

Big data analytics on patents for innovation public policies

Maria José Sousa¹  | George Jamil²  | Cicero Eduardo Walter^{3,4}  |
Manuel Au-Yong-Oliveira⁵  | Fernando Moreira⁶ 

¹Business Research Unit, Instituto
Universitário de Lisboa, Lisboa, Portugal

²Informações em Rede Consultoria e
Treinamento LTDA, Belo Horizonte, Brazil

³Federal Institute of Education, Science and
Technology of Piauí, Teresina, Brazil

⁴GOVCOPP, Department of Economics,
Management, Industrial Engineering and
Tourism, University of Aveiro, Aveiro, Portugal

⁵INESC TEC, GOVCOPP, Department of
Economics, Management, Industrial
Engineering and Tourism, University of Aveiro,
Aveiro, Portugal

⁶REMIT, IJP, DCT, Universidade Portucalense,
Portugal & IEETA, University of Aveiro, Aveiro,
Portugal

Correspondence

Fernando Moreira, REMIT, IJP, DCT,
Universidade Portucalense, Portugal & IEETA,
University of Aveiro, Aveiro, Portugal.
Email: fmoreira@uportu.pt

Abstract

This study seeks to answer the following research question: “What factors can explain the number of patent filing requests made by residents in Brazil at patent offices in Brazil, the United States, Europe, and triadic patent families?”. The methods used in this research are quantitative, using big data from private and public investments in Science and Technology, and about patent deposit numbers in Brazil from 2000 to 2017. A model of linear regression was performed and explains how these investments in Science and Technology influence patent deposit numbers. The results of this research study point towards the importance of universities, up and beyond the traditional training and education aspect of university activity. The importance of public and private innovation investments is also shown to be important. This study shows that the patent registrations in the different regions under analysis are affected by different factors. There is thus no single formula towards the creation of innovation output and governments would do well to continue to invest in higher education while also investing in public research and development activities. Additionally, and not least important, private entities should be continually encouraged to make innovation investments and favourable government policies need to thus exist for this to happen. Finally, the low numbers regarding patent filings in Brazil may be linked to institutional deficiencies in the country. Patent breaches may be difficult to punish, and the judicial system may be slow and untrustworthy, compared to the United States and to Europe—leading to diminished patent registrations in Brazil. A set of implications and recommendations for policy derived from this study and will be strategic for policymakers.

KEYWORDS

analytics, innovation output, patents, public policies, residents in Brazil

1 | INTRODUCTION

This exploratory study aims to analyse patent filings in Brazil, the United States and in Europe, by residents in Brazil. The Research Question (RQ) under analysis is “What factors can explain the number of patent filing requests made by residents in Brazil at patent offices in Brazil, the United States, Europe, and triadic patent families?”

Innovation in public policy has an essential role in influencing the competitive capacity of companies and is strongly associated with their ability to innovate and the way they are organized. The concept of innovation can be translated as “the successful production, assimilation, and exploitation of novelty,” according to the Green Paper on Innovation from the European Commission (1996) and according to the Oslo Manual (OECD, 2005). The concept is then structured around three pillars: the renovation and enlargement of the range of products and services and the

associated markets; the creation of new methods of production, supply and distribution; and the introduction of changes in management, work organization, and skills of the workforce—organizational innovation. All of those types of innovation can be translated into patents, contributing to create added value and differentiation for companies. Furthermore, the political, economic and social context that has been experienced in recent years led to a commitment to innovation in a concerted and integrated way—not only technologically, but, also, at the organizational level. In this context, Public Policies can play an important role by promoting programs that contribute to improving the way companies invest in their capacity for innovation and become more open to patent their products and services, increasing their competitive capacity.

The innovativeness of this research is based on the capture in a single analysis of the complex nature of the relationships between patents, innovation, information dissemination, and competitiveness. The study also focused on the analysis of several patent systems with specificities which affect the innovation policies and innovation processes in organizations. This article bring light to the factors that can explain the number of patent filing requests made by residents in Brazil at patent offices in Brazil, the United States, Europe, and triadic patent families.

The joint analysis of the constructs regarding patents, the big data available in the government systems regarding the requests by the inventors, and the public policies role in the process to facilitate and increment innovation is also an innovative approach of this research, with the goal to create new knowledge for new innovation policy development. These constructs will be defined and analysed in the following sections of the article, which proceeds with a theoretical section (involving big data, public policy and patents), followed by a methodology section. Conclusions and implications are then discussed, as are research limitations and suggestions for future research.

2 | BIG DATA IN INNOVATION IN PUBLIC POLICIES—A THEORETICAL FRAMEWORK

2.1 | Big data analytics conceptualization

Big Data analytics integrates several methodologies to accurately anticipate and make predictions in the context of complex organizational scenarios. Big data is becoming increasingly valuable to research and innovation. Ultimately, big data analytics may help any industry to find out the blind spots in an organization, which departments are doing better and to use the strategy in other departments for improvement.

Big data has been a critical factor to policy, not only in terms of developing new innovation policies, but also in providing structure and analytics for decision making. The implementation of predictive analytic methods for identifying potential innovation program improvements for the challenges that are faced by society, organizations, and citizens is a reality.

Analytics measures organizational results and outcomes, using different methods: descriptive and inductive statistics, and estimations processes to help in defining the future strategy of the innovation policy. Those measurements provide data, indicators, and results which can be analysed by and used to help influence the new policy agenda.

2.2 | Big data analytics for innovation in public policies regarding patent analysis

Establishing new methods and decision-making supporting contexts in the public sector is an actual challenge. As ideological and political matters continue to evolve, with the turbulent emergence of trends like nationalism and protectionism, new problems are to be assessed and addressed, to guarantee the necessary development, evolution, and related strategies, such as education and innovation investments.

Socialist innovation strategies and public policy may aim at curbing financial gains by private enterprise with patents, while on the other hand making more solid investments in education, which is without question a solid bet, as history has shown us. More liberal public policy may want to incentivate private enterprise, and the rents gained from patents, thus encouraging entrepreneurial risk-taking, (Acemoglu and Robinson, 2013), while perhaps limiting public funds for education and relying, once again, on private enterprise for investments in the education sector. We are yet to encounter the perfect mix, thus the need for this and other studies.

Considering specifically innovation, public efforts must be structured towards stable decisions which can foster new social and economic developments (Schumpeter, 1934, 1942). In a paradoxical aspect, innovation strategies can also be viewed by centralized governments and power structures, as a means to defend these protectionist ideologies, aiming to guarantee an eventual short and medium-term international advantage, with a restricted sociological vision (Akcigit, Ates and Impullitti, 2018). The social-political approach is not the objective of our present text, but this limited observation just produces a notion on how the innovation context plays a significant role, even if we take into consideration this actual scenario.

Studying the public sector's role in this strategic game, we can search the detail of emerging technology application by public agents. As discussed in Tadeu et al. (2019), digital transformation maturity emerges as an issue for emerging technologies adoption, provoking the participation of public agents to sponsor infrastructures and the formation of ecosystems which would propel innovation strategies by public and private systems.

According to the observation conducted in Jamil, Rocha and Jamil (2019), when studying the market intelligence process, emerging technologies such as analytics and associated big data methods can be applied to produce better scenario analysis and thus leading to a consequent

improvement in decisions around innovation efforts. In the same publication presented a research study focusing on a similar observation, specifically addressing the healthcare sector and its related productive and value chains.

Finalizing this first level of perception, analytics and big data are concepts widely discussed these days, as their application reached our daily lives, with resources implemented even in simple “apps” designed to be used in a standalone fashion or as a part of a complete information system, through our smartphones or domestic appliances, not needing dedicated computing services to serve a final customer (SAS, 2020; McKinsey, 2020).

As conceptualized by these sources, Big data can be understood as the conjunction of methods, designed processes, mathematical methods, and associated information technologies to produce knowledge from data, collected from an organizational environment (SAS, 2020). Analytics, in turn, is the infrastructure to collect data and optimally analyse trends and patterns while also aiming to produce knowledge-related results, although in a dynamic way, as it is usually associated to programs, modules and applications which are used as background processes in massively used websites, such as electronic commerce, digital auctions or entertainment.

Both services—big data and analytics—will be objectively studied in the text ahead, as an associated set of governmental tools, to produce knowledge for public management decisions.

2.3 | Patent analysis processes

As noticed, innovative market offers can be considered in several different political instances, as observed in the global scenario nowadays, ranging from open democracies to nationalism-supported trends, exerted by some countries, including powerful economies. Hence, it is opportune to address patent analysis processes, mainly to understand how big data and analytics can be applied in these processes for national strategic plans (Patentanalysis, 2020).

Patent analysis processes are actions and associated methods applied to analyse documents, experimental results, technological artefacts, and other materials submitted by interested researchers and entrepreneurs for a patent concession and registration (WIPO, 2016; Patentanalysis, 2020). The patent concession process develops around the analysis of submitted artefacts to defer if the patent is to be effectively registered as a temporary property of the person or entity who is requiring it.

As a complex process, which must identify aspects such as functionality, design, environmental impacts, technical regulations, ethical attributes, pre-regulation demands, health restrictions, among several other critical parameters, patent analysis processes are massively supported by data of different types.

Along with specific taxonomies, this data must also be detailed concerning collection, storage, sharing and usage processes, which must be evaluated by registration institutions, to confirm that there is no risk of falsification or any other kind of misinformation.

2.4 | Data analysis in patent registering processes

Aligning both concepts and practical aspects of the patent analysis and registration processes, it is possible to identify several practices and actions which must be adopted by the registration institutions. It is necessary to also pay attention to the expressive differences among patent processes, regarding different market sectors, albeit recognizing the need for such a differentiation between certain sectors, according to certain authors (Ponta et al., 2020). Additionally, significant differences exist between the quality and power of institutions, in different regions. The above meaning that the approach to innovation public policy needs to be individualized and done on a case-by-case basis—there is no ready-made formula for success; except that perhaps the elite group in power must not solely seek gains for its members (Acemoglu and Robinson, 2013)—for which certain mechanisms must exist, for obvious reasons. One cannot thus simply export innovation public policy across boundaries (North, 2005), due in addition to significant cultural differences—reflected in the institutions and practices in a given region.

The emphasis of innovation public policy needs to be on “making markets more efficient in developed countries” (North, 2005, p.21). Big data has a major role to play.

For example, if we consider patent registration for an electrical application, such as a specific IoT supported, semi-autonomous, domestic outlet, typical data presented by manufacturers must identify quantitative parameters, regarding electrical insulation, protection, performance and, overall, match with specified announced product features such as commands, standardized answers to user requirements, voltage and current measurements, etc. For a medicine drug, however, it is a completely different scenario.

In this last case, it is possible to identify several international requirements, which include, for example, datasets collected from non-human and human patient tests, collateral effects, adherence to safety levels and definitions, packing and conservation specifications among others. Along with quantitative data, it is possible to find diagrams, pictures, text, images and various formats which will demand a heavy overload of analysts' work to determine if all legal demands were fulfilled by patentees.

Thus, upon reflecting, we can consider how analytics and big data services can be applied to improve this complex evaluation process.

For example, applying analytics-type methods, it is possible to address the presented data to examine if these datasets are comparable to those published in scientific studies, institutional requirements, and industrial specifications. Using data modelling techniques, for instance, it is

possible to study if those values are comparable and compatible to patent concession and registering demands, advancing the comprehension towards a potential adherence for the item under evaluation.

Analytics will also allow a performance improvement to analysts, as a reliable way to compare results, especially quantitative data, such as seasonality datasets, levels of patient quantitative reactions, measurable collateral effects, among several other signals and associated phenomena.

Regarding Big Data, it is possible to understand its application when approaching different formats, sizes, and aspects of datasets. As a combination of structured and unstructured sources, it is possible to expect that contents so different such as photos, images, videos and audios (mostly unstructured) can be prospected along with spreadsheets, numerical level demonstrations, measurements and other formatted standards, producing the required information to assess a safe analysis process to grant the intended patent.

Those two data analysis resources can also be designed to operate in conjunction, encompassing data collection, first level analysis, trends in comparative studies and uniformity—so as to be addressed by analytics tools—and for the improvement of in-depth cause-consequence relations, context analysis, adherence to regulatory specifications and analysis and, lastly, aggregation and validation of research field analysis—as to be conducted through Big Data processing. Potential results could serve for collaborative validation, eventually applying modelling techniques and rigorous process studies, bringing a qualified level of perception to analysts, while also making for a more robust patent registering operation.

Interestingly, after a more mature level of combination of these two methods is reached, it is possible also to expect potential automation, so as to implement resources namely machine learning and other artificial intelligence tools and associated methods, presenting a perspective of performance enhancement, without any level of quality degradation, regarding the analysis process. This could lead also to a better standardization tendency, a fundamental aspect sought to integrate governance, transparency and verifiability to these critical steps in order to grant patents to products and other artefacts.

As a sensitive and difficult process, with significant repercussions on market developments, social impacts and industrial relationships, patent analysis can be benefited by an association to data science tools, such as those studied herein. Albeit, the high cost of the software and related tools for big data analysis means that such an analysis will not, unfortunately, be readily available to all. On the other hand, policy makers must be aware that when “institutional changes [such as those involving patent laws] are applied to third world economies they frequently alter income distribution and produce political instability, sometimes leading to downstream consequences that are the very reverse of the intended objective” (North, 2005, p.21). Therefore, innovation public policy makers face a serious challenge, in view of the novel science tools at their disposal.

3 | METHODOLOGY

The application of research instruments such as data mining for the analysis of a vast set of data on patents has been widely used to support the development of the national and business planning processes for Research and Development (Seo et al., 2016), bearing in mind that patent data is a valuable resource for understanding the dynamics and activities of an invention ecosystem (Saheb & Saheb, 2020). Within this perspective, in the same way that data-generating patents are capable of generating a vast set of data that provide information about a particular invention that can later assist in improving the invention itself (Sideri, 2020; Simon & Sichelman, 2017), the identification of which variables affect the number of patent filing requests within a vast data set, contribute to the prioritization and optimization of resources that will lead to an increase in the number of patent applications and, consequently, to a greater dynamism in the invention ecosystem.

In this sense, the number of patents has been used as an indicator for assessing the innovation capacity of companies in a given country or region, as well as their levels of efficiency in terms of production and technology (Fujimoto et al., 2015; Manzini & Lazzarotti, 2016; Stern et al., 2000), using, among other factors, scientific actors such as government agencies and universities as external proxies, in complement to the internal knowledge sources of private organizations (Ponta et al., 2020; Romijn & Albaladejo, 2002). In addition, some investigations (Archibugi, 1992; Comai, 2018) have recommended the use of other information such as scientific production, since data on patents in isolation do not adequately represent the ecosystem of the invention. Accordingly, in the present investigation, it seeks to assess the influence exerted by scientific, governmental and private actors on the number of patent filings.

The methodological approach is quantitative based on the analysis of the patent requests made by residents in Brazil, the United States, Europe, and triadic patent families in patent families. All of the analysis was done with the aid of R Studio in order to respond to the research question: What factors can explain the number of patent filing requests made by residents in Brazil at patent offices in Brazil, the United States, Europe, and triadic patent families?

3.1 | Descriptive and exploratory analysis

The data used in the present investigation were obtained through the 2019 National Science and Technology and Innovation Indicators, made available by the Ministry of Science, Technology, and Innovations of Brazil (available at the electronic address: http://www.mctic.gov.br/mctic/opencms/indicadores/indicadores_cti.html).

The period considered for the present analysis was 2000–2017, operationalized through the following variables:

Y = Number of patent filing requests made at the National Institute of Industrial Property (INPI-Brazil);

Y_2 = Number of patent filing requests made at the American Patent and Trademark Office (USPTO-United States);

Y_3 = Number of patent filings filed with the European Patent Office (EPO-EU);

Y_4 = Number of patent filings filed for triadic patents (European Patent Office; Japanese Patent Office and American Patent and Trademark Office);

X = Public Investments in Science and Technology understood as investments in Research and Development and Related Scientific Activities;

X_2 = Private Investments in Science and Technology understood as investments in Research and Development and Related Scientific Activities;

X_3 = Brazilian scientific production understood as articles indexed in Scopus;

X_4 = Government researchers and research support staff;

X = Private researchers and research support staff;

X = University researchers and research support staff.

As previously mentioned, the present investigation focuses on the analysis period between 2000 and 2017. Table 1 presents a summary of the main descriptive statistics for the independent and dependent variables.

An analysis of Table 1 reveals that in the period considered, patent filing requests made by residents in Brazil have been higher in the European office ($\mu = 4.351$ filing requests, with a variation coefficient of 49.5%), due to requests related to patent families ($\mu = 678.9$ filing requests, with a variation coefficient of 17.7%) and at the American Patent and Trademark Office ($\mu = 410$ filing requests, with a variation coefficient of 43.7%). On the other hand, patent filing applications at the Brazilian patent office presented the lowest average ($\mu = 7.42$ filing requests, with a variation coefficient of 5.6%). Except for orders placed at the Brazilian office, the coefficients of variation for orders from other offices reveal that there was a wide dispersion in the number of orders over the period under review. Additionally, it is possible to notice that public investments in Science and Technology have been higher ($\mu = R\$ 21.93$ million, with a variation coefficient of 58.2%) although closely monitored by private investments ($\mu = R\$ 19.95$ million, with a variation coefficient of 53.2%). Regarding Brazilian scientific production, understood as articles published and indexed in the Scopus Database, the average was 35.4 articles, with a high coefficient of variation of 44.5%. Regarding the total number of researchers, the analysis reveals that there is a predominance of researchers linked to Universities ($\mu = 254$, with a variation coefficient of 43.7%). Nevertheless, with the aid of the Cox-Stuart (Cox & Stuart, 1955) and Mann-Kendall (Mann, 1945) tests, it is possible to note that, except for applications for patent families, the number of applications for deposits in offices in Brazil, the United States, and Europe has assumed a trend over the period considered, assuming a significance level of 5%, according to the results shown in Table 2.

3.2 | Ordinary least square regression

To identify which variables are significant to explain the number of patent filing requests, a regression analysis based on the ordinary least squares method was applied, considering that it is an adequate statistical technique to predict a dependent variable from the knowledge of one or more

TABLE 1 Descriptive summary of dependent and independent variables in the period 2000–2017

Variables	Mean	Coefficient of variation	Median	Min	Max	Range
(Y) Number of patent applications (INPI)	7.42	5.6%	7.46	6.45	7.97	1.52
(Y_2) Number of patent applications (United States)	410.5	43.7%	358	219	769	550
(Y_3) Number of patent applications (EU) ^a	4.351	49.5%	4.910	332	7.348	7.016
(Y_4) Number of patent applications (Triadic)	678.9	17.7%	684	407	875	468
(X) Public Investments ^b	21.93	58.2%	17.76	8.65	47.90	39.26
(X_2) Private Investments ^b	19.95	53.2%	16.24	7.19	37.74	30.55
(X_3) Brazilian scientific production	35.4	44.5%	34.27	15.26	60.71	45.45
(X_4) Government researchers	8.24	17.3%	8.47	5.84	9.89	4.05
(X) Private researchers	64.43	27.9%	59	46.22	99.7	53.5
(X) University researchers	254	43.7%	238.2	113	444	330.7

^aData refer to the period 2000–2013.

^bAmounts in millions of Reais.

Dependent variables	Cox-Stuart test	Mann-Kendall test
(Y) Number of patent applications (INPI)	<i>p</i> -value: .004	<i>p</i> -value: .0006
(Y ₂) Number of patent applications (United States)	<i>p</i> -value: .004	<i>p</i> -value < .01
(Y ₃) Number of patent applications (EU) ^a	<i>p</i> -value < .01	<i>p</i> -value < .01
(Y ₄) Number of patent applications (Triadic)	<i>p</i> -value: 1	<i>p</i> -value: .733

^aData refer to the period 2000–2013.

TABLE 2 Summary of trend tests for dependent variables in the period 2000–2017

TABLE 3 Pearson's correlation matrix between dependent and independent variables

Variables	(Y)	(Y ₂)	(Y ₃)	(Y ₄)	(X)	(X ₂)	(X ₃)	(X ₄)	(X)	(X)
(Y) (INPI)	1									
(Y ₂) (United States)	0.69	1								
(Y ₃) (EU)	0.57	0.87	1							
(Y ₄) (Triadic)	0.60	0.36	0.52	1						
(X) Public Inv.	0.68	0.99	0.87	0.36	1					
(X ₂) Private Inv	0.72	0.99	0.89	0.45	0.98	1				
(X ₃) Scientific production	0.72	0.98	0.92	0.50	0.97	0.98	1			
(X ₄) Gov. Researchers	0.67	0.90	0.87	0.57	0.90	0.93	0.95	1		
(X) Priv. Researchers	0.62	0.93	0.79	0.32	0.94	0.93	0.91	0.84	1	
(X) Univ. Researchers	0.73	0.98	0.89	0.48	0.98	0.99	0.99	0.95	0.93	1

TABLE 4 Bivariate regression models for the number of patent filing requests

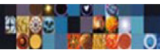
Models	β	SE β	t-value	<i>p</i> -value
(Y) Number of patent applications (INPI)	Intercept: 6.719	0.202	33.319	<i>p</i> -value < .01*
	University researchers: 0.003	0.001	3.757	<i>p</i> -value: .003*
(Y ₂) Number of patent applications (United States)	Intercept: 102.483	6.369	16.09	<i>p</i> -value < .01**
	Public Investments: 14.042	0.253	55.43	<i>p</i> -value < .01**
(Y ₃) Number of patent applications (EU)	Intercept: -113.03	587.28	-0.192	<i>p</i> -value: .851
	Scientific production: 126.29	15.27	8.271	<i>p</i> -value < .01***

Note: Model 1: $R^2 = 54.06\%$; * $p < .01$; Model 2: $R^2 = 99.61\%$; ** $p < .01$; Model 3: $R^2 = 85.08\%$; *** $p < .01$. All assumptions related to the distribution, independence and homoscedasticity of the residues are valid (Shapiro-Wilk: $p > .01$; Durbin-Watson: $p > .05$; Breusch-Pagan: $p > .05$).

independent variables (Hair et al., 2009; Wooldridge, 2012). Accordingly, the Pearson correlation matrix was determined to preliminarily identify the relationship between the variables, as shown in Table 3.

The results of the Pearson's correlations presented in Table 3 indicate that the exception of the relationship between the dependent variable (Y₄) Number of patent applications (Triadic) and the other independent variables, which present a positive relationship between weak and moderate ($r = [0.2; 0.6]$), all other dependent variables (Y) Number of patent applications (INPI); (Y₂) Number of patent applications (United States); (Y₃) Number of patent applications (EU) present a positive relationship between moderate and very strong ($r = [0.6; 1]$). Adding to this, it is possible to notice that all of the independent variables have positive correlations and comprised between strong and very strong ($r = [0.7; 1]$), which indicates the possibility of Multicollinearity, confirmed later through the Klein criteria and the Variance Inflation Factor (VIF) (Imdadullah et al., 2016). In this sense, considering that there is no universal solution to the problem of Multicollinearity (Field, 2009), it was decided to determine Bivariate Regressions, using the determination coefficient as the criterion of choice (Dalgaard, 2008; Härdle & Simar, 2015; Myers et al., 2010). Table 4 presents the Bivariate Regression models for each of the dependent variables.

It appears from the results in Table 4, that the variable university researchers have a significant influence on the number of patent filings at the Brazilian office, being responsible for 54.06% of the variation in the number of applications in the period 2000–2017. Concerning requests for deposits at the American office, the variable that has the most influence is public investments in research and development activities and



related scientific activities, accounting for 99.61% of the variation in the number of patent filing requests. Last, but not least, it is possible to identify that scientific production is the variable that most influences the number of patent filing requests at the European office, accounting for 85.08% of the variation in the number of requests during the period 2000–2013. Regarding patent families, none of the variables was identified as significant for determining a valid model (F -statistic: p -value < .01). The results found follow the studies performed in recent years and are in line with the literature review.

4 | IMPLICATIONS FOR POLICY

From a general point of view, patents boost creativity by incentive effects and diffusion and limit competition by establishing temporary monopolies. Patents tend to increase innovators' costs, especially when they need to combine innovations from a variety of sources, as well as the fact that patents may contribute to competition by encouraging the vertical breakup of knowledge-intensive industries and by supporting new entrants.

Patent structures, powered by global competition in knowledge-intensive sectors and also as a result of technological change, have developed to overcome this context. Nonetheless, patents reflect a trade-off between the costs of granting restricted market power to businesses and the benefits of fostering innovation, in much the same way as competition policy is meant to maintain a balance between the costs of increasing market power from concentration and the benefits of efficiencies on a scale. It is a global perception that patent and competition laws are often misaligned and need to be discussed more and more simultaneously.

It is possible to examine the expansion of patenting operation in several dimensions: coverage of the subject matter, form of patenting organization, and geography. And several governments' general policy stance is to enable companies, citizens, and research organizations to learn about the patent system and apply for patents. In Europe, a number of countries have introduced initiatives to enable universities and government research bodies to patent and licence technology (Geuna and Nesta, 2006). The heterogeneity of responses to the scheme, a heterogeneity firmly grounded in the heterogeneity of technology (Burk and Lemley, 2002) and its growth, is a major problem for policy makers, requiring expensive investments in patent portfolio creation for defensive purposes while using other methods to secure returns on their own inventions.

Finally, some guidelines for policymakers to promote policymaking to serve industry and inventors should be made: (a) more inclusive R&D initiatives and more specific measures should be a priority for governments. The patent regulations are so complicated and have a high degree of technical specificity that they are a very difficult method to enforce. More comprehensive information on invention licensing in this regard should be a priority for increasing the number of registrations of patents; (b) up-to-date research and statistics on patent policy reforms and the impact of their application should be a higher priority. In order to determine the need for improvements in current policy, it is important to know the impact of policy on industries.

5 | CONCLUSIONS AND IMPLICATIONS OF THE RESEARCH

Patents are an important measure of innovation activity and innovation output. Patent registration in itself, however, is not a measure of success in business, as firms with numerous patents have still been considered market failures (Nokia and its mobile telephony division is a recent example). Of note is that certain entities prefer to maintain their discoveries a secret, rather than register them and make them public. Coca-Cola is an example of this, a firm which continues to reap considerable benefits from its secret beverage formula rather than having capitalized from this formula for a short period of time (after which abnormally high rents would have ceased). Albeit, patents still remain an important means to encourage technological progress as they are an exclusive property right granted by the Government to individuals or other inventing entities and therefore make possible monopoly-type rents for a short period of time.

The quantitative methods used in this research resorted to big data from private and public investments in Science and Technology, and about patent deposit numbers in Brazil from 2000 to 2017.

Patent filing requests made by residents in Brazil have been higher in the European office, due to requests related to patent families, and at the American Patent and Trademark Office. On the other hand, patent filing applications at the Brazilian patent office presented the lowest average. Except for orders placed at the Brazilian office, the coefficients of variation for orders from other offices reveal that there was a wide dispersion in the number of orders over the period under review. Additionally, it is possible to notice that public investments in Science and Technology have been higher, although closely monitored by private investments.

Regarding Brazilian scientific production, understood as articles published and indexed in the Scopus Database, the average was 35.4 articles, with a high coefficient of variation of 44.5%. Regarding the total number of researchers, the analysis reveals that there is a predominance of researchers linked to Universities.

To identify which variables are significant to explain the number of patent filing requests, a Regression analysis based on the ordinary least squares' method was applied.

The variable university researchers have a significant influence on the number of patent filings at the Brazilian office. Concerning requests for deposits at the American office, the variable that has the most influence is public investments in research and development activities and related scientific activities. Last, but not least, it is possible to identify that scientific production is the variable that most influences the number of patent filing requests at the European office.

Hence, the results of this research study point towards the importance of universities, up and beyond the traditional training and education aspect of university activity. The importance of public and private innovation investments is also shown to be important. This study shows that the patent registrations in the different regions under analysis are affected by different factors. There is thus no single formula towards the creation of innovation output and governments would do well to continue to invest in higher education while also investing in public research and development activities. Additionally, and not least important, private entities should be continually encouraged to make innovation investments and favourable government policies need to thus exist for this to happen.

Finally, the low numbers regarding patent filings in Brazil may be linked to institutional deficiencies in the country. Patent breaches may be difficult to punish, and the judicial system may be slow and untrustworthy, compared to the United States and to Europe—leading to diminished patent registrations in Brazil.

6 | LIMITATIONS OF THE MODELS AND FURTHER RESEARCH

The models presented in the present research must be seen as preliminary, given the effects of multicollinearity. In this sense, the independent variables function as proxies to each other, making it difficult to estimate a completely valid model (Field, 2009; Hair et al., 2009). Furthermore, it would be interesting to measure to what extent are patents translated into real and practical innovations? In academia, for example, researchers may be rewarded for producing patents, even if those patents lead to no new product sales (and there not existing any real motivation to seek new product development and sales, in most such cases). On the other hand, do patents inhibit or aid innovation output in society? Some firms make the strategic choice to not patent their innovations—for example, Coca-Cola and its secret formula, as mentioned above—opting, instead, to avoid the full disclosure involved in the patenting process, which leads to rents limited in time—and, hence, limited competitive advantage. “In some sectors patents are not widely adopted as mechanisms to protect innovations” (Ponta et al., 2020, p.178). Hence, we advise that additional research be done regarding the effect of the registration of patents on business success—both sales and profits—to thus determine the effectiveness of this specific innovation output in industry. Are we measuring the right innovation indicator in measuring patent production? Certainly, it should not be the sole indicator of the innovation capability of firms. Finally, what is the contribution of absorptive capacity (a concept brought to us in the seminal paper by Cohen and Levinthal, 1990) to patent production, according to sector of activity, and how may public policy improve on the absorptive capacity of organizations? That is, up and beyond simply increasing R&D funding and budgets.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data available on request from the authors

ORCID

Maria José Sousa  <https://orcid.org/0000-0001-8633-4199>

George Jamil  <https://orcid.org/0000-0003-0989-6600>

Cícero Eduardo Walter  <https://orcid.org/0000-0002-5783-6613>

Manuel Au-Yong-Oliveira  <https://orcid.org/0000-0002-2154-6171>

Fernando Moreira  <https://orcid.org/0000-0002-0816-1445>

REFERENCES

- Acemoglu, D., & Robinson, J. A. (2013). *Why nations fail. The origins of power, prosperity and poverty*. London, England: Profile Books.
- Akcigit, U., Ates, S. T., & Impullitti, G. (2018). Innovation and trade policy in a globalized world. In *International finance discussion papers 1230* (Vol. 2018). Board of Federal Reserve System. <https://doi.org/10.17016/IFDP.2018.1230>
- Archibugi, D. (1992). Patenting as an indicator of technological innovation: A review. *Science & Public Policy*, 19(6), 357–368.
- Burk, D. L., & Lemley, M. A. (2002). Is Patent Law Technology-Specific?. *Berkeley Technology Law Journal* (Vol. 17, p. 1155).
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152.
- Comai, A. (2018). Beyond patent analytics: Insights from a scientific and technological data mashup based on a case example. *World Patent Information*, 55, 61–77.

- Cox, D. R., & Stuart, A. (1955). Some quick sign tests for trend in location and dispersion. *Biometrika*, 42(1/2), 80–95.
- Dalgaard, P. (2008). *Introductory statistics with R (Vol. 2)*. Springer.
- European Commission. (1996). *The Green Book on Innovation*. Brussels: EU.
- Field, A. (2009). *Descobrimos a Estatística usando o SPSS (Vol. 2)*. Porto Alegre, Brazil: Artmed.
- Fujimoto, S., Ishikawa, A., Mizuno, T., & Watanabe, T. (2015). Are firms that are awarded more patents more productive? In *Econophysics and data driven modelling of market dynamics* (pp. 129–142). Cham, Switzerland: Springer.
- Geuna, A., & Nesta, L. (2006). University patenting and its effects on academic research: the emerging European evidence. *Research Policy*, 35(6), 790–807.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados* (6ed ed.). Bookman Editora.
- Härdle, W., & Simar, L. (2015). *Applied multivariate statistical analysis (Vol. 4)*. Springer.
- Imdadullah, M., Aslam, M., & Altaf, S. (2016). Mctest: An R package for detection of collinearity among Regressors. *The R Journal*, 8(2), 495–505.
- Jamil, G. L., Santos, L. H. R., & Jamil, C. C. (2019). Market intelligence as an information system element: Delivering knowledge for decisions in a continuous process, Market intelligence as an information system element. In G. L. Jamil (Ed.), *Handbook on expanding business opportunities with information systems and analytics*. Hershey, PA: IGI Global Publishers.
- Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, 13(3), 245–259.
- McKinsey. (2020). Analytics insights. Retrieved from <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights>
- Myers, J. L., Well, A., & Lorch, R. F. (2010). *Research design and statistical analysis*. Routledge.
- Manzini, R., & Lazzarotti, V. (2016). Intellectual property protection mechanisms in collaborative new product development. *R&D Management*, 46(S2), 579–595.
- North, D. C. (2005). *Understanding the process of economic change*. Princeton, NJ: Princeton University Press.
- OECD/Eurostat. (2005). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*, 3rd ed., The Measurement of Scientific and Technological Activities. *Em Communities*. <https://doi.org/10.1787/9789264013100-en>
- Patentanalysis (2020). Patent analysis report and techniques. Retrieved from <http://patentanalysis.org/articles-on-patent-analysis/role-of-patent-analysis/>
- Ponta, L., Puliga, G., Oneto, L., & Manzini, R. (2020). Innovation capability of firms: A big data approach with patents. In *Proceedings of the International Neural Networks Society* (pp. 169–179). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-030-16841-4_18
- Romijn, H., & Albaladejo, M. (2002). Determinants of innovation capability in small electronics and software firms in Southeast England. *Research Policy*, 31(7), 1053–1067.
- SAS. (2020). Big Data: What is it and why it matters. Retrieved from https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html
- Schumpeter, J. A. (1934). *The theory of economic development*. New York, NY: Harper & Brothers.
- Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. London, England: Routledge.
- Saheb, T., & Saheb, T. (2020). Understanding the development trends of big data technologies: An analysis of patents and the cited scholarly works. *Journal of Big Data*, 7(1), 12.
- Seo, W., Kim, N., & Choi, S. (2016). Big data framework for analyzing patents to support strategic R&D planning. In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (pp. 746–753). IEEE.
- Sideri, K. (2020). Prospect patents, data markets, and the commons in data-driven medicine: Openness and the political economy of intellectual property rights. *Science & Public Policy*, 0(0), 1–10.
- Simon, B. M., & Sichelman, T. (2017). Data-generating patents. *Northwestern University Law Review*, 111(2), 377–438.
- Stern, S., Porter, M., & Furman, J. (2000). *The determinants of National Innovative Capacity (working paper 7876)*. National Bureau of Economic Research. <https://doi.org/10.3386/w7876>
- Tadeu, H. B., Duarte, A. L. C. M., Taurion, C., & Jamil, G. L. (2019). Digital transformation: digital maturity applied to study Brazilian perspectives for industry 4.0. In *Best practices for manufacturing processes: experiences in Latin America*, Berlin, Germany: Springer. https://doi.org/10.1007/978-3-319-99190-0_1
- WIPO (2016). World Intellectual Property Organization: WIPO Manual on Open Source Patent Analytics. Retrieved from <https://wipo-analytics.github.io/>
- Wooldridge, J. (2012). *Introductory econometrics: A modern approach*. South-Western College Publishing a Division of Thomson Learning.

AUTHOR BIOGRAPHIES

Maria José Sousa (Ph.D. in Management) is a University Professor and a research fellow at ISCTE/IUL and of CIEO (Algarve University). Her research interests currently are public policies and political science, innovation and business management issues. She is a best seller author in ICT and People Management and has co-authored over 70 articles and book chapters and published in several scientific journals (e.g. *Journal of Business Research*, *UAIS*, *Future Generations Computer Systems*, *Journal of Medical Systems*, *I. J. of Knowledge, Culture and Change Management*, *Knowledge Management*, *Wseas Transactions on Business and Economics*, *Information Systems Frontiers* and others), she has organized and peer-reviewed international conferences. She is also External Expert of COST Association - European Cooperation in Science and Technology.

George Jamil, Bsc. Electric Engineering (UFMG/82), Master in Computing Science (UFMG/99), Doctorate in Information Science (UFMG/05), Post Doctorate student at Port University (conclusion for April/14). Books and papers published in Brazil and abroad. Consultancy expertise in Strategy, Marketing and IT Management. Professor for undergraduation (Production and Civil Engineering), Post-Graduation, Masters and PhD courses.

Cicero Eduardo Walter, Ph.D. Student of Economics and Business Sciences at the University of Aveiro (Portugal). Master and Specialist in Management (Management of Organizations - Business Management Branch) from the School of Technology and Management from Polytechnic Institute of Bragança - ESTIG-IPB - (Portugal). Professor of Business and Management at the Federal Institute of Education, Science, and Technology of Piauí (Brazil). Member of the research unit Governance, Competitiveness, and Public Policy (GOVCOPP) of the University of Aveiro as a doctorate student. Currently researching Innovation Management, Business Management, and Learning Styles.

Manuel Au-Yong-Oliveira did his post-doctoral studies at the University of Aveiro (2016-2019) with the title “Elements that contribute to organisational competitiveness”; Manuel has a PhD in Industrial Engineering and Management, from FEUP (University of Porto, 2012; distinction awarded to the thesis), an MBA from Cardiff Business School (UK, 1993; distinction awarded to the dissertation) (equivalence given to a Master's in Management by Universidade Católica Portuguesa), and an undergraduate degree in Management Information Processing, from Universidade Portucalense (1991). Manuel is an Assistant Professor at the University of Aveiro (DEGEIT) where he lectures at the undergraduate, master's and doctoral levels on marketing, innovation, strategy and research methods. Before lecturing full-time at the University of Aveiro Manuel also lectured for a number of years at the University of Porto - FEUP (Business Construction Project, on the Master's in Innovation and Technological Entrepreneurship), where he also supervised a number of master's dissertations. Manuel is currently a researcher affiliated to GOVCOPP (full researcher) and INESC TEC (collaborator). Previously, Manuel was the Director of the Undergraduate Degree in Management, at the University of Aveiro (14'09'2015 a 7'12'2016). At present, Manuel is the Director of the Master's Degree in Management (since 2016) at the University of Aveiro, in Portugal. Furthermore, Manuel is a member of the Executive Committee of his department - DEGEIT - University of Aveiro (since 2014). Finally, Manuel coordinates the Committee for Ethics and Social Responsibility at DEGEIT - University of Aveiro.

Fernando Moreira is Head of Science and Technology Department. He is graduated in Computer Science (1992), M.Sc. in Electronic Engineering (1997) and PhD in Electronic Engineering (2003), both at Faculty of Engineering of the University of Porto, and Habilitation (2018). He is a member of the Science and Technology Department at Portucalense University since 1992, currently as Full Professor and a visiting professor at the University of Porto Business School. He teaches subjects related to undergraduate and post-graduate studies. He supervises several PhD and M.Sc. students. He is a (co-)author more than 200 scientific publications with peer-review on national and international journals and conferences. He serves as a member of the Editorial Advisory Board for several journals and books. He organized several special issues from JCR journals. He has already regularly served as a member of Programme and Scientific committees of national and international conferences. He was the MSc in Computation coordinator during the last ten years. He holds editorial experience, and he is co-editor of several books. He is associated with NSTICC, ACM and IEEE. His principal research areas are mobile computing, ICT in Higher Education, Mobile learning, Social business and Digital transformation. Awarded with Atlas Elsevier Award, April 2019.

How to cite this article: Sousa MJ, Jamil G, Walter CE, Au-Yong-Oliveira M, Moreira F. Big data analytics on patents for innovation public policies. *Expert Systems*. 2021;e12673. <https://doi.org/10.1111/exsy.12673>