



Citescore of Publications Indexed in Scopus: An Implementation of Panel Data

Carolina Henao-Rodríguez¹(✉), Jenny-Paola Lis-Gutiérrez²,
Carlos Bouza³, Mercedes Gaitán-Angulo², and Amelec Viloria⁴

¹ Corporación Universitaria Minuto de Dios, Bogotá, Colombia
linda.henao@uniminuto.edu

² Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
{jenny.lis, mercedes.gaitana}@konradlorenz.edu.co

³ Universidad de La Habana, La Habana, Cuba
bouza@matcom.uh.cu

⁴ Universidad de la Costa, Barranquilla, Colombia
aviloria7@cuc.edu.co

Abstract. This article is intended to establish the variables that explain the behavior of the CiteScore metrics from 2014 to 2016, for journals indexed in Scopus in 2017. With this purpose, journals with a CiteScore value greater than 11 were selected in any of the periods, that is to say, 133 journals. For the data analysis, a model of standard corrected errors for panel was used, from which a coefficient of determination of 77% was obtained. From the results, it was possible to state that journals of arts and humanities; business; administration and accounting; economics, econometrics, and finance; immunology and microbiology; medicine and social sciences, have the greatest impact.

Keywords: CiteScore · Publications · Journals · Indexing · Scopus

1 Introduction

When considering the impact of publications, two approaches can be identified. The first one associated with the analysis of impact indicators of journals, and the second one related to the importance given by the institutions and researchers regarding the impact assessments. Among the empirical studies included in the first group, the one carried out by [1] shows the results of the implementation of a quantile regression for predicting a probability distribution to set the future number of quotations from a publication. In the same way, [2] reports that the publications of the first quarter of the year had a number of citations higher than the papers published in the last quarter. A similar finding associated with the month of publication was found by [3], for the citations during the two years following the publication.

For their part, [4] reported that self-citations of journals greatly affect the impact factors of publications at meso and micro levels, since it is a way of artificially increasing the impact assessment indicators. This fact severely undermines the authenticity of the indicators. [5] discerned on five types of citations: application, affirmation, negation, review and perfunctory mention. Their results encourage us to go

beyond the citation counts to assess the scholar scientific contribution of a paper, through the implementation of a panel data model with fixed effects. At the same time, [6], propose other impact assessments, based on the journal editor's data for evaluating the scholarly impact of an academic institution.

Within the second approach stands out the study of [7] who identified that researchers do not interpret the number of citations and the perceived impact to the same extent. Also, they stated that, in contexts of expertise in a specific area, there is a bias to prefer their own publications. [8], based on a survey applied to researchers in the United Kingdom and the development of two logit models, concluded that institutional factors, the contexts of intensive research, and non-academic work experience, explain the preference of researchers for high impact of their publications.

It is worth mentioning some other studies of multivariate techniques for measuring the impact of journals, such as [5, 6, 8, 9], among others.

This paper seeks to establish the variables explaining the behavior of the CiteScore from 2014 to 2016, for journals indexed in Scopus with a higher value in this indicator at 11. For this purpose, a panel data model was applied with CiteScore as dependent variable, and with the following explanatory variables: coverage years of the journal, languages in which the journal publishes, type of access to the journal, H-index of the journal origin country, and the dichotomous variables associated with the journal study area.

2 Method

2.1 Data

The data compiled to build the model were obtained from the Scopus Web Page in the following link: https://www.scopus.com/sources.uri?DGCID=Scopus_blog_post_check2015, and the Scopus Source List, available in October 2017. Firstly, a filter was applied to get a list of active publications and leaving the inactive ones out of the search. Secondly, a filter was applied regarding the type of source, choosing the "Journal" and "Trade Journal" options for this study.

2.2 Variables

An econometric exercise was conducted to analyze the CiteScore determinants. To avoid bias due to the heterogeneity of the individuals analyzed, the panel data econometric technique was applied for avoiding issues related to the identification of the models. The study period is between the years 2014 and 2016, and the technique was applied to 133 journals indexed in Scopus during 2017, with a CiteScore value greater than eleven, in any of the years of the mentioned period.

As mentioned above, the dependent variable used was the CiteScore, and the explanatory variables are: (i) The journal coverage years. A positive relationship with the CiteScore is expected since the journal has more years of coverage, so there will be a greater likelihood of citation. (ii) The journal coverage years. A positive relationship with the CiteScore is expected since the journal has more years of coverage, so there

will be a greater likelihood of citation. (iii) The languages in which the journal publishes. It was intended to confirm if there is any empirical evidence if the greater the number of languages of publication, the more citations it has. (iv) A dichotomous variable that specifies whether it is an open-access journal or not. In general, it is expected that open-access journals are most often cited. (v) The H-index of the journal origin country as a control variable. It is expected that the citation shows a positive relationship with the quality and the impact of research in its environment. (vi) Dichotomous variables associated with the study area of the journal: biological and agricultural sciences; arts and humanities; biochemistry; genetics and molecular biology; business administration, and accounting; chemical engineering; chemistry; decision sciences; earth and planetary sciences; economics, econometrics and finance; engineering, environmental sciences; immunology and microbiology; sciences of materials; mathematics; medicine, neuroscience, nursing, pharmacology, toxicology and pharmacy; psychology, social sciences, and health professions. The degree of citation is expected to be affected by the science of study¹.

2.3 Model

The model is specified as follows:

$$\begin{aligned} \text{logciteScore}_{jt} = & \beta_0 + \beta_1 \text{logcobertura}_{jt} + \beta_2 \text{logidiomas}_{jt} + \beta_3 \text{open}_{jt} + \beta_4 \text{loghindex}_{it} + \beta_5 x_{it} \\ & + \beta_6 c_{it} + \beta_7 v_{it} + \beta_8 b_{jt} + \beta_9 n_{jt} + \beta_{10} m_{jt} + \beta_{11} h_{jt} + \beta_{12} f_{jt} + \beta_{13} g_{jt} + \beta_{14} k_{jt} + \beta_{15} d_{jt} \\ & + \beta_{16} s_{jt} + \beta_{17} a_{jt} + \beta_{18} q_{jt} + \beta_{19} w_{jt} + \beta_{20} e_{jt} + \beta_{21} r_{jt} + \beta_{22} t_{jt} + \beta_{23} u_{jt} + \beta_{24} i_{jt} + \beta_{25} qa_{jt} \\ & + \eta_j + \delta t + \varepsilon_{jt}. \end{aligned} \quad (1)$$

j corresponds to the journal; t represents the year; i is the journal origin country; logciteScore is the logarithm of CiteScore; logcobertura is the logarithm of the journal coverage years; logidiomas is the logarithm of the number of languages in which the journal publishes; open is a dichotomous variable which specifies if it is open-access; loghindex is the H-index logarithm; while the dichotomous variables that specify the study area are: agricultural and biological sciences (x); arts and humanities (c); biochemistry; genetics, and molecular biology (v); business; administration and accounting (b); chemical engineering(n); chemistry (m); decision sciences (h); earth and planetary sciences (f); economics, econometrics and finance (g); engineering (k); environmental sciences (d); immunology and microbiology (s); sciences of materials (a); mathematics (q); medicine (w); neuroscience (e); nursing (r); pharmacology, toxicology and pharmacy (t); psychology (u); social sciences (i); health professions (qa); δt represents the effects that vary with time not observed, η_j captures a common deterministic trend and ε_{jt} is a random disturbance that is supposed $\varepsilon_{jt} \sim N(0, \sigma^2)$.

¹ The dichotomous variables specifying the following study areas were omitted: General; Computer Science; Energy; Physics and Astronomy; Veterinary Medicine, and Odontology, since they caused multicollinearity in the proposed model.

3 Results

To establish the best model, some tests were carried out, with the results shown below. When completing the Hausman test, the null hypothesis of the difference between the coefficients of random and fixed effects is rejected. Therefore, the estimation of fixed effects is performed.

Test: Ho: Difference in coefficients not systematic.

Chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 5.92

Prob > chi2 = 0.0149

When completing the Wooldridge test to detect autocorrelation, the null hypothesis that there is no first order autocorrelation at a significance level of 1% is rejected. Wooldridge test for autocorrelation in panel data.

H0: no first-order autocorrelation

F (1, 122) = 17.592

Prob > F = 0.0001

When completing the Wald test to detect if there is heteroskedasticity in the model, the null hypothesis of homoskedasticity at the significance level of 1% is rejected. Modified Wald test for groupwise heteroskedasticity in fixed effect regression model.

H0: $\sigma(i)^2 = \sigma^2$ for all i

Chi2 (127) = 7.4E + 07

Prob > chi2 = 0.0000

When completing the Pesaran test (2015) of cross-sectional dependence for not balanced panel data set; the null hypothesis cannot be rejected. Therefore, the errors of the cross-section are weakly dependent. Pesaran (2015) test for weak cross-sectional dependence unbalanced panel detected, test adjusted.

H0: errors are weakly cross-sectional dependent.

CD = 0,431

P-value = 0,667

The problems of contemporary correlation, heteroskedasticity and autocorrelation detected, can be solved along with estimates of standard corrected errors for panel. For this reason, the estimation was performed, and time dichotomous variables were included, due to the inability to directly capture the fixed effects (see Algorithm 1).

Algorithm 1. The Prais-Winsten regression; heteroskedastic panels corrected standard errors.

logcitescore	
logcobertura	-0.010 (0.16)
logidiomas	1.042 (2.06)*
open	-0.294 (2.00)*
loghindex	0.299 (3.69)**
x	0.127 (1.88)
c	0.208 (3.56)**
v	0.222 (6.07)**
b	0.027 (0.55)
n	0.090 (0.24)
u	(0.04) -0.442 (2.90)**
i	0.215 (5.48)**
qa	-0.262 (2.52)*
_Iyear_2015	0.026 (2.24)*
_Iyear_2016	-0.039 (0.22)
_cons	0.326 (0.56)
R2	0.78
N	375
* p<0.05; ** p<0.01	

The model of standard corrected errors for panel, has a determination coefficient of 78%. The significant variables in the model were the dichotomous variables that specify if the journal is open-access, and the following study areas: arts and humanities (c); biochemistry; genetics and molecular biology (v); engineering (k); economics, econometrics and finance (g); environmental sciences (d); science of materials (a);

medicine (w); neuroscience (e); nursing (r); psychology (u); social sciences (i); health professions (qa), and the dichotomous variable of time associated with the year 2015.

Given the above, empirical evidence shows that journals about arts and humanities; business; biochemistry; genetics and molecular biology, engineering, econometrics, and finance; medicine and social sciences, have the greatest impact, considering the assessment by CiteScore. In addition, the results show a negative relationship between the citation and the journals of environmental sciences; sciences of materials; neuroscience; nursing; psychology and health professions. This is consistent with the results shown by other authors. [10] presented the characteristics of the citations in the text in more than five million full-text articles from two databases (PMC, open-access subset and Elsevier journals) and found that the fields of biomedical and health sciences; life and earth sciences; and physics, science and engineering have similar reference distributions, although they vary in their specific details, while the fields of mathematics and informatics; and social sciences and humanities, have distributions of reference different from the other three.

In this context, [11] conclude that the average values of reference density in some categories of Social Sciences and Arts and Humanities, were equal to or higher than the “hard sciences” since the citations to the references occur at least with the same frequency in these two areas of knowledge, despite the potentially less impact of the journal.

Contrary to what was expected, if a journal is open-access, the number of citations does not increase, since the coefficient of the dichotomous variable “open” presents a negative sign that was not expected. Despite this, the finding is supported by other studies. For instance, [11], and [12] found no significant differences in the average values or the growth rates between Gold open-access and Non-Gold open-access journals, getting bibliometric and bibliographical data collections from 27.141 journals (indexed between 2001 and 2015 in the SCImago Journal & Country Rank (SJR)).

For their part, [13] conducted a study where they documented the growth in number of journals and articles, along with the increased rates of standardized citations of open-access journals listed in the Scopus bibliographic database, from 1999 to 2010. They concluded that the open-access journals and articles had grown faster than the subscription journals, but even so, they represented a very low percentage among the Scopus journals. [14] found that, by using a model of standard corrected errors for panel and a model of feasible generalized least squares, the open-access was not significant when considering the journals with an SJR greater than 8.

The variable “languages” is significant, meaning that the greater the number of languages in which the journal publishes, the greater the likelihood of citation. Finally, the H-index of the journal origin country is significant as a control variable and makes possible to confirm that the citation features a positive relationship with the quality and the impact of the research in its environment. According to the results found by [15], they argued that the number of citations is used to assess the impact of academic research or the quality of an academic department and reported that there are other important factors different from the journal, including the length of the article, the number of references, and the status of the first author’s institution. Therefore, it is expected that the relevance of the research in the journal origin country is greater, due to a higher quality of the educational institutions.

4 Conclusions

The model of standard corrected errors for panel has a determination coefficient of 77%, and the significant variables in the model at 5% of the level of significance are the dichotomous variables that specify if the journal is open-access, and the following study areas: arts and humanities, business administration and accounting; economics, econometrics and finance; immunology and microbiology; sciences of materials; medicine; neuroscience; nursing; psychology; social sciences; health professions, and the dichotomous variable of time associated with the year 2015.

The model shows that open-access does not present the expected sign; therefore, it cannot be said that an open-access journal presents a greater impact indicator, confirming this way the findings of Sánchez, Gorraiz, and Melero (2018). In addition, empirical evidence shows that journals of arts and humanities; business; administration and accounting; economics, econometrics, and finance; immunology and microbiology; medicine and social sciences, have the greatest impact.

A similar analysis is suggested to be applied in future researches on journals indexed in WoS; as well as the use of other indicators as independent variable (like in the case of CiteScore, SNIP, among others); comparative analysis between different areas of knowledge is finally suggested in order to identify whether the findings are maintained.

References

1. Stegehuis, C., Litvak, N., Waltman, L.: Predicting the long-term citation impact of recent publications. *J. Inform.* **9**(3), 642–657 (2015)
2. Donner, P.: Effect of publication month on citation impact. *J. Inform.* **12**(1), 330–343 (2018)
3. Kosteas, V.D.: Journal impact factors and month of publication. *Econ. Lett.* **135**, 77–79 (2015). <https://doi.org/10.1016/j.econlet.2015.08.010>
4. Yu, T., Yu, G., Wang, M.Y.: Classification method for detecting coercive self-citation in journals. *J. Inform.* **8**(1), 123–135 (2014)
5. Stremersch, S., Camacho, N., Vanneste, S., Verniers, I.: Unraveling scientific impact: citation types in marketing journals. *Int. J. Res. Mark.* **32**(1), 64–77 (2015)
6. Wu, D., Li, J., Lu, X., Li, J.: Journal editorship index for assessing the scholarly impact of academic institutions: an empirical analysis in the field of economics. *J. Inform.* **12**(2), 448–460 (2018). <https://doi.org/10.1016/j.joi.2018.03.008>
7. Radicchi, F., Weissman, A., Bollen, J.: Quantifying perceived impact of scientific publications. *J. Inform.* **11**(3), 704–712 (2017)
8. Salter, A., Salandra, R., Walker, J.: Exploring preferences for impact versus publications among UK business and management academics. *Res. Policy* **46**(10), 1769–1782 (2017)
9. Lando, T., Bertoli-Barsotti, L.: Measuring the citation impact of journals with generalized Lorenz curves. *J. Inform.* **11**(3), 689–703 (2017). <https://doi.org/10.1016/j.joi.2017.05.005>
10. Boyack, K.W., Van Eck, N.J., Colavizza, G., Waltman, L.: Characterizing in-text citations in scientific articles: a large-scale analysis. *J. Inform.* **12**, 59–73 (2018). https://ac.els-cdn.com/S1751157717303516/1-s2.0-S1751157717303516-main.pdf?_tid=a31a573c-aa04-43b7-abe8-d38f230004be&acdnat=1523576882_830f710bace5ba4add6778c9ac486eba

11. Sánchez, S., Gorraiz, J., Melero, D.: Reference density trends in the major disciplines. *J. Inform.* **12**, 42–58 (2018). https://ac.els-cdn.com/S1751157717301487/1-s2.0-S1751157717301487-main.pdf?_tid=919a6ae3-2f8c-45f6-b2a2-7d61f4927ab5&acdnat=1523577929_71c9094e2952f12c35958d1d3b02495e
12. Zhang, L., Watson, E.M.: Measuring the impact of gold and green open-access. *J. Acad. Librariansh.* **43**(4), 337–345 (2017). <https://doi.org/10.1016/j.acalib.2017.06.004>
13. Solomon, D., Laakso, M., Björk, B.: A longitudinal comparison of citation rates and growth among open-access and subscription journals. *J. Inform.* **7**(3), 642–650 (2013)
14. Henao-Rodríguez, C., Lis-Gutiérrez, J.-P., Gaitán-Angulo, M., Malagón, L.E., Vilorio, A.: Econometric analysis of the industrial growth determinants in Colombia. In: Wang, J., Cong, G., Chen, J., Qi, J. (eds.) *ADC 2018. LNCS*, vol. 10837, pp. 316–321. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92013-9_26. <https://www.springerprofessional.de/en/data-mining-and-big-data/15834080>
15. Mingers, J., Xu, F.: The drivers of citations in management science journals. *Eur. J. Oper. Res.* **205**(2), 422–430 (2010)