



# Estimating the accuracies of journal impact factor through bootstrap



Kuan-Ming Chen<sup>a</sup>, Tsung-Hau Jen<sup>a,\*</sup>, Margaret Wu<sup>b</sup>

<sup>a</sup> Science Education Center, National Taiwan Normal University, No. 88, 4th Section, Ting-Chou Road, Wen-Shan District, Taipei City 11677, Taiwan, ROC

<sup>b</sup> Work-based Education Research Centre (WERC), Victoria Institute for Education, Diversity and Lifelong Learning, Victoria University, PO Box 14428, Melbourne, VIC 8001, Australia

## ARTICLE INFO

### Article history:

Received 29 May 2013

Received in revised form

18 November 2013

Accepted 20 November 2013

Available online 15 December 2013

### Keywords:

Bootstrap

Journal impact factor

Error estimation

Education

Citation index

## ABSTRACT

The journal impact factor (JIF) reported in journal citation reports has been used to represent the influence and prestige of a journal. Whereas the consideration of the stochastic nature of a statistic is a prerequisite for statistical inference, the estimation of JIF uncertainty is necessary yet unavailable for comparing the impact among journals. Using journals in the Database of Research in Science Education (DoRISE), the current study proposes bootstrap methods to estimate the JIF variability. The paper also provides a comprehensive exposition of the sources of JIF variability. The collections of articles in the year of interest and in the preceding years both contribute to JIF variability. In addition, the variability estimate differs depending on the way a database selects its journals for inclusion. In the bootstrap process, the nested structure of articles in a journal was accounted for to ensure that each bootstrap replication reflects the actual citation characteristics of articles in the journal. In conclusion, the proposed point and interval estimates of the JIF statistic are obtained and more informative inferences on the impact of journals can be drawn.

© 2013 Elsevier Ltd. All rights reserved.

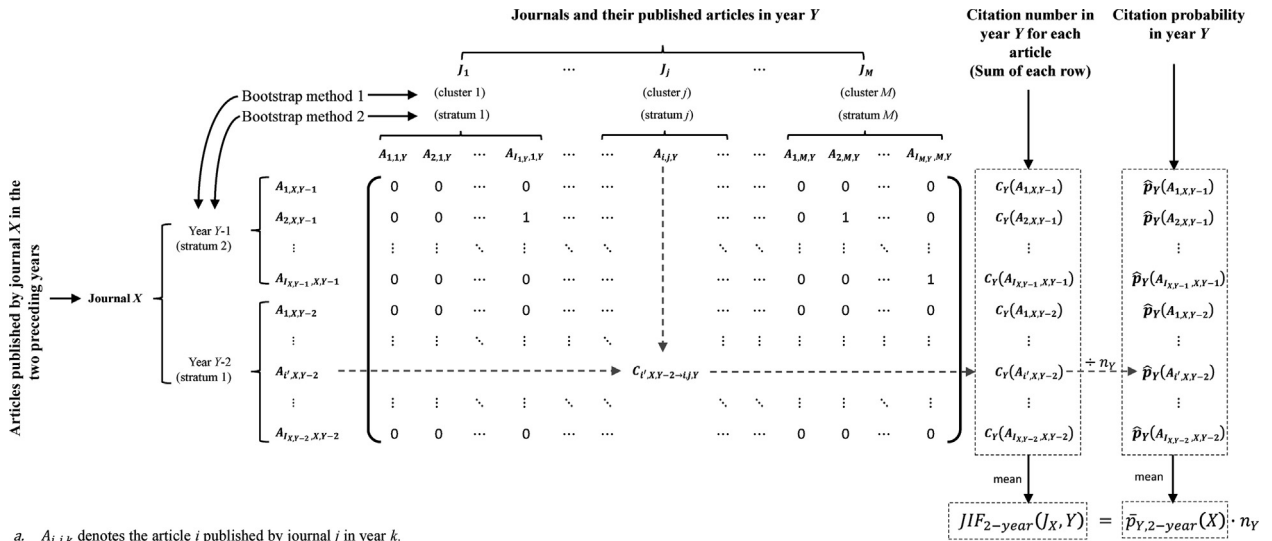
## 1. Introduction

The idea that citation frequency measures the impact of a journal dates back nearly 60 years ago (Garfield, 1955). Since 1975, the journal impact factor (JIF) reported in the Science Citation Index (SCI) has been used to represent the influence and prestige of a journal (Garfield, 1999). The 2-year and 5-year JIFs indicate the mean citation rate across articles of a journal in the two and five preceding years, respectively. The JIF obtained in this way is an average citation rate per article and allows comparison among journals regardless of the total number of articles published by a journal. However, from a statistical perspective, any measurement comes with variability or uncertainty, and this variability must be considered to make comparison or draw inferences (Greenwood, 2007; Leydesdorff, 2013; Schneider, 2013). To date, advanced statistical techniques on high-speed computers are readily accessible; therefore, the current study utilized the bootstrap method (Efron & Tibshirani, 1993) to estimate the average JIF with confidence interval (CI) and standard error (SE) for a journal. Furthermore, taking both the JIF definition and the characteristics of database into consideration, the sources underlying JIF variability are separately examined.

A number of concerns about JIF has been brought about by researchers, which mainly focused on the merits and disadvantages of using JIF (e.g., Abramo, D'Angelo, & Di Costa, 2010; Buela-Casal, Perakakis, Taylor, & Checa, 2006; Fassoulaki,

\* Corresponding author. Tel.: +886 2 77346773; fax: +886 2 29327187.

E-mail addresses: [skmchen@ntnu.edu.tw](mailto:skmchen@ntnu.edu.tw) (K.-M. Chen), [tsunghau@ntnu.edu.tw](mailto:tsunghau@ntnu.edu.tw), [tsunghau.jen@gmail.com](mailto:tsunghau.jen@gmail.com) (T.-H. Jen), [wu@edmeasurement.com.au](mailto:wu@edmeasurement.com.au) (M. Wu).



- a.  $A_{i,j,k}$  denotes the article  $i$  published by journal  $j$  in year  $k$ .
- b.  $C_{i',j',k' \to i,j,k}$  is 1 if  $A_{i',j',k'}$  was cited by  $A_{i,j,k}$ , and 0 if not cited.
- c.  $C_Y(A_{i',j',k'}) = \sum_{j=1}^M \sum_{i=1}^{I_{j,Y}} C_{i',j',k' \to i,j,Y}$  denotes the citation number of  $A_{i',j',k'}$  received in year  $Y$ .
- d.  $\hat{p}_Y(A_{i',j',k'}) = \frac{C_Y(A_{i',j',k'})}{n_Y}$  denotes the observed probability of  $A_{i',j',k'}$  being cited in year  $Y$ , where  $n_Y = \sum_{j=1}^M I_{j,Y}$  is the number of published articles collected in the database in year  $Y$ .

Fig. 1. Citation matrix and resampling framework for the 2-year JIF of journal X in year Y.

Paraskeva, Papilas, & Karabinis, 2000; Hecht, Hecht, & Sandberg, 1998; Holden, Rosenberg, Barker, & Onghena, 2006; Leydesdorff & Amsterdamska, 1990; Opthof & Leydesdorff, 2010; Ramirez, Garcia, & Del Rio, 2000; Sombatsompop & Markpin, 2005). However, there is not a great deal of publication on the uncertainty or precision associated with JIF (Greenwood, 2007; Schubert & Glänzel, 1983; Seglen, 1992; Smith, 2006). As researchers and institutions use the citation rate of an article in a journal as an indicator of the journal's general performance, the citation rate must be estimated along with the uncertainty surrounding it. Our proposed method of computing uncertainties associated with JIF provides more informative data so that better inference about journal impact can be drawn, as compared to a mere point estimate of JIF.

Fig. 1 illustrates how the 2-year JIF is obtained. Let  $J_X$  denote the target journal X for which the JIF is computed. Let  $Y$  denote the year of interest. The reference lists in all articles published in year  $Y$  across journals ( $J_1, J_2, \dots$ , and  $J_M$ , where  $M$  is the total number of journals in the database in year  $Y$ ) are used to check which articles published by  $J_X$  in years  $Y-1$  and  $Y-2$  have been cited. After the citation counts, the 2-year JIF of  $J_X$  in year  $Y$  is defined as

$$JIF_{2-year}(J_X, Y) = \frac{\sum_{k'=Y-2}^{Y-1} \sum_{i'=1}^{I_{X,k'}} \sum_{j=1}^M \sum_{i=1}^{I_{j,Y}} C_{i',X,k' \to i,j,Y}}{\sum_{k'=Y-2}^{Y-1} I_{X,k'}} \tag{1}$$

where  $C_{i',X,k' \to i,j,Y}$  takes values of 1 or 0 to denote whether article  $i'$  in journal X in year  $k'$  is cited by article  $i$  in journal  $j$  in year  $Y$ . The numerator of Eq. (1) stands for the total citation number across articles for  $J_X$ , and the denominator is the total number of articles published by  $J_X$  in years  $Y-1$  and  $Y-2$ . As the number of articles published by a journal varies across volumes, citations of these articles also differ. Thus, the mean citation rate of a journal is subject to chance elements. In statistical terminology, the JIF is a random variable. Any comparison among journals by the JIF statistic should consider this stochastic nature. Therefore, the variability associated with the JIF statistic is estimated to reflect the random nature of JIF in the current study.

Eq. (1) can be separated into two parts to better understand different sources of variability of JIF. One part is the citation probability in year  $Y$  ( $\hat{p}_Y$ ) of each article published in the two preceding years. Given that the observed probability of article  $A_{i',X,k'}$  in journal X being cited in year  $Y$  is

$$\hat{p}_Y(A_{i',X,k'}) = \frac{\sum_{j=1}^M \sum_{i=1}^{I_{j,Y}} C_{i',X,k' \to i,j,Y}}{n_Y} \tag{2}$$

the total citation is thus  $\hat{p}_Y(A_{i',X,k'}) \cdot n_Y$ , where  $n_Y = \sum_{j=1}^M I_{j,Y}$  denotes the number of articles collected in the database in year  $Y$ .

The second part of Eq. (1) is the average citation probability across the articles published in the two preceding years

$$\bar{p}_{Y,2-year}(X) = \frac{\sum_{k'=Y-2}^{Y-1} \sum_{i'=1}^{I_{X,k'}} \hat{p}_Y(A_{i',X,k'})}{\sum_{k'=Y-2}^{Y-1} I_{X,k'}} \tag{3}$$

Download English Version:

<https://daneshyari.com/en/article/523972>

Download Persian Version:

<https://daneshyari.com/article/523972>

[Daneshyari.com](https://daneshyari.com)