An Analysis of Citation Counts of ETRI-Invented US Patents

Yong-Gil Lee, Jeong-Dong Lee, and Yong-Il Song

ABSTRACT—From its foundation until 2004, ETRI has registered over 1,000 US patents. This letter analyzes the characteristics of these patents and addresses the explanatory factors affecting their citation counts. For explanatory variables, research team related variables, invention specific variables, and geographical domain related variables are suggested. Zero-altered count data models are used to test the impact of independent variables. A key finding is that technological cumulativeness, the scale of invention, outputs in the electronic field, and the degree of dependence on the US technology domain positively affect the citation counts of ETRI-invented US patents. The magnitude of international presence appears to negatively affect the citation counts of ETRI-invented US patents.

Keywords—ETRI-invented US patents, citation counts, zero-altered model, technological cumulativeness, US technology domain.

I. Introduction

Since its foundation, ETRI has registered over 1,000 US patents, which is the highest number amongst Korean public research institutes, followed by Korea Institute of Science and Technology (KIST) and Korea Research Institute of Chemical Technology (KRICT).

Prior research shows that highly cited patents are also technologically important [1]. Most patents cited are referenced in patents issued within the same narrowly defined field of innovation as the cited patents. The very existence of those later patents attests to the fact that highly cited patents open up the way to a technologically successful line of innovation [2].

This letter deals with the citation patterns and counts of about nine hundred US patents invented by ETRI and registered from the early 1980s to 2004. The current letter is organized as follows. First, we describe the characteristics of ETRI-invented US patents. Then, we clarify the explanatory variables, mainly extracted from the front page of US patents, and follow this with a short introduction to the patent citation count model we used in this study. We continue with an analysis of the results, with a concluding remark at the end.

II. ETRI-Invented US Patents

1. Statistics Extracted from US Patents

This letter deals with a total of 895 ETRI-invented US patents registered from the early 1980s to 2004. The data for this study is gathered from the front page of US patents. As shown in Table 1, these data fields represent the independent variables of the current study.

Table 1. Data fields extracted from US patents.

Variable name	Measure
Citation	Cited counts
Self-citation	Self-cited counts
INV	No. of inventors
SELF	No. of self-citing counts
COL	Co-assignee
CLA	No. of claims
FAM	No. of family patents
SCI	Scientific linkage, no. of paper citations
USP	No. of US patent citations
JP	No. of Japan patent citations

Manuscript received Feb. 07, 2006; revised May 18, 2006.

Yong-Gil Lee (phone: +82 2 958 6178, email: yonggil@kist.re.kr) and Yong-Il Song (email: yongilsong@kist.re.kr) are with Research Policy Team, KIST, Seoul, Korea.

Jeong-Dong Lee (email: leejd@snu.ac.kr) is with Technoeconomics and Policy Program, SNU, Seoul, Korea.

2. Basic Statistics

In Table 2, basic statistics of ETRI-invented US patents are provided for various technological fields in accordance with international patent classification (IPC). Though US patents of ETRI are mainly from the electronics field, numerous patents are also found to come from the physics field.

Mean citation numbers are highest in electronics, in which ETRI's strategic research field lies. In this field, US patents of ETRI are found to cite more US patents than Japanese patents. This fact will be confirmed in the next section.

Table 2. Basic statistics by technological fields.

Statistics	Physics	Electronics	Others	Total
No. of patents (ratio)	246 (27.5%)	590 (65.9%)	59 (6.6%)	895 (100%)
Mean of citations	3.09	5.15	3.00	4.44
Mean of self-citations	0.07	0.09	0.07	0.08
Mean of INV	3.74	3.66	3.61	3.68
Mean of SELF	0.04	0.03	0.05	0.04
Number of COL	47	149	5	201
Mean of CLA	7.91	8.07	8.07	8.02
Mean of FAM	4.00	4.24	4.24	4.18
Mean of SCI	1.81	1.53	2.37	1.66
Mean of USP	5.74	5.38	5.49	5.49
Mean of JP	0.13	0.29	0.39	0.25

III. Patent Citation Counts Model

1. Explanatory Variables

Explanatory variables, which are expected to affect citation counts, are all extracted from the front pages of US patents. The variables are divided into three categories, including a research team related category, an invention-specific category, and a geographical domain related category, following the classification rule by Lee and others. The specific variables are described in Table 3.

2. Zero-Altered Models

Patent citations show the typical characteristics of count data; therefore, typical count data models such as Poisson and negative binomial can be used. However, the presence of excess zeros in the model due to the lack of citations for many patents, require modification of the original models. According to the studies by Lambert [5] and Heilbron [6], under such circumstances, zero-altered Poisson (ZAP) and zero-altered negative binomial (ZANB) models are preferable to typical count data models.

The estimation of these models starts from the construction of the log-likelihood function. The following four log-likelihood functions represent Poisson, negative binomial (NB), ZAP, and ZANB, respectively, where y_i is the citation count for the i-th patent count:

$$\ln L_P = \sum_{y_i} [-\lambda_i + y_i \ln(\lambda_i) - \ln(y_i!)] ,$$

Table 3. Variables extracted from US patents (slightly different from [3]).

Category	Variables	Initial	Measurement
Research team	Size of research team	INV	Number of inventors
	Technological cumulativeness	SELF	Self-citing counts
	Research collaboration	COL	Single or co-assignee (dummy)
Invention	Scale of invention	CLA	Number of claims
	Magnitude of international presence	FAM	Number of family patents
	Scientific linkage	SCI	Number of non-patent citations
	Physics	PHY	Belonging to physics
	Electronics	ELE	Belonging to electronics
Geographical localization	Degree of dependence on US technology domain	USP	Number of US patent citations
	Degree of dependence on Japan technology domain	JP	Number of Japanese patent citations
	Degree of dependence on other technology domain	OTH	Number of other patent citations
Age	Age of patent	AGE	Years open to the public

Table 4. Test results of model fitness.

Models	Poisson	NB	ZAP	ZANB
Log likelihood	–3591.5	–2116.2	–3273.9	–2097.6
Vuong statistic (ZAP, ZANB)	-	-	6.414**	4.165**
Predicted number of zeros (actual zeros : 233)	25	198	179	208
Dispersion parameter (α)	-	1.088** (0.075)	-	1.029** (0.067)
Zero-altered model (τ)	-	-	–0.977** (0.070)	–5.807** (1.750)

** 1% level significant, ( ): standard error, α =1/*k*

$$\ln L_{NB} = \sum_{y_i} [\ln(1-\phi) + \ln \frac{\Gamma(y_i+k)}{\Gamma(y_i+1)\Gamma(k)} + k\ln(t_i) + y_i \ln(1-t_i)],$$

$$\ln L_{ZAP} = -n\ln(1+e^a) + \sum_{y_i=0} \ln(e^a + e^{-\lambda_i}) + \sum_{y_i>0} [-\lambda_i + y_i \ln(\lambda_i) - \ln(y_i!)],$$

$$\ln L_{ZANB} = \sum_{y_i=0} \ln[\phi + (1-\phi)t_i^k] + \sum_{y_i>0} [\ln(1-\phi) + \ln \frac{\Gamma(y_i+k)}{\Gamma(y_i+1)\Gamma(k)} + k\ln(t_i) + y_i \ln(1-t_i)],$$

where $\ln\lambda_i = \beta' X_i$, $t_i = k/(k+\lambda_i)$, a = logit(ϕ), and ϕ and λ denote respectively the proportion of structural zeros and the mean parameter. More details are specified in [3].

IV. Empirical Result

1. Model Fitness

We first performed various statistical tests for model specification. According to Table 4, the Alpha test verifies that a negative binomial model is more suitable than a Poisson model [4]. With regard to prediction of actual zeros, the fitness of a negative binomial model is ascertained. Meanwhile, the Tau test shows that zero altered models are preferred to the typical count models [5], [6]. Vuong's statistic confirms the result of the Tau test .

2. Estimation Results

In this section, we report and compare the estimation results of the three models, NB, ZAP, and ZANB in Table 5.

As for the research team characteristics, the results show that the patents with more members in the team and deep technological cumulativeness are cited more frequently. The invention-specific characteristics show that, in line with intuition, the scale of the invention positively affects patent citation counts. The patents in the electronics field, in which ETRI's strategic research fields lies, are cited more frequently. Finally, the test results show that the geographical domain related characteristics, such as, the degree of dependence on the US technology domain, positively affect citation counts. This means that the inventiveness in the field of Korean information technology strongly depends on what is going on in the US IT domain.

V. Conclusions

From this study, we first identified the factors affecting citation counts of ETRI-invented US patents. Citation counts of ETRI-invented US patents appear to be positively affected by technological cumulativeness, the scale of invention, outputs in the electronics field, and the degree of dependence upon the US technology domain. In other words, the public IT research of Korea is strengthened by learning and absorbing US information technology and its accumulated knowledge. It is notable that, contrary to intuition, the public IT research of Korea depends more on the US technology domain than that of Japan. It has been assumed that Korea's catching up strategy was mainly based on the imitation of Japanese technology. However, according to this study, at least in the information technology field, Korean research depends more deeply upon US technology.

The magnitude of international presence (number of family patents) seems to negatively affect the citation counts of ETRI-invented US patents. This finding goes against our intuition as well. Since there are a large number of family patents due to the need for wider international protection, we would expect the technological value of the patent to increase with the number of family patents. However, the results show the opposite. It seems that the fact is related to IT characteristics and patent strategies. This fact will be further explored in detail in later studies.

Table 5. Estimation results.

Variables	NB	ZAP	ZANB
	coeff. (std. error)	coeff. (std. error)	coeff. (std. error)
C	–0.708** (0.262)	0.357** (0.049)	–0.747** (0.162)
INV (size of research team)	–0.030 (0.030)	–0.013* (0.006)	–0.017 (0.019)
SELF (technological cumulativeness)	0.331* (0.167)	0.243** (0.019)	0.335** (0.119)
COL (research collaboration)	0.050 (0.097)	0.006 (0.016)	0.036 (0.080)
CLA (size of invention)	0.013* (0.006)	0.010** (0.001)	0.013** (0.005)
FAM (magnitude of international presence)	–0.048** (0.018)	–0.001 (0.003)	–0.018* (0.007)
SCI (scientific linkage)	0.020 (0.020)	0.018** (0.004)	0.013 (0.013)
PHY (physics)	0.153 (0.187)	0.011 (0.037)	0.245* (0.118)
ELE (electronics)	0.443* (0.176)	0.234** (0.032)	0.425** (0.111)
US (degree of dependence on US technology domain)	0.029** (0.011)	0.023** (0.002)	0.025** (0.008)
JP (degree of dependence on Japan technology domain)	0.004 (0.005)	0.002 (0.006)	0.001 (0.025)
OTH (degree of dependence on other technology domain)	0.018 (0.033)	0.052** (0.005)	0.025 (0.024)
AGE (age of patent)	0.290** (0.019)	0.141** (0.003)	0.273** (0.011)

** 1% level significant, * 5% level significant

References

[1] M.P. Carpenter, F. Narin, and P. Woolf, "Citation Rates to Technologically Important Patents," *World Patent Information*, vol. 3, no. 4, 1981, pp. 160–163.

[2] M. Trajtenberg, "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND J. of Economics*, vol. 20, 1990, pp. 172-187.

[3] Yong-Gil Lee, Jeong-Dong Lee, Yong-Il Song, and Se-Jun Lee, "An In-Depth Empirical Analysis of Patent Citation Counts Using Zero-Inflated Count Data Model," to be published in *Scientometrics*.

[4] A. Cameron and P. Trivedi, "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators," *J. of Applied Econometrics*, vol. 1, no. 1, 1986, pp.1-46.

[5] D. Lambert, "Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing," *Technometrics*, vol. 34, 1992, pp.1-14.

[6] D. Heilbron, "Zero-Altered and Other Regression Models for Count Data with Added Zeros," *Biometrical J.*, vol. 36, 1994, pp. 531-547.