Deviation and Correction of Indicator Values and Evaluation Attributes in Science and Technology Evaluation: A Multi-Attribute Evaluation Perspective (Postprint)

Authors: Zhou Juanmei, Guo Qianghua, Wang Zuogong, Yu Liping

Date: 2023-08-27T00:00:00+00:00

Abstract

[Purpose/Significance] This paper analyzes the underlying causes of abnormal indicator discriminability and skewed data distributions in scientific and technological evaluation, positing that this phenomenon essentially represents a divergence between evaluation indicator values and evaluation attributes, wherein indicator values fail to adequately reflect the intrinsic meaning of the attributes. [Method/Process] It proposes a novel approach to mitigate this divergence—logarithmic median standardization—and empirically validates the method using mathematics journals from JCR 2016 as a case study. [Results/Conclusion] The findings demonstrate that: citation indicators are particularly susceptible to the divergence between indicator values and attributes; such divergence can be assessed from multiple perspectives, including indicator connotation analysis, pass rates, coefficients of variation, median-to-maximum ratios, and the Herfindahl-Hirschman Index (HHI); logarithmic median standardization substantially reduces this divergence; it is recommended that evaluations exhibiting indicator-attribute divergence employ data processed via logarithmic median standardization.

Full Text

Preamble

Volume 62, Issue 22, November 2018
ChinaXiv Cooperative Journal

Research on the Deviation and Correction of Technology Evaluation Index Values and Evaluation Attributes: A Multi-Attribute Evaluation Perspective

Zhou Juanmei[1], **Guo Qianghua**[2], **Wang Zuogong**[3], **Yu Liping**[4]
[1] School of Economics and Management, North University of China, Taiyuan 030051
[2] School of Economics, Yunnan University, Kunming 650504
[3] Institute of Green Development Strategy in Western China, Guizhou University of Finance and Economics, Guiyang 550025
[4] School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018

Abstract

[**Purpose/Significance**] This paper analyzes the underlying causes of abnormal index discrimination and skewed data distribution in science and technology evaluation, arguing that these phenomena essentially represent a deviation between evaluation index values and evaluation attributes—meaning that the index values fail to adequately reflect the essential meaning of the evaluation attributes. [**Method/Process**] The study proposes a novel method to reduce this deviation: logarithmic median standardization, and demonstrates its application using JCR 2016 mathematics journals as an empirical case. [**Result/Conclusion**] The findings indicate that citation indices are particularly prone to deviation between index values and evaluation attributes. This deviation can be identified from multiple perspectives, including index connotation analysis, pass rates, dispersion coefficients, median-to-maximum ratios, and the Herfindahl-Hirschman Index (HHI). Logarithmic median standardization significantly mitigates this deviation, and the study recommends adopting this processed data for evaluation when such deviations are present.

Keywords: distinction degree; data distribution; science and technology evaluation; influence index; logarithmic median standardization

2. Theoretical Analysis of Evaluation Distinction Degree

2.1 Causes of Abnormal Data Distribution and Distinction Degree in Evaluation Indicators

Given the complexity of these issues, this paper focuses on academic journal evaluation to explore the underlying reasons for the skewed distribution and abnormal distinction degree observed in many journal evaluation metrics. Several key factors contribute to this phenomenon.

First, the problem of skewed data distribution and distinction degree is partially concealed in practice. The public often finds it difficult to accept severely skewed evaluation results, such as when pass rates fall below 5%. Common evaluation practices—including university rankings, industrial park assessments, journal evaluations, and research performance appraisals—frequently publish only rankings without revealing the actual scores, thereby obscuring the magnitude of gaps between evaluated objects. This practice masks the severity of distribution distortions.

Second, only a very small number of science and technology evaluation indicators should normally exhibit skewed distributions. In most cases, evaluation indicators should follow or approximate a normal distribution, with only rare exceptions legitimately showing skewness. For instance, metrics reflecting genuine innovation levels—such as breakthrough theories like relativity or quantum mechanics—naturally exhibit skewed distributions with abnormal distinction degrees, which the public can accept. However, truly innovative indicators are extremely scarce, as original contributions typically require peer review for assessment and cannot be adequately captured through bibliometric measures. Consequently, when numerous bibliometric indicators display skewed distributions, this signals an underlying problem.

Third, existing bibliometric indicators struggle to measure innovation effectively. In academic journal evaluation, metrics fall into three main categories: (1) influence indicators (e.g., impact factor, 5-year impact factor, cited impact factor) that reflect citation patterns; (2) timeliness indicators (e.g., cited half-life, citing half-life); and (3) source indicators (e.g., funded paper ratio, regional distribution, average author count). While these indicators often correlate positively with journal quality, they cannot directly measure the most critical dimension—paper quality or innovation level. The skewed distribution of bibliometric indicators is therefore abnormal because these metrics fail to capture genuine innovation, yet their data distributions exhibit the skewness that should characterize true innovation metrics.

Finally, this leads to the core concept of "deviation between evaluation index value and evaluation attribute." Based on the above analysis, only original achievements should follow skewed distributions with large distinction anomalies. Since current bibliometric indicators cannot evaluate originality but still exhibit skewed distributions, the only explanation is that the raw bibliometric data contain distorted information that fails to reflect the true attributes being measured. This phenomenon—where bibliometric indicator values do not accurately represent the intended evaluation attributes—manifests as skewed data distributions and abnormal distinction degrees. The raw data for total citations, impact factors, and other metrics exaggerate the true differences in journal influence. For example, a journal with 16,000 total citations is not necessarily 20 times more influential than one with 800 citations. The actual distribution of journal influence should approximate a normal distribution, but the distorted indicator values create the false impression that influence itself is skewed.

2.2 Impact of Indicator Data Distribution and Distinction Degree on Comprehensive Evaluation

Building on the indicator-level analysis, we now examine the issue from a multi-attribute evaluation perspective. Multi-attribute evaluation combines bibliometric indicators through weighting schemes and specific methods to produce overall assessments. Major journal ranking systems—such as Peking University's Chinese Core Journals, CSTPCD, CSCD, and CASS core journals—employ different evaluation frameworks, weighting methods, and specific techniques, but share similar underlying principles. Consequently, the distribution and distinction degree of individual evaluation indicators inevitably affect the final evaluation results. Research has shown that skewed indicator data can bias assessments of average journal performance, with right-skewed distributions leading to underestimated evaluation values. Therefore, selecting indicators with better distribution properties is crucial for accurate evaluation.

2.3 Approaches to Improving Data Distribution and Distinction Degree

The fundamental solution involves non-linear transformation of raw indicators to better express their attribute meanings and improve distinction degree, making distributions more normal. The relationship between indicator attributes and dimensionless processing methods is inseparable; standardization approaches should be designed based on the characteristics of raw data distributions. This study proposes applying natural logarithm transformation to raw bibliometric indicators to improve their distinction degree and distribution. The United Nations Development Programme employs a similar approach when calculating the Human Development Index, applying logarithmic transformation to income indicators to reflect diminishing marginal utility. Logarithmic transformation offers valuable properties: it is strictly monotonic, strongly concave, capable of altering distribution shapes, and effective at compressing value ranges—making it particularly suitable for rapidly growing indicators with large spans. When observations cluster at low levels (right-skewed distributions), concave utility functions are appropriate.

2.4 Research Design

The technical route of this study is illustrated in Figure 1 [Figure 1: see original paper]. The first approach evaluates journals using raw indicator data without transformation. The second approach applies logarithmic transformation combined with a new standardization method to modify the data distribution and distinction degree before evaluation. The two approaches are then compared to assess differences in data distribution and distinction degree.

2.4.1 Assessing Raw Indicator Distribution and Distinction Degree

First, we analyze the connotation of each bibliometric indicator to determine its nature. If an indicator genuinely reflects innovation levels, skewed distribution may be normal and require no adjustment. For indicators measuring influence, timeliness, or general source characteristics, we test their raw data distributions—primarily through normality tests such as the Jarque-Bera test.

Various methods exist for assessing distinction degree. Wang and Zhang proposed using the coefficient of variation, while Yu and Liu suggested applying Gini coefficient principles. Given the complexity of Gini coefficient calculations, this study adopts the Herfindahl-Hirschman Index (HHI) proposed by Hirschman to measure distinction degree:

$$HHI = \sum_{i=1}^{m} \left(\frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \right)^2 \quad (1)$$

where x_{ij} represents the indicator value, i is the evaluation object index, j is the indicator index, m is the number of evaluation objects, and HHI is the sum of squared indicator shares. This is an inverse indicator of distinction degree.

2.4.2 Logarithmic Transformation of Raw Indicators

All selected indicators undergo natural logarithm transformation. Note that some indicators have zero values, which cannot be log-transformed. In such cases, minimal data adjustment is applied, such as adding 1% of the maximum value to all observations.

2.4.3 Implicit Requirements of Evaluation Purpose for Distinction Degree

Evaluation purposes often impose implicit requirements on distinction degree, such as expecting scores above 60 (or standardized values above 60). This suggests indicators should have certain properties—one simple criterion is that the median should preferably exceed half of the maximum value. When this condition is not met, necessary adjustments during standardization are required.

2.4.4 Logarithmic Median Standardization

Standardization aims to linearly transform log-transformed indicators so that median values are appropriately increased (e.g., to half of the maximum) to improve public acceptance of the scores. Two cases are distinguished:

Case 1: For indicators where the median already exceeds half of the range, traditional standardization is applied because the distribution already approximates public expectations.

Case 2: For indicators where the median falls below the midpoint, the relative difference is added before standardization to avoid score clustering in low-value regions. This approach is termed "logarithmic median standardization."

Let D_{ij} represent the relative gap between the median and the range midpoint:

$$D_{ij} = \frac{0.5\{\max[\ln(x_{ij})] - \min[\ln(x_{ij})]\} - \text{median}[\ln(x_{ij})]}{\max[\ln(x_{ij})] - \min[\ln(x_{ij})]} \quad (2)$$

For Case 2, logarithmic median standardization is applied:

$$X_{ij} = D_{ij} + (100 - D_{ij}) \times \frac{\ln(x_{ij}) - \min[\ln(x_{ij})]}{\max[\ln(x_{ij})] - \min[\ln(x_{ij})]} \quad (3)$$

For Case 1, traditional standardization is used to enhance distinction degree:

$$X_{ij} = \frac{\ln(x_{ij}) - \min[\ln(x_{ij})]}{\max[\ln(x_{ij})] - \min[\ln(x_{ij})]} \times 100 \quad (4)$$

2.4.5 Multi-Attribute Evaluation and Analysis

Based on the above processing, this study employs multi-attribute evaluation methods, using TOPSIS as an example to compare results before and after logarithmic transformation, thereby illustrating the impact of indicator distinction degree and data distribution processing on evaluation outcomes.

3. Research Data and Empirical Results

3.1 Research Data

This study uses JCR 2016 mathematics journals as the research sample. JCR 2016 provides 11 main evaluation indicators: total citations, impact factor without self-citations, impact factor, immediacy index, impact factor percentile, 5-year impact factor, Eigenfactor, normalized Eigenfactor, article influence score, cited half-life, and citing half-life. The impact factor percentile is excluded due to its non-parametric nature from ranking transformation. Additionally, cited half-life and citing half-life are excluded for journals exceeding 10 years, as these are only reported as ">10" without specific values. Consequently, all selected indicators reflect journal influence.

Among 310 JCR 2016 mathematics journals, some had missing data and were removed, leaving 294 journals for analysis. Descriptive statistics of the raw indicators are presented in Table 1 .

3.2 Distribution and Distinction Degree of Raw Indicators

Table 2 presents the distribution and distinction degree metrics for the raw indicators. The median-to-maximum ratio reflects the median's position, with the 5-year impact factor showing the highest value at 0.177—meaning that if the maximum score is 100, half of the journals score below 17.7, indicating severe data clustering and poor distinction at low values. Most dispersion coefficients fall within two standard deviations (coefficient of variation < 2.0), except for citing half-life. The HHI values are generally low, though their absolute levels require further interpretation given the index's origin in measuring market concentration.

If each indicator's maximum value is set at 100 and 60 as the passing threshold, among 294 journals, the highest pass count across eight indicators is only 8 journals. The vast majority of journals "fail" across all indicators, suggesting the raw data distort actual journal influence and lack credibility. All indicators reject the null hypothesis of normal distribution in Jarque-Bera tests, with positive skewness indicating right-skewed distributions.

3.3 Distribution After Data Processing

All eight indicators are log-transformed. Analysis of their median values reveals that the immediacy index median is -2.048, while the midpoint of the range is -2.996, placing its median ahead of the midpoint—making it the only indicator suitable for conventional standardization using formula (4). The remaining seven indicators are processed using formula (3). The post-processing distribution and distinction degree are shown in Table 3 .

The average median-to-maximum ratio improves dramatically from 0.090 (raw data) to 0.486 (processed data), meaning 50% of journals now score around the 48.6% level—a much more uniform distribution. The mean coefficient of variation decreases from 1.107 to 0.337, reflecting more uniform data distribution. The average HHI drops from 0.008 to 0.0038, indicating substantially improved distinction degree and more balanced distribution as an inverse indicator.

The average number of passing journals increases from 5 (raw data) to 75.38 after processing—a remarkable improvement. While evaluation results remain non-normal after processing, the Jarque-Bera statistic decreases dramatically from 6374.367 to 75.624, bringing the distribution much closer to normality. Skewness also improves significantly.

3.4 Comparison of Evaluation Results

To compare distinction degree and distribution before and after processing, this study applies both linear weighting and TOPSIS methods with equal weights for simplicity. Partial results (top 30 journals) are shown in Table 4 . Even with identical methods, ranking differences emerge between processed and unprocessed data.

Table 5 summarizes distinction degree and distribution metrics for the evaluation results. Before processing, linear weighting yields only 2 passing journals, while TOPSIS produces just 1—results that are difficult to justify and unlikely to gain acceptance from journal editors. After processing, linear weighting produces 54 passing journals and TOPSIS yields 41, substantially improving pass rates and aligning evaluation results with perceived journal influence.

The median-to-maximum ratio improves from 0.133 (linear weighting) and 0.125 (TOPSIS) before processing to 0.510 and 0.480 respectively after processing. Coefficients of variation decrease from 0.812 to 0.271 (linear weighting) and from 0.906 to 0.289 (TOPSIS). HHI values also improve markedly, decreasing from 0.0056 to 0.0037 for linear weighting and from 0.0062 to 0.0037 for TOPSIS.

5. Research Conclusions

5.1 The Root Cause of Low Distinction Degree: Deviation Between Index Value and Evaluation Attribute

This study analyzes the abnormal distinction degree and skewed data distribution in science and technology evaluation, particularly in bibliometric indicators, where low scores cluster excessively while high scores disperse too widely, creating right-skewed distributions that persist in multi-attribute evaluation results. This leads to questionable scores that undermine evaluation credibility. The fundamental cause is the deviation between evaluation index values and evaluation attributes—citation counts and impact factors cannot objectively reflect true influence. Only indicators genuinely measuring academic innovation can avoid this deviation, but such indicators are currently scarce and difficult to obtain.

5.2 Citation Indicators Are Particularly Prone to Deviation

Bibliometric citation indicators are especially susceptible to deviation between index values and evaluation attributes, including total citations, impact factor, cited impact factor, h-index, 5-year impact factor, immediacy index, Eigenfactor, and normalized Eigenfactor. Other timeliness indicators, source metrics, and editorial indicators require separate analysis.

5.3 Multi-Perspective Assessment of Deviation

Deviation between index values and attributes can be assessed through multiple approaches: (1) analysis of whether the indicator's connotation genuinely reflects "towering above the rest" characteristics typical of major breakthroughs; (2) examination of pass rates, dispersion coefficients, median-to-maximum ratios, and HHI concentration indices.

5.4 Logarithmic Median Standardization Effectively Reduces Deviation

The proposed logarithmic median standardization method first applies natural logarithm transformation to compress the range between extreme values and approximate normal distribution, followed by targeted standardization. For indicators with medians below the midpoint, the relative difference is added before standardization. Empirical results demonstrate that this approach substantially improves median-to-maximum ratios, reduces dispersion coefficients and concentration indices, and brings distributions closer to normality, enabling index values to better reflect attribute values.

5.5 Recommendation: Adopt Logarithmic Median Standardization

When evaluation indicators exhibit abnormal data distribution and distinction degree, analysts should first confirm whether deviation between index values and attributes exists. If confirmed, the study recommends using logarithmically transformed data for both individual indicator evaluation and multi-attribute evaluation to accurately reflect evaluation objects' true attributes and overall performance.

References

[1] Vinkler P. Introducing the current contribution index for characterizing the recent, relevant impact of journals[J]. Scientometrics, 2008, 79(2): 409-420.

[2] Seglen PO. The skewness of science[J]. Journal of the American Society for Information Science, 1992, 43(9): 628-638.

[3] Bornmann L, Mutz R, Neuhaus C, et al. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results[J]. Ethics in Science and Environmental Politics (ESEP), 2008, 8(1): 93-102.

[4] Adler R, Ewing J, Taylor P. Citation statistics[J]. Statistical Science, 2009, 24(1): 1-26.

[5] Van Raan AFJ. Measurement of central aspects of scientific research: performance, interdisciplinarity, structure[J]. Measurement, 2005, 3(1): 1-19.

[6] Su Weihua. Research on multi-indicator comprehensive evaluation theory and methods[D]. Xiamen: Xiamen University, 2002.

[7] Davis PM. Eigenfactor: does the principle of repeated improvement result in better estimates than raw citation counts?[J]. Journal of the American Society for Information Science and Technology, 2008, 59(13): 2186-2188.

[8] Hirsch J. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences, 2005, 102(46): 16569-16572.

[9] Glänzel W. On the h-index: a mathematical approach to a new measure of publication activity and citation impact[J]. Scientometrics, 2006, 67(2): 315-321.

[10] Yu Liping, Pan Yuntao, Wu Yishan. Research on problems in the hierarchical management of university social science academic journals[J]. Journal of Intelligence, 2011(4): 1-4.

[11] Wen Dongmao, Bao Xuming, Fu You. Impact of grade assignment on college entrance exam distinction degree: a simulation analysis based on Zhejiang "Nine-School Joint Exam" data[J]. China Higher Education Research, 2015(6): 17-21, 72.

[12] Ruané F, Tol R. Rational (successive) h-indices: an application to economics in the Republic of Ireland[J]. Scientometrics, 2008, 75(2): 395-405.

[13] Ye Ying. The logarithmic f-index and its evaluation significance[J]. Information Science, 2009, 27(7): 965-968.

[14] Fu Xinjin, Fang Shu, Xu Haiyun. Empirical research on university network academic influence[J]. Library and Information Service, 2013, 57(8): 86-90.

[15] Xu Xinjun. Comparative analysis of journal hc-index and hd-index[J]. Journal of Intelligence, 2015, 34(4): 59.

[16] Liu Xuezhi, Yang Zeyu, Shen Fengwu, et al. Research on non-linear standardization of indicators based on S-curve[J]. Statistics & Information Forum, 2018, 33(2): 17-21.

[17] Guo Yajun, Ma Fengmei, Dong Qingxing. Analysis of the impact of dimensionless methods on the scatter degree method[J]. Journal of Management Sciences in China, 2011(5): 19-28.

[18] Yu Liping, Pan Yuntao, Wu Yishan. Application of TOPSIS in journal evaluation and its extension under high powers[J]. Statistical Research, 2012(12): 96-101.

[19] Yu Xinchen, Pan Youneng. Research on the relationship between academic journal evaluation indicators and evaluation results' data distribution[J]. Journal of Intelligence, 2015, 34(9): 102-106.

[20] Yu Liping, Liu Aijun. Impact of indicator data distribution and internal gaps on academic journal evaluation: a case study of JCR mathematics journals[J]. Library and Information Service, 2014, 58(21): 105-110.

[21] Zhan Min, Liao Zhigao, Xu Jiuping. Comparative study on linear dimensionless methods[J]. Statistics & Information Forum, 2016(12): 17-22.

[22] UNDP. Human development report technical notes 2014[R/OL]. [2018-06-01]. http://hdr.undp.org/sites/default/files/hdr14_{{{technical}}_{{notes}}}.pdf.

[23] Su Weihua. Preliminary exploration of logarithmic efficacy coefficient method[J]. Statistical Research, 1993(3): 63-67.

[24] Feng Ting. A concave exponential efficacy function in comprehensive evaluation[J]. Statistics & Information Forum, 2016, 31(7): 16-21.

[25] Wang Lianfen, Zhang Shaojie. Design of industrial competitiveness measurement index system[J]. Statistics and Decision, 2008(10): 49-50.

[26] Hirschman AO. The political economy of import-substituting industrialization in Latin America[J]. The Quarterly Journal of Economics, 1968(1): 1-32.

Author Contributions

Zhou Juanmei: Standardization method design
Guo Qianghua: Research design
Wang Zuogong: Data analysis, cleaning, and processing
Yu Liping: Overall paper design and writing

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