Do Short Papers Contribute Significantly to Impact Factor?——An Empirical Study Based on Large Samples from the Past Five Years

Authors: Wen Jingxiao, Ning Bi, Ning Bi

Date: 2023-01-13T00:00:00+00:00

Abstract

Purpose/Significance In response to academic controversies surrounding the impact factor calculation formula, this study investigates whether the contribution of non-citable literature to journal impact factors is overly prominent, thereby preventing impact factors from accurately reflecting the academic influence of journals. Method/Process By utilizing citation data from 2017-2021 for various document types published in journals indexed in the Web of Science Core Collection database, we calculate and analyze the impact factor contribution attributable to non-citable literature over the past five years. Results/Conclusion Overall, the contribution of non-citable literature to the impact factor over the past five years is relatively low (approximately 4% on average, median of 0) with minimal variation; there is no evident malicious manipulation of impact factors, and therefore concerns about such manipulation do not warrant questioning the rationality of the impact factor calculation formula.

Full Text

Is the Contribution of Non-Citable Items to the Impact Factor Significant?—An Empirical Study Based on Large Samples from the Past Five Years

Wen Jingxiao[1], Ning Bi[2]

[1]School of Economics and Management, University of Science and Technology Beijing
[2]Clarivate Analytics Information Services (Beijing) Co., Ltd.

Abstract

[**Purpose/Significance**] Based on the academic controversy surrounding the impact factor calculation formula, this study investigates whether the contribution of non-citable items to journal impact factors is disproportionately prominent, thereby preventing impact factors from accurately reflecting journals' academic influence. [**Method/Process**] Using citation data from 2017-2021 for all document types published in journals indexed in the Web of Science Core Collection, we calculated and analyzed the impact factor contributions made by non-citable items over the past five years. [**Results/Conclusion**] Overall, the contribution of non-citable items to impact factors over the past five years has been low (approximately 4% on average, with a median of 0) and relatively stable, with no evident signs of malicious manipulation. Therefore, there is no need to question the rationality of the impact factor calculation formula due to concerns about potential manipulation.

Keywords: Impact Factor; Non-Citable Items; Journal Evaluation; Academic Influence

Introduction

The Impact Factor, proposed by Dr. Eugene Garfield as a crucial metric for measuring journal academic influence, has garnered widespread attention in the academic community since its inception. Among numerous journal evaluation indicators, the impact factor has become an indispensable metric in journal evaluation and one of the most representative indicators in the Journal Citation Reports (JCR). Unlike the Eigenfactor, which focuses on citation networks, or social media-based metrics that emphasize user behavior data, the impact factor' s focus on citation frequency provides valuable insights for developing comprehensive influence indicators through deep integration of various evaluation metrics. However, alongside its prominent status is a "global worship" of impact factors: journals with higher impact factors are perceived as having greater standing in their fields, leading to their priority selection by libraries and research institutions and preferred submission by researchers. For individual researchers, publishing in higher-impact journals can directly affect their professional advancement and academic 话语权.

The impact factor is calculated as the ratio of citations received in a given year to all documents published in the previous two years (the numerator) to the number of "citable items" published in those same two years (the denominator). The numerator includes both citable and non-citable items, creating a potential loophole: journals could theoretically inflate their impact factors by publishing large volumes of non-citable items, which increase the numerator's citation count without affecting the denominator. This has long raised concerns in academia about potential manipulation of impact factors, with some scholars arguing that non-citable items have been exploited as tools for artificially boosting impact factors, while others have proposed modifications to the calculation formula.

Against this backdrop, this study employs a big data approach to analyze the composition of impact factors over the past five years (2017–2021) for journals in the Web of Science Core Collection, aiming to address the question of whether impact factors are being maliciously manipulated. According to current impact factor rules, only Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) journals in the Web of Science Core Collection receive impact factors; therefore, this study focuses exclusively on SCI and SSCI journals.

2. Literature Review

Concerns about potential manipulation of impact factors have existed for many years, with numerous scholars proposing modifications to the calculation formula. Wu Yishan suggested changing the denominator from "citable items published in the past two years" to "all documents published in the past two years" to prevent exploitation of the system. Moed proposed modifications to improve calculation accuracy. Building on these suggestions, Liu Xueli and Gai Shuangshuang developed methods for analyzing and predicting the impact factor structure of SCI-source journals. Liu Xueli also investigated the contribution of non-citable items to science and technology journal impact factors, finding that reviews and letters had the highest per-document citation rates among non-citable items, constituting a non-negligible document type for boosting journal impact factors. Nevertheless, he concluded that most journals' non-citable items contributed normally to impact factors and argued that document type diversity represents an important marker of academic journal maturity and an inevitable choice for enhancing journal influence.

Expanding on this research, Fu Zhongjing analyzed the citation characteristics of non-citable items in ten international authoritative journals, finding that non-citable items both receive citations from subsequent documents and cite previous documents, thereby contributing to impact factors in both ways. He also concluded that editorials, letters, and science news in non-citable items play primary roles in boosting impact factors.

Furthermore, Sheng Lina, Fang Hongling, and Wang Yan conducted similar studies on international ophthalmology journals, SSCI-indexed library and information science journals, and twenty international authoritative journals, respectively, all reaching similar conclusions: non-citable items significantly contribute to impact factor enhancement, with editorials and letters playing particularly important roles. Interestingly, none of these studies expressed concerns about malicious manipulation of impact factors. Instead, they recommended that Chinese academic journals learn from international publishing philosophies by enriching their document types to comprehensively enhance academic influence. Based on this foundation, the present study selects all journals in the Web of Science Core Collection with calculable impact factors, computing their impact factor composition over the past five years to explain these findings and analyze whether large-scale malicious manipulation exists.

3. Research Methods

3.1 Data Acquisition Method

For data acquisition, we referenced Liu Xueli's method for predicting SCI journal impact factors. We accessed the Web of Science platform, selected the Web of Science Core Collection database, chose all citation indexes, and entered "IS=xxxx-xxxx AND PY=20xx" in the advanced search to retrieve all document types published by specific journals in corresponding years. We then used the left-hand filter to refine by document type and employed the "Create Citation Report" function to obtain data on the number of documents published by each journal in a given year and their citation counts in each of the subsequent two years. Following this method, we used Python-based web scraping to collect data for all journals in the Web of Science Core Collection that had sufficient data for impact factor calculation over the five-year period. Note: Data were collected on December 18, 2022. Impact factors calculated using this method may differ slightly from those published annually in the JCR, but these minor differences do not affect the research conclusions in a large-sample context.

3.2 Calculation Formula

Following the impact factor definition— "the number of citations received in the statistical year to all documents published in the previous two years, divided by the number of citable items published in those two years" —we substituted the acquired data to calculate each journal' s impact factor for the past five years.

To better characterize the extent of non-citable items' contribution to journal impact factors, we define the contribution value and contribution rate of non-citable items to the impact factor:

In the formula above, represents the contribution value of a journal's non-citable items to the impact factor in year y, while and represent the citation counts received in year y by non-citable items published in the first and second previous years, respectively. represents the contribution rate of non-citable items to a journal' s impact factor in year y, while is the journal' s impact factor in year y.

The contribution rate of non-citable items can serve as an indicator for analyzing whether a journal' s impact factor is being manipulated. A low value indicates that the impact factor is minimally affected by non-citable items, suggesting low probability of manipulation through massive publication of such items. Conversely, a high contribution rate suggests a higher probability of manipulation.

Based on these formulas, we calculated the contribution values and contribution rates of non-citable items to impact factors for all journals over the past five years.

3.3 Calculation Results

Using this method, we selected 7,715 journals that had continuous impact factor data from 2017–2021. The mean and median contribution rates of non-citable items to impact factors for these journals are shown in Table 1 . The distribution of contribution rates is detailed in Table 2 .

Table 1: Mean and median contribution rates of non-citable items to impact factors for journals indexed in the Web of Science Core Collection, 2017–2021

Table 2: Distribution of contribution rates of non-citable items to impact factors for journals indexed in the Web of Science Core Collection, 2017–2021

Analysis of Tables 1 and 2 reveals three key findings: First, the annual mean contribution rate of non-citable items remains at a low level. As shown in Table 1, the mean contribution rates for all journals in the sample over the past five years are 3.50%, 3.64%, 3.69%, 3.80%, and 4.13%, respectively, with medians of 0%. Specifically, 85% of journals have non-citable item contribution rates below 10%, and 60% have rates below 1%. Overall, non-citable items contribute minimally to impact factors, indicating a low probability of systematic malicious manipulation. Second, the distribution of contribution rates is right-skewed. Each year' s mean contribution rate exceeds its median, indicating a right-skewed distribution where most journals have low contribution rates while a small number have substantially higher rates that pull up the overall average. Third, the annual mean contribution rate shows stable changes. No significant fluctuations exist in the mean contribution rates over the five-year period. At the individual journal level, only 12.8% of journals showed differences greater than 10% in their non-citable item contribution rates over the five years. Since citable items' contributions to impact factors naturally vary annually, deliberate manipulation to achieve satisfactory impact factor values each year would likely cause substantial fluctuations in non-citable item contribution rates. Therefore, the observed stability further suggests a low probability of large-scale malicious manipulation.

3.4 Analysis of High Contribution Rates in a Small Number of Journals

During our investigation, we observed that a small number of journals consistently exhibited high contribution rates from non-citable items each year. We conducted further analysis on these journals. By ranking journals based on their average contribution rates over the past five years, we compiled information on the top ten journals each year, including their contribution rates, impact factors, impact factors contributed by non-citable items, and other relevant data, as shown in Tables 3 through 7 . In these tables, "Contribution Rate" represents the contribution rate of non-citable items to the impact factor, "IF" is the journal impact factor, "IF(NCI)" is the impact factor contributed by non-citable items, "PCI" is the number of citable items published in the previous two years, "PNCI" is the number of non-citable items published in the previous

two years, "CI" is the number of citations received by citable items, and "NCI" is the number of citations received by non-citable items.

Table 3: Top ten journals by contribution rate of non-citable items in 2017

Table 4 : Top ten journals by contribution rate of non-citable items in 2018

Table 5 : Top ten journals by contribution rate of non-citable items in 2019

Table 6 : Top ten journals by contribution rate of non-citable items in 2020

Table 7: Top ten journals by contribution rate of non-citable items in 2021

For journals with high contribution rates from non-citable items, we found that all published substantially more non-citable items than citable items—often several times more. However, examining citation counts reveals that the citation efficiency of these non-citable items remains far lower than that of research articles and review articles, consistent with other journals. Therefore, these journals' high contribution rates stem not from exceptional citation efficiency but from their large volume of non-citable items. Despite their lower citation efficiency, the sheer quantity of these documents generates sufficient citations to contribute meaningfully to impact factors.

Since the impact factor calculation includes non-citable items in the numerator but not the denominator, theoretically, publishing large volumes of non-citable items could exploit this rule to boost impact factors. To understand why these journals publish so many non-citable items, we conducted further research and found that this pattern reflects their inherent nature. For example, *Dissent* publishes reflective essays on American politics, incisive social and cultural commentary, and sophisticated European political reporting beyond Europe. *Journal of Cultural Economy* provides a unique interdisciplinary forum for examining how material cultural practices function in economic and social organization. Compared to the specialized nature of research articles and reviews in citable items, non-citable items better satisfy diverse reader needs, making their extensive publication understandable. Additionally, many high-contribution-rate journals are medical journals, suggesting possible associations with disciplinary characteristics.

4. Conclusions

Based on publication and citation data from journals in the Web of Science Core Collection over the past five years, this study calculated journal impact factors according to the relevant formulas and analyzed the respective contributions of citable and non-citable items. Our findings reveal the composition and trends of impact factors over this period:

1. Among journals in the Web of Science Core Collection, the average contribution rate of non-citable items to impact factors over the past five years ranges from 3.50% to 4.13%. The vast majority of journals exhibit low contribution rates from non-citable items, with approximately half showing

zero contribution. However, a small number of journals have extremely high contribution rates. In terms of publication volume, most journals do publish non-citable items, and many prestigious journals publish substantial numbers that provide considerable impact factor contributions. Overall, except for a minority of journals, non-citable items contribute limitedly to impact factors.

2. The average number of non-citable items published, their average citation counts, and their average contribution rates to impact factors have remained relatively stable over the past five years, without significant upward trends. Statistically, approximately half of all journals show zero contribution from non-citable items, and some journals publish almost none. Consequently, no clear evidence indicates widespread malicious exploitation of the calculation formula to manipulate impact factors.

3. The high contribution rates observed in certain journals likely stem from their specific positioning. These journals' publishing philosophies differ from conventional academic journals, employing non-citable items to better express themes that citable items cannot adequately address, thereby making non-citable items central to their content and resulting in high contribution rates. Such cases should not be automatically interpreted as malicious manipulation based solely on high contribution rates or publication volumes of non-citable items.

5. Summary and Outlook

Scholars have long expressed diverse views on the rationality of the impact factor calculation formula, proposing various modifications and alternative metrics for measuring academic journal influence. This study, grounded in the impact factor calculation formula, examined the composition of impact factors over the past five years for journals in the Web of Science Core Collection, analyzing contributions from both citable and non-citable items to address whether impact factors are being maliciously manipulated and, by extension, to evaluate the formula' s rationality.

The most controversial aspect of the formula is its inclusion of non-citable item citations in the numerator while excluding their publication counts from the denominator. We speculate this design originally aimed to encourage journals to enrich their document types and content—promoting high-quality non-citable items while removing penalties for their typically lower citation rates. Ironically, this design created a loophole for potential exploitation.

Given our conclusion that no clear evidence indicates widespread malicious manipulation, the formula' s vulnerability has not compromised its function in evaluating journal influence. Therefore, concerns about manipulation should not lead us to question the formula' s rationality.

This study did not deeply explore the causes of significant inter-journal varia-

tion in non-citable item contribution rates. We hypothesize that average publication and citation rates of non-citable items vary considerably across disciplines, leading to these differences. Xu Guangkui and Tu Zhifang, analyzing inter-disciplinary differences in journal evaluation metrics, noted that research in natural sciences—particularly life sciences and health—has shorter publication and dissemination cycles than humanities and social sciences, resulting in higher citation probabilities within shorter timeframes. The varying contribution rates of non-citable items across disciplines may relate to this phenomenon, warranting further investigation.

References

[1] Zhang Huiling, Dong Kun, Xu Haiyun. Research Progress on Academic Journal Influence Evaluation Methods[J]. Library and Information Service, 2018, 62(16): 132-143.

[2] Liu Xueli. The Global SCI Phenomenon and Impact Factor Worship[J]. Chinese Journal of Scientific and Technical Periodicals, 2012, 23(02): 185-190.

[3] Petr Heneberg. Parallel Worlds of Citable Documents and Others: Inflated Commissioned Opinion Articles Enhance Scientometric Indicators[J]. Journal of the Association for Information Science and Technology, 2014, 65(3): 635-643.

[4] Wu Yishan. The Impact Factor Definition Should Be Adjusted[N]. China Science Daily.

[5] H. F. Moed, Th. N. Van Leeuwen. Improving the Accuracy of Institute for Scientific Information' s Journal Impact Factors[J]. Journal of the American Society for Information Science, 1995, 46(6): 461.

[6] Liu Xueli. A Method for Predicting SCI Journal Impact Factors Based on Web of Science Database[J]. Science-Technology and Publication, 2014(02): 87-91.

[7] Liu Xueli, Gai Shuangshuang, Zhang Shile, et al. Citation Characteristics of "Non-Citable Items" and Their Contribution to Science and Technology Journal Impact Factors[J]. Acta Editologica, 2015, 27(05): 495-499.

[8] Fu Zhongjing. Citation Characteristics of Non-Citable Items in International Authoritative Journals and Their Contribution to Impact Factors[J]. Chinese Journal of Scientific and Technical Periodicals, 2016, 27(03): 324-329.

[9] Fang Hongling. Citation Characteristics of "Non-Citable Items" in SSCI-Indexed Library and Information Science Journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(12): 1301-1305.

[10] Sheng Lina. Analysis of Contributions of Different Document Types to Impact Factors in International Ophthalmology Authoritative Journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(01): 106-111.

[11] Wang Yan. Citation Characteristics of Non-Citable Items in International Authoritative Surgery Journals and Their Contribution to Impact Factors[J]. Information Research, 2017(10): 26-29.

[12] Xu Guangkui, Tu Zhifang. Comparative Study of Two Types of Academic Evaluation Indicators—From the Perspective of Comparing Impact Factor and Google Scholar Metrics for Journal Evaluation[J]. Library and Information Service, 2017, 61(03): 109-117.

Author Contributions

Wen Jingxiao: Literature collection, data collection, manuscript writing and revision;
Ning Bi: Conceived the research idea, guided manuscript revision.

Note: Figure translations are in progress. See original paper for figures.