A BAYESIAN CITATION INDICATOR OF INDIVIDUAL SCIENTIFIC PERFORMANCE COMBINING IMPACT FACTOR AND CITATION RATE

Afschin Gandjour

ABSTRACT

Purpose: The rise of New Public Management has created a need for indicators to evaluate scientific performance. This study introduces a novel metric that combines impact factor and citation rate to measure individual scientific output.

Design/methodology/approach: The methodology used in this study is Bayesian shrinkage estimation.

Findings: The proposed Bayesian citation indicator combines impact factor and citation rate into a single metric, providing a weighted-average measure of scientific impact over the remaining lifespan of a published article. It also considers the uncertainty associated with future citation rates.

Conclusions: This study presents a new approach to assessing individual scientific output. By incorporating multiple metrics, it addresses the limitations and assumptions inherent in using a single metric.

Keywords *- citation rate, impact factor, performance indicator*

INTRODUCTION

New Public Management (NPM) refers to the implementation of management ideas from the business and private sector into public services (Haynes 2003).[1] NPM reforms have prompted public universities and research institutions to prioritize performance, competition, and efficiency in resource allocation. Various mechanisms have been developed to put these new governance structures into practice, including goal agreements, performance-oriented budgeting by indicators, and performance-oriented payment schemes (Schubert 2009). These management control systems have been applied in teaching, research, and outreach domains.

To assess research output within the context of NPM, bibliometric analyses can be conducted at different levels: micro (individual author), meso (e.g., university department), and macro level (e.g., institution or country). The citation impact factor (IF) is a widely

Corresponding Author: A.Gandjour@fs.de

used metric to evaluate scientific output, particularly at the micro and meso levels. The IF represents the average number of citations per paper in a journal during the preceding two years. A common practice in assessing researchers' performance is to aggregate journal IFs across all their publications. This approach assumes that the impact of a journal article remains constant regardless of the publication year, implying a linear increase in the number of citations over time. Additionally, it assumes negligible uncertainty in measuring journal IFs, implying that IFs do not vary from year to year.

In contrast to the IF, the Field-Weighted Citation Impact (Snowball Metrics Steering Group 2017) is primarily designed as a bibliometric indicator to evaluate research impact at a macro level, such as institutions. It calculates the ratio of actual citations received to the expected number of citations based on citation patterns within a specific field. This metric includes citations received up to three years after publication.

Since only a small percentage of total citations occurs in the first two years after publication (Van Nierop 2009), the cited half-life of a journal provides additional information beyond what the IF captures. The cited half-life represents the median age of articles in a journal that received citations in a given year. Kuo and Rupe (2007) propose combining the IF and cited half-life into a "reliability-based citation impact" (R-impact) factor by multiplying the cited half-life with the IF. This approach allows consideration of both the short- and long-term impact of an article. However, when summing R-impact factors across all publications to assess individual performance, it still assumes that the number of citations per year does not change (i.e., citations increase linearly with time).

Another metric that was introduced in 2005 and now is well-established is the h-index (Hirsch 2005). It strictly refers to an individual's performance (as opposed to a journal's performance). The h-index is based on the distribution of citations received by a researcher's publications. A scholar with an index of h has published h papers each of which has been cited in other papers at least h times. The h-index has been critiqued because its correlation with awards that indicate recognition by the scientific community has substantially declined (Koltun 2021).

An alternative approach to evaluate the scientific output of researchers is the use of expert-based journal lists. These lists evaluate scientific performance based on a subset of journals that belongs to a specific field or discipline, typically selected subjectively. Within a journal list, journals may also be ranked, taking into account factors such as journal reputation, IF, quality of citations, or the perceived contribution of a journal to the accumulated knowledge in the field (Bernick 2010). Expert-based journal rankings are particularly popular in the field of business administration/economics. However, a key challenge is that raters may assign higher scores to journals that align with their area of expertise due to familiarity (Serenko 2018).

According to Adler et al. (2009) and Molléri et al. (2018), the most commonly used indicators for assessing journals, papers, and individual scientists are the journal IF and the h-index. When evaluating the performance of individual scientists, the journal IF is often used as a proxy for the number of citations received by their papers.

The purpose of this study is to present a novel measure of individual scientific impact that combines IFs, citation rates, and expert opinion into a single metric. While numerous other indicators for assessing scholarly impact exist (see Cai et al., 2019, for a recent overview), this study focuses on the IF and the individual citation rate due to their popularity. For the same reason, expert-based opinion is considered. The study employs a Bayesian average calculation as the underlying methodology, resulting in a composite metric dubbed the 'Bayesian citation indicator' (BCI). To maintain simplicity, the study does not modify the underlying metrics' simplifying assumptions. The goal is not to create a "perfect" metric but to provide more information than a single metric alone. For instance, if a scientist ranks highly based on one metric but low based on another, the BCI allows for weighing and aggregating this information. By considering multiple metrics, this approach also mitigates the limitations and assumptions associated with relying solely on a single metric.

METHODS

The proposed individual research performance indicator combines past citations and the IF of an article by calculating the weighted-average scientific impact over the remaining life of the published article. This indicator is based on a parametric Gaussian-Gaussian Bayesian shrinkage estimator that takes into account the level of uncertainty associated with both past citations and the IF. The resulting composite measure provides a summary of the citation impact of an article by utilizing information from both measures. It should be noted that other metrics of scientific impact can also be included.

The parametric Gaussian-Gaussian Bayesian model is derived from Bayes theorem (Lynch 2007, p. 62ff.). In this model, both the prior distribution θ and the likelihood $Y|\theta$ follow Gaussian (normal) distributions, which can be expressed as:

$$\theta \sim N(\mu, \tau^2) \quad (1)$$

$$(Y|\theta) \sim N(\mu, \sigma^2). \quad (2)$$

The posterior distribution is also Gaussian with a mean given by:

$$E(\theta|Y) \sim B\mu + (1 - B)Y, \quad (3)$$

where $B = \sigma^2/(\sigma^2 + \tau^2)$ (Carlin 2008). As B lies between 0 and 1, the posterior mean represents a weighted average of the prior mean μ and the direct estimate Y (Carlin 2008). This Bayesian estimation process leads to a shrinkage effect, pulling the estimate towards the prior mean when σ^2 is relatively large compared to τ^2 (i.e., prior knowledge is more precise than the available data information) (Carlin 2008).

In our analysis μ represents the IF, and Y represents the average number of citations per year (the citation rate) for a specific article. As mentioned earlier, the IF serves as a proxy for individual scientific output and assumes that the number of citations increases linearly with time. Based on this assumption, an article published k years ago in a journal with a cited life of m years would have accumulated k/mth of its lifetime number of citations. The uncertainty surrounding the annual number of citations (Y) can be determined using the formula for the variance of the sample mean $\left(\frac{\sigma^2}{n}, \text{where } n \text{ is the number of independent observations}\right)$. Therefore, we express the variance (var) in year k after publication as follows:

$$\operatorname{var}(k) = \frac{\sigma^2}{k}. \tag{4}$$

Over time, the citation rate Y (Eq. (3)) will receive increasing weight. The additional weight obtained by Y is directly proportional to the reduction in uncertainty of the number of citations over time, represented by the decrease in $\operatorname{var}(k)$.

In contrast to the citation rate, the IF is assumed to have negligible uncertainty based on its real-world usage as a metric. At time m, the variance of the citation rate is also infinitesimal small, which is equivalent to the variance of the IF. This relationship is utilized in calculating the Bayesian average based on Eq. (3):

$$E(\theta|\bar{c}) = \frac{\frac{a}{k}}{a+\frac{a}{k}} IF + \left(1 - \frac{\frac{a}{k}}{a+\frac{a}{k}}\right)\bar{c}. \tag{5}$$

Here, a represents the level of variance of the citation rate and the IF at time m, while $\bar{c}$ is the average number of past citations per year. To operationalize the model, a is set to 1. After the first year ($k = 1$) we have $\frac{a}{k} = a$, indicating that the IF and $\bar{c}$ receive equal weight, i.e., $\frac{1}{2} IF + \frac{1}{2}\bar{c}$.

The weighted-average citation impact over the lifetime is calculated as follows:

$$c = (m-k)\left[\frac{\frac{a}{k}}{a+\frac{a}{k}} IF + \left(1 - \frac{\frac{a}{k}}{a+\frac{a}{k}}\right)\bar{c}\right] + k\bar{c}, \tag{6}$$

where $k\bar{c}$ is the accumulated number of citations over k years. The first summand in the equation represents the scientific impact beyond year k after publication.

Finally, evaluators have the option to introduce a factor β that measures the quality of citations, as well as a factor δ that defines the belonging to a specific field or discipline:

$$c^* = \beta\delta\left((m-k)\left[\frac{\frac{a}{k}}{a+\frac{a}{k}} IF + \left(1 - \frac{\frac{a}{k}}{a+\frac{a}{k}}\right)\bar{c}\right] + k\bar{c}\right), \tag{7}$$

where c^* is the quality-adjusted citation impact over the lifetime of a publication. The quality factor β can be measured within the interval [0,1] where zero represents no quality and one represents optimal quality. Similarly, the factor δ can be measured within the interval [0,1], or alternatively, it may be binary coded as zero or one.

To account for the quantity of research, the quality metric c^*, which refers to a single publication i, is aggregated over the number of publications n by an individual scholar:

$$\sum_{i=1}^{n} c_i^* . \qquad (8)$$

This aggregate measure provides the impact-adjusted output of an individual scholar. When using the h-index to assess performance across all articles, our formulas allow for calculating two h-indices: one related to past citations ($k\bar{c}$) and the other related to the predicted citation impact ($c - k\bar{c}$). It is important to note that h-indices may not be additive. For instance, if a scholar has published k papers, each of which has been cited k times, the h-index based on past citations already reaches the maximum value over their lifetime.

ILLUSTRATIVE APPLICATION EXAMPLE

Let's consider an individual who has published a total of four publications in a single journal, zero, one, two, and three years ago, respectively (zero years refers to just published). These publications have received zero, one, two, and three citations, respectively (i.e., one citation per year). Assuming an IF of one and a cited life of ten years, we can calculate the weighted-average citation impact over the remaining lifetime of each article using the first summand in Eq. (6). The results are as follows: ten, nine, eight, and seven citations for each of the four articles, respectively. Thus, the weighted-average citation impact over the lifetime is ten citations for each article. The h-index related to past and future citations is two and four, respectively.

DISCUSSION

The rise of NPM has generated an increasing demand for indicators to measure scientific performance at the micro, meso, and macro levels. This study presents a novel measure of individual scientific performance at the micro-level (BCI) that combines the IF, citation rate, quality of citations, and expert opinion into a single metric. The underlying methodology employed in this study is a Bayesian shrinkage estimator. It is worth noting that in the Bayesian paradigm uncertainty is epistemological, reflecting degrees of belief in states of nature (Gandjour 2003).

A field that has utilized rankings based on IFs, citation rates, and expert-based journal rankings is business administration/economics. One prominent expert-based list in the field is the "FT-50" list, consisting of 50 journals, which is utilized for the Financial Times business school rankings. Other rankings have also incorporated expert opinion alongside IFs, such as the former Handelsblatt ranking in German speaking countries and the Academic Journal Guide from the Chartered Association of Business Schools. Additionally, researchers in the field have been ranked based on their citation rates and h-index, as demonstrated by the German economist ranking by the Frankfurter Allgemeine Zeitung and Dilger and Müller (2012). Our metric allows for the aggregation of information provided by these various rankings.

There have been numerous criticisms regarding IFs, citation rates, and expert-based journal lists. For example, expert-based lists like the FT-50 faced criticism for excluding domains within business administration, such as business and management communication (Rogers 2007). Additionally, journal editors face challenges in distinguishing the top 5% or 10% of submissions from the next 5% or 10% (Conley 2012), and even top-ranked journals may publish works that receive no citations in the Web of Science (cf. Serenko 2011). Furthermore, pure citation-based rankings lack adjustments for the number of co-authors and do not exclude papers written on unrelated topics. For example, the German economist ranking by the Frankfurter Allgemeine Zeitung does not exclude papers on non-economic topics. While our metric was not specifically designed to address these concerns, it reduces their significance by incorporating a limited amount of information from each metric. This approach may be more acceptable than directly modifying one of the existing metrics.

As stated, our metric does not introduce new assumptions. However, our combined measure may be criticized for requiring information that is not readily available. One challenging piece of information to obtain is the cited lifetime m of an article. When such information is lacking, one may use the cited half time as an alternative. In fact, this is the approach adopted by Kuo and Rupe (2007) to calculate the so-called R-impact factor. With a shorter time horizon resulting from the use of the cited half time, individual citations are given less weight compared to the IF.

Another limitation of our metric is that it only applies to publications in journals with an IF or a similar indicator of the average number of citations over a specific time period, such as the SCImago Journal Rank (SJR indicator). Journals without an IF (or a similar indicator), as well as working papers, book chapters, and other types of publications are not considered.

It is important to note that the BCI does not aim to predict the future citation count of an article (this question has been addressed, e.g., by Acuna et al. (2012) and Yu et al. (2014)). Predicting future citations would require the inclusion of explanatory variables that have been shown to influence citation counts, such as the number and reputation of co-authors. However, from an assessor's perspective, it may not be desirable to assign weight to these additional variables. The question of construct validation and re-test reliability of the proposed Bayesian citation indicator should be approached in a similar manner. Only if the goal were to forecast the future citation count of a paper would the prediction made by the BCI need to be validated against empirical data on the actual number of citations in the long run.

As mentioned previously, expert-based journal lists or rankings often implicitly consider factors such as the IF, quality of citations, and relevance to a specific field or discipline. However, these factors can become intertwined in expert-based appraisals, making it challenging to disentangle their individual relevance. Some suggestions have been made on how to quantify the quality of citations. For instance, in 1976, it was proposed to assign higher weight to citations from highly cited journals compared to those from less-cited ones (Pinski 1976). Recent analyses have built upon this idea (Kodrzycki 2006). In fact, the SJR indicator captures the importance or prestige of the journals from which citations originate, in addition to considering the citation rate (both indicators are reported separately). Therefore, to avoid double counting, incorporating the SJR indicator implies excluding expert-based opinion from the performance evaluation.

The BCI can also be applied to evaluate scientific impact at the meso-level, particularly as bibliometric metrics tend to be more robust when applied to administrative units (meso level) rather than individuals (micro level) (Gauthier 1998). While the BCI already incorporates the most commonly used academic metrics of scientific impact, future research could enhance the BCI by integrating additional elements. These elements might include journal-specific functions that consider changes in citations over time or even non-academic impact, such as social media posts, press releases, news articles, and political debates sparked by academic work (Ravenscroft 2017).

Conclusions

In conclusion, this study introduces the Bayesian Citation Indicator (BCI) as a novel and comprehensive metric for assessing scientific performance at the micro-level. By combining factors like the Impact Factor, citation rate, citation quality, and expert opinion through Bayesian shrinkage estimation, the BCI offers a holistic approach to evaluating individual contributions to academia. It addresses limitations associated with existing metrics, providing a unified solution for assessing the impact of research. While the BCI's applicability is primarily limited to publications in journals with citation metrics, it offers a valuable tool for understanding the value and influence of scientific work. Additionally, the BCI can be extended to the meso-level, offering assessment capabilities for administrative units.

Notes

[1] Acknowledgement: The author would like to thank Amirhossein Sadoghi and Alexander Libman for valuable comments on an earlier version. The usual disclaimer applies.

Afschin Gandjour is Professor of Health Management at Frankfurt School of Finance & Management, Germany. E-mail: A.Gandjour@fs.de.

References

Acuna, D. E., Allesina, S., and Kording, K. P. (2012), “Future impact: Predicting scientific success”, Nature, Vol. 489 No. 7415, pp. 201-2.

Bernick, E., and Krueger, S. (2010), “An assessment of journal quality in public administration”, International Journal of Public Administration, Vol. 33 No. 2, pp. 98-106.

Cai, L., Tian, J., Liu, J., Bai, X., Lee, I., Kong, X., and Xia, F. (2019), “Scholarly impact assessment: a survey of citation weighting solutions”, Scientometrics, Vol. 118 No. 2, pp. 453-478.

Carlin, B. P., and Louis T. A. (2008), “Bayesian methods for data analysis”, Third edition, Chapman & Hall.

Conley, J. P. (2012), “Low acceptance rates, commercial publishing, and the future of scholarly communication”, Economics Bulletin, Vol. 32 No. 4, p. A37.

Dilger A., and Müller, B. A. (2012), „Ein Forschungsleistungsranking auf der Grundlage von Google Scholar“, Zeitschrift für Betriebswirtschaft, Vol. 82 No. 10, pp. 1089-1105.

Gandjour, A., and Lauterbach, K. W. (2003), “Inductive reasoning in medicine: lessons from Carl Gustav Hempel's 'inductive-statistical' model”, Journal of Evaluation in Clinical Practice, Vol. 9 No. 2, pp. 161-9.

Gauthier, É. (1998), “Bibliometric Analysis of Scientific and Technological Research: A User’s Guide to the Methodology, Science and Technology Redesign Project”, Statistics Canada.

Haynes, P. (2003), “Managing complexity in the public services”, Open University Press.

Hirsch, J. E. (2005), “An index to quantify an individual's scientific research output”, Proceedings of the National Academy of Sciences of the United States of America, Vol. 102 No. 46, pp. 16569-72.

Kodrzycki, Y. K., and Yu, P. D. (2006), “New approaches to ranking economics journals”, B E Journal of Economic Analysis & Policy, Vol. 5 No. 1.

Koltun V, Hafner D. The h-index is no longer an effective correlate of scientific reputation. PLoS One. 2021 Jun 28;16(6):e0253397.

Kuo, W., and Rupe, J. (2007), “R-impact: reliability-based citation impact factor”, IEEE Transactions on Reliability, Vol. 56 No. 3, pp. 366–367.

Lynch, S. M. (2007), “Introduction to applied Bayesian statistics and estimation for social scientists”, Springer Science & Business Media.

Molléri, J. S., Petersen, K., and Mendes, E. (2018), “Towards understanding the relation between citations and research quality in software engineering studies”, Scientometrics, Vol. 117 No. 3, pp. 1453-1478.

Pinski, G., and Narin, F. (1976), "Citation influence for journal aggregates of scientific publications: Theory with application to literature of physics", Information Processing & Management, Vol. 12 No. 5, pp. 297–312.

Ravenscroft, J., Liakata, M., Clare, A., and Duma, D. (2017), "Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements", PLoS One, Vol. 12 No. 3, pp. e0173152.

Rogers, P. S., Campbell, N., Louhiala-Salminen, L., Rentz, K., and Suchan, J. (2007), "The impact of perceptions of journal quality on business and management communication academics", Journal of Business Communication, Vol. 44 No. 4, pp. 403-426.

Schubert, T. (2009), "Empirical observations on new public management to increase efficiency in public research—Boon or bane?". Research Policy, 38(8), 1225-1234.

Serenko, A., and Dohan, M. (2011), "Comparing the expert survey and citation impact journal ranking methods: example from the field of artificial intelligence", Journal of Informetrics, Vol. 5 No. 4, pp. 629-48.

Serenko, A., & Bontis, N. (2018). A critical evaluation of expert survey-based journal rankings: The role of personal research interests. Journal of the Association for Information Science and Technology, 69(5), 749-752.

Snowball Metrics Steering Group. Snowball Metrics Recipe Book. December 8, 2017. https://www.elsevier.com/__data/assets/pdf_file/0006/53169/Snowball_Metrics_Recipe_Book.pdf

Van Nierop, E. (2009). Why do statistics journals have low impact factors?. Statistica Neerlandica, 63(1), 52-62.

Yu, T., Yu, G., Li, P. Y., and Wang, L. (2014), "Citation impact prediction for scientific papers using stepwise regression analysis", Scientometrics, Vol. 101 No. 2, pp. 1233-1252.

About IPMR

IPMR The International Public Management Review (IPMR) is the electronic journal of the International Public Management Network (IPMN). All work published in IPMR is double blind reviewed according to standard academic journal procedures.

The purpose of the International Public Management Review is to publish manuscripts reporting original, creative research in the field of public management. Theoretical, empirical and applied work including case studies of individual nations and governments, and comparative studies are given equal weight for publication consideration.

IPMN The mission of the International Public Management Network is to provide a forum for sharing ideas, concepts and results of research and practice in the field of public management, and to stimulate critical thinking about alternative approaches to problem solving and decision making in the public sector.

IPMN includes over 1300 members representing about one hundred different countries, both practitioners and scholars, working in all aspects of public management. IPMN is a voluntary non-profit network and membership is free.

ISSN 1662-1387