The Relationship Between Source Metrics and Citation Metrics and Their Impact on Journal Evaluation: A Case Study of JCR Mathematics Journals (Postprint)

Authors: Yu Liping

Date: 2017-11-08T00:00:00+00:00

Abstract

[Purpose] To analyze the relationship between source indicators and target indicators of academic journals and the resulting impact on journal evaluation. [Method] Taking mathematics journals from JCR 2015 as an example, canonical correlation analysis was employed to investigate the relationship between journal source indicators and impact indicators. Results The experimental results demonstrate that: Eigenfactor is the primary indicator of journal impact; the correlation between source indicators and impact indicators is significant, with impact indicators dominated by Eigenfactor showing the highest correlation with publication volume, followed by citing half-life, and then article selection rate; total citations, cited half-life, impact factor, and impact factor percentile contribute substantially to Eigenfactor, while other metrics such as non-self-citation impact factor, 5-year impact factor, and immediacy index contribute relatively less; impact indicators contain more information than source indicators. [Limitations] The relationship between source indicators and impact indicators requires further verification. [Conclusion] From the perspective of multi-attribute journal evaluation, the weight of impact indicators should be greater than that of source indicators; among impact indicators, it is necessary to increase the weights of Eigenfactor Score and Article Influence Score, and to allocate weights reasonably for total citations, cited half-life, impact factor, and impact factor percentile, while appropriately reducing the weights of non-self-citation impact factor, 5-year impact factor, and immediacy index.

Full Text

Abstract

[**Objective**] This paper evaluates academic journals with the help of their source indexes and impact indicators. [**Methods**] We collected data from the Journal Citation Report (JCR) in 2015 for Mathematics ones, and then conducted canonical correlation analysis with the data. **Results** Firstly, eigenfactor was the major indicator of the influence of journals. Secondly, the journal source indexes and impact indicators were significantly correlated with each other. Thirdly, total citation numbers, citation half-life, journal impact factors and journal impact factor percentile had higher contribution to the eigenfactor. Fourthly, impact indicators contain more information than the source indexes. [**Limitations**] More research is needed to investigate the relationship between the source indexes and impact indicators. [**Conclusions**] The impact indicators are more important than the source indexes. We need to increase the eigenfactor score and the weight of normalized eigenfactor. We should also decrease the weights of impact factor without journal self citations, 5-year impact factor and immediacy index.

Keywords: Academic journals, Source indexes, Impact indicators, Canonical correlation analysis, Evaluation

Classification Numbers: G203, G250

Introduction

In academic journal evaluation, there are two major categories of indicators: source indexes and impact indicators. Source indexes include metrics such as citable items, funded paper ratio, citing half-life, regional distribution count, and international paper ratio. Impact indicators encompass metrics like journal impact factor, 5-year impact factor, H-index [**?**], eigenfactor [**?**], among others. Impact represents the focal point and essential content of journal evaluation. Consequently, compared to source indexes, impact indicators are more numerous, more sophisticated in design, and more widely applied, whereas source indexes are relatively limited in number and lack variation. Investigating the relationship between journal source indexes and impact indicators helps analyze the characteristics and patterns of their interrelationship and holds significant importance for journal evaluation.

Literature Review

Since source and impact indicators constitute two distinct categories involving numerous metrics, only a few scholars have examined their relationship. For instance, Yu Liping et al. [**?**] used medical journals from the Chinese Institute of Scientific and Technical Information as examples and found that high source indexes do not necessarily indicate substantial journal impact. While the correlation between source indexes and impact indicators is generally low, it varies across different data segments. More scholars have focused on the relationship between specific source indexes and impact indicators, which can be categorized into three main groups:

First, studies examining temporal factors such as publication frequency and citation half-life in relation to impact indicators. Tsay et al. [?] demonstrated that journals with higher publication frequencies tend to have higher citation frequencies, impact factors, and immediacy indices, often accompanied by shorter citation half-lives. Sangam [?] found that faster literature growth leads to more rapid obsolescence and shorter half-lives. Zhou Zhizhong [?] compared the differences between cited half-life and citing half-life for library and information science journals in CSSCI and JCR databases, revealing that foreign journals in this field have longer cited and citing half-lives that are positively correlated.

Second, research on the relationship between journal citable items and impact indicators. Huang Mingrui [?] found that journal citable items depend on journal positioning and the size of the reader-author community, showing significant positive correlations with both discipline diffusion indicators and discipline impact indicators. Zhao Jinyan [?] employed Spearman rank correlation and discovered that literature source quantity, regional distribution count, and institutional distribution count are significantly positively correlated with total citation frequency, while literature source quantity and average citation count are significantly positively correlated with comprehensive evaluation scores. Yan Meijuan [?] conducted statistics on the relationship between citable items and impact factor/5-year impact factor for 239 neuroscience journals in the 2010 JCR, finding a positive correlation within a certain range. An Mei [?] studied high-impact engineering journals in Web of Science and found that journals publishing more papers have relatively stable rankings for both impact factor and eigenfactor, whereas journals publishing fewer papers often experience significant drops in eigenfactor ranking compared to impact factor ranking. Yu Jianqing [?] concluded that for agricultural university journals, source literature quantity largely determines the cumulative ranking of source indexes, with journals having more source literature generally showing higher reference counts, regional distribution counts, and institutional distribution counts. Liu Yan [?] argued that journals with earlier establishment dates, larger source literature volumes, and higher academic quality necessarily have greater academic impact.

Third, studies on the relationship between funded paper ratio and other source indexes with impact indicators have yielded inconsistent conclusions. Liu Ruiyuan et al. [?] found that funded papers have significantly higher citation rates and average citations per paper than non-funded papers, and that journal funded paper ratio is significantly positively correlated with impact factor, 5-year impact factor, eigenfactor, and reader survey scores. However, other scholars have reached different conclusions. Wang Qian et al. [?] discovered no significant correlation between funded paper ratio and impact factor ($P > 0.05$) for Chinese medical core journals, nor complete correlation with other journal evaluation indicators. Yan Yan et al. [?] argued that the funded paper ratio indicator inverts the positions of evaluation subject and object, confuses logical relationships, creates unfairness for non-funded paper authors, and suppresses valuable human and research resources, making it unsuitable as a journal evaluation metric.

Current research shows that academia has begun to examine the relationship between source indexes and impact indicators, focusing primarily on the relationship between citable items and impact indicators. Studies on the relationship between citing half-life, funded paper ratio, and impact remain limited, with substantial disagreement regarding the funded paper ratio' s relationship with impact. Additionally, findings vary due to different research objects and databases, necessitating further in-depth investigation. Methodologically, existing studies primarily employ regression analysis, correlation coefficients, and simple statistical methods, but not canonical correlation analysis (CCA), which analyzes correlations between multiple variable sets. The impact of source-impact indicator relationships on multi-attribute journal evaluation requires further academic investigation. This paper uses 2015 JCR mathematics journals as examples to examine the relationship between source indexes and impact indicators and discusses implications for multi-attribute journal evaluation.

Research Methods: Canonical Correlation Analysis

This study examines the relationship between academic journal source indexes and impact indicators, both involving more than two variables, thus addressing multivariate correlation between two variable sets. Traditional research typically examines relationships between single source indexes and multiple impact indicators, which lacks systematic perspective. Additionally, multicollinearity among journal impact indicators compromises the robustness of findings.

Canonical correlation analysis provides a novel approach to addressing source-impact indicator relationships. Hotelling [?] proposed canonical correlation analysis to solve correlation problems between multiple variable sets. Early limitations due to extensive matrix calculations without computer assistance have been overcome by computational advances since the 1970s.

Let random matrices be $X = (x_1, x_2, ..., x_p)$ and $Y = (y_1, y_2, ..., y_q)$. To investigate canonical correlations between X and Y, we construct linear combinations as follows: $by + +by$ (text appears fragmented). We seek coefficients a and b that maximize the correlation coefficient r between U and V, calculated as:

[Mathematical formulas appear fragmented in original]

Standardizing random matrices U and V yields: $var(U) = var(V) = 1$, $var(a'X) = 1$, $var(b'Y) = 1$, and $cov(U, V) = a' cov(X, Y)b$. The problem thus transforms into finding coefficients a and b that maximize correlation coefficient r under the constraints of formula (3). This is solved by constructing Lagrange functions and conducting statistical tests:

$\chi^2 = -[(n-1) - (p+q+1)/2] \ln \Lambda$, where $\Lambda = \prod_{i=1}^{k} (1 - \hat{\lambda}_i^2)$, $Q_i = -[(n-1) -$

$(p+q+1)]\ln[...]$ (formula appears incomplete). If $Q_i > \chi^2$, the corresponding canonical correlation coefficient is considered significant.

Data

This study uses data from the 2015 JCR database for mathematics journals. The 2015 JCR includes ten impact indicators: Total Cites, Journal Impact Factor, Impact Factor without Journal Self Cites, 5-Year Impact Factor, Immediacy Index, Cited Half-life, Eigenfactor Score, Article Influence Score, Average Journal Impact Factor Percentile, and Normalized Eigenfactor. The three source indicators are: Citable Items, Citing Half-life, and Articles in Citable Items. The journal impact factor percentile and normalized eigenfactor were newly introduced in 2015.

The 2015 JCR mathematics collection contains 312 journals. After removing journals with missing data, 281 journals remained. Descriptive statistics for each indicator are shown in Table 1 .

Results

Overall Correlation Analysis

Canonical correlation coefficients and significance tests are presented in Table 2 . U represents impact indicators and V represents source indicators. Three sets of canonical correlations are denoted as U1-V1, U2-V2, and U3-V3. The first and second sets pass Wilks' Lambda test at the 1% level (p = 0.000), while the third set passes at the 10% level. The first correlation coefficient is highest at 0.830, followed by the second at 0.575, and the third at 0.234, indicating a very close overall relationship between journal impact and source indicators.

Canonical coefficients for each variable are shown in Table 3 . In the first canonical relationship, eigenfactor score and normalized eigenfactor have the largest absolute coefficients (-27.828 and 27.181 respectively) among impact indicators, while citable items shows the largest absolute coefficient (-1.029) among source indicators, indicating this relationship primarily reflects the correlation between eigenfactor and citable items.

In the second canonical relationship, eigenfactor score and normalized eigenfactor again show the largest absolute coefficients (157.547 and -157.942) among impact indicators, while citing half-life has the largest absolute coefficient (-1.006) among source indicators, suggesting this relationship primarily reflects the correlation between eigenfactor and citing half-life.

In the third canonical relationship, eigenfactor score and normalized eigenfactor maintain the largest absolute coefficients (1051.399 and -1051.220) among impact indicators, while articles in citable items shows the largest absolute coefficient (0.997) among source indicators, indicating this relationship primarily reflects the correlation between eigenfactor and articles in citable items.

Canonical Structure Analysis

Canonical structure analysis calculates canonical loadings, equivalent to weights, which measure correlations between original and canonical variables while avoiding multicollinearity. Larger absolute loadings indicate greater importance in explaining canonical variables. Table 4 presents canonical and cross-loadings for impact and source indicators.

To further analyze each canonical relationship and indicator loadings, we constructed canonical factor structure diagrams shown in Figures 1 [Figure 1: see original paper]-3 [Figure 3: see original paper].

In the first canonical relationship (Figure 1), total cites (-0.818), eigenfactor score (-0.801), and normalized eigenfactor (-0.801) show the largest absolute coefficients among impact indicators, while citable items (-0.989) has the largest absolute coefficient among source indicators, further confirming the eigenfactor-citable items relationship and demonstrating substantial explanatory power of total cites.

In the second canonical relationship (Figure 2), cited half-life (-0.431) and impact factor (0.291) show the largest absolute coefficients among impact indicators, while citing half-life (-0.986) has the largest absolute coefficient among source indicators. This primarily reflects the eigenfactor-citing half-life relationship, with cited half-life and impact factor contributing substantial explanatory power.

In the third canonical relationship (Figure 3), cited half-life (0.678) and impact factor percentile (-0.281) show the largest absolute coefficients among impact indicators, while articles in citable items (0.977) has the largest absolute coefficient among source indicators. Although this reflects the eigenfactor-articles in citable items relationship, cited half-life and impact factor percentile provide substantial explanatory power.

Redundancy Analysis

Redundancy analysis assesses the overall explanatory power of the canonical correlation model, as shown in Table 5 . For impact indicators, the three canonical relationships explain 19.8%, 4.8%, and 6.1% of variance respectively (totaling 30.7%), with source indicators explaining 13.6%, 1.6%, and 0.3% respectively. The self-explanation level is relatively low, with the first canonical relationship showing the highest explanatory proportion.

For source indicators, the three canonical relationships explain 33.1%, 34.6%, and 32.3% of variance respectively (totaling 100%), with impact indicators explaining 22.8%, 11.5%, and 1.8% respectively. The self-explanation level is very high across all three relationships.

Thus, journal impact indicators contain substantial information not reflected by source indicators or themselves, while source indicators contain all their own information plus limited information from impact indicators.

Discussion and Conclusions

1. **Eigenfactor is the primary indicator of journal impact.** Our empirical study demonstrates that across all three canonical relationships, impact indicators primarily reflect contributions from eigenfactor score and normalized eigenfactor, indicating eigenfactor' s crucial position among impact indicators. This fundamentally stems from eigenfactor' s characteristics: it uses a PageRank-like iterative algorithm to calculate journal weight impact values, ensuring synchronization between citation quantity and quality. Journal evaluation should increase the weight of eigenfactor score and normalized eigenfactor.

2. **Significant correlation exists between source and impact indicators.** Journal impact and source indicators show significant correlations, with eigenfactor-based impact indicators correlating most strongly with citable items (0.830), followed by citing half-life (0.575), and finally articles in citable items (0.234). Many studies have found strong correlations between citable items and impact indicators, which our research supports. A newer citing half-life indicates more current information and faster knowledge updating, facilitating more citations. Higher articles in citable items ratio means a higher proportion of effective papers as citation sources, naturally leading to greater citation impact.

3. **Certain impact indicators contribute substantially to eigenfactor.** Across the three canonical relationships, total cites, cited half-life, impact factor, and impact factor percentile contribute significantly to eigenfactor, whereas impact factor without journal self cites, 5-year impact factor, and immediacy index contribute minimally. Therefore, journal evaluation should emphasize eigenfactor weight while appropriately allocating weights to total cites, cited half-life, impact factor, and impact factor percentile, and reducing weights for impact factor without journal self cites, 5-year impact factor, and immediacy index.

4. **Impact indicators contain more information than source indicators.** Redundancy analysis reveals that impact indicators contain substantial information not captured by source indicators or themselves, while source indicators contain all their own information plus limited information from impact indicators. This demonstrates that impact indicators carry greater information volume, source indicators carry less, and despite their correlation, neither can substantially explain the other, indicating they are not interchangeable. Therefore, impact indicators should carry greater weight than source indicators in journal evaluation.

This study provides in-depth analysis of the relationship between journal source and impact indicators using JCR mathematics journal data, advancing understanding of their effects on journal evaluation. Methodologically, it offers a new approach based on canonical correlation analysis, with practical implications for weight assignment in journal evaluation metrics. Due to variations across databases and indicators, further research is needed to explore relationships between source and impact indicators for different academic journals. This paper provides one research paradigm.

References

[1] Hirsch J E. An Index to Quantify an Individual' s Scientific Research Output [J]. Proceedings of the National Academy of Sciences, 2005, 102(46): 16569-16572. DOI: 10.1073/pnas.

[2] Bergstrom C T, West J D, Wiseman M A. The Eigenfactor Metrics [J]. The Journal of Neuroscience, 2008, 28(45):

[3] Yu Liping, Pan Yuntao, Wu Yishan. Study on Source Indexes and Influence of the Academic Journal Evaluation [J]. Science Research Management, 2010, 31(6): 173-179.

[4] Tsay M Y, Chen Y L. Journals of General & Internal Medicine and Surgery: An Analysis and Comparison of Citation [J]. Scientometrics, 2005, 64(1): 17-30.

[5] Sangam S L. Obsolescence of Literature in the Field of Psychology[J]. Scientometrics, 1999, 44(1): 33-46.

[6] Zhou Zhizhong. Half-life Period Analysis of Library and Information Journals [J]. Journal of Southwest University for Nationalities: Humanities and Social Science, 2013(9): 233-240.

[7] Huang Mingrui. The Effect of Paper Quantity in Evaluation of Scientific Journals' Influence [J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(7): 749-757.

[8] Zhao Jinyan. Study on Source Indexes and Influence of the Academic Journal [J]. Journal of Huanggang Normal University, 2015, 35(5): 115-117.

[9] Yan Meijuan. The Relationship of Paper Amounts, Impacting Factors and 5-year Impacting Factors Among 239 Kinds of Neuroscience Journals [J]. Journal of Northeast Agricultural University: Social Science Edition, 2012, 10(5): 83-85.

[10] An Mei. Research on Impact Index Wave and Citable Items [J]. Chinese Journal of Scientific and Technical Periodicals, 2011, 22(5): 711-714.

[11] Yu Jianqing. Statistic Analysis of the Citing Source Indicators of the Journals of Agricultural University[J]. Journal of Library and Information Sciences in Agriculture, 2009, 21(9): 63-65.

[12] Liu Yan. Comprehensive Evaluation and Analysis on the Academic Influence Index of Editing and Publishing Journals [J]. Science and Publish, 2013(10): 108-112.

[13] Liu Ruiyuan, Liu Xueli, Wang Bu, et al. The Rationality of Ratio of Fund-subsidized Papers on Academic Journals Evaluation [J]. Chinese Journal of Scientific and Technical Periodicals, 2013, 24(3): 473-476.

[14] Wang Qian, Lin Ping, Sun Changpeng. The Relationship and Influencing Factors of Medical Journals' Funding Ratio, Impact Factor, and Other Evaluation Indexes [J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(6): 634-638.

[15] Yan Yan, Gu Guanhua. Ratio of Fund-subsidized Papers: An Unreasonable Evaluation Index of Journals [J]. Journal of Southeast University: Philosophy and Social Science, 2011, 13(6): 123-125.

[16] Hotelling H. Relations Between Two Sets of Variates[J]. Biometrics, 1936, 28(3-4): 321-377.

Conflict of Interest Statement

The authors declare no conflict of interest.

Supporting Data

Supporting data is available in the online version of the journal at http://www.infotech.ac.cn.

Publication History

Received: March 21, 2016 Revised: April 16, 2016

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