the relationship between the direction of patent technology announced by other companies may affect the company's patent portfolio, the relationship with other companies' patent development direction was examined. In addition, the impact of patents on the performance of a patent portfolio is analyzed according to whether it is developed through various technologies, or whether it focuses on a particular minority technology, which affects a company's patent portfolio.

Despite the fact that many previous literatures have studied the performance of individual patents or the performance of patent portfolios, this research attempts to provide new insights by focusing more detailed relationships on the underlying skill level data contained in each patent data. At the same time, complex tests of these unknown relationships can investigate not only the relationship between a single image and patent performance, but also unknown relationships.

Therefore, this study provides insight into how the company's patent technology development direction and portfolio characteristics affect overall patent performance. In addition, each relationship or unknown characteristic studied has been discussed with evidence provided in previous studies.

Previous literatures have also attempted to map network effects using sub-descriptive data or to map patent data to patent data, but very little research has been done to directly use the investment characteristics of the underlying technology to perform the portfolio.

Existing specific studies have focused on specific industries and investigated the unknown parts of the patent portfolio by company. It is also interesting that, despite the existence of persuasive theories that argue for a close relationship between persuasive theory of technology development and portfolio diversity and performance, there is no recent study of empirical studies of these relationships.

It is interesting to note that despite recent explosive investment in patent issuance by all firms, there is still lack of empirical research on this relationship. Finally, while the diversification of firms' technology development or patent issuance appears to have a positive impact on the special portfolio, the relationship between patent issuance and firm performance is not yet clear. Some researchers have studied and suggested that the development of technology that focuses on diversification and resulting patent publishing has had a noticeable effect, but other researchers claim that this apparent relationship does not exist. This study attempts to resolve the gaps that have not been addressed in many previous studies. This research topic has been empirically tested for 16 years using panel data from the 12 most active companies in the printer manufacturing industry. The global printer manufacturing market is characterized by a small number of companies dominating market share and fewer manufacturing companies. Each company's patent portfolio presents a model with a fixed effect on corporate variables, taking into account the nature of the panel data released annually.

Additional information on this study follows. More detailed information on hypotheses and theoretical frameworks that have been verified in this study is presented along with previous related literature. Discuss data collection and sample and research methodologies. Finally, we present the limitations and implications of this study, while drawing meaningful conclusions and conclusions.

1.2 Research Motivation

All of corporations are needed to be perform best using limited amounts of sources. When they decide which technological areas to invest and get results, every corporations' decision would be different because of difference from source, own strength and weakness, foresights at investments and others. Such investment direction for technological development and actual source used in this development are decided by several strategical considerations from each corporation. Patent information can help to identify the source of the company's R&D portfolio management and potential technology development (Holger Ernst, 2003). Since their patent data is always open to public, using this R&D investments' results can help us to getting clues about corporations' strategical base. Using the data based on the patent IPC, clustering can be used to obtain information on which patents belong to the trendy patent, the classic patent, and the dated patent (Dereli, 2009). Based on the facts that these patent data can provide technological trends, it is necessary to study how the similarities between the technical trends in the industry and the portfolio of the corporation affect the portfolio performance.

Using the classification codes in the patent classification system, both the technical similarities and the technical dissimilarities between classification codes are shown at the same time. The patent issued by the company and the patent classification code assigned by the patent classification system describe the specific sub-technology group of the technology. Using the information on the technology groups of the individual technology patents that the patent classification system shows, creating a technology group portfolio of the technologies issued by the companies, it is possible to see the investment of the technological development of the technology group by the companies. The printing industry is largely divided into inkjet product segments and laser product segments. This study explores how actively companies participating in the printing industry account for these two segments and the more specific individual subclasses that make up the segments in their technology patent portfolios. As presented in previous studies, we measured the value of each patent by using forward citation, which can express the value and importance of the patent, while measuring how the patent subclass portfolio is changing over time.

2. Literature Review

Past studies have been explored patents data in orders to get insights about strategical behavior of each patent's assignees. Patent data is an internationally public source and contains detailed descriptions of each technology and backward citation, forward citation and technology classification etc. In previous studies linking patent data with corporate R & D activities, they used patent data to measure knowledge flows between industry and industry (Evenson, R E, 1993). It is possible to study the innovative power and patent strength in the technology segment by constructing a patent portfolio that can obtain the results of the company's patent activity and the quality of the patent (Bernd Fabry, 2006). Using patent information to assess the soundness and compatibility of the company's technological development will help to evaluate M & A value (Anthony et al., 2002) It is meaningful to combine the two elements of patenting strategy and economic feasibility to evaluate value creation through patents (Michele Grimaldi, 2015).

Patents that are developed within the boundaries of technology that companies already own are classified as exploitative technology rather than explorative technology (Gilsing et al., 2008; Vanhaverbeke et al., 2006). By contrast, a technology class is viewed as exploratory if it is present in the organization's stock of technology classes in the observation year, but absent during the preceding five years. In the Dietary Supplements market segments, it was suggested that the patent portfolio is a useful tool for analyzing R & D and business opportunities. These results provide interesting information about the innovative potential of a company. Past studies also provided various variables that could be used as indicators of patent portfolios (Bernd Fabry, Holger Ernst 2006). Several studies have also conducted numerical studies on more specific indexes used in patent portfolios such as patent portfolio size, patented market coverage, and technical relevance (Holger Ernst, Nils Omland 2011). The composition of the patent portfolio is going to aggregate the characteristics of the individual patents, and the past studies to quantify the characteristics of the individual patents have been actively carried out. Previous studies have presented a structural concept for evaluating the value of a patent through information extracted from a patent portfolio and provide characteristics for this conceptual framework. This framework has been shown to be useful for technology strategy management and technology strategy development (Michele Grimaldi, Livio Cricelli, 2015). Previous research on the utilization index of the patent portfolio has been done by many researchers in the past and has been fully recognized for its application potential. There have been many attempts to measure the similarity between portfolio and portfolio using the concept of patent portfolio. The method of measuring the similarity between documents is generally presented by a general method such as cosine similarity or Euclidean distance measurement, and some more advanced measurement methods are presented to date. One study proposed two hybrid models to calculate similarities between patent portfolios. There are two ways to calculate semantic similarity in categorical similarity and textual data using international patent classification criteria. In addition, the result of the study suggests that the hybrid similarity measurement method using a proper combination of the categorical model and the semantic model is better than the conventional similarity measure (Yi Zhang, Lining Shang 2016).

Taken together, the findings of the past patent portfolio study show that companies can quantify various features of their respective patent portfolios. A patent portfolio owned by a corporation is an aggregated form of data such as citation of individual patents, international technical classification data, and textual data, and this type of patent portfolio data consists of patents that are the result of R&D , And implications of the firm's technology development strategy.

Therefore, research has been done in the past to relate the patent portfolio data to the company's technological position and enterprise strategy. Patent portfolios and portfolio citation data can be used as a basis for a company's technical position and technical group. This study conducted research on technological position and technical strategy through k-means clustering analysis of patent citation index and patent portfolio (Shann-Bin Chang 2012). Among the data that the patent document contains, the international patent classification system such as IPC or CPC is one of the most actively used data of many studies. There an exists number of patent classification systems which can help effectively classifies all of patent depending on technological areas patents belong. USPC developed by USPTO (United States patent and trademark office) and IPC based on international patent classification law have been preferred in either academical patent recognition works and commercial patent classification by law.

However, jointly developed patent classification system CPC (cooperative Patent classification) is adapted by EPO and USPTO from 2013. Although many past patent studies have studied patent classification or categories based on the IPC, the Cooperative Patent Classification (CPC) has been found to be better at identifying related patents in the industry (Rahul Kapoor, Matti Karvonen 2015). Taken together, the previous studies did little to empirically investigate the relationship between quantified portfolio variables. Our research focuses on the characteristics of patent portfolios that affect the performance improvement by using the characteristic data of the patent portfolio presented in the previous studies.

3. Hypotheses

3.1 Forward Citation for patent performance and Industry averaged technology development

Although many previous studies have studied the performance measurement of individual patent units or patent portfolios, these studies did not study the independent variables that affect performance. The primary objective of a company's patent portfolio is to ensure that the performance measured by a variety of methods is superior. Research on the variables needed for technology development and overall R & D process management to achieve this goal and how those variables affect will ultimately contribute to effective strategic direction.

Citation frequency data have been used as a tool to measure the value of patent because more forward citation has a meaning that this patent well performed as a foundation for other citing patents. In other words, this citation count or frequency data measure the quality of patents (Harhoff, 2006). Based on existing studies that measure the quality or value of the patents in individual patents, the aggregated average quoted index may be considered as the overall performance or value of the portfolio when multiple patents are collected and evaluated on a portfolio basis Previous research has shown that. However, forward citation records are become higher by time. Same forward citation value of 10 at 1999 year and 2010 year do not have same implication. Over time, forward citation clearly increase and this citation change rate is not predictable. Adapting forward citation without consideration of technological field patents belong or issued year can distort the significance of usage. Approaches based on fixed effect calibrate this time effects and field effects on forward citation data. Because of the documented nature of a patent that must be legally valid at issue, a patent must be cited for another patent that overlaps other existing patents and, in other words, a forward citation indicating that the patent holder owns the common technical base that other patents own. In addition, since companies' technology strategies are also influenced by the direction of patent or technology development issued by other companies, it is necessary to study the effect of portfolio development on technology development direction of other companies.

Backward Citation represents the total number of times the patent cited. Unlike Forward Citation, Backward Citation was mainly used to study knowledge flows or localness of patents, rather than being used as a performance measurement variable. Higher number of backward citations implie that this patent came from a lot of background knowledges and it can further sustain a meaning of stabilities or maturities of that technologies. It can evoke us a both of contrary ideas that patents having more backward citation ensures stability of patent issuing or patents having more backward citation are biased to the direction of exploitation rather than exploration. The attractiveness of technology is seen in the market, and the R & D strategy should correspond to the needs of the market (H. Ernst, 2003).

As the technology development direction of the company progresses in accordance with the demand in the market, the company can take various advantages by developing appropriate technology according to the demand. It can be said that the more attractive the technology development direction is, the more attractive the technology development can be. Therefore, it is necessary to pay close attention to the similarity or difference between the overall technology direction of the industry and the technology development direction of the single company formed by other competing companies.

Based on the patent issuance behavior of companies, it can be confirmed that companies participating in technological development in a certain industry have different directions of their own technological development. Likewise, we can identify the technology subclass where the technology is most actively developed in the industry, and we can confirm the similarity between the average technology development direction of the industry and the technology development direction of the enterprise unit. Local learning effects are also valid in citation between patents. Almeida (1996) showed that citation between documents has localization effects, which makes it more frequent to cite nearby patents. In addition, the technology invented in the industry is likely to be cited by other participants in the industry, so the higher the industry average technology development direction and the similarity of the company's patent portfolio, the higher the performance value measured by forward citation Can be expected.

Hypothesis 1 (H1) Higher the degree of similarity between the industry average patent development direction and the company's patent portfolio direction, the higher the average forward citation of the company.

3.2 Patent classification: CPC Classification

There an exists number of patent classification systems which can help effectively classifies all of patent depending on technological areas patents belong. USPC developed by USPTO (United States patent and trademark office) and IPC based on international patent classification law have been preferred in either academical patent recognition works and commercial patent classification by law. However, jointly developed patent classification system CPC (cooperative Patent classification) is adapted by EPO and USPTO from 2013. This CPC classification system primarily provides 9 main streams of technological section which can clearly describe the patents' technological area. This first classification can be confirmed by first capital string part of CPC classification codes.

Followed by first capital string in CPC code, added characters can explain detail technological area of patents. This CPC classification can be hierarchically narrowed down into section, classes, sub-classes, groups and sub-groups by orders. In this research, I used CPC classification information in the range from section to only main group not until subgroup since only including till the range of subclasses cannot represent concrete technological concepts of that classes. However, extending this patent classification scopes beyond subclasses to main groups and sub-groups contain excessive huge number of technological areas that cannot be separated into meaningful technological areas which are needed to be studied in this research. As we identified printing industry participants firms by filtering abstract information contained in patent documents, CPC codes included in all of patents which is having words *'printing' or 'printer'* not *'3D printer'* are acquired. This filtered CPC codes are most frequently appearing technological sub-segments to represent core printing technologies performing in this industry. Based on the above subclasses classification system, we analyzed which technology segment corresponds to the patent technology group in the printer industry of each company by year.

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Xerox Corporation	406	537	438	582	518	398	512	503	440	370	314	354	388	385	438	612
Brother Kogyo	72	90	104	138	165	151	134	80	83	127	124	187	270	386	449	648
Canon Kabushiki Kaisha	592	892	801	1,115	1,132	1,191	1,213	1,325	1,299	1,178	1,204	1,551	1,321	1,441	1,452	1,704
Fuji Xerox	122	116	128	201	185	134	110	142	130	149	148	163	222	235	305	423
Hewlett-Packard	189	199	208	307	342	431	512	608	1,368	1,619	1,602	1,647	1,119	1,013	910	1,146
Lexmark International	12	19	27	48	75	99	115	97	110	93	78	101	139	124	112	155
Ricoh Company	176	198	229	281	294	268	242	232	277	367	305	447	470	545	606	827
Seiko Epson	79	85	130	168	199	234	316	424	473	570	561	732	779	775	872	908
SHARP	147	160	197	276	286	325	278	262	251	231	221	335	322	300	353	487
Sony	198	244	261	349	398	442	414	458	384	443	375	581	456	491	636	830
Toshiba	268	242	258	375	390	417	365	339	457	467	456	632	562	575	576	773
Samsung Electronics	119	155	171	316	398	386	404	488	523	604	577	838	910	1,116	1,296	1,665

Table 2. Firms list in printer manufacturing industry and their number of patents in this industry

3.3 Co-occurrence within printer industry subclasses

In information science research area, suggesting symmetrical co-occurrence matrix based on co-citation, co-word and co-link is often used to help understand the structures of documents. In this research, we tried to understand the feature of patent documents by inspecting the co-occurrence among patent classification subclasses. Breschi (2003) showed knowledge relatedness between technological fields using the IPC classification codes co-occurrence matrix. His research measured the cosine similarities in the co-occurrence matrix and calculated the angle separation between field and field, but in this study, we use the co-occurrence frequency.

Almost of patents are recorded with several number of subclasses at the same time, not identified with only one subclasses in general. Existing number of papers studied this co-occurrence of subclasses focusing on the network effects. Appearing more than one of subclasses in each patent document's classification system can imply that these subclasses are having synergy effects with together. Additionally, more number of subclasses are linked to the specific subclass, technological significances of that subclasses are more meaningfully fused with other technological areas. Therefore, number of links connected with target subclasses and their boundary until which classes are covering and related with target subclass can help us to understanding the extension characteristics of subclasses. Co-occurrence matrix can be used to easily find out how many kinds of CPC classes appear in a patent, and the patent scope measured using the numbers of subclasses appearing in the patent has a positive effect on firm's value (Lerner, 1994).

Typically, one or more subclasses are assigned to each patent document. The number of subclasses included in this document varies by patent. By identifying whether the same sub-classes appear simultaneously in the same document, you can identify the technical scope or complexity to which the patent applies. Examining the relationship between scope, complexity, and performance requires careful attention. However, in industries where the company actively participates, synergies between various subcategory activities and absorption capacity of the company are expected to improve the company's technological development investment performance. Patent scope has a significant impact on the value of each firms (Joshua Lerner,1994). In this context, the performance of the portfolio represented by the company's forward citation will also be positively impacted by the patent scope.

Hypothesis 2 (H2) The higher the patent scope within the firm's patent portfolio, the higher the average forward citation of the company.

Using this form of co-occurrence matrix, I can easily calculate the ratio of co-occurred frequencies to correspond subclasses by dividing diagonal element with sum of elements below the diagonal element.

3.4 Cosine Similarity between two vectors

Cosine similarity between two patent vectors are calculated using below equation. Cosine similarity calculation is a method of measuring the similarity between two vectors. The two vectors are expressed with respect to the angle difference theta between the direction and the direction indicated in the n-dimensional space. The cosine similarity can be calculated from the angular difference with respect to the direction indicated by each vector, rather than by the angle difference induced by magnitude.

$$\cos(\theta) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{1}$$

3.5 Technological concentration at each subclass segments

The Herfindahl-Hirschman Index measures market share dispersion and variability using market shares of individual firms (Richard, 1982). HHI also measures technological concentration in corporate portfolios (Yu-Shan Chen, 2010) and is used for relative dispersion or concentration in various sectors. In our study, we used HHI to measure the degree of concentration of how patents are distributed across firms within the technology segments that each subclass represents.

$$H = \sum_{i=1}^{N} s_i^2 \tag{2}$$

In this calculation, s_i is the share of number of patents belongs to specific technology subclass.

3.6 Technological Diversification within firm’s patent portfolio

Unlike the Herfindahl-Hirschman Index, which shows the degree of concentration of subclasses in the industry, we can observe the diversification of a company's patent portfolio to see which technology segment the company is investing heavily in. Technological diversification has been found to have a positive impact on firm's innovative competence and influence exploratory innovation rather than exploitation innovation (Garc´ıa, 2008).

Within the same printer industry, companies are developing a variety of technologies to match their competitive capacity and technology development strategy. As we can see from below table 3, As can be seen in Table 3 below, patent segments developed by each company during the year are distributed across various subclasses. Technology diversification have strong impact upon R&D expenditures and firm’s sales (Granstrand 1994). However, there has been no study of whether technological diversification within a narrower unit of patent subclasses within a specific industry would result in risk reduction or performance improvement which every enterprise wants, as in the case of previous enterprise studies.

Hypothesis 3 (H3) The higher the patent portfolio diversification within the printer industry, the higher the forward citation. However, too much diversification has an adverse effect on performance.

4. Methodology

The importance of each company's investment in technology development and the issue of patents is becoming increasingly important. However, the benefits of technologies and patents developed by each company in a highly competitive situation can bring temporary benefits, but in most cases do not lead to a long lasting superiority. Therefore, as time goes on, companies are investing in technology, the number of patent issuance is increasing, and the management of R & D process for a huge amount of technology development investment is not succeeding efficiently. This situation can be a major concern for many companies, and empirical research for effective R & D management is needed.

In this study, companies that actively participate in technological development in the printer manufacturing industry are screened and their patent application activities are studied. Quality of patents affect the variation of market value of firms, which has different effects for different industries (Jean, 2004). There was relatively little activity in the specific industry to identify the main technological development participants with only patent data and to evaluate the research activities of the companies in the industry based on the patent issuing activities of the companies. This study uses the CPC classification, which describes each patent issued by a company, to monitor a company's patent issuance activity.

4.1 Data Collection

The figure below shows a visual representation of the data collection and processing. We collected patent data from 1995 to 2010 and focused on the data on assignee, CPC classification and citation included in the patent.

A total of 90,948 printer industries were used during the 16 years from 1995 to 2010, out of patent data registered with the United States Patent and Trademark Office (USPTO). Of these, 12 were the most actively selected companies in the printer industry and 90,948 patents were used in the company. Figure 1 and Table 1 below show the number of patents issued by the printer industry each year. The number of patents issued in the industry from 1995 to 2010 is steadily increasing.

Since the financial data of every printer manufacturing industry of the company is not released every year, it is necessary to know indirectly the companies participating in the industry. Using a text mining-based approach, we identified the most active printer manufacturing companies. First, patent abstracts containing USPTO patented abstract data are used to filter patents that contain the word 'print' or 'printer' rather than '3D printing'. In the beginning of these patent documents, the assignee data is organized to identify the companies that have issued the most patents in this printer manufacturing industry.

All of the 12 most active companies in the printer industry have been identified, including the most common CPC classification codes in the printer industry patents. At the end of the process, feature engineering of data was conducted to complete the CPC-based patent portfolio for each company.

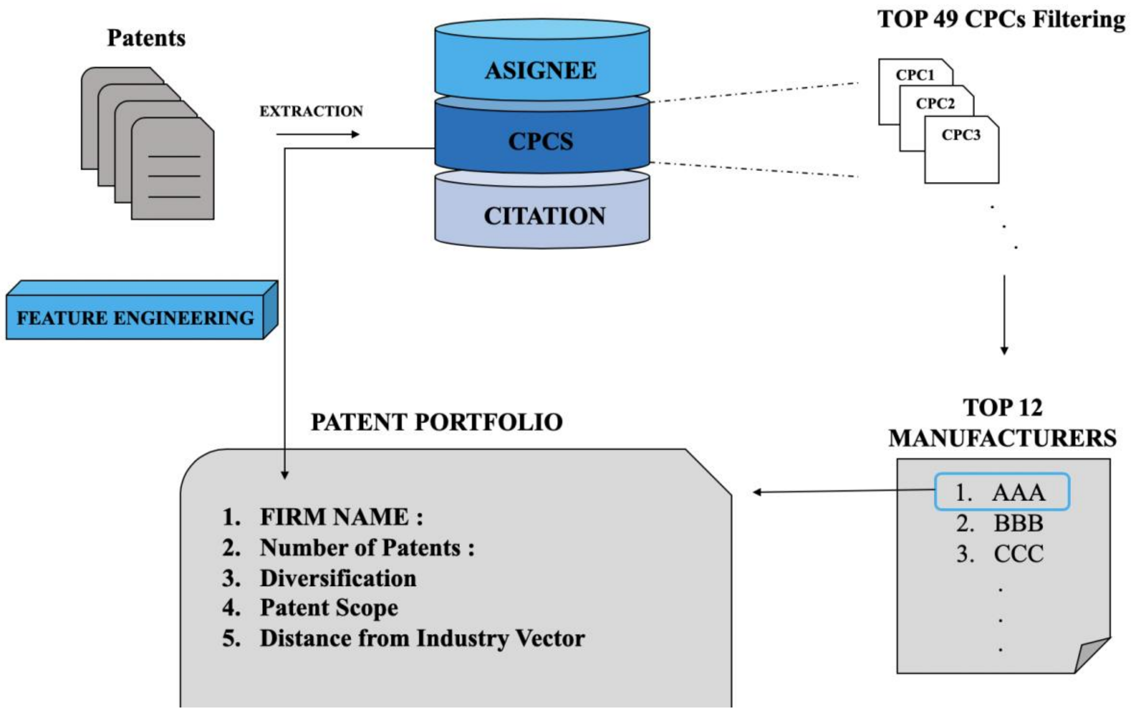

Figure 1. Data Collection and Preprocessing

This completed data collection process and portfolio creation will show a total of 192 observations over the 16 years of 12 companies. The process of engineering the data contained in a patent document into individual variables is described in more detail in the next section. As can be seen from the table below, from 1995 to 2010, 12 companies are most actively issuing patents in the USPTO. Most companies are participating in the printer manufacturing industry by increasing the number of patents every year. Over the years, the absolute number of patents issued by each company is generally increasing. However, even a simple number of patents cannot provide reasonable evidence of a company's technology management decision base because all patents are at different levels of performance and goals. It is difficult to identify technology groups that are important to the printer industry with issued patents alone. To compensate for the weakness of the absolute number of patents, it is necessary to use other variables.

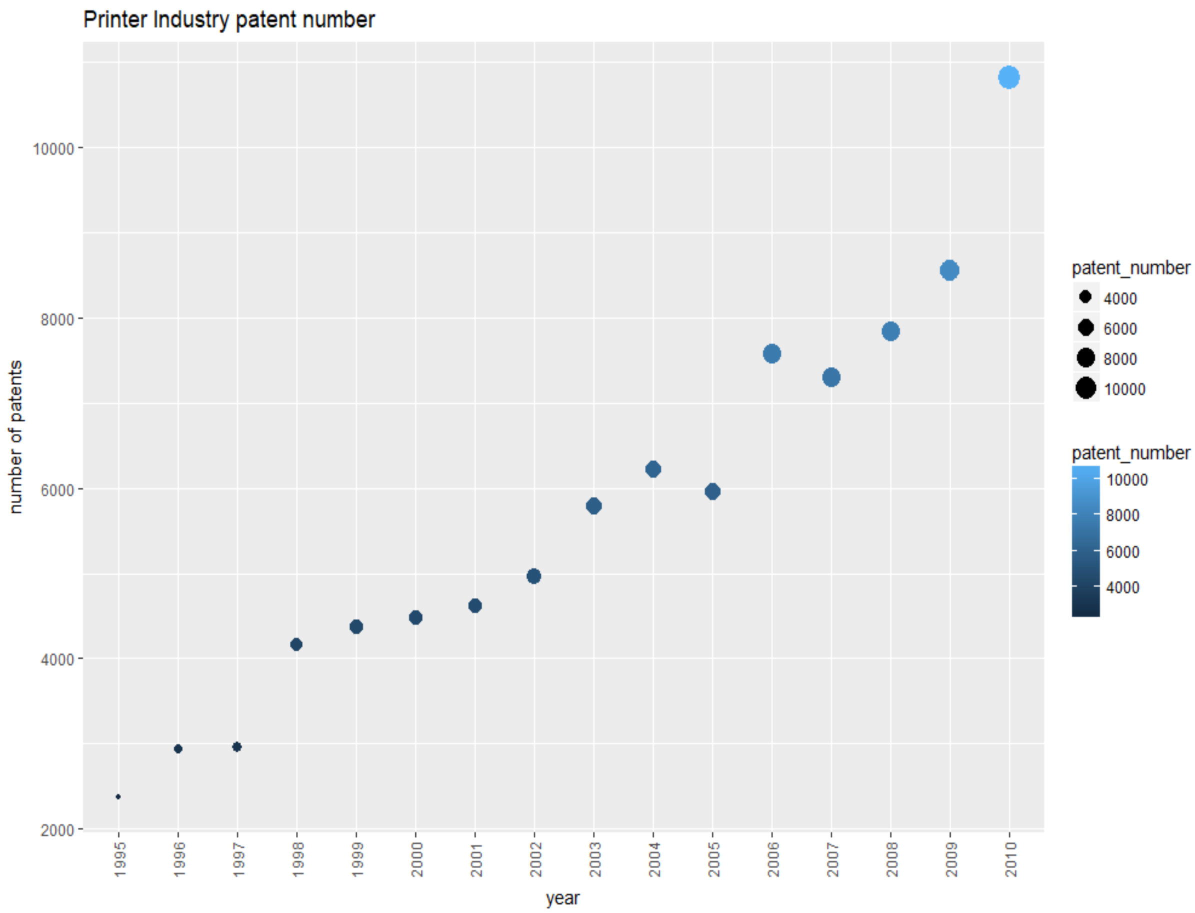

Figure 2. Printer industry patent number

Year	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Number of patents	2,374	2,937	2,952	4,157	4,383	4,480	4,618	4,959
Year	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Number of patents	5,797	6,222	5,970	7,577	7,309	7,844	8,551	10,818

Table 1. Printer industry patent number

year	firm	G03F7	G06F15	G06F3	H05K3	B65H29	G03G21	B41J2	H04N1
1995	Xerox Corporation	3	4	14	2	16	29	40	83
1995	Brother Kogyo	1	0	3	0	0	2	17	6
1995	Canon Kabushiki Kaisha	36	2	14	2	0	39	75	147
1995	Fuji Xerox	0	1	1	0	1	2	13	30
1995	Hewlett-Packard	1	3	13	7	0	0	69	19
1995	Lexmark International	0	0	1	0	0	0	1	1
1995	Ricoh Company	0	1	10	0	0	10	8	57
1995	Seiko Epson	4	2	0	3	0	1	20	1
1995	SHARP	6	9	15	6	1	5	3	18
1995	Sony	8	2	7	4	0	0	2	8
1995	Toshiba	8	3	24	4	0	3	5	25
1995	Samsung	5	0	0	3	0	2	7	10

Table 3. Subclasses developed by each company for the year

4.2. Measurements

4.2.1 Dependent variable

Average forward citation

The average forward citation of patents issued by companies in the corresponding year is calculated. Forward citation is accumulated over the years, so forward citation is naturally high for old patents. To prevent time-dependent aspects of forward citation, the time dependent effect was suppressed by the forward citation value received for 5 years after the grant year of the patent. Figure 2 and Figure 3 below show the boxplot of 5-year forward citation by company and 5-year forward citation by year, respectively. In addition, as shown in Figure 2, the average value and distribution range of this performance variable is different for each company, and 5-year averaged forward citation can confirm heterogeneity for each company.

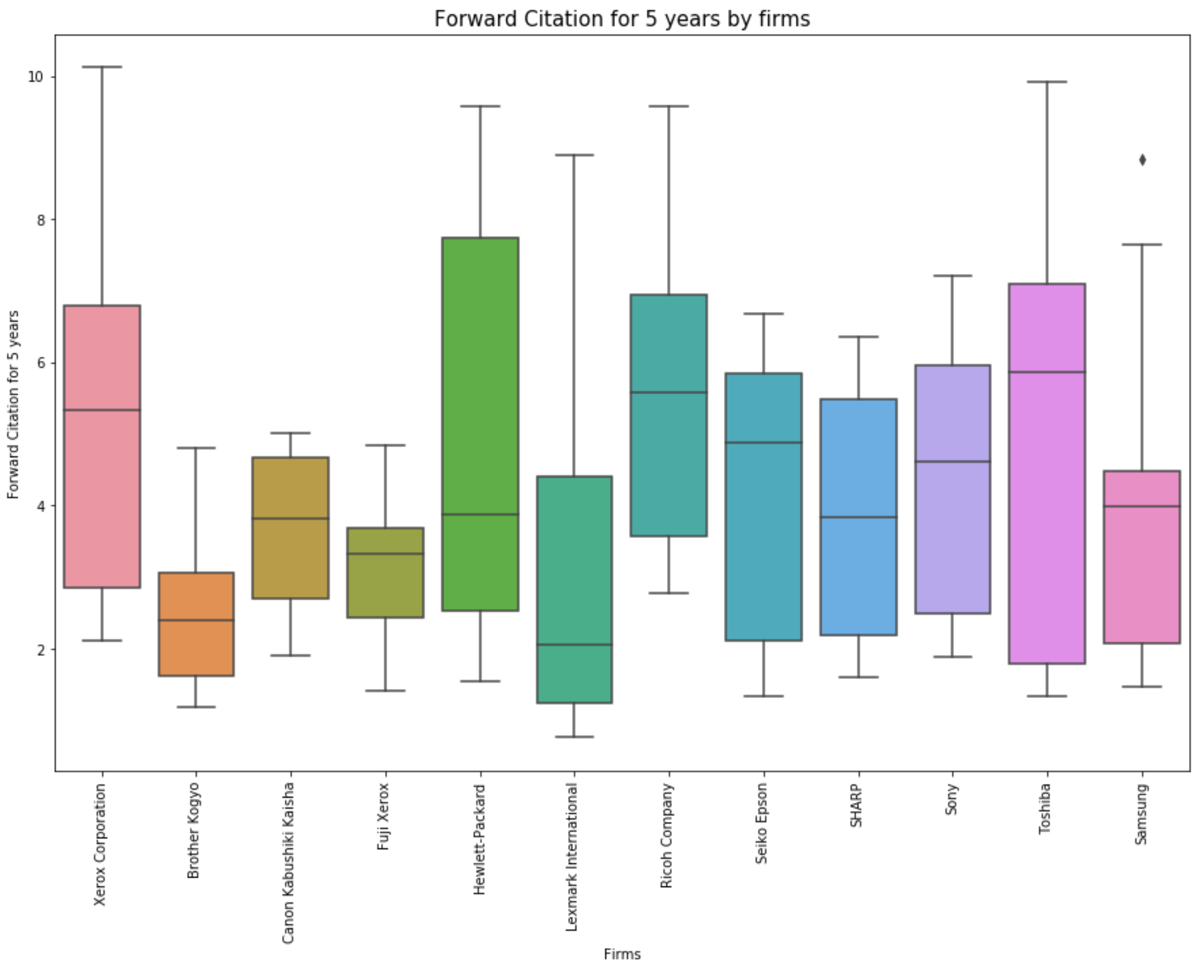

Figure 3. Forward Citation Boxplot by Firms

Just as the dependent variable above is shown in boxplot for each company, the graph below is a boxplot showing the distribution of dependent variables over the years. As can be seen, there is a difference in the distribution of dependent variables depending on the year. Also, the later published patents are likely to have lower average citation values for patents received over the next five years. This is because the number of patents issued by corporations increased rapidly as the year progressed, and the size of the entire citation did not increase so much, and as a result, the value of the average forward citation increased. Panel data regression is appropriate when we consider the difference of annual index of citation and the characteristics of dependent variables where heterogeneity exists in each company.

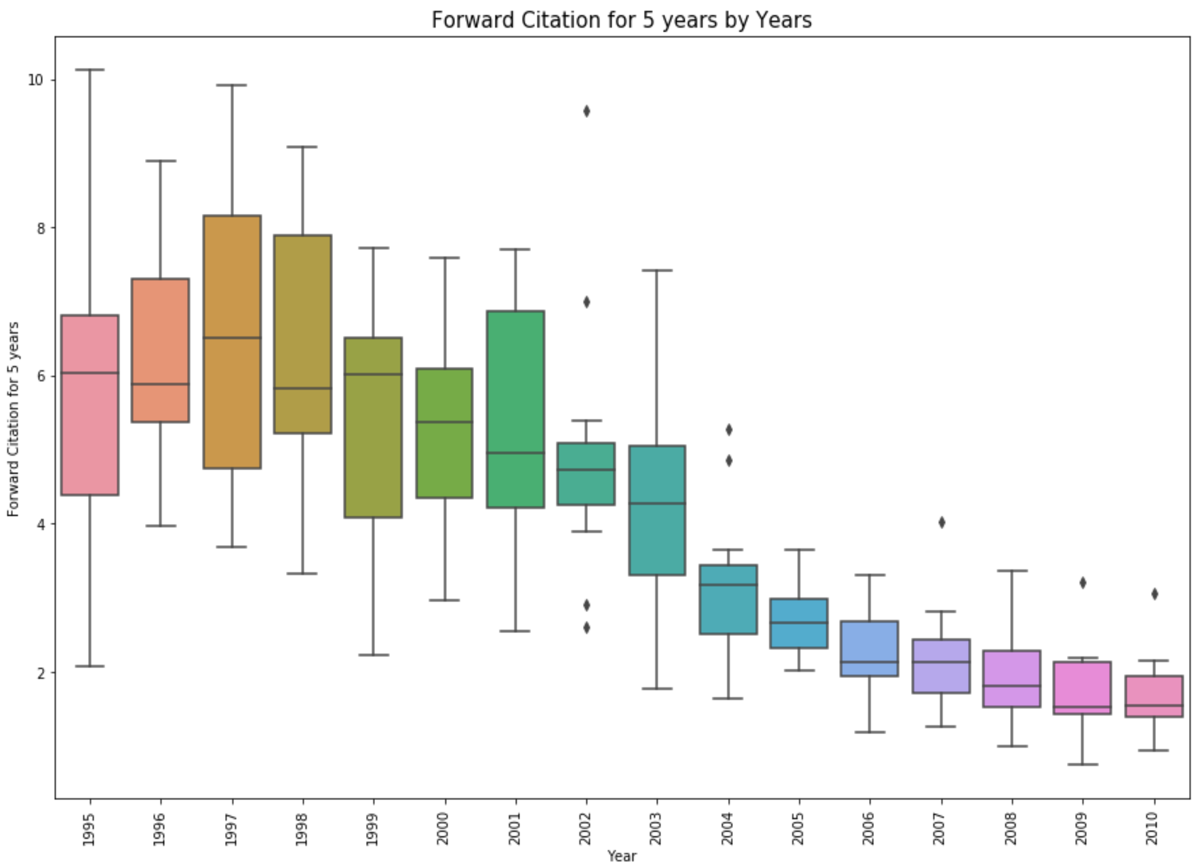

Figure 4. Forward Citation boxplot by Years

4.2.2 Independent variables

Technology Portfolio Diversification

A number of studies on past R & D management have used several methods to measure how many different technology groups the company is investing in. The Herfindahl index, which is used to measure the market share of the market, is used to measure market concentration depending on how many companies have market share in that market. This measure was used to measure portfolio diversity in portfolio management literature. Other studies have measured entrepreneurial portfolio diversification through entropy calculations. In this study, entropy index was used to measure portfolio diversity. The formula for portfolio diversification (PD) using this entropy is as follows.

$$PD = \sum_{i=1}^{49} P_{it} \ln \frac{1}{P_{it}} \tag{3}$$

P_{it} is proportion of the number of ith subclasses in the year t. Using this equation, we can see how companies are building their technology portfolios across various subclasses. The higher the PD value, the higher the technology diversity of the portfolio.

Similarities between industry average technological trends and firm's portfolio

The 12 companies participating in the printer industry have their own patent portfolios each year and have calculated the industry averaged technological direction by aggregating all of these portfolios. Industry averaged technological vector IV_t is calculated by the following equation (3). v_{jt} is portfolio vector of firm j in year t.

$$IV_t = \frac{1}{n} \sum_{j=1}^{12} v_{jt} \tag{4}$$

Using this Industry averaged technological vector, $\cos(\theta)_{jt}$ (cosine similarity between IV_t and firm's technology vector v_{jt}) is calculated using equation (4). Among the various similarities measurement methods, cosine distance measurement is used rather than Euclidean distance measurement because it is less sensitive to magnitude and more focused on direction of vectors.

$$\cos(\theta)_{jt} = \frac{IV_t \cdot v_{it}}{\|IV_t\| \cdot \|v_{it}\|} \tag{5}$$

Co-occurrence between patent subclasses and Patent Scope

The patents issued by the corporation include at least one classification subclass, and the co-occurrence matrix shows how each subclass within a patent coincides with which subclasses. This co-occurrence matrix shows how often each subclass, together with its various subclasses, co-occur within the patent document. Below table5 provides example of subclasses co-occurrence frequencies matrix.

Since diagonal elements are showing frequencies of each subclass itself, these elements are presenting highest value in every column. Elements under the diagonal are co-occurrence frequencies between subclass to subclass. The entire matrix takes the form of a 49 by 49 co-occurrence, which identifies the co-occurrence tendency of the patents issued by each company. Inside this co-occurrence matrix, the patent scope measures how many patents owned by the firm are represented along with the number of subclasses. Of the three variables we tried to verify, patent scope and portfolio diversification seem to be similar at first glance. However, the patent scope is a variable that captures aspects of the extent to which R&D covers the CPC-based coverage of corporate patent data. The latter, diversification, is a count-based arrangement of CPC data, even if R&D is performed over a wide range, so diversification may be low even with high patent scope, and vice versa.

Table 4. Variables Description

Variables	count	mean	std	min	25%	50%	75%	max
Similarity with Ind.Vector	192.00	0.75	0.11	0.46	0.68	0.77	0.83	0.97
Diversification	192.00	2.44	0.25	1.41	2.30	2.43	2.60	3.04
Diversification^2	192.00	6.00	1.21	1.99	5.31	5.91	6.78	9.25
Patent Scope	192.00	40.81	24.33	1.00	22.75	34.00	54.00	109.00
Performance	192.00	4.11	2.23	0.76	2.18	3.63	5.74	10.12

	B41F13	B41F31	G06K9	B41N1	G06F17	B41F33	B41F15	B41F27	B41F7	B41N3	H01L23	B41L13	B41F21	B65H5	H05K1
B41F13	4														
B41F31	0	4													
G06K9	0	0	2,243												
B41N1	0	0	0	39											
G06F17	0	0	239	0	4,614										
B41F33	0	0	0	0	0	17									
B41F15	0	0	0	0	0	0	12								
B41F27	0	0	0	0	0	0	0	4							
B41F7	0	0	0	0	0	1	0	0	2						
B41N3	0	0	0	7	0	0	0	0	0	25					
H01L23	0	0	0	0	41	0	0	0	0	0	3,709				
B41L13	0	0	0	0	0	0	0	0	0	0	0	4			
B41F21	1	0	0	0	0	0	0	0	0	0	0	0	11		
B65H5	0	0	0	0	1	0	0	0	0	0	0	0	2	456	
H05K1	0	0	0	0	21	0	1	0	0	0	439	0	0	0	1654

Table 5. Example of Co-occurrence among subclasses

5. Analysis Results

Our data is a listing of the characteristics of the patent portfolio of each year's companies. Considering the heterogeneity of the performance of each firm and considering the change of the trend according to the year, the model considering the fixed effect of companies in the Least Square Dummy variable (LSDV) model is appropriate. It can be seen that the coefficients of each model are constant regardless of the corporate variables due to the structure of the model that observes the change of the dependent variable due to the change of the independent variable with each company variable included in the equation as dummy variable.

In other words, when the dependent variable changes due to each independent variable in the graph, a total of twelve companies display different intercepts but tend to be parallel to the same slope. Fixed effects model detects changes within a group over time. This change is a change that fully reflects omitted variable bias, so changes between groups and groups are not used.

We analyzed the relationship between dependent variables and independent variables of each patent portfolio calculated in the manner described in the previous measurement session. The Hausman test did not conclude that the fixed effects model is significantly better than the random effects model. Therefore, we decided to interpret the model by focusing on the random effects model.

Table 6. Analysis model result

	Dependent variable:				
	Avereage Citation				
	(1)	(2)	(3)	(4)	(5)
Similarities.ind	-5.950^{***}	-5.950^{***}	-6.228^{***}	-5.593^{***}	6.052
	(1.144)	(1.144)	(1.132)	(1.090)	(11.290)
Entropy	1.136^{**}	1.136^{**}	12.700^{***}	0.419	0.909
	(0.567)	(0.567)	(4.592)	(0.563)	(0.606)
Patent Scope	0.015^{**}	0.015^{**}	0.021^{***}	0.094^{***}	0.020^{***}
	(0.006)	(0.006)	(0.007)	(0.019)	(0.007)
entropy^2			-2.490^{**}		
			(0.982)		
PatentScope^2				-0.001^{***}	
				(0.0002)	
$\text{Similarities.ind}^2$					-8.343
					(7.808)
Observations	192	192	192	192	192
R^2	0.152	0.152	0.183	0.238	0.158
Adjusted R^2	0.064	0.064	0.092	0.154	0.065
F Statistic	10.351^{***} (df = 3; 173)	10.351^{***} (df = 3; 173)	9.617^{***} (df = 4; 172)	13.447^{***} (df = 4; 172)	8.055^{***} (df = 4; 172)
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

First, if we look at the results of the analysis, as shown in the results table below, the higher the similarity between the average vector direction of the companies participating in this industry and the orientation of each company portfolio, the lower the portfolio performance.

The p-value of the coefficient for this variable is very low, below 0.01. This is the opposite of the H1 hypothesis and therefore the H1 hypothesis was rejected. The first hypothesis rejection by this variable implies that the performance is worse for firms with a patent portfolio that is similar to the average vector direction of the 12 top companies participating in the printer manufacturing industry. In other words, it is necessary to construct a patent portfolio in a different direction that is different from the overall average patent portfolio of the industry. This is consistent with the fact that technological diversification positively affects performance, as has been demonstrated in previous studies, and can also be interpreted as the need for a heterogeneous competitive capability that a firm possesses (McEvily, Zaheer 1999).

Second, the impact of the patent scope in the above model appears to be consistent with the hypothesis of H3. However, the absolute value of the coefficients describing this independent variable is very small, which is understandable given that the average value of the patent scope variables is higher than the other variables.

As shown in the descriptive table for each of the above variables, since the average value and the standard deviation value of the patent scope variable are higher than other variables, the corresponding absolute value of the coefficient value appears to be recorded low.

Third, the results of the technological diversification in entropy measured by entropy showed that the improvement of portfolio performance by entropy was positively related to each other. It also meets our H2 hypothesis. However, if we include a quadratic term for this variable in the model, the sign is negatively opposite to that of the first term. The negative value of the quadratic term is smaller than that of the first term, but its implication cannot be ignored.

In general, when the sign of the first order and the sign of the second order are opposite, the dependent variable graph by this variable becomes a curve shape. In particular, in our case, the dependent variable graph by this independent variable shows the inversed-U shape because the first-order is positive and the second-order is negative. A graph of this effect can be seen in additional figures. To illustrate this inversed-U shape entropy pattern, we have plotted a variation of the performance by the single entropy parameter.

Although the graph is a single variable of entropy, since the graph is a quadratic function including both entropy first order and second order, it increases until a certain period and then decreases again. In the technology diversification range within the patent portfolio corresponding to less than 2.55

entropy, the performance of the portfolio increases with diversification. However, from the point above 2.55, that is, when the diversification within the portfolio is too much, the performance decreases as the entropy increases can be confirmed. Among the 192 patent portfolios of the 12 companies, the 58 portfolios were found to have higher degree of entropy than 2.55, and the remaining 134 portfolios were less than 2.55, which was not caused by the excessive technological diversification.

In addition, in practice, the average performance value was significantly higher than 4.18 in the reverse case, including 4.08 in the reverse case. In addition to numerical interpretations, the significance of these results implies that R & D over an extremely diverse range of technologies, even when developing technologies within a single industry of printer manufacturing, can have a negative impact.

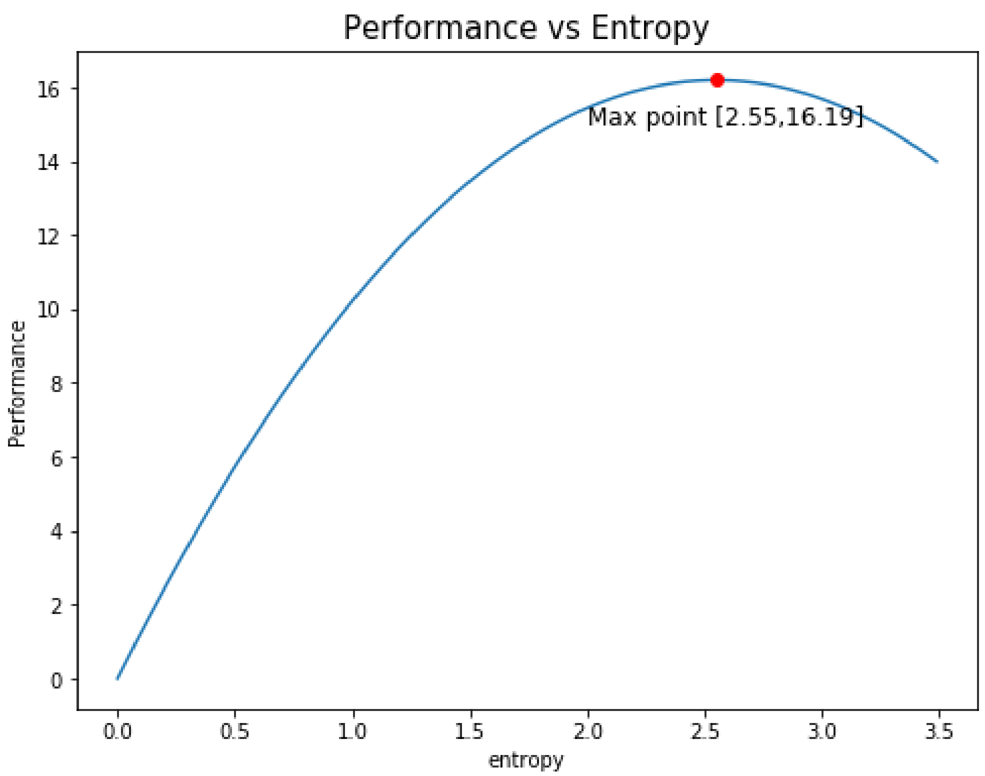

Figure 5. Performance and Entropy Plot

6. Conclusion and Discussion

In this study, we have further explored the applicability of the patent portfolio that has been illuminated by past research, identifying the factors that affect patent portfolio performance and the numerical effects of those factors. The first hypothesis of our three hypotheses did not agree with the analysis, the second hypothesis was found to be correct, and the last hypothesis was found to be partially accepted. Of the three variables we tried to verify, patent scope and portfolio diversification seem to be similar at first glance. However, the patent scope is a variable that captures aspects of the extent to which R&D covers the CPC-based coverage of corporate patent data. The latter, diversification, is a count-based arrangement of CPC data, even if R&D is performed over a wide range, so diversification may be low even with high patent scope, and vice versa.

In the hypothesis proposal stage, we suggested that the degree of accordance with the industry average technology development direction would have a positive impact on the performance of the patent portfolio. It is naïve idea that industry averaging technology development will be in the direction of pursuing market and consumer demand, and processes through patent development and product development through corresponding technology development will have a positive impact on the rise of the patented index.

However, as the analytical model results suggested, this hypothesis showed a degree of correlation in the opposite direction and showed very high confidence (low p-value). As we have already guessed above, since the industry average technology development direction vector is constructed by the portfolio of printer manufacturing companies that have been actively developing technology, this may mean that there is no peculiar competitive direction that distinguishes it from 11 companies.

In order for a real enterprise to expect higher performance improvements in patent portfolio management, it is necessary to focus more on technology development that possesses distinctive competitiveness in the printer manufacturing industry, in addition to the technical features common to other companies. The results of the model analysis on the range of patent technology within the patent portfolio show that the logic of the second hypothesis we proposed is correct. The result is that the more appropriate the combination of the lower technology group represented by the CPC subclasses, the more patents issued by the firm in that year, the better the performance.

It is not important how much technology development is applied by how many sub-technology groups are frequently applied, but if you publish a patent by combining new technologies at least once, it can be said that it gives high performance. In the establishment of technological development strategy of companies, implication implies that a lot of attention should be paid to creative patent issuance by combining various technologies. Especially, in case of companies that do not have such a technical strategy, better performance I can expect a positive message.

The most interesting part of the analysis model results is related to the third hypothesis. Our third hypothesis, which predicted a positive relationship between diversification and portfolio performance, was partially accepted. The second-order parameter of the diversification results in a negative effect on the portfolio performance due to excessive diversification of technology in the portfolio. As a result of the analysis, it can be seen that the direction of increase in performance due to entropy changes toward the decreasing direction with the entropy index of 2.55 as a reference point. Diversification within an excessive technology portfolio affects performance to some extent understandable in the context of actual technology development. It is interesting to note that although the results of this analysis do not fully support our third hypothesis, entrepreneurial enterprises are adversely affected by excessive technological diversification in part by entropy.

As shown in our data samples, 58 of the 192 portfolio samples, except for 134, show an entropy value above 2.55 and are in the middle of a performance down trend due to excessive diversification in the entropy and performance plots. The effect of this excessive technology diversification on the performance side effect is interesting for real companies in establishing technology development strategy. If a company participating in the printer manufacturing industry develops the patents and products required in the industry, it may have a negative performance impact if the effort is too much across a large number of technologies.

When we think about this in combination with the effect of the patent scope variable that we have already demonstrated, it is explained that combining patented technology in the patent publication has a positive effect. Patents that combine new technologies and subclasses that were not previously relevant to the company have a positive impact, but it is best not to create a patent portfolio by working equally among all technology groups, including unfamiliar technologies. Patent scope and diversification These conjectures about the outcome of two variables are very interesting. Companies do not make patents using only the technology groups that they already have.

The absolute number of patents issued by companies each year is increasing exponentially, and in this situation, even a little more creative technology combinations, technologies and patents issued are needed from the sub-technologies which have not been exploited by corporation before. Therefore, in these situations, it is recommended that companies expand their knowledge using technologies that have been carefully approached and utilized when developing technology. Although this researcher is not a study of the expansion of new technology, however, has not been able to study the world's counterparts, the mere implication of the findings is that investments in technology strategies should not be similar to investments in technologies owned by existing companies.

Our research is not yet known whether this is a possible application within the patent portfolio of companies participating in other industries, as they have done research on the performance of the patent portfolio for companies participating in the printer manufacturing industry. Also, our data do not include information on recent patents since 2010 because we had to study patents that were at least five years old since its publication. In addition, since the company does not use financial data at all, but uses patent data only, it may be different from the actual situation if the relationship with financial performance is seen in actual situation.

REFERENCES

Richard A.Miller (1982). The Herfindahl-Hirschman Index as a market structure variable: and exposition for antitrust practitioners, 27 *Antitrust Bull*. 593 (1982)

Anthony F. Breitzman and Mary Ellen Mogee (2002). The many applications of patent analysis. *Journal of Information Science*, 28 (3) 2002, pp. 187–205

H. Ernst. Soll JH. (2003) Integrating Market and Patent Portfolios for Market-Oriented R&D Planning. *Management of Engineering and Technology.*

H. Ernst. (2003) Patent information for strategic technology management. *World Patent Information* 25 (2003) 233–242

Evenson, R. E. (1993). Patents, R&D, and invention potential: International evidence. *The American Economic Review*, 83(2), 463.

Bernd Fabry (2006). Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry. *World Patent Information* 28 (2006) 215–225

Michele Grimaldi (2015). The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning. *Technological Forecasting & Social Change* 94 (2015) 286–302

Anthony Breitzman & Patrick Thomas (2002) Using Patent Citation Analysis to Target/Value M&A Candidates, *Research-Technology Management*, 45:5, 28-36

David Nawrocki & William Carter (2010) Industry competitiveness using Herfindahl and entropy concentration indices with firm market capitalization data, *Applied Economics*, 42:22, 2855-2863

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research policy*, 32(1), 69-87.

Tu¨ rkay Dereli, Alptekin Durmus¸ og˘lu (2009) Classifying technology patents to identify trends: Applying a fuzzy-based clustering approach in the Turkish textile industry*, Technology in Society*, 31 263–272

Joshua Lerner (1994) The importance of patent scope: an empirical analysis., *The Rand Journal of Economics*; Summer, 25, 2

E. J. Kleinschmidt, R. G. Cooper(1991) The Impact of Product Innovativeness on Performance, *Product Innovation Management*, 8 240-251

Cristina Quintana-Garc´ıa and Carlos A. Benavides-Velasco (2008) Innovative competence, exploration and exploitation: The influence of technological diversification, *Research Policy*, y 37, 492–507

A. Pakes, Z. Griliches, PATENTS AND R&D AT THE FIRM LEVEL: A First Report, *Economics Letters* 5, 377-381

ChoTa-Shun Cho and Hsin-Yu Shih (2011) Patent citation network analysis of core and emerging technologies in Taiwan: 1997–2008, *Scientometrics*, 89:795–811

Jean (2004) Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators, *Economic Journal,* Vol.114(495), pp.441-465

Bela Balassa (1977) ‘revealed’ comparative advantage revisited: an analysis of relative export shares of the industrial countries, 1953–1971, *Manchester School*, Vol.45(4), pp.327-344

C.L. Bas, C. Sierra (2002) ‘Location versus home country advantages’ in R&D activities: some further results on multinationals’ locational strategies, *Research Policy* 31 589–609

Patel, P., & Pavitt, K. (1997) The technological competency of the world’s largest firms: Complex path -dependent, but not much variety. *Research Policy*, 26, 141–156.

Granstrand and Oskarsson (1994) Technology Diversification in “MUL-TECH” Corporations, *ieee transactions on engineering management*. vol. 41, no. 4

Harhoff (2006) Citation Frequency and the Value of Patented Inventions, *Review of Economics and Statistics Volume* 81 511-515

Bill McEvily, Akbar Zaheer (1999) Bridging ties: a source of firm heterogeneity in competitive capabilities, *Strategic Management Journal Volume* 20 12 1133-1156

Bernd Fabry, Holger Ernst (2006) Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry, *World Patent Information*, vol.28, Issue 3

Holger Ernst, Nils Omland (2011) The Patent Asset Index – A new approach to benchmark patent portfolios, *World Patent Information*

Michele Grimaldi, Livio Cricelli (2015) The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning, *Technological Forecasting & Social Change*

Shann-Bin Chang (2012) Using patent analysis to establish technological position: Two different strategic approaches, *Technological Forecasting & Social Change*

Yi Zhang, Lining Shang (2016) A hybrid similarity measure method for patent portfolio analysis, *Journal of Informetrics*

Rahul Kapoor, Matti Karvonen (2015), Patent portfolios of European wind industry: New insights using citation categories, *World Patent Information*