Original Research Paper

Political Science

ANALYSIS OF FACTORS ASSOCIATED WITH PATENT APPLICATIONS (2000 - 2010): INDICATORS OF PATENTABILITY SUCCESS

Pablo Álvarez*	Bio-Health Research Institute ibs. GRANADA, Foundation for Bio-Health Research in Eastern Andalusia: FIBAO. Granada *Corresponding Author
Arturo Argüello	Office of Technological Transference of the Public Health System of Andalusia. Public Foundation of Andalusia Progress and Health. Seville
Julia Sánchez Cantalejo	Andalusian School of Public Health. Granada.

ABSTRACT We developed a multivariate linear regression model to analyze patent applications in European countries in order to identify patentability success indicators.

Information was gathered from the Eurostat and World Intellectual Property Indicators databases (period 2000-2010). Three regression models were constructed using as response variables: the total number of applications for PCT patents in the national phase (M1), patent applications to EPO per year at national level (M2), and patent applications to EPO per year and economic activity (M3), and considering 10 variables related to R&D funding and research personnel as predictor variables

Multivariate linear regression models were estimated using the Bayesian Information Criterion (BIC).The most influential predictive variables were: the total number of R&D research personnel by sector, qualification, and sex (determination coefficient of 44.9 %) in the M1 model; the total R&D research personnel by sector, qualification, and sex as well as the business R&D expenditure by economic activity in both the M2 model (coefficient of 79.2%) and M3(coefficient of 78.8%). In conclusion, the mathematical models show that the predictors with greatest effect on patentability are qualified R&D personnel and business R&D funds, and they reveal the distribution of European countries as a function of these variables.

KEYWORDS : Patents, Intellectual Property, Innovation and dissemination of Knowledge

Introduction

Increasing interest in the identification of indicators of R&D performance has focused attention on patent data [1] [2] , considering the protection of scientific production by patents as a marker of professional and institutional success and as a means of evaluating the impact of innovations [3].

There has been a strategic shift towards knowledge-based industrial innovation over the past two decades and various researchers have analyzed the value of patents in relation to the innovative capacity of institutions [4] (2). Analysis of patent data is useful to determine the factors that favor patentability [5] and thereby promote the transformation of scientific achievements into instruments of socioeconomic revitalization [6].

Increasing interest in the identification of indicators of R&D performance has focused attention on patent data [3] [4] [5] [6], considering the protection of scientific production by patents as a marker of professional and institutional success and as a means of evaluating the impact of innovations [7]

Analysis of patent data is useful to determine the factors that favor patentability [8] [9] and thereby promote the transformation of scientific achievements into instruments of socioeconomic revitalization [6].

The correct interpretation of patent data requires statistical studies in which information from different countries are normalized, establishing effective and reliable models for their meaningful analysis [10] and for the evaluation of technological excellence [6], identifying in a precise manner the factors that influence the effectiveness of patent application and differences among countries.

The objective of this study was to use multivariate regression models to identify variables associated with the increase in patent applications in European countries and to identify the factors related to this increase. The most influential factors were those related to investment in high-quality specialist R&D personnel and to business expenditure on R&D

Material and Methods

Search strategy

A search was conducted of patents registered in the Eurostat and World Intellectual Property Indicators (WIPO) Statistics databases from 2000 through 2010.

Criteria for variable selection

Response and explicative variables were selected for predictive models of patent application activity following previously reported criteria (2). Response variables included those previously associated with patentability and other influential factors, including the number and trend of citations in the literature.

The following patent application outcome (response) variables were selected:

1.- *tot_cou_pct*: Total count by patent cooperation treaty (PCT) application in national phase

2.- *pat_ep_nipc:* Patent applications to the European patent office (EPO) by priority year (year of application) at national level by IPC sections and classes.

3.- *pat_ep_nnac:* Patent applications to the EPO by year and economic activity.

The following ten predictive or independent variables were selected:

1.- *rd_pers_qual*: Total R&D personnel and researchers by performance sector, sex, and qualification

2.- *rd_p_pers_sci*: Total R&D personnel and researchers by performance sector, sex, and scientific field

3.- *rd_e_berdfundr2* Business enterprise R&D expenditure (BERD) by economic performance and source of funds

4.- *rd_e_gerdact*: Total gross domestic expenditure on R&D by performance sector and type of R&D activity

5.- *humresour_rd_tot:* Human resources in science and technology

6.- *gr_do_rd_soufun:* Gross domestic expenditure on R&D by source of funds

7.- *tot_rd_exp:* Total R&D expenditure as % of gross domestic product (GDP)

8.- *rd_exp_sec_higed:* R&D expenditure by performance sector

9.- *bus_ent_rd_exp:* Business enterprise R&D expenditure by performance sector

10.- *hieduc_exp_rd_sec-ejec*: Higher education sector – R&D expenditure by performance sector

This set of variables was constructed by calculating their mean value during the study period for each country. Descriptive analysis was performed, calculating means, medians, percentiles, and standard deviations. The effect of each predictive variable on each response variable was determined by multivariate linear regression analysis, obtaining the most parsimonious model in accordance with the Bayesian Information Criterion (BIC). A diagnostic study was then performed to evaluate the linearity of predictor variables, using graphic procedures to smooth clouds of residual points in the adjusted model. The variance inflation factor (VIF) was calculated to study interdependence among predictors, defining multicollinearity as a score above 10 points. This was followed by evaluation of the presence of outliers, defined as a studentized residual above 3. Finally, the Cook distance was calculated to detect any influential observations, and models were constructed with and without these observations when found. Once the most parsimonious model was obtained, its explanatory capacity was measured by establishing the coefficient of determination. Software R version 3.1.1 was used for data analysis.

Results and Discussion

Descriptive analysis of variables

Table 1 exhibits the descriptive results, showing the right-skewed distribution of the three response variables and the predictive variables *rd_p_persqual* and *rd_p_perssci*, explained by the extreme values for some countries. Thus, half of the countries had a value of < 173 for the *pat_ep_nnac* response, and three-quarters of them had a value <1963, whereas values > 5,000 were found for the United Kingdom, France, and Germany, with a marked effect on the mean value. Only a slight skewness to the right was observed for the predictive variables *rd_e_berdfundr2*, *rd_e_gerdact*, *gr_do_rd_soufun*, *tot_rd_exp*, *rd_exp_sec_higed*, *bus_ent_rd_exp*, while *hieduc_exp_rd_sec-ejec* and *humresour_rd_tot* were virtually symmetrical, with a mean value of 38.42 and median of 38.38.

Table 1.- Description of response and explicative variables

	1st Quartile	Median	3rd Quartile	Mean	Standard dev.
Response Variables					
tot_cou_pct	41.53	235.20	1681.00	5581.00	14350.07
pat_ep_nipc	29.20	173.00	1963.00	3085.00	6898.24
pat_ep_nnac	28.08	170.30	1764.00	2975.00	6655.73
Explicative Variables					
rd_p_persqual	18290.00	72520.00	121900.00	249400.00	635667.00
rd_p_perssci	18290.00	72520.00	121900.00	249400.00	635667.00
rd_e_berdfundr2	26.71	129.40	446.70	267.50	280.72
rd_e_gerdact	79.60	296.30	868.60	473.30	436.10
humresour_rd_tot	31.61	38.38	45.83	38.42	8.61
gr_do_rd_soufun	39.66	46.22	60.46	48.32	14.45
tot_rd_exp	0.71	1.28	2.17	1.50	0.87
rd_exp_sec_higed	0.23	0.36	0.47	0.38	0.20
bus_ent_rd_exp	0.30	0.88	1.44	0.99	0.76
hieduc_exp_rd_sec-ejec	0.23	0.36	0.47	0.38	0.20

Regression models

Three regression models were obtained (Table 2), the first (M1) with response variable *tot_cou_pct*, the second (M2) with *pat_ep_nipc*, and the third (M3) with *pat_ep_nnac*. No multicollinearity was detected among predictive variables. Residual analysis identified Norway as an outlier in the M1 model and Germany in M2 and M3 models, and these countries were therefore eliminated from the respective models. The optimal model for the three response variables included the predictive variables *rd_p_persqual* and *rd_e_berdfundr2*. In both M2 and M3, the response variable was increased ($p<0.01$) by each increment in total R&D research personnel by sector, qualification, and sex (*rd_p_persqual*) and by each increment in R&D business enterprise expenditure by economic performance and source of funds (*rd_e_berdfundr2*). These two models also had high adjusted coefficients of determination (79.2 and 78.8 %, respectively). In the case of M1, the response variable was increased ($p<0.05$) by each increment in total R&D research personnel by sector, qualification, and sex (*rd_p_persqual*) and the determination coefficient was 44.9 %. Table 2 exhibits the estimations corresponding to the optimal model selected for each response variable.

Table 2.- Regression models for response variables

	M1 R2=44.9 %			M 2 R2=79.2 %			M3 R2=78.9 %		
	B	SE	p	B	SE	p	B	SE	P
Constant	183.4	139.40	0.203	500.900	230.6	0.39	-4.77	224.9	0.042
rd_pers_qual	2.923	6.716	0.000	1.210	1.188	6.30	1.169	1.158	7.840
rd_e_berdfundr2	0.003	0.001	0.000	1.693	5.608	50.0	1.611	5.469	0.006

R^2: adjusted coefficient of determination

B: variable coefficient

SE: standard error

Distribution by country of the predictive variables influencing the regression models

According to mean values of the predictive variable *rd_pers_qual*, which significantly influenced all three mathematical models, the five top positions in Europe are held by Germany, the UK, France, Spain, and Italy, with the highest mean value (20 % of the total) for Germany, explaining its exclusion from M2, followed by 14% for the UK, 13% for France, 13% for Spain, and 9% for Italy. Mean values of around 2-3% were obtained for Sweden, Switzerland, Austria, Belgium, and Finland, with lower values for the remaining countries.

According to mean values of the predictive variable *rd_e_berdfundr2*, which also significantly influenced M2 and M3, the top countries in business R&D expenditure by economic activity and source of funds were Switzerland, Finland, Sweden, Denmark, and Luxembourg, with mean values ranging between 11 and 9 % of the total, followed by Austria, Belgium, France, Ireland, Holland, the UK, Italy, and Spain, with means ranging between 5 and 2%, while values in the remaining countries ranged between 1.9 and 0.4%.

According to the present results, the total number of personnel involved in R&D, including laboratory technicians and specialized and qualified researchers, predicts patent success in terms of the following indicators: total number of PCT patent applications, patent applications to the EPO by year of priority at national level by section and class, and EPO patent applications by year and economic activity (Table 2). These findings underscore the major importance of highly qualified human capital for the production of science and the development of patents for scientific achievements. This predictive variable (rd-p-per.qual) also indirectly includes the number of teaching staff and doctoral researchers, in agreement with previously reports33, confirming the interdependence between research quality and patentability. Scientific fields have a specific role in patent development, but only in relation to the scientific and technical quality of researchers or in relation to scientific activity and sex. Besides this predictive variable, response variables were also predicted by the variable related to increased business R&D funds by economic activity and source of funds. Differences were observed in the relative impact of the different factors found to influence patentability, which included research group size, research funding/development, business R&D expenditure, level of higher education, and sector of performance. Thus, the predictor variable rd_p_pers_qual was found to be significant in all three multivariate linear regression models constructed around patent applications, and the variable rd_e_berdfundr2 was also significant in the M2 and M3 models, with these two variables explaining 79.2 % and 78.9 % of their variability, respectively.

A direct relationship with funding was also observed, in agreement with previous reports [4].Among the European countries in our study, the highest mean values for the predictive variable (rd_pers_qual) were obtained for the UK, France, Spain, and Italy, while the greatest synchrony between the predictive variables influencing the M2 and M3 models (rd.p.per-qual and rdeberdfundr2) was observed for Sweden, Switzerland, Austria, and France.

Conclusions

The mathematical models developed reveal that patentabity increase is most influenced by the number of qualified R&D personnel and the amount of business R&D funds and show the distribution of different European countries as a function of these variables. These data confirm the importance of highly qualified and specialized research personnel for success in obtaining patents for scientific achievements. This new approach to predictive analysis provides reliable and robust support to policy makers in the worlds of business, science, technology, and education and offers a basis for the development of guidelines to promote patentability.

Funding

R&D contract with Saepro S.L.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper

Acknowledgments

The authors thank M Olvera and Marta Reyes for critical review and intellectual contributions.

References

(1) Reitzig, M., & Puranam, P. (2009). Value appropriation as an organizational capability: The case of IP protection through patents. Strategic Management Journal, 30(7), 765-789.
(2) Moser, P. (2016). Patents and innovation in economic history. Annual Review of Economics, 8, 241-258.
(3) Alvarez, P., Arguello, A., E Aranega, A., Vélez, C., Aránega, A., González, C., & C Prados-Salazar, J. (2015). Cancer Stem-Cells Patents in the Context of their Therapeutic Purposes: Exploring the Latest Trends (2011-2015). Recent Patents on Regenerative Medicine, 5(1), 55-64.
(4) Hottenrott, H., Hall, B. H., & Czarnitzki, D. (2016). Patents as quality signals? The implications for financing constraints on R&D. Economics of Innovation and New Technology, 25(3), 197-217.
(5) Zhou, H., Sandner, P. G., Martinelli, S. L., & Block, J. H. (2016). Patents, trademarks, and their complementarity in venture capital funding. Technovation, 47, 14-22.
(6) Tetko, I. V., Lowe, D. M., & Williams, A. J. (2016). The development of models to predict melting and pyrolysis point data associated with several hundred thousand compounds mined from PATENTS. Journal of Cheminformatics, 8(1), 2-18.
(7) Álvarez, P., Marchal, J. A., Boulaiz, H., Carrillo, E., Vélez, C., Rodríguez-Serrano, F, & Aranega, A. (2012). 5-Fluorouracil derivatives: a patent review. Expert opinion on therapeutic patents, 22(2), 107-123.
(8) Kyebambe, M. N., Cheng, G., Huang, Y., He, C., & Zhang, Z. (2017). Forecasting emerging technologies: A supervised learning approach through patent analysis. Technological Forecasting and Social Change, 125, 236-244.
(9) Li, W. Q., Li, Y., Chen, J., & Hou, C. Y. (2017). Product functional information based automatic patent classification: Method and experimental studies. Information Systems, 67, 71-82.
(10) Miyata, Y. (2000). An empirical analysis of innovative activity of universities in the United States. Technovation, 20(8), 413-425.