Journal of Informetrics 6 (2012) 169–176

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Skewed citation distributions and bias factors: Solutions to two core problems with the journal impact factor

Rüdiger Mutz*, Hans-Dieter Daniel

ETH Zurich Professorship for Social Psychology and Research on Higher Education, Mühlegasse 21, 8001 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 26 August 2011
Received in revised form
20 December 2011
Accepted 21 December 2011

Keywords:
Journal impact factor
Normalization
McCall's area transformation
Rubin Causal Model

ABSTRACT

The journal impact factor (JIF) proposed by Garfield in the year 1955 is one of the most prominent and common measures of the prestige, position, and importance of a scientific journal. The JIF may profit from its comprehensibility, robustness, methodological reproducibility, simplicity, and rapid availability, but it is at the expense of serious technical and methodological flaws. The paper discusses two core problems with the JIF: first, citations of documents are generally not normally distributed, and, furthermore, the distribution is affected by outliers, which has serious consequences for the use of the mean value in the JIF calculation. Second, the JIF is affected by bias factors that have nothing to do with the prestige or quality of a journal (e.g., document type). For solving these two problems, we suggest using McCall's area transformation and the Rubin Causal Model. Citation data for documents of all journals in the ISI Subject Category "Psychology, Mathematical" (Journal Citation Report) are used to illustrate the proposal.

1. Introduction

Scientific journals differ with respect to their position and prestige within the scientific community. One of the most commonly used and prominent indicators of a journal's position and prestige is the journal impact factor (JIF), which was introduced in 1955 by Garfield (1999):

> A journal's impact factor is based on 2 elements: the numerator, which is the number of citations in the current year to any items published in a journal in the previous 2 years, and the denominator, which is the number of substantive articles (source items) published in the same 2 years. (p. 979)

At the very beginning the JIF was to aid selection of highly cited and large journals for the Science Citation Index (Garfield, 1955, 2006). Nowadays, the JIF is used to generate rankings of journals to help scientists find important journals with potential excellent (in the sense of highly cited) contributions (Todorov & Glänzel, 1988). The JIF profits much from the fact that this measure can be easily reproduced from data provided by Thomson Reuters ISI, for instance, and it is available fast in connection with other journal impact measures (e.g., immediacy index) in the Journal Citation Reports (JCR). Additionally, the JIF is comprehensible, simple, and clearly defined, and comparable over time (Glänzel & Moed, 2002). However, there are some clear flaws, which have led to controversial discussions about the correctness of using the JIF to compare and evaluate journals (e.g., Boor, 1982; Leydesdorff & Bornmann, 2011a; Moed, Van Leeuwen, & Reeduk, 1999). In their state-of-the-art report, Glänzel and Moed (2002) listed several serious flaws of the JIF: among others, normalization for reference practices in different disciplines is missing; the merits of the citing journals are not taken into consideration; the peak in citations is not

* Corresponding author.
E-mail address: mutz@gess.ethz.ch (R. Mutz).

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2011.12.006

always 2 years; the citation frequency is affected by an age bias; one single measure might not represent the prestige and the position of a scientific journal (p. 174). For instance, Neuhaus, Marx, and Daniel (2009) found in their comparative analysis (Thomson Reuters Scientific versus Chemical Abstracts Service) for wide-scope journals (*Angewandte Chemie*, *Journal of the American Chemical Society*) that the literature databases offer only a rather unreliable indicator of the document type. Further, their findings showed that the composition of the journals in terms of length of the citation windows and thematic focus of the journals have a considerable impact on the overall JIF.

In this contribution we focus on two core problems with the JIF: first, citations of articles are generally not normally distributed, and what is more, the distribution of citations is affected by extreme values or outliers (Bornmann & Mutz, 2011; Bornmann, Mutz, Neuhaus, & Daniel, 2008). This fact has serious consequences for the use of the mean value in the calculation of the JIF, because mean values react very sensitively to outliers in general. A few extremely highly cited papers suffice to result in a strongly positive bias of the JIF. The mean value is of great importance in statistics, because it is arithmetically defined. The quadratic differences between each data point $\Sigma(x-\alpha)^2$ from a parameter α is a minimum, if α is the mean value. In other words, if one does not know anything about a single data point of a distribution (e.g., number of citations), the arithmetic mean is the best and most informative value with the smallest average (quadratic) residual to the real data point, the so-called expected value. However, due to the quadratic term the mean value is also strongly influenced by skewed distributions and outliers. Alternatively, the median (50% of the data are below the median) can be used, but it is not arithmetically defined. Additionally, robust statistic (Huber, 1981) offers several methods to investigate the stability of statistical procedures if the assumptions of statistical tests are violated. Instead, we favor in our contribution an approach (McCall‘s area transformation), which not only keeps in line with the original JIF definition based on mean values, but also considers current discussion on this problem (e.g., Bornmann & Mutz, 2011; Leydesdorff & Bornmann, 2011b).

Second, “there is a wide spread belief that the ISI Impact Factor is affected or ‘disturbed’ by factors that have nothing to do with (journal) impact” (Glänzel & Moed, 2002, p. 173). Glänzel and Moed (2002, p. 178) named the following five factors that may influence and bias the JIF: document type, subject matter, the paper’s age, the paper’s social status (due to the author’s institution, for instance), and the observation period (i.e., the citation window).

In the following we propose one solution for each of the two core problems and illustrate our proposal using journal citation data for the ISI Subject Category “Psychology, Mathematical” of the JCR.

2. Skewed citation distribution: McCall’s area transformation

As mentioned above, the distributions of citations are skewed, and the mean value-based JIF is strongly affected by extreme values or outliers. As a solution to this problem we suggest *McCall’s area transformation procedure* (Krus & Kennedy, 1977; McCall, 1922), which is quite close but not redundant to the percentile approach suggested by Bornmann and Mutz (2011), Bornmann, Mutz, Marx, Schier, and Daniel, (2011), and Leydesdorff and Bornmann (2011b) with similar objectives. In contrast to an ordinary linear transformation of a scale (e.g., $5 \times X + 10$), McCall’s (nonlinear) area transformation not only transforms the skewed citation distribution into a standard normal distribution (*z*-distribution) but also standardizes its third moment, that is, the skewness of the distribution (Dekking, Kraaikamp, Lopuhaö, & Meester, 2005).

Similar to Leydesdorff and Bornmann (2011b) percentiles are used in the first step. Given a distribution of rank ordered citations, for each citation category a percentile rank $(100 \times p)$% can be calculated, which is the percent proportion that $(100 \times p)$% of the citations fall below this value. For instance, a percentile rank of $p = 0.20$ for articles with 50 citations means, that 20% of all articles have less than 50 citations. Whereas Leydesdorff and Bornmann (2011b) stop here and base their journal impact approach on these percentiles, we go one step further. The percentile ranks or cumulative frequencies of citations scores are transformed into *z*-values of the standard normal distribution. For each percentile rank a certain *z*-value can be accurately assigned. For instance, a percentile rank or proportion of 0.5 corresponds to a *z*-value of zero, a proportion of 0.975 to a *z*-value of 1.96. This procedure has some advantages over the pure percentile approach, suggested by Leydesdorff and Bornmann (2011b). The standard normal distribution is defined precisely by its bell curve as shape, by its area under the curve of 1.0, by its mean value of 0 and its standard deviation of 1.0. *z*-values can be used as an ubiquitous currency, they can be added and averaged, they can be linearly transformed ($a \times z + \mathrm{b}$) into any other scale. Outliers of citations are also considered, but due to their low proportion their *z*-value is shrunken towards the mean. Whereas percentiles are uniformly distributed, *z*-values are per se normally distributed as presumed by most statistical procedures. In the end a new scale for citations is generated with all necessary properties (e.g., normal distribution) not only for calculating the JIF, but also for any further statistical analysis.

In detail, regarding citation data the JIF calculation procedure consists of five steps: first, the citation data for each journal of a certain ISI-subject category or field are collected according to Garfield’s definition of JIF and then pooled. Second, absolute frequencies of each citation category (0, 1, 2,. . .) are calculated for each journal.

For example, journal “A” has 30 articles with no citations, 20 articles with one citation, and 10 articles with two citations. Additionally, within each citation category the journals are ranked in ascending order of the number of articles (see Table 1, column 1). Journals with more papers in a certain citation category (e.g., 40 articles with 1 citations) are ranked higher than journals with fewer papers in the same category (e.g., 20 articles with 1 citation). Second, the number of articles in a citation category per journal is converted to proportions (column 4 in Table 1). Third, the cumulative proportions (column 5 in Table 1) are transformed into *z*-values of a standard normal distribution using *z*-Tables, which are included in most introductory statistic books. If two or more journals have the same citation frequencies or rank ties (e.g., 10 articles with