The Measurement of Productive Efficiency in Scientific Journals through SFA models: Application to Quantitative Economics Journals

Authors:
Francisco J. Ortega
Department of Applied Economics I. University of Seville. Faculty of Economics and Business Studies. Avenida Ramón y Cajal nº 1. 41018. Seville (Spain). E-mail: fjortega@us.es

Jose M. Gavilan*
Department of Applied Economics I. University of Seville. Faculty of Economics and Business Studies. Avenida Ramón y Cajal nº 1. 41018. Seville (Spain). E-mail: gavi@us.es. Telephone +34 954 556 970. Fax: +34 954 551 636.

Abstract:

The importance of a scientific journal is usually established by considering the number of citations received by the papers that the journal publishes. In this way, the number of citations received by a scientific journal can be considered as a measure of the total production of the journal. In this paper, in order to obtain measures of the efficiency in the production process, the approach provided by Stochastic Frontier Analysis (SFA) is considered, and econometric models are proposed. These models estimate a frontier production which is the maximum achievable number of citations to the journal based on its resources. The efficiency can then be measured by considering the difference between the actual production and the estimated frontier. This approach is applied to the measurement of the productive efficiency of the journals of the JCR Social Sciences Edition database, which belong to both areas of "Economics" and "Social Sciences, Mathematical Methods".

Key words: Production, productivity, efficiency, scientific production, frontier production models, panel data models.

MSC Codes: 62J99, 62F10, 90B30.
JEL Codes: C46, C51, D24.

* Corresponding Author.

1. Introduction

The content of a scientific document is commonly supported through its references to other previously published scientific documents. In general, the importance of a specific written document is established on the principle that the more times the document is cited by the scientific community, the more important it becomes. As a consequence, the scientific journals receiving a high quantity of citations are considered to be the most significant.

It is clear that the evaluation of a journal based on the number of citations received presents certain issues, derived from the fact that citations to a paper are not always associated to the usefulness of its content, but can be motivated by other reasons (Callon et al. 1995; Ortega, 2003). In practice, however, this is the criterion most commonly used.

Generally speaking, a productive process involves the use of a series of resources (called *inputs*) in order to obtain another series of products (called *outputs*) which constitute the *production*. The *productivity* is defined as the ratio between the obtained production and the used resources.

Taking into consideration that one of the main objectives of a scientific journal is to obtain a high number of citations, the number of citations received during a specific period of time can be considered as a measure of its production. The measurement of productivity in relation to one single factor or input, in the form of the total number of published papers, is also widely employed. To this end, the Impact Factor is used (Basulto & Ortega, 2005; Ortega, 2003), which is calculated by dividing the total number of citations received by a journal in a 2-year period (production) by the total number of papers that have been published in the journal during that time (input).

When analysing the productivity, more determinant factors of the production can be considered. In that case, a journal could, compared to another journal, present greater or lesser productivity according to the input taken into account. One possible solution to this situation is to construct aggregate indices of productivity (Coelli et. al, 1998).

In the fields of economics, mathematics and econometrics, models have been developed in order to study the aforementioned problem from a different perspective: from a series of observations of several companies (items or firms) concerning their total production (through one or more outputs) and their resources (through several productive factors or inputs), the purpose of these types of models is to identify which firms make better use of their available resources, that is to say, which companies carry out the productive process with a higher *efficiency*.

There exist two alternative approaches to the problem of measuring the efficiency: Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). In Coelli et al. (1998), a detailed exposition of the two methodologies is offered. In an informal way it can be stated that the aim of these models is to establish *a frontier production* from the observed set of data that determines the maximum attainable output using the given inputs. This goal provides the reason for the generic name *frontier production models*. In this way, for each firm, the value of the maximum attainable production is estimated on the basis of “its current possibilities”. The determination of the difference between the actual production and the maximum possible production enables indicators of the efficiency of the productive process (in the sense that the nearer the actual production is to the estimated maximum attainable production, the more efficient a firm is).

It is important to point out that these efficiency measures have to be understood in relative terms with respect to the group analysed. A firm may well appear to be highly efficient when analysed among one group of companies, however, if this same firm is analysed among another different group of firms, it could appear to have a much lower level of efficiency.

In the framework of scientific documentation, the DEA approach has been widely utilised. As a matter of fact, all the papers mentioned below use this approach.

Abbott and Doucoubiagos (2003) carry out an analysis of the efficiency of Australian universities. Similar studies include Abramo and D'Angelo (2009) and Abramo et al. (2011), where Italian universities are analysed, and Bonaccorsi and Daraio (2003), who consider institutes of the French INSERM and biomedical research institutes of the Italian CNR.

Ruiz et al. (2010) examine the efficiency in the scientific production of a sample of Colombian research groups. Wang and Guan (2005) also study the efficiency of research groups, in this case from China. Agasisti et al. (2012) analyse the production of 69 academic departments located in Italy.

The DEA approach has also been applied to studies of efficiency in scientific production in a number of countries and regions. In Rousseau and Rousseau (1998), this approach is applied to a sample of 18 countries of the world, while in Guan and Chen (2010), 30 Chinese provinces are considered.

In relation to the analysis of the efficiency applied to a group of journals, Lozano and Salmerón (2005) show the results of a DEA analysis applied to a group of journals of Operations Research/Management Sciences in two aspects: the duration of the process of revision/publication and the relation between the impact and the length of the papers. Petridis et al. (2013) provide an evaluation of 54 forestry journals.

The main objective of this paper is to use the SFA approach in order to make an analysis of the production efficiency of a set of scientific journals (all of which belong to a homogeneous area) and to establish which journals produce at a higher level of efficiency (that is, making the most of their available inputs). Therefore, the goal is to identify which journals of the group, within their capabilities, obtain production close to their maximum, and which journals are currently far from such a maximum. The main innovation of this paper in relation to the aforementioned work is the use of the SFA approach.

To this end, in Section 2, the journals included in the present study and the factors selected to establish the frontier production (estimated maximum number of attainable citations for the journal) are presented. In Section 3, the statistical model and the set of data used in this study are described. In Section 4, the estimated model and the interpretation of the results obtained are presented. Finally, in Section 5, the main conclusions drawn from the study are discussed.

2. Journals included and variables selected

In order to select a homogeneous set of journals, the JCR Social Sciences Edition database has been used. In this database, the journals belonging simultaneously to the areas of "Economics" and "Social Sciences, Mathematical Methods" that appear indexed from 2008 to 2011 have been chosen, resulting in a total of 21 journals. The list of the selected journals is given in Table 1.
The selected output (variable Y) is the total number of citations to each journal in the database for 2011. This variable measures the total production of each journal in that year.

The selection of suitable inputs is crucial for the establishment of the level of efficiency in the productive process. Moreover, it is always important to take into account that the indicator of efficiency obtained depends on the selected factors, since the determination of the maximum frontier wholly relies on these factors.

One of the factors to take into consideration is the number of papers published by the journal. As previously pointed out, this input is considered when calculating the Impact Factor as a measure of productivity, which is based on the total number of citations and the total number of published articles (although the data of two consecutive years is aggregated). However, the total number of citations received by a journal in a specific year (2011, in this case) is a set of data retrospectively obtained (Gupta, 1997; Basulto & Ortega, 2005), that is, the citations refer to articles that have been published in previous years. Therefore, we consider that it is more suitable to designate the input as the average number of articles published by the journal in the past, instead of using the number of articles published in a specific year. As proxy variable to this measure, the average number of articles published during the three previous years (that is, 2010, 2009 and 2008) has been used; this variable is denoted NA.

In order to obtain the maximum potential of production of a journal, we believe that the quantity of researchers that have read the journal in the past constitutes a further determining factor. Although it is not standard practice, as proxy variable to this input, the average of the Impact Factor obtained by the journal in the period 2008-2010 is used, and this variable is denoted IF. The reason for this election is due to the fact that the Impact Factor can be understood as a measure of the visibility reached by a journal (Callon et al. 1995; Basulto & Ortega 2005). A high Impact Factor indicates that the journal has reached a large quantity of readers, who in turn have used the read contents for their own publications. It is also known that, given the significance that this indicator has gained in the evaluation of scientific activity, researchers have a greater interest in publishing in those journals with the highest Impact Factors, and therefore they are far more interested in the contents published in these journals. It is important to point out that, although the impact factor could be considered as an output, the fact of using data from the past, makes it an input for the future; in other words, what a journal has achieved with regard to visibility to date, constitutes a determinant factor of what the journal could achieve in the future. The fact that the group of journals is reasonably homogeneous, in the sense that all of these journals belong to the area of quantitative methods in Economics, contributes towards the interpretation of the Impact Factor as measure of the visibility of the journal, since the potential number of readers of each journal of this group can be considered similar.

The data about the number of citations, the total number of published articles and the Impact Factor are all items included in the JCR Social Sciences Edition database of the Web of Science.

Finally, an input corresponding to the human factor, otherwise known as the labour factor, has been included in the model. Although information about the total number of people involved in the production of the journals remains unavailable, we have obtained the total number of members (Editors, Co-Editors, Associate Editors…) of the editorial boards, and named this variable ED. Despite the variations in the structure of these boards from one journal to another, this fact remains insignificant since only the total number of members is taken into account and not the mission of each board member. This data has been obtained from the information published on the webpages of each of the journals in March 2013.

A further input, related to the capital factor of the journals, such as their budgets, would prove interesting as an inclusion in the model. However, we have no access to this information.

3. Model and set of data used

Using the output and the inputs described in the previous section, the econometric model proposed for the determination of the efficiency indicators is a Cobb-Douglas production model with stochastic frontier (Aigner et al., 1977, and Meeusen & van den Broeck, 1977), where all the variables are considered in logarithmic terms. Therefore, the proposed model is

$$log(Y_i) = \beta_0 + \beta_1 log(IF_i) + \beta_2 log(NA_i) + \beta_3 log(ED_i) + \underbrace{v_i - u_i}_{\varepsilon_i}, \quad i = 1, \dots, 21. \tag{1}$$

The random perturbations ε_i are composed of two parts (for this reason, a number of authors say that this is a composed error model): $v_i \in \Re$, which represents the random sources of variation; and $u_i > 0$, which corresponds to the inefficiency in the productive process. The perturbations v_i, as usual, are assumed to follow a Normal distribution, specifically, $v_i \sim N(0, \sigma_v^2)$. With regard to the perturbations u_i representing the effect of the inefficiency in the production, the most common hypothesis is also assumed, that $u_i = \left|u_i^*\right|$, where $u_i^* \sim N(0, \sigma_u^2)$. By definition, it is stated that the perturbations u_i follow a Half-Normal distribution, which is represented by $u_i \sim HN(0, \sigma_u^2)$, this distribution is a truncated normal to its positive values. Additionally, it is supposed that all the perturbations (both v_i and u_i) are independent.

After having estimated the parameters of the model, estimations of the inefficiency (or efficiency) of each journal in the productive process can be obtained: the main goal of this study. In the proposed model, where the production is expressed as a logarithm, the efficiency is usually defined as $EF_i = \exp\{-u_i\}$, thereby obtaining a variable ranging in (0,1]. The value 1 corresponds to the case $u_i = 0$, that is, a maximum degree of efficiency. In contrast, the smaller $\exp\{-u_i\}$ represents the lowest degree of efficiency of the journal.

It should be borne in mind that, in order to obtain the maximum likelihood estimator, a reparametrisation of the initial model is considered. Specifically, instead of using the parameters σ_v^2 (variance of the perturbation associated to the random effects) and σ_u^2 (variance of the perturbation associated to the inefficiency), the parameters $\sigma_\varepsilon^2 = \sigma_v^2 + \sigma_u^2$ (total variance of the perturbation, since it assumed that all the perturbations are independent) and $\gamma^2 = \sigma_u^2 / \left(\sigma_v^2 + \sigma_u^2\right)$ are utilised. This last parameter represents the proportion of the total variance of the perturbation term due to inefficiency. It is clear that $0 \le \gamma^2 \le 1$.

The analysis of the parameter γ^2 is crucial, since it determines the presence or absence of differences in the efficiency of the journals. As a matter of fact, if this parameter is not significantly different from zero, then it is implied that the whole error term is due to random effects, in other words, there exist no differences with respect to the level of production efficiency in the journals of the analysed group.

The proposed model has been estimated by the method of maximum likelihood by using the software R, specifically version 0.997-14 of the package *frontier*, which utilises the source code FORTRAN of the software FRONTIER 4.1. In Coelli (1996), both a description of the implemented method and the various types of models that can be estimated are presented.

The selected journals and the set of data obtained for the estimation of the proposed model are listed in Table 1.

Table 1. Journals selected and set of data obtained.

Abbreviated Title of Journal	Y	IF	NA	ED
ASTIN BULL	408	0.779	33.00	19
ECONOMET J	510	0.725	32.00	57
ECONOMET REV	650	1.351	25.67	20
ECONOMET THEOR	1507	0.842	67.67	52
ECONOMETRICA	19659	3.683	57.67	57
EMPIR ECON	783	0.571	67.67	55
INSUR MATH ECON	1567	1.205	123.00	41
INT J GAME THEORY	780	0.527	43.67	54
J APPL ECONOMET	2315	1.417	45.00	72
J BUS ECON STAT	2919	1.701	34.33	82
J ECONOMETRICS	8523	1.836	133.00	61
J MATH ECON	893	0.466	88.33	50
J PROD ANAL	1101	0.643	33.33	55
JAHRB NATL STAT	113	0.239	30.67	21
MATH FINANC	1098	1.168	28.67	44
OXFORD B ECON STAT	1501	0.993	37.67	18
QME-QUANT MARK ECON	240	1.078	15.00	68
QUANT FINANC	712	0.701	77.00	67
REV ECON STAT	7639	2.557	66.00	49
SOC CHOICE WELFARE	725	0.615	71.00	68
THEOR DECIS	511	0.650	43.00	69

4. Estimated Model. Interpretation of the results

The estimated parameters of the model considered in (1) together with their standard errors, the values of its z-statistics and the p-values for their significance are presented in Table 2.

In the last column, the parameters have been marked with their resulting significance at 10% (*), 1% (**) or 0.1% (***). All of these parameters, except the slope corresponding to the variable ED, are significant.

Table 2. Estimated Model.

Parameters	Estimations	Standard Errors	Z-values	P-values
β_0	3.9968	1.1860	3.3699	0.0008***
β_1	1.5181	0.1414	10.7336	0.0000***
β_2	0.8075	0.2319	3.4824	0.0005***
β_3	0.1255	0.2129	0.5895	0.5555
σ_ε^2	0.3895	0.2303	1.6911	0.0908*
γ	0.8125	0.2961	2.7445	0.0061**

It is important to emphasize that, given the selected inputs, this model does not estimate the average value of the production of the journals, but instead estimates the maximum frontier of attainable production. As expected, all three estimated slopes are positive. Therefore, the estimated model confirms that an increase in any of the resources of a journal involves an increment in its frontier production, in other words, the journals with the greatest resources can reach the highest levels of received citations. The estimation of the parameter γ together with its standard error and its p-value indicate that the hypothesis $\gamma = 0$ must be rejected, which means that there exist significant differences in the efficiency levels of the group of journals analysed. From these results, measures of the efficiency of the journals under consideration can be obtained, which constitutes the main goal of this kind of model, as explained in the previous section. For each of the analysed journals, such estimations are shown in descending order in Table 3 and are graphically represented in Figure 1.

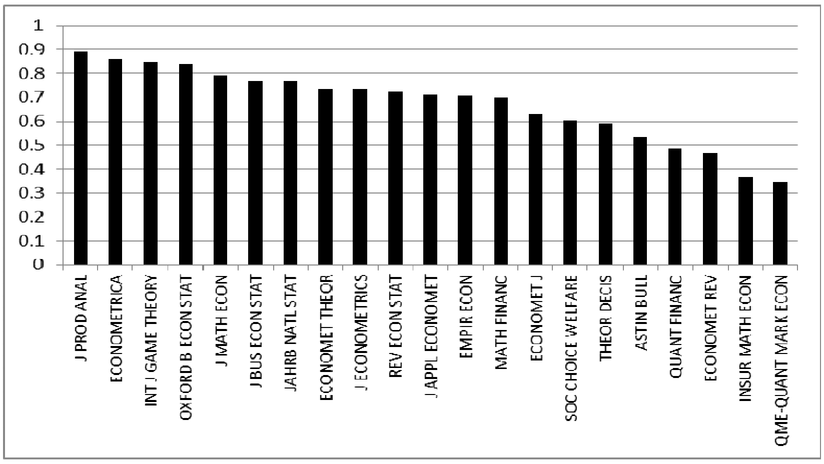

Figure 1. Estimated Efficiency of the journals.

The journal with the highest level of efficiency is J PROD ANAL, whose total production (number of received citations in 2011), as seen in Table 3, is not especially outstanding. However, such production is reached with “moderate” resources with respect to the average visibility in the three previous years (IF), and particularly in relation to the average number of published articles (NA). This journal is closely followed by ECONOMETRICA, INT J GAME THEORY and OXFORD B ECON STAT. In the case of ECONOMETRICA, this journal presents by far the greatest production of the group analysed, while it is true that its initial resources are highly superior, particularly in visibility (IF). On the other hand, INT J GAME THEORY has a low number of citations in 2011 and at the same time presents a high level of efficiency. This is due to the fact that, despite its scarcity of resources, it reaches, with these limited resources, a production close to its estimated maximum.

The journals presenting the lowest levels of efficiency are QME-QUANT MARK ECON and INSUR MATH ECON. The situation of the latter journal is especially worth mentioning; despite its relatively high production in the group, its level of efficiency is one of the lowest, owing to fact that it possesses high levels of inputs, particularly in the variable NA.

Table 3. Estimated Efficiency of the journals.

	Abbreviated Title of Journal	**Efficiency**
1	J PROD ANAL	0.8905
2	ECONOMETRICA	0.8581
3	INT J GAME THEORY	0.8496
4	OXFORD B ECON STAT	0.8396
5	J MATH ECON	0.7924
6	J BUS ECON STAT	0.7671
7	JAHRB NATL STAT	0.7645
8	ECONOMET THEOR	0.7377
9	J ECONOMETRICS	0.7317
10	REV ECON STAT	0.7244
11	J APPL ECONOMET	0.7104
12	EMPIR ECON	0.7100
13	MATH FINANC	0.6998
14	ECONOMET J	0.6307
15	SOC CHOICE WELFARE	0.6024
16	THEOR DECIS	0.5887
17	ASTIN BULL	0.5356
18	QUANT FINANC	0.4853
19	ECONOMET REV	0.4668
20	INSUR MATH ECON	0.3655
21	QME-QUANT MARK ECON	0.3450
	Mean Efficiency	**0.6712**

With regard to the joint analysis of the estimated efficiencies, in Figure 2 the density fitted to all 21 journals under consideration is presented. This density has been fitted while taking into consideration that the efficiencies are bounded between 0 and 1 (Silverman (1986)). The average efficiency of the group is 0.67, the mode is approximately 0.72, and a high percentage of journals have efficiency above 0.8. All these figures indicate that the considered journals operate, jointly and in absolute terms, at a high level of efficiency, since the mean efficiency obtained takes a high value (taking into account that the maximum reachable value is 1). This type of analysis can be used to compare the level of efficiency between a number of groups of journals and/or several periods of time.

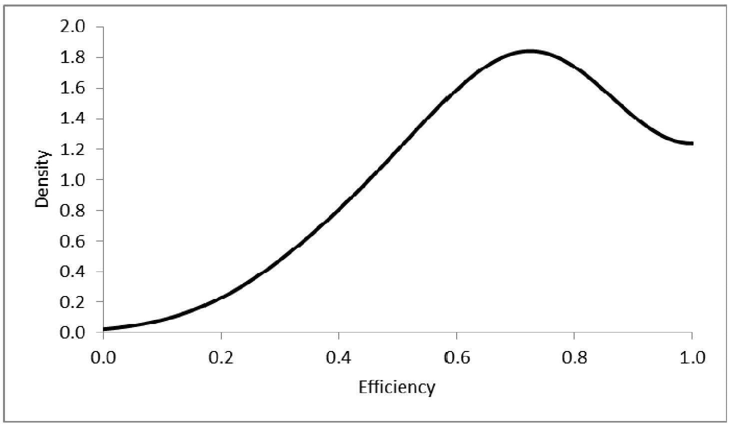

Figure 2. Density fitted to the estimated efficiencies.

On the other hand, in Table 2, it can be seen that the inputs IF and NA are highly significant in the determination of the frontier production, whereas the variable employed to quantify the labour factor (ED) remains insignificant at the usual levels. In other words, the estimated model indicates that the journals that have had in the recent past a higher Impact Factor and that have published a greater number of papers, are the journals that, potentially, are able to reach a greater number of citations in the analysed period. This result is corroborated by the coefficient of linear correlation between the values of the estimated frontier and all the three explanatory variables included in the model and by carrying out the corresponding significance tests of such coefficients. The results are presented in Table 4, where it can be observed that the coefficients of correlation corresponding to IF and NA differ significantly from 0.

Table 4. Linear correlation between the explanatory variables and: (i) the estimated frontier; and (ii) the efficiency.

	FRONTIER		**EFFICIENCY**	
	Correlation	**p-value**	**Correlation**	**p-value**
IF	0.9231	0.0000	0.1543	0.5042
NA	0.4545	0.0385	-0.0605	0.7944
ED	0.1560	0.4996	0.0117	0.9597

Despite these observations, it cannot be concluded that the level of efficiency in the production is fundamentally determined by the Impact Factor and/or the number of published papers (in the sense that those journals with a higher Impact Factor and/or those that have recently published more articles, produce with greater efficiency). In fact, no direct relation between the level of efficiency and any of the explanatory variables exists, as the correlation coefficients show in Table 4. Hence, the inputs or resources determine the frontier production, but they do not dictate the level of efficiency. Neither is there any direct relation between the efficiency and the actual production, since the correlation coefficient between the variable Y (number of received citations in 2011) and the efficiency takes the value 0.3440, with a p-value for the significance test equal to 0.1267.

5. Conclusions

In this paper, a new concept in scientific documentation is addressed: the measurement of the efficiency in the production of scientific journals. In this framework, the objective, instead of being the assessment of the volume of production or total number of citations received, becomes the production achieved in relation to the starting possibilities of the journals.

As indicated in the introduction, the Impact Factor takes account of the total production (citations) in relation to a single input, which is the number of published papers, thereby representing a productivity index. Since more than one determinant factor has been taken into account in the production process, econometric models previously developed in the field of the applied economics are used. These models estimate the frontier production or maximum number of attainable citations for a scientific journal on the basis of its initial resources, thus attaining an indicator of the efficiency that depends on the difference between the obtained citations and their maximum possible value.

One major inherent issue is the choice of the inputs and the way in which they are measured, since these decisions are determinant in the results. In this paper, the Impact Factor (IF) as a measure of visibility of the journal, the average number of articles published in the period 2008-2011 (NA), and the number of members of the editorial boards taken as the labour factor (ED) are the inputs selected. It would be interesting to be able to include an input related to the capital factor, such as the budget of the journals, but to date, no access to this information has been made available. The results obtained also depend on the group of journals being analysed, since although the indicators calculated cannot be understood in absolute terms, they do offer a comparison between the selected journals. In the case analysed in this paper, from among the 21 journals belonging to the area of quantitative methods in Economics, the group journals with the highest level of efficiency are J PROD ANAL, ECONOMETRICA, INT J GAME THEORY, and OXFORD B ECON STAT. On the other hand, the journals with the lowest levels of efficiency are QME-QUANT MARK ECON and INSUR MATH ECON. It is also worth noting that, as a group, the selected journals operate at a high level of efficiency with an average equal to 0.67.

With regard to the results stemming from the estimation of the model, it should be borne in mind that the most influential factors in the determination of the frontier production (the maximum attainable number of received citations) are the Impact Factor and the number of published papers. Moreover, the level of efficiency does not depend on any specific input, but on a combination of all inputs involved.

References

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis approach. *Economics of Education Review*, vol. 22, 89-97.

Abramo, G., & D'Angelo, C.A. (2009). Assesing technical and cost efficiency of research activities: a case study of the Italian university system. *Research Evaluation*, vol. 18, 61-70.

Abramo, G., Cicero, T., & D'Angelo, C.A. (2011). A field-standardized applicaction of DEA to national-scale research assessment of universities. *Journal of Informetrics*, vol. 5, 618-628.

Agasisti, T., Catalano, G., Landoni, P., & Varganti, R. (2012). Evaluating the performance of academic departments: an analysis of research-related output efficiency, *Research Evaluation*, vol. 21, 2-14.

Aigner, D.J., Lovell, C.A.K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, vol. 6, 21-37.

Basulto, J., & Ortega, F.J. (2005). Modelling citation age data with right censoring. *Scientometrics*, vol. 62 (3), 329-342.

Bonaccorsi, A., & Daraio, C. (2003). A robust nonparametric approach to the analysis of scientific productivity, *Research Evaluation*, vol. 12, 47-69.

Callon, M., Courtial, J.P., & Penan, H. (1995). *Cienciometría. La medición de la actividad científica: de la Bibliometría a la vigilancia tecnológica*. Gijón, Ediciones Trea (in Spanish).

Coelli, T.J. (1996). A guide to FRONTIER version 4.1: a computer program for frontier production function estimation. CEPA Working Paper 96/07, Department of Econometrics, University of New England, Armidale, Australia.

Coelli, T.J., Rao, D.S.P., & Battese, G.E. (1998). *An introduction to efficiency and productivity analysis*. Boston, Kluver Academic Publishers.

Guan, J., & Chen, K. (2010). Modeling macro-R&D production frontier performance: an application to Chinese province-level R&D, *Scientometrics*, Vol. 82 (1), 165-173.

Gupta, B.M. (1997). Analysis of distribution of the age of citations in theoretical population genetics. *Scientometrics*, Vol. 40 (1), 139-162.

Lozano, S., & Salmerón, J. L. (2005). Data Envelopment Analysis of OR/MS journals, *Scientometrics*, Vol. 64 (2), 133-150.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, vol. 18, 435-444.

Ortega, F.J. (2003). Propuesta de mejora del índice agregado de impacto. Aplicación a la valoración de sexenios de investigación. *Revista Española de Documentación Científica*, vol. 26 (4), 403-417, (in Spanish).

Petridis, K., Malesios, C., Arabatzis, G., & Thanassoulis, E. (2013). Efficiency analysis of forestry journals: Suggestion for improving journals' quality, *Journal of Informetrics*, vol. 7, 505-521.

Rousseau, S., & Rousseau, R. (1998). The scientific wealth of European nations: Taking effectiveness into account, *Scientometrics*, Vol. 42 (1), 75-87.

Ruiz, C. F., Bonilla, R., Chavarro, D., Orozco, L. A., Zarama, R., & Polanco X. (2010). Efficiency measurement of research groups using Data Envelopment Analysis and Bayesian networks, *Scientometrics*, Vol. 83 (3), 711-721.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Wang, J., & Guan, J. (2005). The Analysis and Evaluation of Knowledge Efficiency in Research Groups, *Journal of the American Society for Information Science and Technology*, vol. 56, 1217-1226.