Journal of Informetrics 8 (2014) 181–196

Contents lists available at ScienceDirect

Journal of Informetrics

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

Estimating the accuracies of journal impact factor through bootstrap

Kuan-Ming Chen[a], Tsung-Hau Jen[a,*], Margaret Wu[b]

[a] Science Education Center, National Taiwan Normal University, No. 88, 4th Section, Ting-Chou Road, Wen-Shan District, Taipei City 11677, Taiwan, ROC
[b] 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.

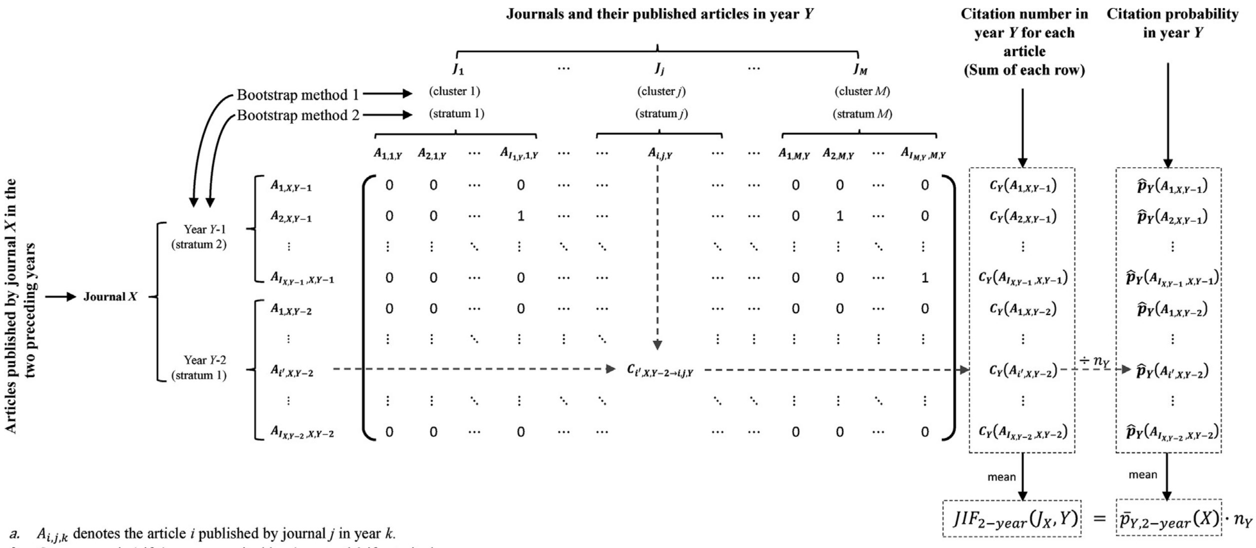
1. Introduction

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 (K.-M. Chen), tsunghau@ntnu.edu.tw, tsunghau.jen@gmail.com (T.-H. Jen), wu@edmeasurement.com.au (M. Wu).

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.11.007



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. $\hat{p}_Y(A_{i',j',k'}) = {}^{C_Y(A_{i',j',k'})}/_{n_Y}$ 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 ($J_1, J_2, \ldots$, 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

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

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

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

and

$$\mathrm{JIF}_{2\text{-year}}(J_X, Y) = \bar{p}_{Y,2\text{-year}}(X) \cdot n_Y. \quad (4)$$

Therefore, $\hat{p}_Y(A_{i',X,k'})$ forms the kernel of $\bar{p}_{Y,2\text{-year}}$ estimation, and $\bar{p}_{Y,2\text{-year}}$ further considers the collection of articles in the preceding years. Accordingly, different sources of variability can have effects on the values of $\hat{p}_Y$ and $\bar{p}_Y$.

Historical data could be used to estimate the probability distribution of JIF for a journal so that the variability is obtained. However, it is impractical to collect such a large amount of data of a database for several decades, especially for a newly established database or for a new journal. Alternatives have been developed to estimate the uncertainty of bibliometric indices. Greenwood (2007) fitted a random effects Poisson model within a Bayesian framework using Markov chain Monte Carlo method to estimate the 95% credible intervals of JIFs and JIF ranks. Baccini, Barabesi, Marcheselli, and Pratelli (2012) used a survival function to simulate individual authors' expected h-index and their associated variability on the basis of likelihood function and the empirical distribution of citation rates of articles published by the authors. With both studies, the parametric distributions about the citation per article published by the target journal or by the author must be made a priori. In addition, the citation variability caused by using different databases and the nested structure of articles in their publishing journals or in publishing years were overlooked. Nevertheless, these studies all have the common view that the measure of uncertainty along with the point estimate should be considered. Therefore, in the current study an empirical resampling method known as *bootstrap* (Efron & Tibshirani, 1993) is adopted to estimate the uncertainty of JIF. The bootstrap method relies less on the parametric assumptions. In addition, our study takes into account the stochastic nature of the JIF in relation to the nested structure of articles in journals and the mechanisms of collecting journals in a database.

With the bootstrap method, the articles of a journal published in year Y and in the preceding years can be regarded as a sample of the candidate articles of the journal because "*Any journal in effect takes a small, biased sample (biased in that subjective selection criteria are involved) of articles from a finite but large pool of articles*" (Amin & Mabe, 2000, p. 4). The process of publishing an article, based on different publication criteria among journals, has chance elements, depending on who submits which article to which journal to be reviewed by whom and when articles are published. The wait-time for publication for accepted articles varies and is generally unrelated to the impact of an article. Thus, the mean citation rate of a journal is subject to sampling error to a large extent (Greenwood, 2007). In addition, in the current study the nature of '*biased sampling*' is taken care of by treating articles in a journal as a cluster or as stratified samples rather than as simple random samples. Accordingly, journals and their published articles in a database are repetitively sampled with replacement, and each sample simulates the collection of journals and their published articles for a year. The JIF is then computed for each journal within each sample. Through resampling multiple times, a distribution of JIFs for each journal is obtained; thereby the variability of JIF of the journal can be estimated.

The Database of Research in Science Education (DoRISE, http://dorise.sec.ntnu.edu.tw/main/), including 79 journals of education and science education published in Taiwan, ROC, was used for demonstrating the bootstrap methods. DoRISE, similar to the SCI and SSCI databases, was established with informatics techniques suitable for computerized analysis. Therefore, the method demonstrated with DoRISE in the current study is applicable to other similar databases.

2. Method

To compute confidence intervals for JIF estimates, sampling distributions of the JIF statistics need to be established. To build these sampling distributions, we use a resampling procedure called bootstrap. There are two separate parts in simulating the citation rates of journal articles. The first part involves the sampling of articles that cite previously published articles. The second part involves the sampling of articles that are being cited.

The first part of the sampling process begins with the choice of a database to determine the sample set of journals and their articles in year Y (i.e., n_Y articles) that cite previously published articles. The second part of the sampling procedure involves sampling articles of the target journal in the preceding years and determining whether these sampled articles are cited by n_Y articles sampled in the first part of the sampling procedure.

Different samples of n_Y articles lead to differences in the citation rate of the articles of the target journal. Similarly, different samples of articles sampled from the target journal have different rates of being cited by n_Y articles. Therefore, through the two sampling procedures, the sources of variability of the JIF statistic come from sampling the articles in year Y and the articles of target journal in the preceding years. These two sampling parts are simulated using the bootstrap method.

2.1. Data source

Articles of 79 journals in DoRISE from years 2006 to 2011 were analyzed (see Appendix A for a complete listing of journals). As some journals did not publish over the whole period under study, the number of journals varied between the 2-year and 5-year JIFs. For the 2011 2-year JIFs, 41,905 references of 1201 articles across 73 journals in year 2011 were used to count the citation among 2541 articles across 73 journals published in years 2009 and 2010. For the 2011 5-year JIFs, the identical 41,905 references were used to count the citation among 5641 articles across 62 journals published in the five preceding years (2006–2010). In the subsequent analysis, a random ID was assigned to each journal for the anonymity of journals.

2.2. Method: bootstrap

In statistics, we are often interested in population parameter $\theta = t(\mathbf{P})$, where $\mathbf{P}$ denotes the population data of interest and t a function of the population data. However, population data are typically inaccessible; therefore, representative samples are drawn from the population. Through sampling, $\hat{\theta} = s(\mathbf{x})$ is obtained, where $\hat{\theta}$ is the sample statistic and $\mathbf{x}$ is the sample data. The central limit theorem is generally used to obtain the sampling distribution of $\hat{\theta}$ and the accuracy in estimating the population parameter θ, assuming that observations are independently and identically distributed random variables. However, in the case of JIF, the assumptions are not satisfied because self-citations within research groups or mutual citations among articles in similar disciplines exist, suggesting a dependency among observations. The bootstrap method requires no such parametric assumptions and, therefore, is suitable for estimating the uncertainty of JIF, especially with complex sampling procedures in that articles are published by a journal and there is a nested structure between articles and journals.

According to the plug-in principle (Efron & Tibshirani, 1993; Efron, 2003), the CI of $\hat{\theta}$ is obtained by the sampling distribution of $\hat{\theta}^* - \hat{\theta}$ instead of the distribution of $\hat{\theta} - \theta$, where $\hat{\theta}^*$ is the statistic from each bootstrapped sample. In our study, $\hat{\theta}$ denotes a reported JIF of any journal in DoRISE. Essentially, the bootstrap samples of JIF should follow the sampling frame (Efron, 2003; Pons, 2007) to retain the link between articles and their publishing journals in a database. Because the citation rates of an article differ from year to year (Marx & Cardona, 2003), this year-to-year fluctuation at the article-level further adds to the JIF variability at the journal-level. Accordingly, the articles of a journal are sampled by treating years as strata. Assume that journal X remains in the database from year Y-n to year Y; the variability of n-year JIF for journal X in year Y is estimated by resampling both the articles published by journal X in the n preceding years year-by-year and the articles within each journal in year Y. Every time after an article of journal X in the n preceding years was sampled, the articles in year Y were also sampled. It should be noted that different databases have different procedures for recruiting journals. Therefore, in the current study we propose two bootstrap methods for the two cases of journal inclusion in a database. The first case assumes that the collection of journals in a database is a random sample of journals from a population of journals. The second case assumes that the collection of journals represents the population of journals and remains the same from one year to another. As a matter of fact, the inclusion of journals in a database is likely to be in-between these two cases.

2.2.1. Bootstrap method 1

If journals in a database are a random part of the population of journals and the articles of a journal can be treated as a sample of candidate articles to be published in the journal, the JIF variability comes from both the samplings of the articles published by the target journal and the articles used to count the citation of articles published by the target journal. Therefore, not only the articles published by the target journal in the n preceding years need to be resampled by treating the publishing years as strata, but also the articles and journals published in year Y need to be resampled based on a two-stage cluster sampling procedure to retain the nested structure of articles in a journal. First, the journals were randomly sampled with replacement, and the sample sizes were the total numbers of journals collected in the database in year Y. Second, articles were randomly sampled with replacement from the articles of each previously sampled journal, and the sample size was the total number of articles of that journal.

The sampling procedure of articles in year Y was repeated for each sampled article of journal X in the n preceding years. The citation number was computed based on whether an article sampled in the n preceding years was cited by the sampled articles of each sampled journal in year Y. The sampling procedure was then repeated until the number of articles sampled year-by-year in the n preceding years was equal to the total number of articles in journal X, rendering one bootstrap replication. For each target journal, 1500 bootstrap replications were obtained.

2.2.2. Bootstrap method 2

If journals in a database are treated as the population per se, the JIF variability will only arise from the articles within each journal and not from the collection of journals since the collection does not change. Therefore, the procedure of bootstrap method 2 is the same as in method 1 except that all the journals in the database in year Y are used without sampling, whereas the articles within each journal are resampled with replacement. For each target journal, 1500 bootstrap replications were obtained.

2.3. Data analysis

Two sampling distributions of JIF resulted from 1500 replications from each of the two bootstrap methods for each journal in DoRISE. The 2-year and 5-year JIFs with their SEs and 95% CIs for 2011 were reported. The 95% CI was taken as the lower and upper endpoints where 95% of the distribution lies in-between (Efron & Tibshirani, 1993).

The percentile CI ensures non-negative endpoints of CI, as JIF is between 0 and n_Y, being the citation rates of journal articles. Consequently, the bootstrapped values of JIF are not normally distributed. However, the nonparametric percentile method provides confidence intervals for a population mean (i.e., average citation number) with coverage error to the order of n_s^{-1}, where n_s is the sample size (Efron, 1981; Gross & Lai, 1996). Therefore, even though the bootstrapped JIFs are not normally distributed the 95% percentile intervals could still be adequate estimates of the 95% confidence intervals due to large sample size.

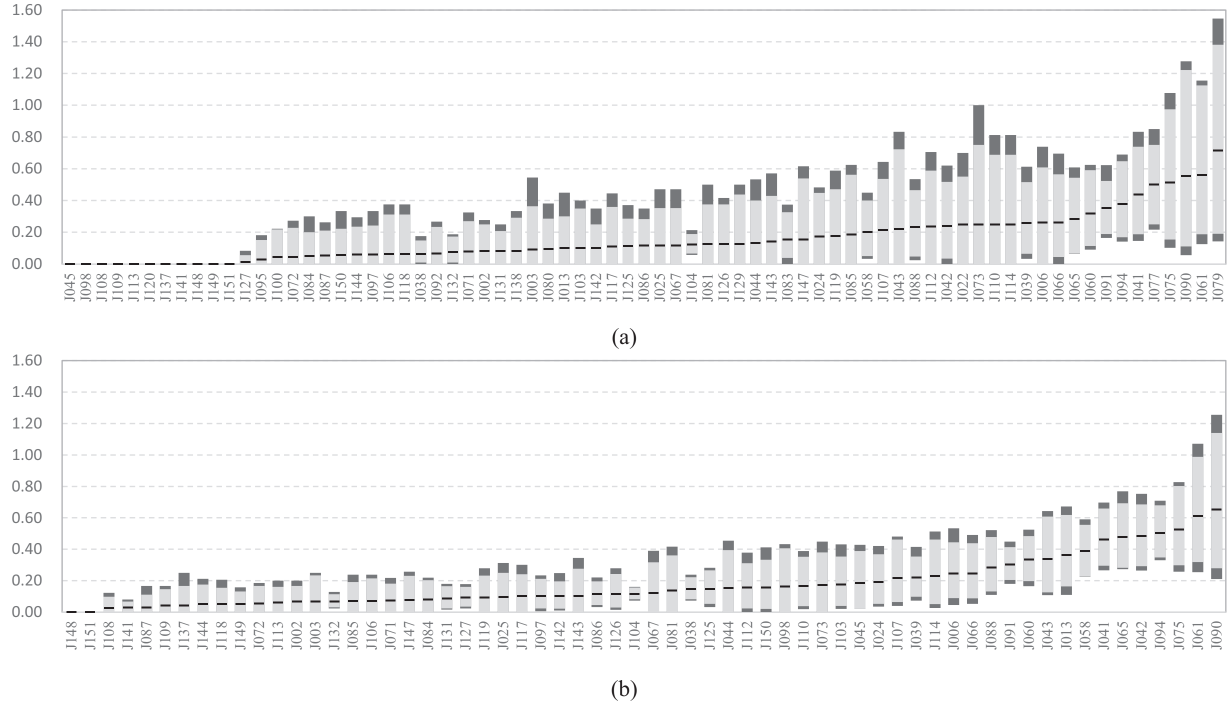


Fig. 2. (a) 2011 2-year and (b) 5-year JIF*s (bold line), with 95% CIs estimated with bootstrap method 1 (dark gray bar) and bootstrap method 2 (light gray bar) for journals in DoRISE.

In addition to providing SEs and CIs, the means of the sampling distribution were compared to the JIFs calculated based on the original citation data matrix without bootstrap.

3. Results and discussion

For each journal, the original JIF without bootstrap and the average JIFs (JIF*) by two bootstrap methods, with the corresponding SEs and 95% CIs, are presented in Tables B.1 and B.2 (Appendix B). The JIF and JIF*s demonstrate similarity for every journal, indicating that the expected value of the JIF sampling distribution obtained by bootstrap and the published JIF represent exactly the same statistic. However, the bootstrap method can estimate the variability of JIF when the sampling procedures are complex. Fig. 2 shows the 2011 (a) 2-year and (b) 5-year JIF*s with 95% CIs by two bootstrap methods for the journals in DoRISE. The range of 95% CI correlates positively with the value of JIF*. For each journal, the range of 95% CI by bootstrap method 2 is smaller than the CI by bootstrap method 1 since less sampling variability is expected for bootstrap method 2 as the journals in the database represent the population of journals. However, an actual database comprises neither the whole population nor a random selection of journals; thus, we suggest that the range of 95% CI for published JIF lies in-between those obtained by these two bootstrap methods. Finally, the difference in JIF*s between any two journals in DoRISE is likely to be statistically non-significant because of their overlapping 95% CIs.

Based on sampling theory (Kish, 1965; Lohr, 2010), the sources contributing to the variability of the JIF estimate are discussed separately as follow.

3.1. Variability of $\hat{\mathbf{p}}_{\mathbf{Y}}$

The distribution of citation number is binomial as an article is either cited or not cited. Accordingly, as a database with n_Y articles in year Y is a random sample from the population with N_Y articles, the SE of $\hat{p}_Y$ (citation probability of an article published in the preceding years) is

$$\mathrm{SE}(\hat{p}_Y) \approx \sqrt{\frac{\hat{p}_Y\left(1-\hat{p}_Y\right)}{n_Y}}. \qquad (5)$$

However, in bootstrap method 1 when the journals in DoRISE are randomly sampled from a population of journals, the selected articles need to be treated as being nested within a selected journal rather than being randomly assigned to the journal. This is a cluster sampling of articles so that the SE of $\hat{p}_Y$ is approximately estimated as (Kish, 1965).

$$\mathrm{SE}_{\mathrm{cluster}}(\hat{p}_Y) \approx \sqrt{D_{\mathrm{eff}} \cdot \left[\frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}\right]}. \tag{6}$$

D_{eff} denotes the design effect for a cluster sampling, and with equal cluster sampling it is estimated as (Hansen, Hurwitz, & Madow, 1953)

$$D_{\mathrm{eff}} \approx 1 + \rho(I_Y - 1), \tag{7}$$

where I_Y is the cluster size (i.e., journal size). Mostly, the journal sizes in year Y are unequal, and I_Y is replaced by the weighted averaged cluster size I'_Y (Donner & Klar, 1994)

$$I'_Y = \frac{\sum_{j=1}^{M} I_{j,Y}^2}{\sum_{j=1}^{M} I_{j,Y}}. \tag{8}$$

The intra-cluster correlation coefficient ρ is established through dividing the between-cluster variance by the sum of the within- and between-cluster variances. If different journals have different rates of citing an article, the ρ will be large so that the variability of $\hat{p}_Y$ will increase. For example, an article in medical education is more likely cited by articles of journals in medical education, but less likely by articles of journals in other disciplines, such as reading education or science education. Therefore, a larger ρ for this article would be expected than for another article related to all the disciplinary areas in education. Similarly, high self-citation within a journal also leads to a larger ρ as well as a larger SE of $\hat{p}$.

Eq. (6) illustrates how the SE of $\hat{p}$ for each article is influenced. First, the probability of an article to be cited ($\hat{p}_Y$) is generally smaller than .5; therefore, the $\hat{p}_Y(1-\hat{p}_Y)$ is an increasing function of the citation rate (e.g., for articles in DoRISE the $\hat{p}_Y$ ranged between 0.000 and 0.007). Consequently, the more probable an article of a journal is cited, the larger is its SE. Second, the SE of $\hat{p}_Y$ is proportional to the inverse of the square root of the number of articles in a database (n_Y) and, when more articles are included, SE becomes smaller. Third, articles with higher intra-journal correlation of citations (larger ρ) have larger SE of $\hat{p}_Y$ and vice versa. Finally, according to Eq. (7), given the same number of articles in year Y in the database, more journals (i.e., fewer articles in each journal) included in the database will lead to a smaller SE of $\hat{p}_Y$ than fewer journals (i.e., more articles in each journal).

If the journals in the database are the population per se (bootstrap method 2), the journals would not contribute to sampling variability. However, since articles are published in journals, a journal forms a sampling stratum when articles are sampled. Assuming that the sample size of a stratum is proportional to the population size of that stratum, the SE of $\hat{p}_Y$ is thus (Cochran, 1977; Jen, Tam, & Wu, 2011)

$$\mathrm{SE}_{\mathrm{stratified}}(\hat{p}_Y) \approx \sqrt{D'_{\mathrm{eff}} \cdot \left[\frac{\hat{p}_Y\left(1-\hat{p}_Y\right)}{n_Y}\right]}, \tag{9}$$

where the design effect D'_{eff} for a stratified sampling procedure is

$$D'_{\mathrm{eff}} \approx 1 - \varphi \tag{10}$$

and φ is the ratio of between-stratum variance to the total variance. Therefore, in addition to the effects of $\hat{p}_Y$ and n_Y on the SE as in bootstrap method 1, a larger φ indicates a larger between-stratum variance and a smaller within-stratum variance. In this case, for a journal in year Y, the citation status is more consistent across articles and a smaller variance of $\hat{p}_Y$ is expected. Although journals of a database are treated as clusters in bootstrap method 1 and as strata in bootstrap method 2, the value of ρ is equal to the value of φ. However, their effects on the variability of $\hat{p}_Y$ differ depending on the mechanism of selecting journals to include in a database.

3.2. Variability of $\bar{p}_Y$

Although the articles published in the preceding years can be treated as a sample to represent the citation of the target journal, the citation of articles fluctuates from one year to another. Therefore, the effect of stratification by year as opposed to random selection needs to be considered. The SE of average citation probability in year Y is (Cochran, 1977; Jen et al., 2011)

$$\mathrm{SE}_{\mathrm{stratified}}(\bar{p}_Y) \approx \sqrt{D''_{\mathrm{eff}} \cdot \frac{\mathrm{Var}(\hat{p})}{n_I}}, \tag{11}$$

where $\text{Var}(\hat{p})$ is the variance of citation probability in year Y across all the articles published by journal X and n_I is the number of articles published in the preceding years as

$$n_I = \sum_{k'=Y-2}^{Y\text{-}1} I_{X,k'}. \tag{12}$$

Similar to Eq. (10), D''_{eff} is the design effect for a stratified sample:

$$D''_{\text{eff}} \approx 1 - \varphi' = 1 - \frac{\text{Var}(\bar{p}_h)}{\text{Var}(\hat{p})} \approx \frac{\text{Var}_w(\hat{p})}{\text{Var}(\hat{p})}, \tag{13}$$

where φ' is the ratio of between-stratum variance ($\text{Var}(\bar{p}_h)$) to the variance of citation probability ($\text{Var}(\hat{p})$). If the between-stratum and within-stratum variances add up to the total variance, Eq. (13) can be simplified and the design effect can be approximated by the ratio of within-stratum variance ($\text{Var}_w(\hat{p})$) to the variance of citation probability. Therefore,

$$\text{SE}_{\text{stratified}}(\bar{p}_Y) \approx \sqrt{\frac{\text{Var}_w(\hat{p})}{n_I}}. \tag{14}$$

According to Eq. (14), the variability of 5-year JIF is likely to be smaller than the 2-year JIF because more articles (n_I) are published over five years than over two years. In addition, a journal in which articles are evenly cited demonstrates a smaller within-stratum variance of citation probability (i.e., $\text{Var}_w(\hat{p})$), hence a smaller SE compared to a journal in which only a few articles are highly cited and the rest are not cited at all.

3.3. Combining $\hat{\boldsymbol{p}}_{\boldsymbol{Y}}$ and $\bar{\boldsymbol{p}}_{\boldsymbol{Y}}$

Both the sources of variability in estimating $\hat{p}_Y$ and $\bar{p}_Y$ were incorporated in our study so that the articles of journals in year Y and the articles of journals in the two or five preceding years were separately resampled. Two journals, J041 and J090, are used as exemplars to demonstrate the typical effects, shown in Figs. 3 and 4 respectively.

In both figures, a set of six panels illustrates sources of JIF variability for each exemplar journal: (a)'s present the frequency distribution of 2- and 5-year JIFs (JIF*) while simultaneously considering both the sources of variability in estimating $\hat{p}$ and $\bar{p}$ based on bootstrap method 1. (b)'s show the distribution of JIFs (JIF†) as the journals and their articles in year Y were bootstrapped, whereas the target journals and their published articles in the preceding years were kept intact. In contrast, (c)'s demonstrate the distribution of JIFs (JIF‡) as the journals and their articles in year Y were kept intact, whereas the articles in the target journal in the preceding years were bootstrapped.

The sampling distributions of 2011 5-year JIF in the right-side three panels show narrower ranges than the ones of 2011 2-year JIF in the left-side three panels, indicating that a larger number of articles (n_I) leads to smaller variability. In addition, for every journal the variance of n-year JIF* approximates to the sum of variances of n-year JIF† and JIF‡. For example, the variance in estimating the 2-year JIF* of journal J041 (Fig. 3(a)-1) is equal to the sum of variances in estimating JIF† (Fig. 3(b)-1) and JIF‡ (Fig. 3(c)-1) ($0.17^2 \approx 0.14^2 + 0.10^2$). Similarly, by comparing the left three panels in Fig. 4, the variance in estimating the 2-year JIF* of journal J090 is equal to the sum of variances in estimating $\hat{p}$ and $\bar{p}$ ($0.34^2 \approx 0.26^2 + 0.23^2$). The same conclusion can be made based on the bootstrapped results of the 5-year JIF of journals J041 and J090.

For comparison, the correlation of JIFs between two target years was also computed by using longitudinal data. Appendix C lists the overall reliabilities of the measurement for both the 2-year and 5-year JIFs in DoRISE based on the true score model in classical test theory and the correlations between the two sets of JIFs in 2010 and 2011. Accordingly, the average SE of JIF based on the reliabilities might be underestimated and the proposed bootstrap methods provide more reasonable estimations of JIFs for the journals in DoRISE.

4. Implications and limitations of this study

One of the essential ideas in the current study is to treat articles as samples nested in journals and to treat publishing years as strata, representative of the citation rate of articles of a journal at population level, and, using sampling theory, the sources of JIF variability are examined. Accordingly, the generalizability of JIF and its variability is closely related to how well the samples represent their population. For a database which collects journals and articles through random sampling or systematic sampling (e.g., cluster sampling or two-stage stratified cluster sampling) the samples could represent the whole population. DoRISE attempted to collect all educational journals in Taiwan, although some journals were excluded when authorization was not obtained. Nevertheless, the articles in DoRISE could be regarded as a sample representing the population of educational articles published in Taiwan. In general, when the selection of journals and their articles for a database is not a representative sample of the population of journals and articles, the generalizability of JIF through bootstrap is confined to the database per se.

Researchers or institutions may want to compare journals across databases. Comparison among JIFs from different databases is feasible when the databases contain representative samples of journals from the same population and the

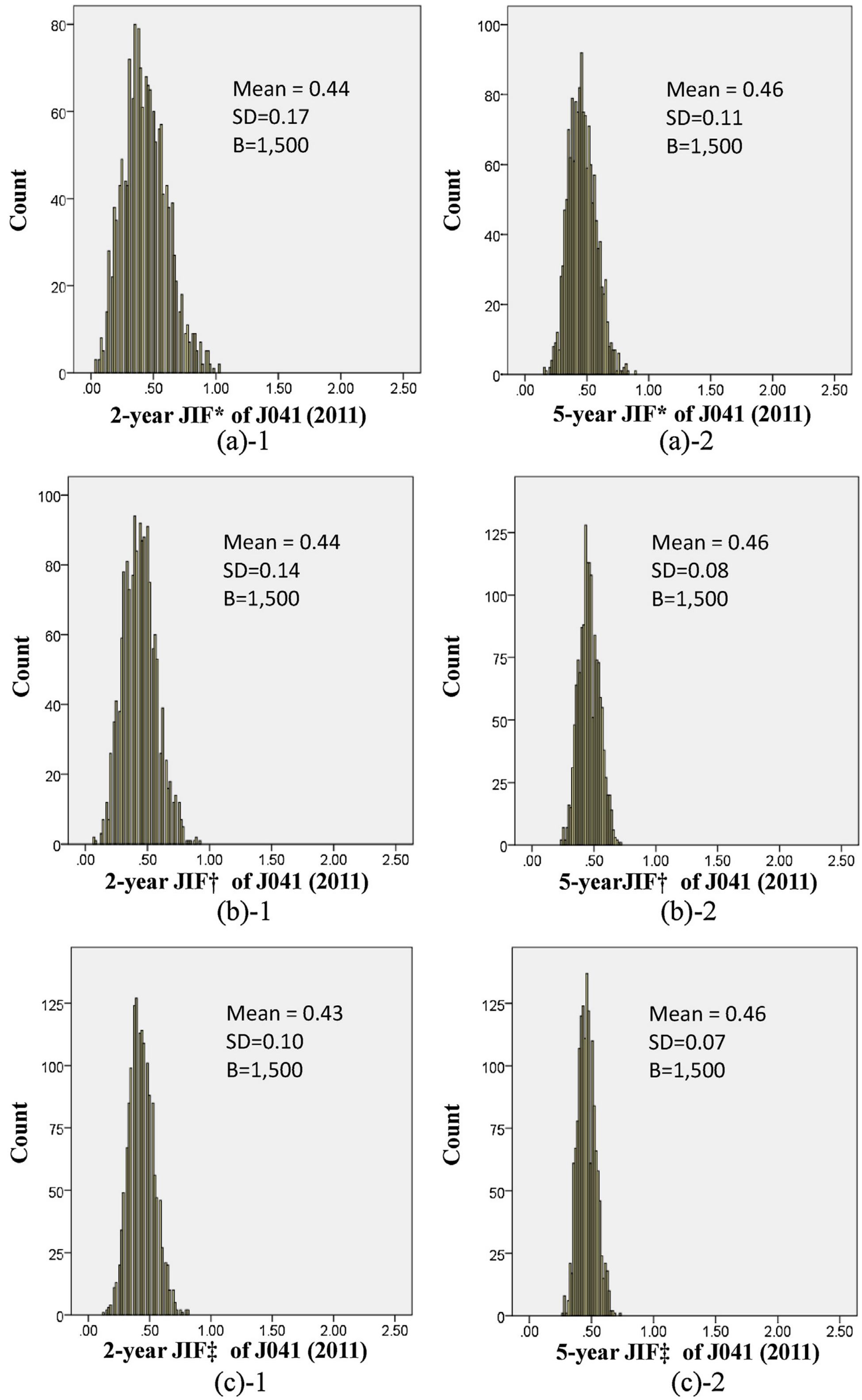


Fig. 3. Frequency of sampling distributions of 2-year and 5-year JIFs for journal J041 based on 1500 replications by bootstrap method 1. (a) Taking both the error sources in estimation of $\hat{p}$ and $\bar{p}$ into consideration (JIF*); (b) taking only the error source in estimation of $\hat{p}$ into consideration (JIF†), and (c) taking only the error source in estimation of $\bar{p}$ into consideration (JIF‡).

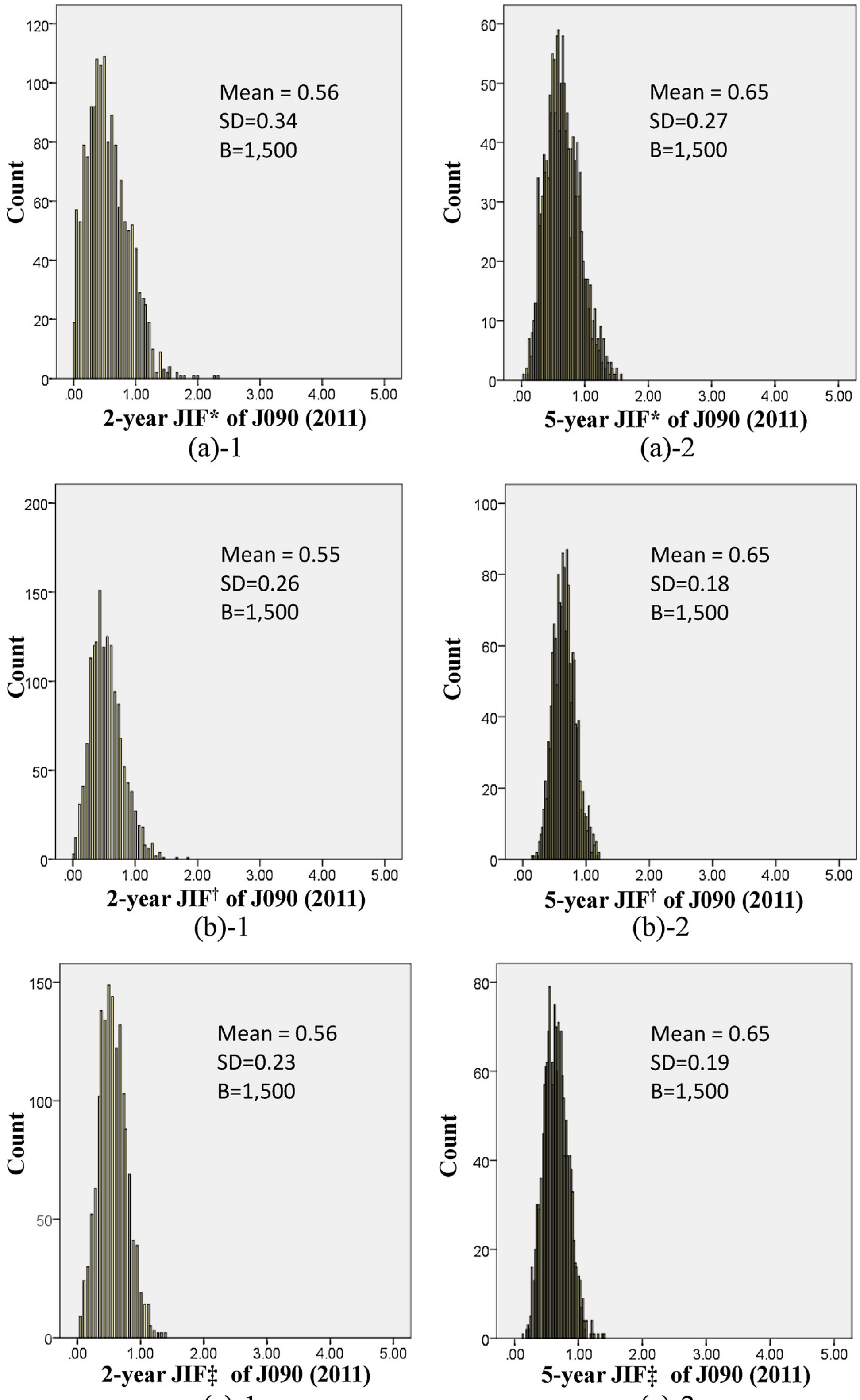


Fig. 4. Frequency of sampling distributions of 2-year and 5-year JIFs for journal J090 based on 1500 replications by bootstrap method 1. (a) Taking both the error sources in estimation of $\hat{p}$ and $\bar{p}$ into consideration (JIF*); (b) taking only the error source in estimation of $\hat{p}$ into consideration (JIF†), and (c) taking only the error source in estimation of $\bar{p}$ into consideration (JIF‡).

JIFs are normalized to the same scale. Accordingly, the JIF of a journal from a database in Taiwan is not necessarily comparable with journals from outside of Taiwan. In addition, as a multiplier of $\bar{p}_Y$ in Eq. (4), the number of articles published in the target year in the database (n_Y) could be a scaling factor that varies across databases. Therefore, $\bar{p}_Y$ as the potential citation probability could be a better index than JIF for the comparison of impact among journals across databases.

Various publishing practices render differences in citation patterns, and therefore journals with identical JIF differ in impact regarding their subject fields (Dorta-González & Dorta-González, 2013; Garfield, 1979; Waltman, Van Eck, Van Leeuwen, & Visser, 2013). Some revised indices have been proposed to overcome such differences in citation patterns across subject fields (Zitt & Small, 2008). One method uses the ratio of a journal's JIF and the average citation rate for all the journals in the same subject field (Marshakova-Shaikevich, 1996; Van Leeuwen & Moed, 2002). Another method calculates the average frequency of citations in an article in the target journal's subject field (Moed, 2010), as an indicator to normalize the impact of the target journal. However, citation practices across subject fields may differ qualitatively from each other, hence comparing JIFs across subject fields calls for caution, and more research should be carried out in the future.

Researchers have criticized that the JIF of a journal may not reflect the overall quality of articles published by the journal (Abramo et al., 2010; Hecht et al., 1998; Holden et al., 2006; Leydesdorff & Amsterdamska, 1990), or that the growing percentages of self-citation within a journal or research group might change the implication of JIF (Fassoulaki et al., 2000). However, the issues addressed in the current study are more about the reliability rather than the validity of JIF, as the focus is on the stochastic nature of JIF.

How journals and their articles are selected in a database influences the variability estimation and the generalizability of a journal's JIF. The bootstrap method makes the assumption that articles in the database are samples representative of the population of journal articles. It is important that resampling should be conducted according to the sampling framework based on which the original sample was selected. With DoRISE the articles are selected by journals, and the journals are selected for inclusion in the database on a yearly basis. Therefore, the nested structure of journals in a year as strata and the nested structure of the articles in a journal as clusters or strata are necessary for estimating the JIF variability. If a database selected articles according to authors or institutions, the authors or institutions will need to form additional sampling clusters while bootstrapping. That is, the nested structure embedded in a database should not be overlooked in order to reflect actual variability of the JIF statistic.

The method proposed in this study requires detailed information of all articles, such as the citations of each paper over a period of time. Such databases of information are not easily accessible for individual researchers. Therefore, enterprises such as Elsevier (Scopus) and Thomson Reuters (Web of Knowledge) or large institutions responsible of constructing journal citation databases are encouraged to report more indices, rather than a mere report of a point estimate. Finally, the uncertainty of JIF could be estimated through different approaches such as using longitudinal data or assuming the parametric probability distribution function. The differences among these approaches can be examined in future studies.

5. Conclusion

The use of JIFs has become increasingly high-stakes, such as in making appointment and promotion decisions of academic staff. Therefore, it is critical that decision makers realize the limitations of JIFs. We sought to explore additional statistics to make journal impact data more informative. While the estimated JIF is important in assessing journals, the variability associated with the estimated impact is also crucial to making statistical inference. Therefore, in the current study two different bootstrap methods are proposed to estimate the CIs and SEs of JIF by (a) appropriately resampling the journals and the articles in the database and (b) examining the sources of variability associated with the estimation. In addition to reporting a single value for JIF, reporting the uncertainty surrounding JIF is helpful for researchers and policy makers to assess the impact of journals. Based on point and interval estimates, informative inferences about the influence and prestige of journals can be drawn. According to the results presented in the current study, the JIF have large standard errors. It suggests that less confidence should be placed on the comparison among journals' impact based on the JIF.

Acknowledgements

This study was made possible through generous grants from the Aim for the Top University project of National Taiwan Normal University funded by the Ministry of Education and the project (NSC97-2511-S-003-045-MY5) funded by the National Science Council of Taiwan. The authors also appreciate the mentoring and editorial assistance from Dr. Larry Yore and Mrs. Shari Yore.

Appendix A. Journals and years of publication in DoRISE.

Table A.1.

Table A.1
Journals published in DoRISE for the 2006–2011 period.

Journal title	Publishing year[b]					
	2006	2007	2008	2009	2010	2011
1. Area and social studies education research	n/a	n/a	n/a	n/a	n/a	○
2. Bulletin of Association for the History of Science	○	○	○	○	○	○
3. Bulletin of Chung Hwa College of Medical Technology	○	○	○	○	○	○
4. Bulletin of Civic and Moral Education	n/a	n/a	n/a	n/a	n/a	○
5. Bulletin of Eastern-Taiwan Special Education	○	○	○	○	○	○
6. Bulletin of Educational Entrepreneurship and Management	○	○	○	○	○	○
7. Bulletin of Educational Psychology	○	○	○	○	○	○
8. Bulletin of Educational Research	○	○	○	○	○	○
9. Bulletin of Educational Research (Hsinchu County)[a]	○	○	○	○	○	○
10. Bulletin of Special Education	○	○	○	○	○	○
11. Bulletin of Special Education and Rehabilitation	○	○	○	○	○	○
12. Chengshiu General Education Journal[a]	○	○	○	○	○	○
13. Chinese Journal of Guidance and Counseling	n/a	n/a	n/a	○	○	○
14. Chinese Journal of School Health	○	○	○	○	○	○
15. Chinese Journal of Science Education	○	○	○	○	○	○
16. Chinese Physics Education	n/a	n/a	n/a	○	○	○
17. Chung Cheng Educational Studies	○	○	○	○	○	○
18. Comparative Education	○	○	○	○	○	○
19. Contemporary Educational Research Quarterly	○	○	○	○	○	○
20. Curriculum & Instruction Quarterly	○	○	○	○	○	○
21. Educational Journal of NHCUE	n/a	n/a	n/a	○	○	○
22. Educational Policy Forum	○	○	○	○	○	○
23. Educational Review	○	○	○	○	○	○
24. English Teaching & Learning	○	○	○	○	○	○
25. Formosan Education and Society	○	○	○	○	○	○
26. Formosan Journal of Sexology	○	○	○	○	○	○
27. Gifted Education	○	○	○	○	○	○
28. Guandu General Education Journal	○	○	○	○	○	○
29. Guidance Quarterly	○	○	○	○	○	○
30. Instructional Technology & Media	○	○	○	○	○	○
31. International Journal of Science and Engineering	n/a	n/a	n/a	n/a	n/a	○
32. Journal of Adult Lifelong Education	○	○	○	○	○	○
33. Journal of Child Care	○	○	○	○	○	○
34. Journal of Counseling & Guidance	○	○	○	○	○	○
35. Journal of Disability Research	○	○	○	○	○	○
36. Journal of Early Childhood Education	n/a	n/a	n/a	○	○	○
37. Journal of Education	n/a	n/a	n/a	○	○	○
38. Journal of Education & Psychology	○	○	○	○	○	○
39. Journal of Education National Changhua University of Education	○	○	○	○	○	○
40. Journal of Education of Taipei Municipal University of Education	n/a	n/a	n/a	n/a	n/a	○
41. Journal of Education Practice and Research	n/a	n/a	n/a	○	○	○
42. Journal of Education Studies	○	○	○	○	○	○
43. Journal of Educational Measurement and Statistics	○	○	○	○	○	○
44. Journal of Educational Research	○	○	○	○	○	○
45. Journal of Educational Research and Development	○	○	○	○	○	○
46. Journal of Educational Theory and Practice	n/a	n/a	n/a	○	○	○
47. Journal of Elementary Education	○	○	○	○	○	n/a
48. Journal of Environmental Education Research	○	○	○	○	○	○
49. Journal of General Education: Concept & Practice	○	○	○	○	○	n/a
50. Journal of Gifted Education	○	○	○	○	○	n/a
51. The Journal of Guidance & Counseling	○	○	○	○	○	○
52. Journal of Health Promotion and Health Education	n/a	n/a	n/a	○	○	○
53. Journal of Medical Education	○	○	○	○	○	○
54. Journal of National Pingtung University of Education: Education	n/a	n/a	n/a	○	○	○
55. Journal of National Taichung University: Education	○	○	○	○	○	○
56. Journal of National Taichung University: Mathematics, Science & Technology	○	○	○	○	○	○
57. Journal of Research in Education Sciences	n/a	n/a	n/a	○	○	○
58. Journal of Research on Elementary and Secondary Education	○	○	○	○	○	n/a
59. Journal of Scientific and Technological Studies	○	○	○	○	○	n/a
60. Journal of Special Education	○	○	○	○	○	○
61. The Journal of Study in Child and Education	○	○	○	○	○	n/a
62. Journal of Taipei Municipal University of Education: Education	○	○	○	○	○	○
63. Journal of Technological and Vocational Education	○	○	○	○	○	○
64. Journal of the Speech–Language–Hearing Association of Taiwan	n/a	n/a	n/a	○	○	○
65. Kaohsiung Normal University Journal: Education and Social Sciences	○	○	○	○	○	○
66. Kaohsiung Normal University Journal: Sciences and Technology	○	○	○	○	○	○
67. National Chiayi University Journal of Educational Research	n/a	n/a	n/a	n/a	n/a	○

Table A.1 (*Continued*)

Journal title	Publishing year[b]					
	2006	2007	2008	2009	2010	2011
68. NTTU Educational Research Journal	○	○	○	○	○	○
69. Psychological Testing	○	○	○	○	○	○
70. Research and Development in Science Education Quarterly	○	○	○	○	○	○
71. Research in Arts Education	○	○	○	○	○	○
72. School Health Nursing[a]	n/a	n/a	n/a	n/a	n/a	○
73. Science Education Monthly	○	○	○	○	○	○
74. Secondary Education	○	○	○	○	○	○
75. Special Education Quarterly	○	○	○	○	○	○
76. Taiwan Journal of Mathematics Teachers	○	○	○	○	○	○
77. Taiwan Journal of Sociology of Education	○	○	○	○	○	○
78. Technology Museum Review	○	○	○	○	○	○
79. Tzu-Chi University Journal of Educational Research	○	○	○	○	○	○

[a] Journal title is in Chinese and translated into English by authors.
[b] 'n/a' denotes 'not available' for the current analysis.

Appendix B. Descriptive statistics for the journals in DoRISE.

Tables B.1 and B.2.

Table B.1
2011 2-year JIF and 95% CI estimation through two bootstrap methods.

Journal ID	No. of citable articles	JIF	Bootstrap method 1			Bootstrap method 2		
			JIF*	SE	95% CI (2.5%, 97.5%)	JIF*	SE	95% CI (2.5%, 97.5%)
J002	36	0.08	0.09	0.08	(0.00, 0.28)	0.08	0.07	(0.00, 0.25)
J003	11	0.09	0.10	0.15	(0.00, 0.55)	0.09	0.12	(0.00, 0.36)
J006	23	0.26	0.27	0.19	(0.00, 0.74)	0.26	0.15	(0.00, 0.61)
J013	20	0.10	0.11	0.13	(0.00, 0.45)	0.10	0.09	(0.00, 0.30)
J022	20	0.25	0.25	0.18	(0.00, 0.70)	0.25	0.14	(0.00, 0.55)
J024	29	0.17	0.18	0.14	(0.00, 0.48)	0.17	0.11	(0.00, 0.45)
J025	17	0.12	0.12	0.14	(0.00, 0.47)	0.12	0.11	(0.00, 0.35)
J038	108	0.06	0.07	0.05	(0.00, 0.18)	0.06	0.04	(0.01, 0.15)
J039	31	0.26	0.26	0.15	(0.03, 0.61)	0.26	0.12	(0.06, 0.52)
J041	48	0.44	0.44	0.17	(0.15, 0.83)	0.44	0.14	(0.19, 0.74)
J042	29	0.24	0.25	0.16	(0.00, 0.62)	0.24	0.12	(0.03, 0.52)
J043	18	0.22	0.23	0.24	(0.00, 0.83)	0.23	0.20	(0.00, 0.72)
J044	15	0.13	0.13	0.16	(0.00, 0.53)	0.13	0.12	(0.00, 0.40)
J045	18	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J058	60	0.20	0.20	0.11	(0.03, 0.45)	0.20	0.09	(0.05, 0.40)
J060	44	0.32	0.32	0.14	(0.09, 0.62)	0.32	0.12	(0.11, 0.59)
J061	32	0.56	0.56	0.26	(0.13, 1.16)	0.56	0.24	(0.19, 1.13)
J065	46	0.28	0.28	0.14	(0.07, 0.61)	0.28	0.12	(0.07, 0.54)
J066	23	0.26	0.26	0.18	(0.00, 0.70)	0.26	0.14	(0.04, 0.57)
J067	17	0.12	0.11	0.13	(0.00, 0.47)	0.11	0.11	(0.00, 0.35)
J071	37	0.08	0.08	0.09	(0.00, 0.32)	0.08	0.07	(0.00, 0.27)
J072	22	0.05	0.04	0.08	(0.00, 0.27)	0.05	0.06	(0.00, 0.23)
J073	8	0.25	0.25	0.27	(0.00, 1.00)	0.25	0.22	(0.00, 0.75)
J075	39	0.51	0.52	0.25	(0.10, 1.08)	0.52	0.22	(0.15, 0.97)
J077	60	0.50	0.50	0.17	(0.22, 0.85)	0.50	0.13	(0.25, 0.75)
J079	21	0.71	0.71	0.35	(0.14, 1.55)	0.71	0.29	(0.19, 1.38)
J080	21	0.10	0.10	0.12	(0.00, 0.38)	0.09	0.09	(0.00, 0.29)
J081	16	0.13	0.13	0.14	(0.00, 0.50)	0.12	0.11	(0.00, 0.38)
J083	52	0.15	0.15	0.10	(0.00, 0.37)	0.15	0.08	(0.04, 0.33)
J084	20	0.05	0.05	0.09	(0.00, 0.30)	0.05	0.06	(0.00, 0.20)
J085	16	0.19	0.18	0.17	(0.00, 0.63)	0.19	0.15	(0.00, 0.56)
J086	60	0.12	0.12	0.09	(0.00, 0.35)	0.12	0.07	(0.00, 0.28)
J087	19	0.05	0.05	0.09	(0.00, 0.26)	0.06	0.07	(0.00, 0.21)
J088	43	0.23	0.24	0.13	(0.02, 0.53)	0.23	0.11	(0.05, 0.47)
J090	18	0.56	0.56	0.34	(0.06, 1.28)	0.56	0.28	(0.11, 1.22)
J091	85	0.35	0.36	0.12	(0.16, 0.62)	0.35	0.09	(0.19, 0.52)
J092	30	0.07	0.07	0.08	(0.00, 0.27)	0.07	0.06	(0.00, 0.23)
J094	71	0.38	0.38	0.14	(0.14, 0.69)	0.38	0.12	(0.17, 0.65)
J095	33	0.03	0.03	0.05	(0.00, 0.18)	0.03	0.04	(0.00, 0.15)
J097	33	0.06	0.06	0.10	(0.00, 0.33)	0.06	0.07	(0.00, 0.24)
J098	12	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)

Table B.1 (*Continued*)

Journal ID	No. of citable articles	JIF	Bootstrap method 1			Bootstrap method 2		
			JIF*	SE	95% CI (2.5%, 97.5%)	JIF*	SE	95% CI (2.5%, 97.5%)
J100	23	0.04	0.04	0.07	(0.00, 0.22)	0.04	0.06	(0.00, 0.22)
J103	20	0.10	0.09	0.12	(0.00, 0.40)	0.10	0.10	(0.00, 0.35)
J104	318	0.12	0.12	0.04	(0.06, 0.21)	0.12	0.03	(0.07, 0.19)
J106	16	0.06	0.06	0.11	(0.00, 0.38)	0.06	0.09	(0.00, 0.31)
J107	28	0.21	0.22	0.17	(0.00, 0.64)	0.21	0.13	(0.00, 0.54)
J108	21	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J109	19	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J110	16	0.25	0.25	0.23	(0.00,0.81)	0.25	0.18	(0.00, 0.69)
J112	17	0.24	0.23	0.20	(0.00, 0.71)	0.23	0.15	(0.00, 0.59)
J113	18	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J114	16	0.25	0.25	0.22	(0.00, 0.81)	0.25	0.19	(0.00, 0.69)
J117	18	0.11	0.11	0.13	(0.00, 0.44)	0.11	0.10	(0.00, 0.36)
J118	16	0.06	0.06	0.10	(0.00, 0.38)	0.06	0.09	(0.00, 0.31)
J119	17	0.18	0.18	0.17	(0.00, 0.59)	0.18	0.13	(0.00, 0.47)
J120	25	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J125	35	0.11	0.12	0.10	(0.00, 0.37)	0.11	0.07	(0.00, 0.29)
J126	24	0.13	0.13	0.12	(0.00, 0.42)	0.12	0.10	(0.00, 0.38)
J127	72	0.01	0.01	0.02	(0.00, 0.08)	0.01	0.02	(0.00, 0.06)
J129	16	0.13	0.12	0.14	(0.00, 0.50)	0.13	0.12	(0.00, 0.44)
J131	48	0.08	0.08	0.07	(0.00, 0.25)	0.08	0.06	(0.00, 0.21)
J132	107	0.07	0.07	0.05	(0.00, 0.19)	0.08	0.04	(0.01, 0.17)
J137	10	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J138	24	0.08	0.08	0.10	(0.00, 0.33)	0.08	0.08	(0.00, 0.29)
J141	61	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J142	40	0.10	0.10	0.10	(0.00, 0.35)	0.10	0.07	(0.00, 0.25)
J143	14	0.14	0.14	0.16	(0.00, 0.57)	0.14	0.13	(0.00, 0.43)
J144	17	0.06	0.06	0.10	(0.00, 0.29)	0.06	0.08	(0.00, 0.24)
J147	13	0.15	0.15	0.18	(0.00, 0.62)	0.15	0.15	(0.00, 0.54)
J148	28	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J149	27	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J150	18	0.06	0.06	0.09	(0.00, 0.33)	0.05	0.07	(0.00, 0.22)
J151	38	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)

Table B.2
2011 5-year JIF and 95% CI estimation through two bootstrap methods.

Journal ID	No. of citable articles	JIF	Bootstrap method 1			Bootstrap method 2		
			JIF*	SE	95% CI (2.5%, 97.5%)	JIF*	SE	95% CI (2.5%, 97.5%)
J002	60	0.07	0.07	0.06	(0.00, 0.20)	0.07	0.04	(0.00, 0.17)
J003	30	0.07	0.06	0.08	(0.00, 0.25)	0.07	0.07	(0.00, 0.23)
J006	45	0.24	0.25	0.13	(0.04, 0.53)	0.25	0.09	(0.09, 0.44)
J013	55	0.36	0.36	0.14	(0.11, 0.67)	0.36	0.12	(0.16, 0.62)
J024	57	0.19	0.20	0.10	(0.04, 0.42)	0.19	0.08	(0.05, 0.37)
J025	32	0.09	0.10	0.09	(0.00, 0.31)	0.09	0.07	(0.00, 0.25)
J038	248	0.15	0.14	0.04	(0.07, 0.24)	0.15	0.04	(0.08, 0.22)
J039	82	0.22	0.22	0.09	(0.07, 0.41)	0.22	0.07	(0.10, 0.35)
J041	132	0.46	0.46	0.11	(0.27, 0.70)	0.46	0.09	(0.30, 0.66)
J042	89	0.48	0.48	0.13	(0.26, 0.75)	0.48	0.10	(0.29, 0.69)
J043	56	0.34	0.34	0.14	(0.11, 0.64)	0.34	0.12	(0.13, 0.61)
J044	33	0.15	0.15	0.12	(0.00, 0.45)	0.15	0.10	(0.00, 0.39)
J045	49	0.18	0.18	0.10	(0.02, 0.43)	0.18	0.09	(0.02, 0.39)
J058	144	0.39	0.39	0.10	(0.23, 0.59)	0.39	0.08	(0.23, 0.56)
J060	122	0.34	0.33	0.09	(0.16, 0.52)	0.34	0.07	(0.20, 0.48)
J061	85	0.61	0.62	0.20	(0.25, 1.07)	0.62	0.18	(0.32, 0.99)
J065	117	0.48	0.48	0.13	(0.27, 0.77)	0.48	0.11	(0.28, 0.69)
J066	57	0.25	0.25	0.12	(0.05, 0.49)	0.24	0.09	(0.09, 0.44)
J067	41	0.12	0.12	0.10	(0.00, 0.39)	0.12	0.08	(0.00, 0.32)
J071	83	0.07	0.07	0.06	(0.00, 0.22)	0.07	0.05	(0.00, 0.18)
J072	54	0.06	0.06	0.05	(0.00, 0.19)	0.06	0.04	(0.00, 0.17)
J073	29	0.17	0.17	0.13	(0.00, 0.45)	0.17	0.10	(0.00, 0.38)
J075	101	0.52	0.52	0.14	(0.26, 0.83)	0.53	0.13	(0.30, 0.80)
J081	36	0.14	0.14	0.11	(0.00, 0.42)	0.14	0.09	(0.00, 0.36)
J084	64	0.08	0.08	0.06	(0.00, 0.22)	0.08	0.05	(0.00, 0.20)

Table B.2 (*Continued*)

Journal ID	No. of citable articles	JIF	Bootstrap method 1			Bootstrap method 2		
			JIF*	SE	95% CI (2.5%, 97.5%)	JIF*	SE	95% CI (2.5%, 97.5%)
J085	42	0.07	0.07	0.07	(0.00, 0.24)	0.07	0.05	(0.00, 0.19)
J086	159	0.11	0.11	0.05	(0.03, 0.22)	0.11	0.04	(0.04, 0.19)
J087	36	0.03	0.03	0.05	(0.00, 0.17)	0.03	0.04	(0.00, 0.11)
J088	92	0.28	0.28	0.10	(0.11, 0.52)	0.28	0.09	(0.13, 0.48)
J090	43	0.65	0.65	0.27	(0.21, 1.26)	0.65	0.22	(0.28, 1.14)
J091	201	0.30	0.30	0.07	(0.18, 0.45)	0.30	0.05	(0.20, 0.41)
J094	173	0.50	0.50	0.10	(0.33, 0.71)	0.50	0.08	(0.35, 0.68)
J097	90	0.10	0.10	0.06	(0.01, 0.23)	0.10	0.05	(0.02, 0.21)
J098	37	0.16	0.16	0.12	(0.00, 0.43)	0.16	0.10	(0.00, 0.41)
J103	51	0.18	0.18	0.10	(0.02, 0.43)	0.17	0.08	(0.04, 0.35)
J104	855	0.12	0.11	0.02	(0.07, 0.16)	0.12	0.02	(0.08, 0.16)
J106	42	0.07	0.07	0.07	(0.00, 0.24)	0.07	0.06	(0.00, 0.21)
J107	78	0.22	0.22	0.12	(0.04, 0.48)	0.22	0.10	(0.06, 0.46)
J108	37	0.02	0.02	0.04	(0.00, 0.12)	0.02	0.03	(0.00, 0.10)
J109	48	0.04	0.04	0.05	(0.00, 0.17)	0.04	0.04	(0.00, 0.15)
J110	54	0.17	0.17	0.10	(0.02, 0.39)	0.17	0.08	(0.04, 0.35)
J112	45	0.16	0.16	0.10	(0.00, 0.38)	0.16	0.07	(0.02, 0.31)
J113	50	0.06	0.06	0.06	(0.00, 0.20)	0.06	0.05	(0.00, 0.16)
J114	39	0.23	0.23	0.13	(0.03, 0.51)	0.23	0.11	(0.05, 0.46)
J117	50	0.10	0.10	0.08	(0.00, 0.30)	0.10	0.06	(0.00, 0.24)
J118	39	0.05	0.05	0.06	(0.00, 0.21)	0.05	0.05	(0.00, 0.15)
J119	43	0.09	0.09	0.08	(0.00, 0.28)	0.09	0.06	(0.00, 0.23)
J125	96	0.15	0.14	0.06	(0.03, 0.28)	0.14	0.05	(0.05, 0.27)
J126	70	0.11	0.11	0.07	(0.01, 0.28)	0.11	0.06	(0.03, 0.24)
J127	174	0.09	0.09	0.04	(0.02, 0.18)	0.09	0.03	(0.03, 0.16)
J131	128	0.09	0.09	0.04	(0.02, 0.18)	0.08	0.04	(0.02, 0.16)
J132	298	0.07	0.07	0.03	(0.02, 0.13)	0.07	0.02	(0.03, 0.11)
J137	24	0.04	0.04	0.07	(0.00, 0.25)	0.04	0.05	(0.00, 0.17)
J141	150	0.03	0.03	0.02	(0.00, 0.08)	0.03	0.02	(0.00, 0.07)
J142	97	0.10	0.10	0.06	(0.01, 0.25)	0.10	0.05	(0.02, 0.20)
J143	29	0.10	0.10	0.10	(0.00, 0.34)	0.11	0.08	(0.00, 0.28)
J144	40	0.05	0.05	0.06	(0.00, 0.21)	0.05	0.05	(0.00, 0.18)
J147	39	0.08	0.08	0.07	(0.00, 0.26)	0.08	0.06	(0.00, 0.23)
J148	69	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)
J149	76	0.05	0.05	0.04	(0.00, 0.16)	0.05	0.03	(0.00, 0.13)
J150	51	0.16	0.16	0.11	(0.00, 0.41)	0.16	0.08	(0.02, 0.33)
J151	95	0.00	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)

Appendix C. Overall reliability for measuring the journal impact factor in DoRISE

The true score model (Lord, 1965) in classical test theory can be applied to estimate the overall reliability of JIF measurement on the journals in DoRISE; then based on the reliability, we can roughly estimate the average SE of the measurement. The average SE estimates given by this method were compared to the average SE estimates based on the two bootstrap methods.

Assuming that the measured JIF reflects a relatively stable feature of a journal, one way to evaluate the reliability of JIF measurement is to calculate the correlation between two sets of impact factors in two consecutive years. Therefore, to estimate the overall reliability of JIF measurement on the journals in DoRISE, the 2-year and 5-year JIFs of 2010 and 2011 were calculated according to the definition of JIF; the correlation between the two sets of JIFs could be seen as the overall reliability of the measurement. If the random error assumption is sustained and the true impact factors in two consecutive years for all journals measured through the database are the same, the reliability coefficient (R) of the measurement is defined as

$$R \equiv \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2}. \tag{C.1}$$

Then we have

$$\sigma_E^2 = (1 - R)\sigma_X^2, \tag{C.2}$$

and

$$\sigma_E = \sigma_X \sqrt{1 - R}. \tag{C.3}$$

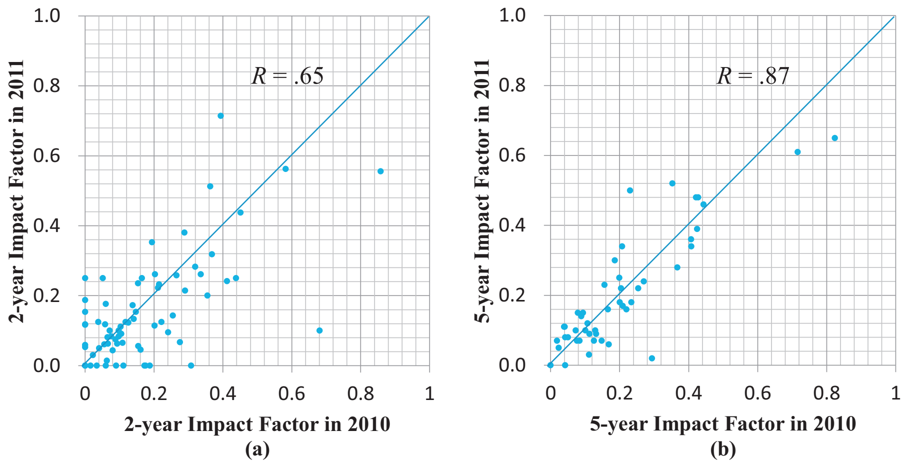


Fig. C.1. Correlations between the (a) 2-year and (b) 5-year JIFs estimated based on two consecutive years 2010–2011.

Table C.1
Average standard errors of 2011 2-year and 5-year JIFs across journals in DoRISE.

Estimation method	Average SE	
	2-year JIF	5-year JIF
Based on bootstrap method 1	0.14	0.10
Based on bootstrap method 2	0.12	0.08
Based on the correlation of two consecutive years	0.09	0.06

In Eq. (C.1), σ_T, σ_X, and σ_E denote the standard deviations of true score, observed score, and error, respectively. The σ_E can be seen as the average SE in measuring the JIFs for all journals while assuming an equal SEfor all target journals. If the assumption of equal SE is not sustained, then we can easily prove mathematically that

$$\sigma_E^2 = \frac{\sum_{M'}^{j=1} \sigma_j^2}{M'}, \tag{C.4}$$

where σ_j denotes the SE in estimating the JIF for the jth journal and M' is the number of target journals.

Fig. C.1 illustrates the correlations between the two sets of JIFs of 2010 and 2011. By accepting the assumptions of true score model, we can approximately estimate the average SE of JIF. According to Eq. (C.3), the SEs of measurement are about $0.57\sigma_X$ for 2-year JIF estimation and $0.35\sigma_X$ for 5-year JIF, where σ_X is the standard deviation of the observed JIFs across journals in DoRISE. According to the JIFs given in Tables B.1 and B.2 in Appendix B, both the standard deviations of the 2-year JIFs and 5-year JIFs are equal to 0.15. Therefore, the SEMs of the 2-year JIFs and 5-year JIFs are about 0.09 and 0.06, respectively.

According to Eq. (C.4), we calculated the average SEs of JIFs across journals in DoRISE based on the two bootstrap methods. Table C.1 shows the two average SEs and the correlation between the two sets of 2010 and 2011 JIFs.

The overall SE of JIFs based on the correlation between the two sets of JIFs from two consecutive years is underestimated. Two possible reasons are: (a) the variance due to the variety of journals included in the database is not take into consideration, and (b) about one-half of the citable articles are overlapped when two sets of 2-year JIFs are estimated in two consecutive years and about four-fifths of the citable articles are overlapped when two sets of 5-year JIFs are estimated in two consecutive years.

References

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2010). Citations versus journal impact factor as proxy of quality: Could the latter ever be preferable? *Scientometrics*, *84*(3), 821–833.

Amin, M., & Mabe, M. (2000, October). *Impact factor: Use and abuse*. Perspectives in Publishing.

Baccini, A., Barabesi, L., Marcheselli, M., & Pratelli, L. (2012). Statistical inference on the *h*-index with an application to top-scientist performance. *Journal of Informetrics*, *6*(4), 721–728.

Buela-Casal, G., Perakakis, P., Taylor, M., & Checa, P. (2006). Measuring internationality: Reflections and perspectives on academic journals. *Scientometrics*, *67*(1), 45–65.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley & Sons.

Donner, A., & Klar, N. (1994). Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *American Journal of Epidemiology*, *140*(3), 279–289.

Dorta-González, P., & Dorta-González, M. I. (2013). Comparing journals from different fields of science and social science through a JCR subject categories normalized impact factor. *Scientometrics*, *95*(2), 645–672.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, *76*(374), 312–319.

Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science*, *18*(2), 135–140.
Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.
Fassoulaki, A., Paraskeva, A., Papilas, K., & Karabinis, G. (2000). Self-citations in six anaesthesia journals and their significance in determining the impact factor. *British Journal of Anaesthesia*, *84*(2), 266–269.
Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, *122*, 108–111.
Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*(4), 359–375.
Garfield, E. (1999). Journal impact factor: A brief review. *Canadian Medical Association Journal*, *161*(8), 979–980.
Greenwood, D. C. (2007). Reliability of journal impact factor rankings. *BMC Medical Research Methodology*, *7*, 48. http://dx.doi.org/10.1186/1471-2288-7-48
Gross, S. T., & Lai, T. L. (1996). Bootstrap methods for truncated and censored data. *Statistica Sinica*, *6*, 509–530.
Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample survey methods and theory*. New York, NY: John Wiley & Sons.
Hecht, F., Hecht, B. K., & Sandberg, A. A. (1998). The journal "impact factor": A misnamed, misleading, misused measure. *Cancer Genetics and Cytogenetics*, *104*(2), 77–81.
Holden, G., Rosenberg, G., Barker, K., & Onghena, P. (2006). An assessment of the predictive validity of impact factor scores: Implications for academic employment decisions in social work. *Research on Social Work Practice*, *16*(6), 613–624.
Jen, T.-H., Tam, H.-P., & Wu, M. (2011). An estimation of the design effect for the two-stage stratified cluster sampling design. *Journal of Research in Education Sciences*, *56*(1), 33–65.
Kish, L. (1965). *Survey sampling*. London, England: John Wiley & Sons.
Leydesdorff, L. (2013). Does the specification of uncertainty hurt the progress of scientometrics? *Journal of Informetrics*, *7*(2), 292–293. http://dx.doi.org/10.1016/j.joi.2012.11.007
Leydesdorff, L., & Amsterdamska, O. (1990). Dimensions of citation analysis. *Science, Technology, & Human Values*, *15*(3), 305–335. http://dx.doi.org/10.1177/016224399001500303
Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). Boston: Brooks/Cole.
Lord, F. M. (1965). A strong true-score theory, with applications. *Psychometrika*, *30*(3), 239–270.
Marshakova-Shaikevich, I. (1996). The standard impact factor as an evaluation tool of science and scientific journals. *Scientometrics*, *35*(2), 283–290.
Marx, W., & Cardona, M. (2003). The impact of solid state communications in view of the ISI citation data. *Solid State Communications*, *127*(5), 323–336.
Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, *4*(3), 265–277.
Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics*, *4*(3), 423–430.
Pons, O. (2007). Bootstrap of means under stratified sampling. *Electronic Journal of Statistics*, *1*, 381–391. http://dx.doi.org/10.1214/07-Ejs033
Ramirez, A. M., Garcia, E. O., & Del Rio, J. A. (2000). Renormalized impact factor. *Scientometrics*, *47*(1), 3–9.
Schneider, J. W. (2013). Caveats for using statistical significance tests in research assessments. *Journal of Informetrics*, *7*(1), 50–62.
Schubert, A., & Glänzel, W. (1983). Statistical reliability of comparisons based on the citation impact of scientific publications. *Scientometrics*, *5*(1), 59–74.
Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628–638.
Smith, R. (2006). Commentary: The power of the unrelenting impact factor – Is it a force for good or harm? *International Journal of Epidemiology*, *35*, 1129–1130.
Sombatsompop, N., & Markpin, T. (2005). Making an equality of ISI impact factors for different subject fields. *Journal of the American Society for Information Science and Technology*, *56*(7), 676–683.
van Leeuwen, T. N., & Moed, H. F. (2002). Development and application of journal impact measures in the Dutch science system. *Scientometrics*, *53*(2), 249–266.
Waltman, L., van Eck, N. J., van Leeuwen, T. N., & Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics*, *7*(2), 272–285.
Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, *59*(11), 1856–1860.