Singly Authored Papers Contribute the Most to Scientists' Impact

Wenhua Ai[1], Zhenyue Zhao[2] and Jiang Li[2]

[1] School of Management, Nanchang University, Nanchang 330000, China
2 School of Information Management, Nanjing University, Nanjing 210023, China
lijiang@nju.edu.cn

Abstract. Utilizing citation data for 100,000 most-cited scientists in the Scopus database, this paper investigated how citations received by an author in different authorship affect his/her academic impact differently. Using a linear regression model as an estimation, it shows that the citations received as the single author of a paper elevates the academic impact the most, followed by that as the first (but not single) author, last author, and middle author. Differences also emerged when we probed into different research fields separately as in some fields citations in the four types of authorship do not differ a lot, and also in some fields, the last-authored citations could 'outweigh' the first-authored ones.

Keywords: Authorship, Academic Impact, Linear Regression.

1 Introduction

It greatly satisfies humans' curiosity to make comparisons between relative individuals such as sportsmen engaged in the same sport, movie stars active in the same country and scientists dedicated to the same research field (or even different fields). A recent study by Ioannidis et al. [1] introduced a ranking of 100,000 most-cited global scientists across different research fields by their academic impact. Utilizing the data from the Scopus database, Ioannidis et al. proposed a new indicator 'composite index' (c),

$$c = \frac{ln(nc9617+1)}{ln(nc9617max+1)} + \frac{ln(h17+1)}{ln(h17max+1)} + \frac{ln(hm17+1)}{ln(hm17max+1)} + \frac{ln(ncs+1)}{ln(ncsmax+1)} + \frac{ln(ncsf+1)}{ln(ncsfmax+1)} + \frac{ln(ncsfl+1)}{ln(ncsflmax+1)}, \quad (1)$$

where *nc*9617 is the total number of citations received throughout the time of data collection (from 1996 to 2017), *h*17 and *hm*17 are respectively the *h*-index [2] and the Schreiber coauthorship-adjusted *hm*-index [3] up to 2017. *ncs*, *ncsf* and *ncsfl* are respectively the number of citations to singly-authored papers; the number of citations to first-author papers; and the number of citations to singly-authored, first-author, or last-author papers. Those with a 'max' suffix are the maximum value for all the scientists and are fixed as: *nc9617max* = 259,310, *h17max* = 222, *hm17max* = 103.981, *ncsmax* = 135,334, *ncsfmax* = 149,125 and *ncsflmax* = 163,476. The number of self-citations was all excluded in these citation metrics.

One novel idea in this ranking method is that it treats the citations a scientist obtained in different authorship differently, which well responds to such understanding that authorship provides an implication on how much a scientist contributed to a particular work [4]. The general idea is that the first author is the main contributor of a particular research work, who should be responsible for the idea, data, design and the writing; the last author is usually the corresponding author, who is generally in charge of the whole research project and should receive the updates of the status of the submission [5]. This poster tries to address the following research question:
Which one has the most contribution to a scientist's academic impact among the four types of authorship, i.e., single author, first author, last author, or the middle author?

2 Data and methodology

This study utilized the data files made public by Ioannidis et al. [1]. Data contains six citation metrics for each of the 100,000 most-cited scientists in the Scopus database. We used Equation (1) to calculate a composite score *c*, which measures the scientific impact of scientists and ranks all the scientists from the highest to the lowest.

A regression model was built by taking *c* (composite) as the dependent variable, and four variables were selected as the independent variables, i.e., the number of citations a scientist received as a single author, first author, middle author, and last author, respectively (these four variables are later referred to as nc_{single}, nc_{first}, nc_{middle}, and nc_{last}). Since Equation (1) clearly shows that there is nonlinear relationship between the four variables and *c*, we transform the four variables into $ln(nc_{single}+1)=X_1$, $ln(nc_{first}+1)=X_2$, $ln(nc_{middle}+1)=X_3$ and $ln(nc_{last}+1)=X_4$ following the construction of *c*, then, we build the linear regression model as follows,

$$c = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u \,. \quad (2)$$

By comparing the estimated coefficients of the four variables, an understanding of their significance to the academic impact (represented by the *c* score) could be uncovered. The same linear regression model was used after we separated the scientists into 21 groups by their scientific fields.

3 Results

Table 1 displayed some descriptive statistics of our variables in use. We found that the values of *c* score amass in a small range. The mean values of the four types of citations vary significantly, where the number of citations as a middle author is the biggest and that as the single author is the smallest, which is largely attributed to the variation of the number of publications in different types of authorship.

Table 1. Descriptive statistics of the variables

Variable	Min.	Max.	Mean	Variance	Skewness	kurtosis
c	2.1	5.7	3.7	.1	1.2	1.8
nc_{single}	.0	135,334	415.8	1,312,195.4	45.4	4179.6
nc_{first}	.0	131,396	1,396.7	3,555,232.6	13.7	517.3
nc_{middle}	.0	151,860	3,649.3	27,795,444.3	5.0	48.8
nc_{last}	.0	131,114	2,454.8	13,438,682.7	7.1	101.1

We drew four scatter diagrams to derive some information about the relationship between the number of four types of citations (in a logarithmic form) and the c score. There are similar patterns in the four sub-diagrams in which the dots distributed densely within a triangle zone, indicating a positive relationship between the c score and the logarithmic number of citations. However, if we fit the data with a linear function using OLS, the coefficients of the linear term can provide rough estimations for contribution brought by a unit citation in different authorship. As is indicated, if a scientist receives one citation, the c score may increase in varying degrees if he takes different authorship in this paper.

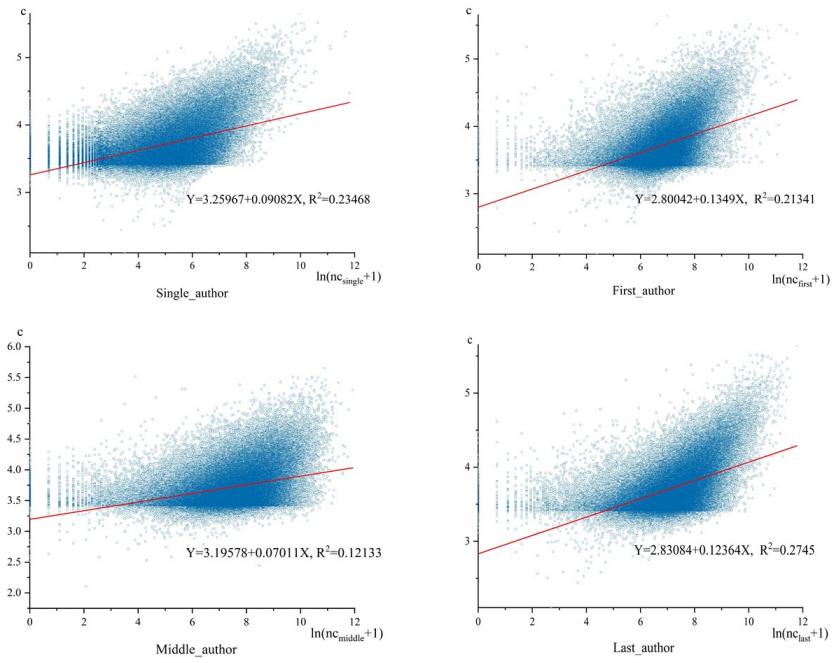

Fig. 1. The relative distribution of the composite c and the citations received as the four types of authorship

To measure the different extents to which the c score may increase after a citation, a more accurate estimation using linear regression was implemented. The estimated

model (R-square=0.793) and the coefficients show that one citation received as single author contributes the most (β_1=0.119, p<0.01), followed by the first author (β_2=0.111, p<0.01), last author (β_3=0.103, p<0.01) and middle author (β_4=0.047, p<0.01).

Table 2 displayed detailed results when we run the regression separately using data of scientists in different disciplines. Generally, a citation received as a single author would elevate the c score the most compared with other authorship while one as a middle author does the least. There are fields where the first author would 'outweigh' the last author (e.g., Economics & Business) and also fields that have the opposite situation (e.g., Agriculture, Fisheries & Forestry). Furthermore, the effect of the four types of authorship varies mildly in some fields (e.g., Communication& Textual Studies) but extensively in others (e.g., Mathematics & Statistics). The negative contribution of middle authors to c values was uniquely found in General Arts, Humanities & Social Sciences.

Table 2. Results of linear regressions in 21 disciplines

Field	R2	nc_{single}	nc_{first}	nc_{middle}	nc_{last}	Variance
Agriculture, Fisheries & Forestry	0.816	0.705 (61.8)***	0.422 (38.2)***	0.188 (15.0)***	0.518 (42.9)***	0.046
Built Environment & Design	0.660	0.757 (16.9)***	0.348 (8.1)***	0.198 (3.9)***	0.377 (7.6)***	0.022
Enabling & Strategic Technologies	0.767	0.694 (75.8)***	0.415 (45.6)***	0.275 (26.0)***	0.466 (49.3)***	0.038
Engineering	0.712	0.677 (68.2)***	0.387 (39.7)***	0.168 (14.0)***	0.487 (44.3)***	0.041
Information & Communication Technologies	0.696	0.704 (66.5)***	0.431 (41.7)***	0.198 (16.4)***	0.449 (39.3)***	0.038
Communication & Textual Studies	0.619	0.775 (10.3)***	0.346 (4.0)***	0.233 (3.0)***	0.329 (3.8)***	0.017
Historical Studies	0.528	0.829 (11.7)***	0.248 (3.6)***	0.319 (3.7)***	0.318 (3.9)***	0.013
Philosophy & Theology	0.586	0.693 (7.6)***	0.250 (2.1)**	0.327 (3.0)***	0.318 (3.0)***	0.018
Economics & Business	0.596	0.685 (46.3)***	0.441 (29.2)***	0.165 (9.8)***	0.302 (19.4)***	0.029
Social Sciences	0.617	0.771 (41.8)***	0.341 (17.2)***	0.185 (8.3)***	0.293 (14.4)***	0.015
General Science & Technology	0.534	0.712 (8.8)***	0.193 (2.1)**	0.026 (0.2)	0.489 (5.5)***	0.042
General Arts, Humanities & Social Sciences	0.981	0.584 (4.2)	0.095 (0.5)	-2.367 (0.3)	5.732 (0.1)	38.526
Biomedical Research	0.872	0.621(199.5)***	0.327(105.8)***	0.266(75.6)***	0.471(142.4)***	0.035
Clinical Medicine	0.889	0.591(347.9)***	0.372(217.3)***	0.286(146.8)***	0.408(221.7)***	0.031
Psychology & Cognitive Sciences	0.837	0.628 (84.9)***	0.378 (46.4)***	0.141 (15.4)***	0.386 (44.1)***	0.030
Public Health & Health Services	0.848	0.657 (67.4)***	0.434 (43.2)***	0.225 (18.5)***	0.378 (33.8)***	0.032
Biology	0.865	0.640(118.5)***	0.391 (70.2)***	0.177 (27.9)***	0.445 (74.6)***	0.036
Chemistry	0.790	0.605(100.3)***	0.442 (74.7)***	0.200 (29.5)***	0.490 (75.8)***	0.046
Earth & Environmental Sciences	0.850	0.658(110.4)***	0.428 (70.8)***	0.195 (27.8)***	0.427(65.7)***	0.037
Mathematics & Statistics	0.659	0.640 (33.2)***	0.469 (22.2)***	0.107 (4.3)***	0.381 (17.3)***	0.042
Physics & Astronomy	0.749	0.665(133.6)***	0.401 (83.9)***	0.248 (46.1)***	0.421 (85.4)***	0.033

4 Summary

This study sheds light on the extents to which citations received in different authorship affect the academic impact of scientists. By probing into the relationship between the

four types of authorship and the composite *c* using linear regressions, several conclusions were drawn. Overall, if a scientist received a citation to his/her academic paper, his/her academic impact would increase the most if this paper was singly authored and would increase the least if s/he was merely a middle author.

It has been proved that multi-author collaboration is the trend of science and benefits participants in the numbers of publications and citations, as well as the likelihood of publishing in top journals [6]. However, there are still some authors writing alone. Among the 100,000 high impact scientists' publications in Ioannidis et al.'s dataset, approximately 10% on average were singly authored [1]. Vafeas [7] found that junior authors or the authors affiliated with highly ranked institutions are more likely to write alone, and they are less likely to write empirical articles when they write alone. Further studies should look into the reasons why singly authored papers affect scientists' impact most significantly.

Acknowledgement. The authors acknowledge the National Natural Science Foundation of China (NSFC Grant No. 71673242) for financial support.

References

1. Ioannidis, J. P., Baas, J., Klavans, R., & Boyack, K. W.: A standardized citation metrics author database annotated for scientific field. PLoS biology, 17(8), e3000384 (2019).
2. Hirsch, J. E.: An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences, 102(46), 16569-16572 (2005).
3. Schreiber, M.: A skeptical view on the Hirsch index and its predictive power. Physica Scripta, 93(10), 102501 (2018).
4. Vinkler, P.: Research contribution, authorship and team cooperativeness. Scientometrics, 26(1), 213-230 (1993).
5. Liu, X. Z., & Fang, H.: Scientific group leaders' authorship preferences: An empirical investigation. Scientometrics, 98(2), 909-925 (2014).
6. Burke, K. G., Fender, B. F., Taylor, S. W.: Swinging for the fences: Human capital and workplace characteristics of accountants publishing in top journals. The Accounting Educators' Journal, 18(1), 29-46 (2008).
7. Vafeas, N.: Determinants of single authorship. EuroMed journal of business, 5(3), 332-344 (2010).