Analysis of technological knowledge flows in the Basque Country

Gavilanes-Trapote, J. [a*], Etxeberria-Agiriano, I. [b], Cilleruelo, E. [c], and Garechana, G. [d]

[a] Industrial Organization and Management Engineering Department. Faculty of Engineering - Vitoria-Gasteiz. University of the Basque Country (UPV/EHU). Nieves Cano 12, 01006 Vitoria-Gasteiz. Spain

[b] Languages and Computer Systems Department. Faculty of Engineering - Vitoria-Gasteiz. UPV/EHU

[c] Industrial Organization and Management Engineering Department. Faculty of Engineering - Bilbao. UPV/EHU

[d] Industrial Organization and Management Engineering Department. Faculty of Economics and Business - Bilbao. UPV/EHU

[a] *javier.gavilanes@ehu.eus*, [b] *ismael.etxeberria@ehu.eus*, [c] *ernesto.cilleruelo@ehu.eus*, [d] *gaizka.garechana@ehu.eus*

Abstract: Knowledge flow of technology is important for continuous growth and extension of science. Patent data analysis has facilitated this knowledge acquisition. The available patent information crosses borders, corresponds and interacts with new inventions to give new strength and dimension to the technology. Therefore, the patent citation information functions as a key indicator of the knowledge flow providing relevant information. It can be identified to which extent a region is a relevant technological knowledge generator to other regions. As an illustrative case, we present a study to determine the role played by the Basque Country region as a generator of technological innovation during the period 1991-2011.

Key words: Patent, Citation, Technology, Diffusion, Innovation.

1. Introduction

Knowledge is an intangible strength that has a definite economic importance when it is well utilized and commercialized. Knowledge spillover is something that occurs, imaginable but difficult to measure effectively.

The tendency to locate knowledge in certain areas together with the effects derived from agglomeration economies are responsible for the strong concentration process and regional specialization observed in a growing way in the economy (Krugman, 1995). From the interaction between infrastructures and the built environment, accessible natural resources, the institutional endowment and the knowledge and skills available in the territory, localized capacities are developed, difficult to imitate and of cumulative nature, which lead to competitive advantages of the territory (Malmberg and Maskell, 1997).

The competitiveness of the regions depends to a large extent on the characteristics of their "Innovation and Development Network", the analysis of the knowledge transfers being an output of competitiveness of the region.

There exists a debate about the suitability of considering patents as a proxy for technological knowledge, which has highlighted both its limitations and its benefits.

In principle, we could consider that the optimal measure of technological knowledge is given by the number of innovations, which in turn are understood as those novelties that have come to be marketed. The main limitation of this measure is determined by the almost total unavailability of data, although the annual surveys of EUSTAT (Basque Institute of Statistics) include questions about the "Economic impact of product innovations on turnover". Even so,

this measure has a series of drawbacks. For example, data come from surveys sensitive to the response rate and to the interpretation of the term innovation by the companies, together with the average life cycle of the products of the companies consulted. On the other hand, patents and their evaluation process are objective. In addition, being official documents increases their rigor with less typographical errors than in other types of documents. Another disadvantage is that the introduction of a new product in the market takes place in the last phase of the innovation process, which may be very far from the moment in which the R&D effort is made (Schmoch, 2003). However, in the case of patents, this lapse of time is much shorter.

Another considerable advantage of patents is that they provide a wide temporal, geographical and technological scope as a measure of technological innovation. Patent information is collected since the mid-nineteenth century, in most countries of the world and practically covers all technological fields. A remarkable exception is software, which is usually protected through property rights, and can only be patented if integrated into a product or production process, OECD (2009).

Particular advantage of great importance for the present work lies in the accessibility of patent information, highlighting three key aspects: the level of structure presented by patent documents, information technologies and the role played by patent offices. The degree of structuring presented by patent documents and the advancement in information technologies have made possible to create patent databases that are easy to use and allow quick retrieval of documents searched at each moment. On the other hand, the resources that patent offices invest in creating, maintaining and disseminating these databases, bringing them closer to users, are fundamental in this process. In this regard, in recent years, most patent offices have created various databases available to all users through the Internet.

Finally, regarding the cost of information, it is interesting to distinguish between the services supplied by patent offices and private entities. The former are making great efforts to disseminate patent information and reduce the cost of their services, some of which are free. Besides, private entities are mainly characterized by offering services adapted to their customers' needs, like the creation of specialized databases in sectors such as pharmaceutical and chemical.

In short, patents are far from being a perfect measure of technological output and its application in products and services, but so far we consider they are the best and most complete measure available for the analysis of the technological knowledge of a region.

Patents provide a detailed description of how inventions have been made at the earlier state of the art, thus constituting a reliable measure of the transfer of technological knowledge (OECD, 2009; Higham *et al.,* 2017). Patent citations indicate the use of prior inventions, which makes it possible to identify the influence of a particular invention or set of inventions and mapping its diffusion in the economy (Acs *et al.,* 2002; You *et al.,* 2017). Citations from other patents or non-patent literature (NPL) or knowledge between science and technology are useful in quantifying the level of knowledge transfer between organizations, geographical regions and/or technology sectors (Gavilanes-Trapote *et al.,* 2015, 2017).

There are basically two types of citation. On the one hand, patent references are citations of relevant technology previously protected by other patents applied for anywhere in the world, at any time and in any language (*backward* and *forward* citations, technological knowledge flows). Conversely, references classified as NPL are scientific publications, conference proceedings, books, database guides, technical manuals, standard descriptions, and so on.

The potential offered by patent citation measurements for policy elaboration is immense. The main methods utilized in literature on innovation are three: i) the measurement of flows or the effects of propagation of knowledge (for example, Jaffe and others, 2000); ii) measurement of the quality of the patent (for example, Harhoff *et al.,* 2004); and iii) the strategic behavior of the companies (for example, Podolny *et al.,* 1996). The present work will focus on analyzing the knowledge flow between patents.

The degree of technological accumulation is defined as the frequency with which a society cites its own previous investigations. The identification of the self-citation (applicant or holder) has important implications, among other things for the study of the propagation effect of technological knowledge. It can be assumed that citations of patents belonging to the same owner represent transmissions of knowledge that are mostly internalized, while citations of patents of "others" are closer to the pure notion of the diffusion effect.

A customary measure of the degree of technological accumulativeness of a region is the sum of retrospective citations of patents of that same region among the total number of patents it has, at a given time. According to Malerba and Orsenigo (1995), the degree of technological accumulation affects the extent to which innovative leaders build a competitive advantage over followers and the preservation of leadership in the future.

2. Objectives

The present work has as main objective to determine the role played by the Basque Country as a generator or receiver of technological knowledge in the world. For this purpose, countries receiving and emitting technological knowledge in the studied region will be identified.

Finally, the technological accumulation of the region will be calculated.

3. Methods

The sample includes all patents from January 1, 1992 to December 31, 2011. There are several reference dates for selecting patents over a period of time. The most important are the application date and the grant date. Patent studies can be found where both dates are used indistinctly. In this study, the date of application has been chosen because it is the closest date to the innovative activity. The disadvantage of using this date as an element of selection is that it does not guarantee that the patent has been granted. Therefore, it is necessary to identify which patents have been granted and eliminate the remaining patents from the sample. The use of this date presents another inconvenience due to the waiting time from the time a patent is applied for until its concession is published (up to 6 years), making it impossible to work with a recent sample without suffering a significant truncation effect. The final sample contains 3,503 patents.

The databases used to access the patent information in the sample are INVENES (http://invenes.oepm. es/InvenesWeb/faces/busquedaInternet.jsp), of the Spanish Patent and Trademark Office (SPTO) and PATSTAT (https://www.epo.org/searching-for-patents/business/patstat.html), of the European Patent Office (EPO). The PATSTAT database contains all fields including patent citations, but lacks the possibility of discriminating Spanish patents by province. Then again, INVENES allows one to search by provinces, but it lacks access to certain important fields such as forward or backward patents.

The followed process can be described as follows. All patents of the Autonomous Community of the Basque Country (CAPV for its Spanish acronym) requested between the years 1992-2011 were downloaded. Queried fields were those of province "PROV" and publication number "NPUB". This publication number uniquely identifies the patents requested in Spain and is equivalent to the field "publn_nr" of the PATSTAT database. In this field, duplicates were found due to the fact that throughout the process of querying for a patent in the Spanish office various documents are being generated for the same registry. After the elimination of these duplicate records, the PATSTAT database was accessed to extract all the fields necessary for the study of the CAPV patents, see Table 1.

The PATSTAT database is a payment service but allows a free trial version for 2 months. SQL queries are carried out online through a graphical interface, see Figures 1 and 2.

Table 1. Bibliographic fields downloaded from the PATSTAT database. Source: self elaboration.

Field	Description
appln_auth	Nationality of the office from where the patent was requested
appln_id	Number assigned by the EPO to each patent for its univocal identification
han_name	Name of patent applicants
person_address	Address of applicants
person_ctry_code	Nationality of applicants
person_id	Number that uniquely identifies each applicant
prior_earliest_date	Priority date or date of the first request
publn_nr	Number that identifies the patent in the national office

The EPO website provides a manual to facilitate the formulation of queries in SQL language. Once the search is done, the selected records can be downloaded in ".csv" (comma separated values) format.

Query results were downloaded into ".csv" files for later importation into the Vantage Point (https://www.thevantagepoint.com) text mining
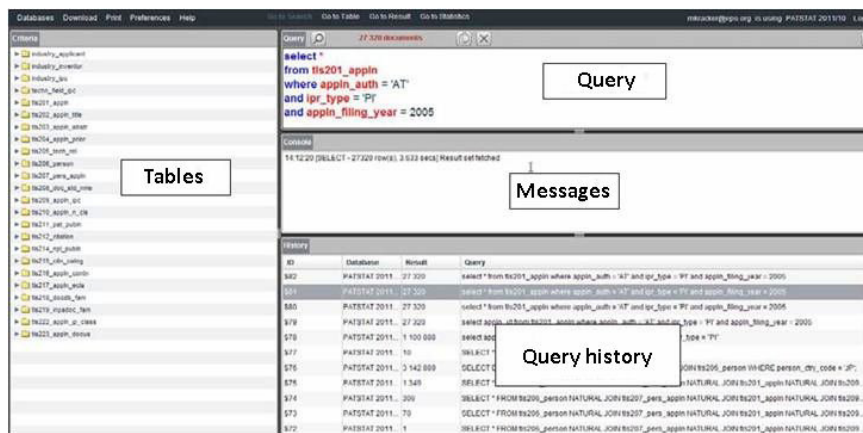
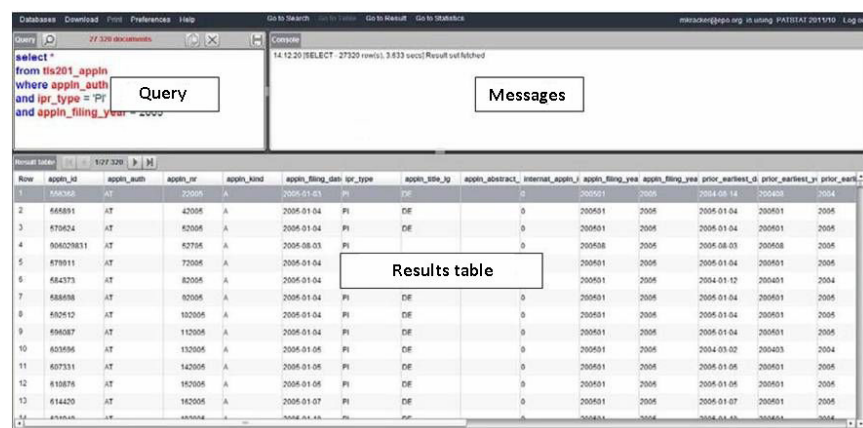**Figure 1.** Graphical PATSTAT database interface. Querying window. Source: self captured.

Figure 2. Graphical PATSTAT database interface. Results window. Source: self captured.

software. Through this program data from different imported files were merged through the key field "appln_id". Subsequently, they were checked for duplicate records. This task, like that of the merged, is not complicated since the database contains a field that uniquely identifies each patent and the program has a specific function to perform these tasks. Once all this information is merged into a single file we move on to the arduous and tedious task of data cleaning.

The cleaning stage aims at solving one of the most notable problems in bibliographic databases: errors and lack of data consistency. For both database producers and researchers downloading data for scientific purposes, the lack of standardization and errors suppose the loss of information that forces the development of corrective systems, almost always personalized, that guarantee the rigor of the research, so dependent on the quality of the data (Gálvez and Moya-Anegón, 2007).

Out of the fields used for the present study, "har_name" has been the one that has required the most cleaning work. The normalization of the organizations and researchers applying for patents contained in this field requires facing two main problems: homonymy (two applicants with the same name) and synonymy (different variants of name referring to a single applicant).

Homonymy is presented only in the names of the researchers. Organizations do not present this problem because they are obliged to have a unique denomination to allow them joining the mercantile registry. The solution to this bias is to extract information from the context. In the case of patents, it could be resolved by analyzing the researcher address in field "person_address". The problem in our database is that this field is only completed in 8% of the records, which makes it far too complicated to affirm that this homonymy bias does not affect any record of the study.

The synonymy can be the source of serious problems in any type of bibliographic recount. Variations of the names of applicants can be classified as follows:

- **Invalid variations**: mainly caused by spelling, phonetic or typographical errors; incorrect use of capitals, nicknames or abbreviations; or accentuation problems.

- **Valid variations**: caused by the exchange of word order; separation of words; use or absence of punctuation marks; use of initials or absence of any part of the name, see examples in Figure 3.
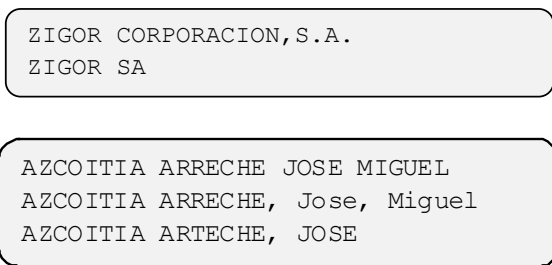
```
ZIGOR CORPORACION,S.A.
ZIGOR SA
```

```
AZCOITIA ARRECHE JOSE MIGUEL
AZCOITIA ARRECHE, Jose, Miguel
AZCOITIA ARTECHE, JOSE
```

Figure 3. Examples of applicants transcribed in different ways. Source: self elaboration.

The solution to synonymy, unlike homonymy, is simple but costly in terms of resources. The Vantage Point program has grouping algorithms (fuzzy clustering) that try to solve this problem. In practice, it ends up being necessary to review applicants' names one by one and manually perform the clustering through an interface provided by the program within the "List cleanup..." command. The result of this work is the thesaurus, a file that contains these groups.

With data selected, correctly merged and cleaned, we are in a position to perform the analyses.

To determine to which extent the Basque Country region is a recipient of technological knowledge generated in other countries, we will look at the "person_ctry_code" field in the *backward* patent citations.

As the "person_ctry_code" citations field only provides information about the country and not about the region to which the patent belongs, a manual treatment, document by document, identifying the location of each applicant company or organization has been necessary, having to discard patents where only researchers are listed, as we cannot know if they belong to the CAPV or to another community.

To determine to which extent the Basque Country region generates the technological knowledge used by other countries as a basis for their innovations, we will look at the "person_ctry_code" field of *forward* patent citations.

Finally, to determinate the degree of technological accumulation we shall calculate what percentage of forward citations of CAPV are self-citations.

4. Results

First of all, we will answer the question: "where do we get the knowledge of patents in the Basque Country?"

To do this, it is necessary to analyze the nationality of the *backward* citations. 81% of patents in the Basque Country present one or more such citations, accounting for a total of 12,752 *backward* citations. 22% out of these have no nationality, reducing the number of citations to 9,947. Finally, the total number of nationalities to be analyzed is 12,810. This number is higher than the number of citations due to the fact that the chosen count type is complete and more than one of these patent citations has several applicants with different nationalities.

Figure 4 shows how the United States, with 29%, is the country that contributes most technological knowledge to patents in the Basque Country. Second, Germany appears with 17%, followed by France and Spain, with 9% each.

Figure 5 shows only patent citations of Spanish nationals (1,194) discriminating against those corresponding to the Basque Country. The final number of citations of Spanish nationality with an organization as applicant is 538, being 61% (324) organizations located in the region of the Basque Country.
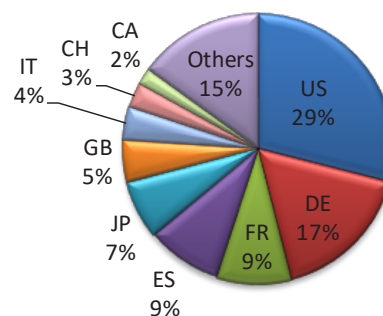

Figure 4. Nationality of patents supporting inventions in the Basque Country, *forward* patent citations.
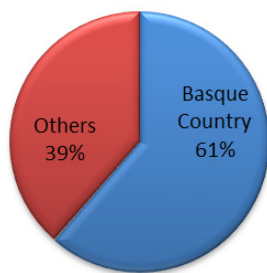
Figure 5. Percentage of *backward* citations of national origin belonging to the Basque Country

Secondly, we will answer the question: "where is the knowledge of patents in the Basque Country transferred to?"

To this end, the nationalities of *forward* citations are analyzed. Only 20% of patents in the Basque Country are cited by other patents, receiving a total of 1,563 citations. Only 4% out of these have no nationality assigned to any of their applicants, so the final number of subsequent appointments is 1,496.

Figure 6 shows how Spain, with 46%, is the main country that receives technological knowledge from patents in the Basque Country. Second, Germany appears with 10% and then, in a smaller percentage, other countries such as France, the United States and Italy, among others.

Figure 7 shows all the *forward* citations of Spanish patents (688) discriminating those corresponding to the Basque Country. The final number of *forward* citations of Spanish nationality with an organization as applicant is 378, being 61% (231) of companies located in the Basque Country region.

These percentages are the same as those mentioned above, so the Basque Country maintains the self-citation level in the region.
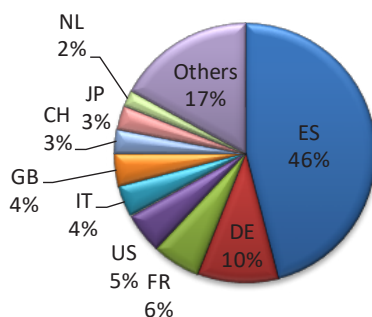

Figure 6. Nationality of patents that are relying on Basque inventions to develop their technological knowledge, *forward* patent citations
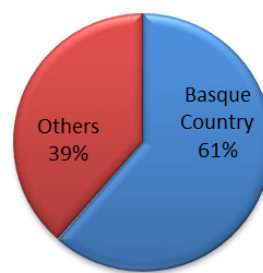

Figure 7. Nationality of patents that are relying on Basque inventions to develop their technological knowledge, *forward* patent citations

Finally, to calculate the degree of technological accumulation in the Basque Country, we have 1,496 *backward* citations with nationality that have received Basque patents, with 231 from Basque companies. This implies at least a 15% degree of technological accumulation, that is, almost 1 out of 6 patents serves as a support to generate new technological knowledge in the region.

5. Conclusion

The present work identifies which regions or applicants have contributed with their innovations to the development of new patents, defining the flow of technological knowledge from a geographical or authoring point of view of the inventions. This information allows us to identify the regions that have contributed with the most technological knowledge, which will help in the design of regional policies.

Through the origin of the *backward* citations, we can determine which countries contribute the most to the technological development of studied patents. In the case of the Basque Country, the United States generates the greatest flow of technological knowledge. In second place comes Germany, followed by France and Spain. This knowledge serves to develop innovations in the studied region.

On the contrary, through the origin of the *forward* citations we can determine how the region of study shares its technological knowledge with other regions. In the case of the Basque Country, it shares almost half of its knowledge with the region of Spain (46%), followed by Germany (10%).

These data clearly indicate that the role played by the Basque Country region within Spanish territory is mainly to generate technological knowledge as a base for future innovations.

The degree of technological accumulation in the Basque Country is relatively high (15%). This means that 15% of the innovative knowledge generated in the region is used to generate new knowledge. This level of internal transfer helps innovative leaders to build a competitive advantage over their followers and maintain their leadership in the future.

Results from this study indicate the Basque Country is an important region from the point of view of knowledge generation within the Spanish territory.

Acknowledgements

References

Acs, Z.J., Anselin, L., and Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge *Res. Policy, 31*(7), 1069-1085. https://doi.org/10.1016/S0048-7333(01)00184-6

Balland, P., and Rigby, D. (2017). The geography of complex knowledge. *Economic Geography, 93*(1), 1-23. https://doi.org/10.1080/00130095.2016.1205947

Galvez, C., and Moya-Anegon, F. (2007). Standardizing formats of corporate source data. *Scientometrics, 70*(1), 3-26. https://doi.org/10.1007/s11192-007-0101-0

Gavilanes-Trapote, J., Cilleruelo-Carrasco, E., Etxeberria-Agiriano, I., Garechana, G., Rodríguez Andara, A. (2019). Qualitative Patents Evaluation Through the Analysis of Their Citations. Case of the Technological Sectors in the Basque Country. In: Ortiz, Á., Andrés Romano, C., Poler, R., García-Sabater, J.P. (eds) *Engineering Digital Transformation. Lecture Notes in Management and Industrial Engineering*. Springer, Cham. https://doi.org/10.1007/978-3-319-96005-0_28

Gavilanes-Trapote, J., Río-Belver, R. M., Cilleruelo, E., Garechana G., and Larruscain J. (2015). Patent overlay maps: Spain and the basque country. *International Journal of Technology Management (IJTM), 69*(3/4), 261. https://doi.org/10.1504/IJTM.2015.072976

Harhoff, D., Scherer, F., and Vopel, K. (2004). Citations, family size, opposition and the value of patent rights (vol 32, pg 1343, 2003). *Research Policy, 33*(2), 363-364. https://doi.org/10.1016/j.respol.2003.10.001

Higham, K.W., Governale, M., Jaffe, A.B., and Zülicke, U. (2017) Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. *Phys. Rev. E 95*, 042309. https://doi.org/10.1103/PhysRevE.95.042309

Jaffe, A., Trajtenberg, M., and Fogarty, M. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review, 90*(2), 215-218. https://doi.org/10.1257/aer.90.2.215

Krugman, P. (1995). *Development, geography, and economic Theory*. Cambridge-Massachusetts: The MIT Press.

Malerba, F., and Orsenigo, L. (1995). Schumpeterian patterns of innovation. *Cambridge Journal of Economics, 19*(1), 47-65. https://doi.org/10.1093/oxfordjournals.cje.a035308

Malmberg, A., and Maskell, P. (1997). Towards an explanation of regional specialization and industry agglomeration. *European Planning Studies, 5*(1), 25-41. https://doi.org/10.1080/09654319708720382

Murata, Y., Nakajima, R., Okamoto, R., and Tamura, R. (2014). Localized knowledge spillovers and patent citations: A distance-based approach. *Review of Economics and Statistics, 96*(5), 967-985. https://doi.org/10.1162/REST_a_00422

Organisation for Economic Co–operation and Development (OECD). (2009). OECD patent statistics manual OECD.

Podolny, J., Stuart, T., and Hannan, M. (1996). Networks, knowledge, and niches: Competition in the worldwide semiconductor industry, 1984-1991. *American Journal of Sociology, 102*(3), 659-689. https://doi.org/10.1086/230994

Schmoch, U. (2003). Service marks as novel innovation indicator. *Research Evaluation, 12*(2), 149-156. https://doi.org/10.3152/147154403781776708

Simmie, J. (2003). Innovation and urban regions as national and international nodes for the transfer and sharing of knowledge. *Regional Studies, 37*(6-7), 607-620. https://doi.org/10.1080/0034340032000108714

Stek, P.E., and van Geenhuizen, M.S. (2016). The influence of international research interaction on national innovation performance: A bibliometric approach. *Technological Forecasting and Social Change, 110*, 61-70. https://doi.org/10.1016/j.techfore.2015.09.017

Thompson, P., and Fox-Kean, M. (2005). Patent citations and the geography of knowledge spill-overs: A reassessment. *American Economic Review, 95*(1), 450-460. https://doi.org/10.1257/0002828053828509

Thompsori, P. (2006). Patent citations and the geography of knowledge spillovers: Evidence from inventor- and examiner-added citations. *Review of Economics and Statistics, 88*(2), 383-388. https://doi.org/10.1162/rest.88.2.383

You, H., Li, M., Hipel, K.W., Jiang, J., Ge, B., and Duan, H. (2017) Development trend forecasting for coherent light generator technology based on patent citation network analysis. *Scientometrics 111*(1), 297-315. https://doi.org/10.1007/s11192-017-2252-y