Industrial Engineering
& Management Systems
Vol 15, No 1, March 2016, pp.99-109
ISSN 1598-7248 | EISSN 2234-6473 |
http://dx.doi.org/10.7232/iems.2016.15.1.099

A Novel Journal Evaluation Metric that Adjusts the Impact Factors across Different Subject Categories

Sujin Pyo, Woojin Lee, Jaewook Lee*
Department of Industrial Engineering, Seoul National University, Seoul, Korea

(Received: February 26, 2016 / Revised: March 8, 2016 / Accepted: March 9, 2016)

ABSTRACT

During the last two decades, impact factor has been widely used as a journal evaluation metric that differentiates the influence of a specific journal compared with other journals. However, impact factor does not provide a reliable metric between journals in different subject categories. For example, higher impact factors are given to biology and general sciences than those assigned to other traditional engineering and social sciences. This study initially analyzes the trend of the time series of the impact factors of the journals listed in Journal Citation Reports during the last decade. This study then proposes new journal evaluation metrics that adjust the impact factors across different subject categories. The proposed metrics possibly provides a consistent measure to mitigate the differences in impact factors among subject categories. On the basis of experimental results, we recommend the most reliable and appropriate metric to evaluate journals that are less dependent on the characteristics of subject categories.

Keywords: Adjusted Impact Factor, Field Normalization, Journal Evaluation

* Corresponding Author, E-mail: jaewook@snu.ac.kr

1. INTRODUCTION

Impact factor has been widely used as a journal evaluation metric to indicate the value of the effectiveness of a specific journal compared with other journals. Impact factor, a journal effectiveness index provided by the database of Journal Citation Reports (JCR) from Thomson Reuters, can be defined as the average number of citations of each journal in recent two years to the articles published in that journal. One distinctive feature of the impact factor is that this metric provides a simple numerical value as descriptive statistics, such as GPA in schools, batting hit rate in a baseball game, and value-at-risk in finance, to show the comparative importance of a specific journal compared with other journals. The factor has also been used as a quantitative measure to evaluate the performances of researchers with respect to their published articles. In many countries, academic performances of researchers have been largely influenced by the impact factors of the journals in which their articles are published.

Since impact factor is widely used indicator to evaluate journal and researchers' academic performance, many researchers have focused on this resource. Bornmann criticized about 2-year interval and definition of citable items (Bornmann and Daniel, 2009). Simons (2008) insisted that impact factor could be misused, because it could be manipulated by increasing review articles. Vanclay also discussed about inaccuracies and errors in the Thomson Reuters impact factor (Vanclay, 2011). However Brody agreed that impact factor is imperfect but not yet replaceable, because substantial alternatives has no clear improvements over IF as a single measure (Brody, 2013). In response to critics, Thomson Reuters supple-

mented more indicators: Five year impact factor, Eigenfactor score, Article Influence Score in the online version of JCR (Dorta-Gonzalez and Dorta-Gonzalez, 2012). Also there has been approach to apply rank normalization to solve Vanclay's dilemma with impact factor (Pudovkin and Garfield, 2012).

However, the impact factor is characterized by several drawbacks that limit its reliability as an index to evaluate the superiority of journals. For instance, the impact factor can be manipulated by increasing self-citation and cross-citations among collaborated journals. Nevertheless, this problem has been addressed by a new policy of Thomson Reuters to exclude such journals in the database of JCR. Another problem is the comparison between journals with different subject categories. Althouse (2008) researched that category variation in the fraction of citations to literature indexed by Thomson Scientific's Journal Citation Reports contributes to differences among subject categories. On the one hand, impact factor can be reasonably used to compare journals with the same subject category. On the other hand, impact factor cannot be reasonably or desirably used to compare journals with different subject categories because deviations usually exist between different research areas as a consequence of various natures of their academic environments. For example, articles in the field of "Biology" are generally higher than those in "Mathematics," "Social Sciences," and "Computer Science" in terms of the impact factor because mathematicians normally require many years to publish a paper after submission, social science researchers mostly prefer to publish books rather than journals, and computer scientists prefer to present their results in conference proceedings (Chen and Konstan, 2010). Field normalization is necessary to evaluate cross field evaluation because impact factor is a field-dependent indicator (Garfield, 1979). The normalization of the effects of impact factors has been investigated to consider the characteristics of research areas. For example, the relativity of citation performance has been evaluated to normalize academic fields (Leydesdorff, 2012). The proportion of the most highly cited papers has been proposed as an alternative to impact factor (Zhang *et al*., 2014). Two field normalization methods, namely, mean-based method and z-score method, have been compared (Zitt *et al*., 2005).

This study aims to introduce new robust journal evaluation metrics that normalize the differences in impact factor among various categories. We provide 11 metrics as representative values of subject categories, normalize impact factors between different categories, and adjust the impact factor variance of categories. Conducting experiments, we suggest the most appropriate journal evaluation metric by using a criterion that minimizes the ordering error count between the adjusted impact factors and the impact factors. Our proposed metric and its corresponding adjusted impact factor can be used to evaluate journals in different subject categories.

The rest of this paper is organized as follows. In Section 2, the trends of the journals and their impact factors from 2005 to 2014 are analyzed. In Section 3, a new journal evaluation metric is introduced. In Sections 4 and 5, analysis measure comparison and experimental results are presented. In Section 6, the conclusion is provided.

2. TRENDS OF JOURNALS

To compare and analyze the categorical impact factor, we used the JCR data, which are available from http://www.webofknowledge.com. We gathered 74,274 journal impact factor data registered in Science Citation Index (SCI) from 2005 to 2014.

In Figures 1 and 2, the overall trends in all of the SCI journals in recent 10 years are depicted. The total number of journals and articles has increased constantly in the observation period. A distribution of the impact factors of journals is shown in a box plot (Figure 3). The upper whisker indicates the maximum value of the impact factor but disregards outliers. Both ends of the box represent quartile1 and quartile3 of the plot, and the bar inside the box indicates the median value. The most remarkable increase in impact factor was observed in 2007.

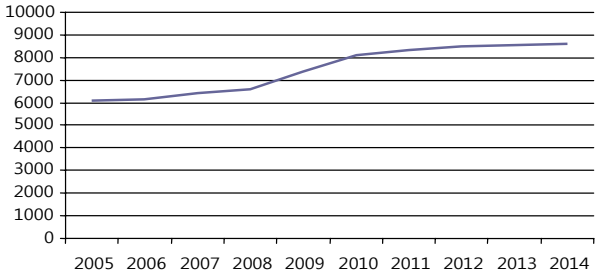

Figure 1. Number of journals.

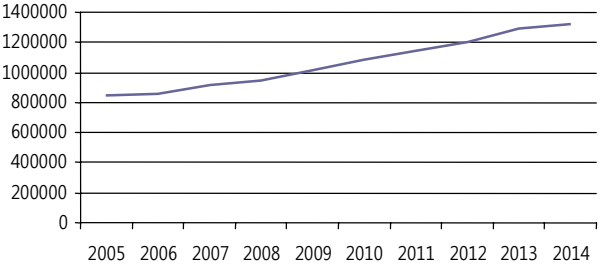

Figure 2. Number of articles.

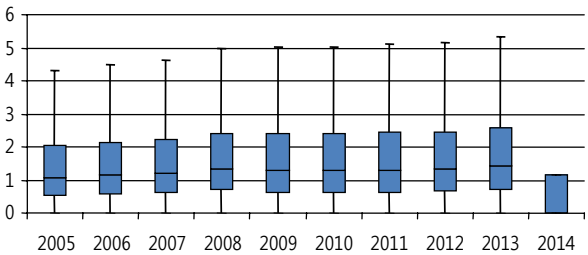

Figure 3. Box-plot of impact factor.

According to JCR classifications, all of the journals listed in According to JCR classifications, all of the journals listed in JCR were grouped into 170 subject categories representing specific academic fields in 2005, and this number increased to 176 subject categories in 2011; each journal can be further included in several subject categories. Aggregate impact factor (AIF) is provided in JCR to represent the average impact factor of the journals in a specific subject category. AIF is a weighted mean of impact factors by articles. We calculated the difference of AIFs between 2005 and 2014 and showed the 10 top and least performing subject categories in Table 1. The categories "NANO SCIENCE AND NANO TECHNOLOGY" and "ENERGY AND FUELS" showed an increase in AIF of more than 3 corresponding to more than 200% increases compared with the average impact factors of all the journals; by contrast, "MULTIDISCIPLINARY SCIENCES" and "DEVELOPMENTAL BIOLOGY" showed a decrease in AIF of more than -1 corresponding to relatively more than 50% decreases. AIF is a measure of backward citation; thus, we could interpret that the related works of these categories are increasing recently and are popular.

We also calculated the average of the impact factor of top 20% journals (Top 20% Avg IF) in each category to identify the variance of the impact factors of the representative journals in subject categories not disturbed by the low impact factor-rated journals. The result is shown in Table 2. Six of the top performing categories were the same as those of the AIFs, whereas four were identical to those in the least performing group of categories.

Table 1. Top and least performing categories (Aggregate Impact Factor)

Top performing 10 categories		Least performing 10 categories	
Difference	Categories	Difference	Categories
3.07	NANOSCIENCE AND NANOTECHNOLOGY	-0.20	MEDICINE, RESEARCH AND EXPERIMENTAL
3.00	ENERGY AND FUELS	-0.20	MATHEMATICS, INTERDISCIPLINARY APPLICATIONS
2.61	CHEMISTRY, MULTIDISCIPLINARY	-0.20	GENETICS AND HEREDITY
2.11	CHEMISTRY, PHYSICAL	-0.20	BIOCHEMISTRY & MOLECULAR BIOLOGY
2.06	AGRICULTURAL ENGINEERING	-0.30	STATISTICS AND PROBABILITY
2.06	PHYSICS, CONDENSED MATTER	-0.30	RHEUMATOLOGY
2.03	MATERIALS SCIENCE, MULTIDISCIPLINARY	-0.50	PERIPHERAL VASCULAR DISEASE
1.90	NEUROIMAGING	-0.80	HEMATOLOGY
1.73	ENGINEERING, ENVIRONMENTAL	-1.10	DEVELOPMENTAL BIOLOGY
1.71	MATERIALS SCIENCE, BIOMATERIALS	-3.20	MULTIDISCIPLINARY SCIENCES

Table 2. Top and least performing categories (Top 20% Average Impact Factor)

Top performing 10 categories		Least performing 10 categories	
Difference	Categories	Difference	Categories
7.45	PHYSICS, CONDENSED MATTER	-0.24	COMPUTER SCIENCE, THEORY AND METHODS
6.42	NANOSCIENCE AND NANOTECHNOLOGY	-0.24	PHYSIOLOGY
6.13	CHEMISTRY, MULTIDISCIPLINARY	-0.310	GENETICS AND HEREDITY
5.35	ENERGY AND FUELS	-0.34	GERIATRICS AND GERONTOLOGY
4.71	CHEMISTRY, PHYSICAL	-0.44	PERIPHERAL VASCULAR DISEASE
4.54	PHYSICS, APPLIED	-0.47	BIOCHEMISTRY AND MOLECULAR BIOLOGY
4.48	MYCOLOGY	-0.48	IMMUNOLOGY
4.24	MATERIALS SCIENCE, MULTIDISCIPLINARY	-0.60	CHEMISTRY, MEDICINAL
3.92	ANATOMY AND MORPHOLOGY	-0.63	CHEMISTRY, INORGANIC AND NUCLEAR
3.74	PHYSICS, PARTICLES AND FIELDS	-4.48	DEVELOPMENTAL BIOLOGY

3. PROPOSED JOURNAL EVALUATION METRICS

An impact factor of a specific journal indicates the average rate by which the articles are cited in the journal. The higher the impact factor of the journal is, the more articles are cited in the journal. Accordingly, an impact factor is used as an evaluation metric of a specific journal in JCR. The impact factor also provides an indicator of the most cited academic fields in researchers. The significant increase in the impact factors of journals in a specific field shows that many researchers have been interested in the research area and have written papers related to the study field. Although the impact factor provides several benefits, the use of impact factors as an evaluation metric also exhibits several defects. First, the impact factors of most journals likely increase with time. Academic fields have been varied and sizes have been larger than before. The number of articles has increased in recent years, and new journals have also emerged largely, and these journals correspond to the significant increase in the articles. In addition to these situations, the increased self-citations in some journals to increase impact factors cause the impact factor to be inevitably overstated and unreliable. Second, an imbalance exists in each category. The distribution of impact factors in each category is totally different. Several categories show the balanced distribution, whereas most categories tend to be skewed because of a few journals; furthermore, the mean slightly differs from the median. This situation distorts the meaningfulness of statistical analysis because of the outliers in each category. Finally, the most important concern in the impact factor is that journals across subject categories are necessary to be compared carefully. JCR has classified all journals into hundreds of subject categories, and each journal may be included in several categories based on its characteristics. For example, Aci Structural Journal is included in three categories: Construction and Building Technology; Engineering, Civil, and Materials Science; and Multidisciplinary. The impact factor can be a meaningful metric to compare journals in the same subject category. However, impact factor cannot be a reasonable metric to compare a journal in a subject category with other journals in a different subject category because the characteristics of each academic field, such as the number of researchers and articles and the publication frequencies of articles, are totally different across academic fields. If the impact factor of a journal is lower than the average of the impact factors of journals in its subject category, then the journal may not be superior to other journals included in other academic fields whose impact factors are relatively lower than that of the journal. One important criterion will be the relative status of the journal included in the subject category when the journal is compared with other journals in different categories. Therefore, a new evaluation metric is necessary to supplement the disadvantages of raw impact factors across different subject categories. If a new evaluation can show the characteristics of each subject category, then a journal can be compared with other journals across different subject categories with reliability and reasonability. In the following subsections, three new metrics are proposed to adjust the impact factors across the subject categories.

3.1 Using AIF

In addition to the impact factors of all journals, JCR has provided users with various statistical data of each subject category. Among the data, an AIF, similar to the impact factor of a journal, indicates the average number of articles cited in a subject category. The AIF is calculated in the same way as the impact factor, that is, the number of citations is divided by the number of articles in a subject category. A category whose AIF is 2 means that the articles in the subject category published one and two years ago have been cited two times on average. The AIF contains the characteristics of the subject category; thus, the metric can be used to normalize the categories to compare journals across subject categories.

Our new evaluation metric, namely, adjusted IF (A-IF), can reflect the information of each category and is defined by an average of impact factors divided by the AIFs of the included subject categories. Specifically, let $j^k \in J$ be the *k*th journal where $k \in \{1, 2, \cdots, n\}$ in the alphabetical order and IF^k be the impact factor corresponding to journal j^k. Let $Ca \in C$ be the ath subject category where $a \in \{1, 2, \cdots, m\}$ in the alphabetical order and AIF_{C_a} is an aggregate impact factor corresponding the subject category. $j^k_{c_a,c_b}$ represents, for example, the *k*th journal included in subject categories c_a and c_b. C_k can be defined as a set of subject categories including the *k*th journal. A-IF of a journal $j^k_{c_a,c_b,c_c}$ is the average value of the impact factor divided by each aggregate impact factor for a subject category included.

$$\text{A-IF}(j^k_{c_a,c_b,\cdots,c_l \in C_k}) = \text{Average}\left(\frac{IF^k}{AIF_{c_a}}, \frac{IF^k}{AIF_{c_b}}, \cdots, \frac{IF^k}{AIF_{c_l}}\right) = \frac{1}{\text{card}(C_k)}\left(\sum_{n=c_a,c_b,\cdots,c_l \in C_k} \frac{IF_k}{AIF_n}\right)$$

where card (C_k) is the cardinality of C_k. A-IF is also the average value of the normalized impact factors considering the characteristics of the included subject categories. Impact factor can be normalized to reflect the subject category; as a result, the normalized impact factor can be used to compare journals across subject categories.

3.2 Using a Quantile for the Journals Listed in a Subject Category

The characteristics of subject categories can be de-

scribed in terms of the quantile of the impact factors of the journals listed in a subject category instead of using an AIF. Only a few journals with higher impact factors tend to have a significant contribution to represent an impact factor of a subject category compared with other journals in the same category. Thus, a quantile in the decreasing order of impact factors in a subject category can be used to prevent journals with low influences on the category from undermining the normalization effect. The first new evaluation metric, namely, QAVG-IF, is a metric that corresponds to a quantile for each subject category included. Specifically, let $\mathrm{Quan}_q(c_a)$ be a top q% quantile in the order of impact factors of the journals in category c_a and $\mathrm{AVG}\left[\mathrm{Quan}_q(C_a)\right]$ be the average impact factor for the journals included in top q% quantile in category c_a. QAVG-IF is similar to A-IF but uses the average impact factor of top q% quantile for journals instead of an AIF.

$$\begin{aligned}&\mathrm{QAVG\text{-}IF}_q\left(j^k_{c_a,c_a,\cdots c_l\in C_k}\right)\\&=\mathrm{Average}\left(\frac{\mathrm{IF}^k}{\mathrm{AVG}\left[\mathrm{Quan}_q(c_a)\right]},\frac{\mathrm{IF}^k}{\mathrm{AVG}\left[\mathrm{Quan}_q(c_b)\right]},\cdots,\frac{\mathrm{IF}^k}{\mathrm{AVG}\left[\mathrm{Quan}_q(c_l)\right]}\right)\\&=\frac{1}{\mathrm{card}(C_k)}\left(\sum_{n=c_a,c_b,\cdots c_l\in C_k,}\frac{\mathrm{IF}_k}{\mathrm{AVG}\left[\mathrm{Quan}_q(c_n)\right]}\right)\end{aligned}$$

The second evaluation metric, namely, QMAX-IF, is a maximum value instead of using the average value, which has also been used by National Research Foundation (NRF) of Korea to evaluate the academic performances of researchers across various academic fields. Especially, NRF has used QMAX-IF using top 20% quantile.

$$\begin{aligned}&\mathrm{QMAX\text{-}IF}_q\left(j^k_{c_a,c_a,\cdots c_l\in C_k}\right)\\&=\mathrm{MAX}\left(\frac{\mathrm{IF}^k}{\mathrm{AVG}\left[\mathrm{Quan}_q(c_a)\right]},\frac{\mathrm{IF}^k}{\mathrm{AVG}\left[\mathrm{Quan}_q(c_b)\right]},\cdots,\frac{\mathrm{IF}^k}{\mathrm{AVG}\left[\mathrm{Quan}_q(c_l)\right]}\right)\end{aligned}$$

QAVG-IF for a journal considers all subject categories included by averaging the impact factor divided by the characteristics of each category, whereas QMAX-IF only considers the maximum value. Thus, QMAX-IF is equal to the impact factor divided by the smallest quantile average value for a subject category. Given that only a few journals have significant impact factors in most categories, QMAX-IF can be high if the distribution of the journals in the category is balanced but not deeply skewed.

Table 3. The number of subject categories and Journals in 2013 JCR

	# of subject categories	# of journals
SCI (2013)	176	8539
SSCI (2013)	56	3080

4. COMPARISON MEASURES AND RESULTS

In our experiment, recent impact factors and statistics for subject categories provided by JCR in 2013 are used to compare the performances of the proposed evaluation metrics. We used JCR database, which contains all the information about the journals, their impact factors, and their corresponding subject categories. Journals in JCR consist of those listed in SCI and Social SCI (SSCI). In Table 3, the journals have been classified into subject categories according to the academic contributions annually. A journal can be included in several subject categories. In 2013, 8539 SCI journals and 3080 SSCI journals are listed in JCR.

Some journals are in SCI and SSCI journal lists and can also be classified into SCI and SSCI subject categories. For example, a journal, Acta Bioethica, is in SCI category (Medical Ethics) and two SSCI categories (Ethics and Social Sciences, Biomedical). The journals usually have different statuses in each category. Although the impact factor of a journal is higher than the average impact factor in one subject category it belongs to, the impact factor can be lower than the average impact factor in another subject category it belongs to. This situation is the reason why the journal evaluation metric should display the varying characteristics across the different subject categories.

4.1 Experimental Results

This work aims to reflect the varying characteristics across subject categories when a journal is compared with another journal listed in a different subject category, because comparing the impacts of journals in various academic fields directly from their impact factors only without adjustment is unreasonable.

In Table 4, the mean values of journals in each subject category are used to calculate the mean and standard deviation. Contrary to the impact factor and A-IF suggested, how the quantile is fixed plays an important role in determining QAVG-IF and QMAX-IF. The quantiles are set up by 5 cases; 20%, 30%, 50%, 75%, and 100%. As the quantile ratio increases, journals for a subject category belong to the quantile in a descending order. All of the suggested metrics yield lower mean values than the impact factor does. The standard deviation and difference between the maximum and minimum

Table 4. Statistics of all evaluation metrics

	Mean	Standard Deviation	Skewness	Max - Min
IF	1.816	0.8914	0.8148	4.620
A-IF	0.818	0.1191	-0.1203	0.667
$QAVG_{20}-IF$	0.412	0.0546	-0.7413	0.354
$QAVG_{30}-IF$	0.484	0.0593	-0.8529	0.396
$QAVG_{50}-IF$	0.610	0.0698	-0.9067	0.482
$QAVG_{75}-IF$	0.762	0.0855	-0.8578	0.619
$QAVG_{100}-IF$	0.950	0.1051	-0.7781	0.789
$QAVG_{20}-IF$	0.505	0.0622	0.3752	0.410
$QAVG_{30}-IF$	0.584	0.0685	0.7001	0.423
$QAVG_{50}-IF$	0.729	0.0823	1.2030	0.459
$QAVG_{75}-IF$	0.904	0.1025	1.4670	0.580
$QAVG_{100}-IF$	1.127	0.1259	1.6642	0.660

values of 11 metrics are reduced significantly compared with that of the impact factor. This finding means that the difference between subject categories is reduced. This finding also shows the effect of normalization among the categories. The table also reveals that the mean and standard deviation of QAVG-IF likely increase as the proposition (%) increases. By definition of the metric, QAVG-IF is divided by the larger $AVG\left[Quan_q(C)\right]$ as the quantile ratio q increases. The values unsurprisingly increase as the quantile ratio increases because $E(k\times x) = k\times E(x)$ and $Var(k\times x) = k^2\times Var(x)$, where k is a constant. In QMAX-IFs, the statistics tends to be similar to the results of QAVG-IFs for the same reasons. The mean value of QMAX-IF is always larger than that of QAVG-IF in the same quantile. This result is not surprising because QMAX-IF is the maximum value whereas QAVG-IF is the average value in the same operands. However, the difference between the maximum and minimum values is reversed at the 50% quantile.

Figure 4 illustrates the distributions of the subject categories in terms of impact factor and A-IF. The impact factors do not exhibit any convergence between subject categories, whereas A-IFs likely converge to 0.9-1 by normalization effect. In Figures 5 and 6, QAVG-IFs and QMAX-IFs show similar results as for A-IF. Therefore, the 11 metrics satisfy the necessary essential condition to compare journals across subject categories. In the figure 5 and 6, the consistency in subject categories tend to be weak as the quantile ratio increases.

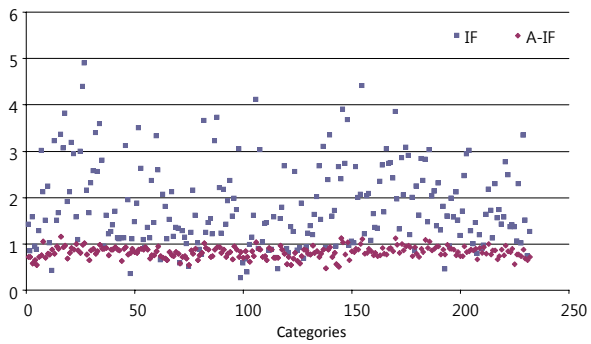

Figure 4. Distributions of impact factors and A-IFs.

4.2 Ordering Count Measure

4.2.1 Unweighted Ordering Error Count

In addition to convergence and reduction of variance, another important condition that should be satisfied by an A-IF is that the orders between journals based on the impact factor within each subject category should be preserved as much as possible. Using the impact factor as a standard, a journal whose impact factor is higher than that of the other journal in the same subject category is desired to also be higher in their A-IFs. Without the order consistency in the same category between impact factors and their A-IFs, the metrics used for A-IFs cannot be a good standard to evaluate journals. To implement this comparison measure for order consistency, we first define the heavisidedistance $H_{Adj\text{-}IF}(j^k, j^l)$ between a pair of journals j^k and j^l in the same category C_j to be 1 if the order between the paired two journals with the impact factor is different from the order between the same journals with the adjusted IFs using a compared metric and to be 0 otherwise. Then we can define an unweighted ordering error count (UOEC) for an adjusted IF, say Adj-IF, as follows.

$$UOEC(Adj\text{-}IF) = \sum_{c_j} \sum_{j_l, j_k \in C_j} H_{Adj\text{-}IF}(j^k, j^l)$$

In the JCR 2013 data, the total possible number of ordering error count is 1,074,080 because it can occur approximately n(n-1)/2 in each subject category with n journals. In Figure 7, red bars indicate QAVG-IFs and green bars refer to QMAX-IFs. The numerical values are shown in Appendix-A. The figure shows that the number of ordering error countings of $QMAX_{20}-IF$ is the largest among the compared 11 metrics. The error counting ratio of $QMAX_{20}-IF$ is nearly 8.17%, which is the largest in all of the metrics. Contrary to $QMAX_{20}-IF$, $QMAX_{100}-IF$ shows the best order consistency with the impact factor with an error ratio of 5.88%. The number of error countings with AVG-IF is always smaller than that with MAX-IF in the same quantile. MAX-IFs are inevitably affected by the significant journals more than AVG-IFs because many subject categories include only a few significant journals compared with other journals. In the same sense, the error counting of both QAVG-IF and QMAX-IF tends to decrease as the quantile increases because the two metrics can contain more information about other journals and are not biased by only a few significant journals. A-IF also shows a remarkable performance with an error ratio of 6.24% compared with QMAX-IFs.

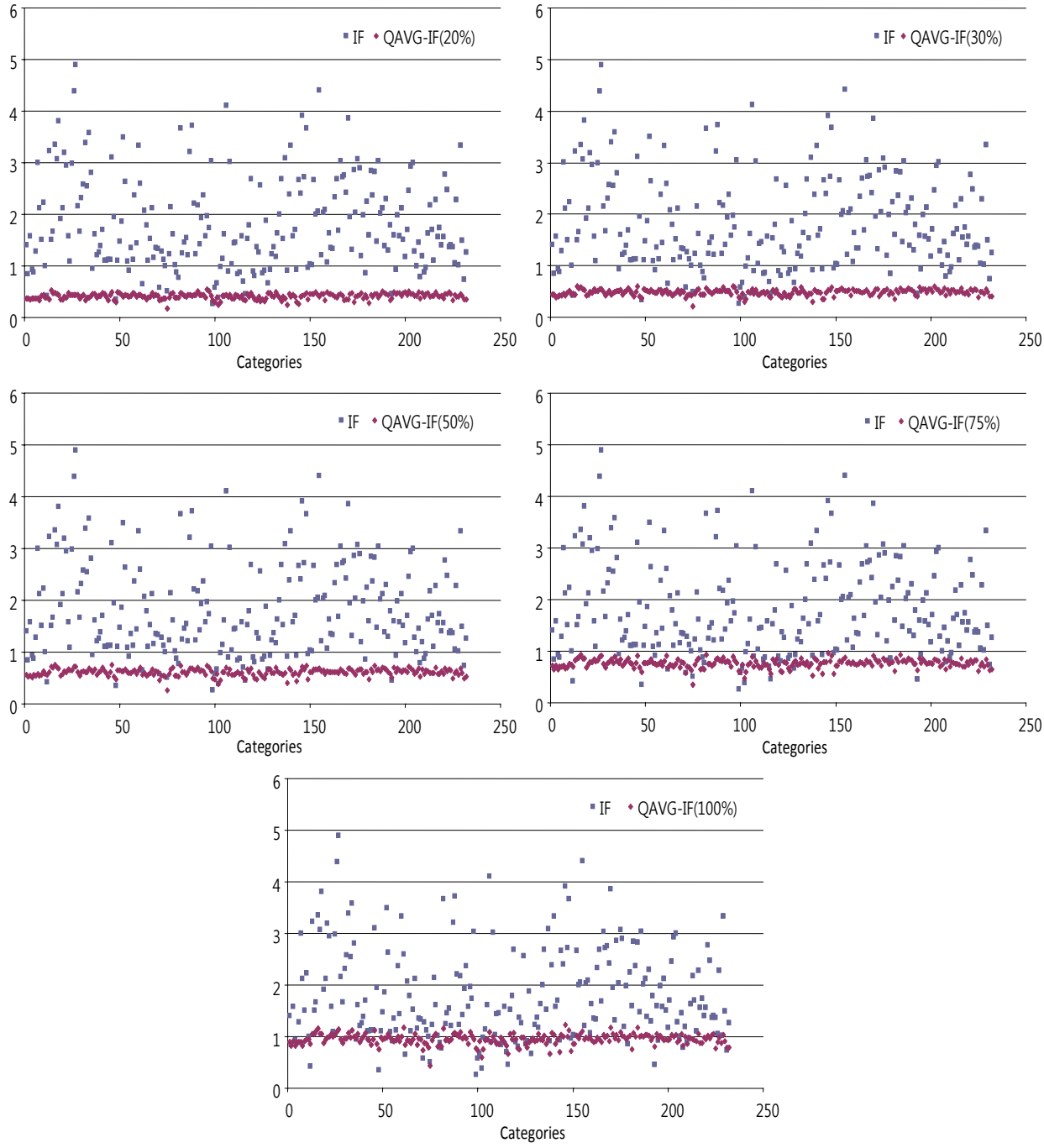

Figure 5. Distributions of impact factor and QVAG-IFs.

4.2.2 Article-Weighted Ordering Error Count

The effect of the number of articles can also be considered in a journal when the ordering error is counted because the number of articles across journals varies. If the order of a journal is reversed by the other journal in terms of A-IFs, then all of the articles in both journals are also influenced and order reversed. The more articles a journal has, the stronger the effect of ordering error counting to the journal is. This point can be incorporated by define an article-weighted ordering error count (WOEC) for an adjusted IF, say Adj-IF, as follows

$$\mathrm{WOEC(Adj\text{-}IF)} = \sum_{C_j} \sum_{j_l, j_k \in C_j} \left(w\left(j^k\right) + w\left(j^l\right) \right) (j^k, j^l) \times H_{\mathrm{Adj\text{-}IF}}(j^k, j^l)$$

where $w(j^k)$ is the number of articles in journal j^k.
In the JCR 2013 data, the total possible number of arti-

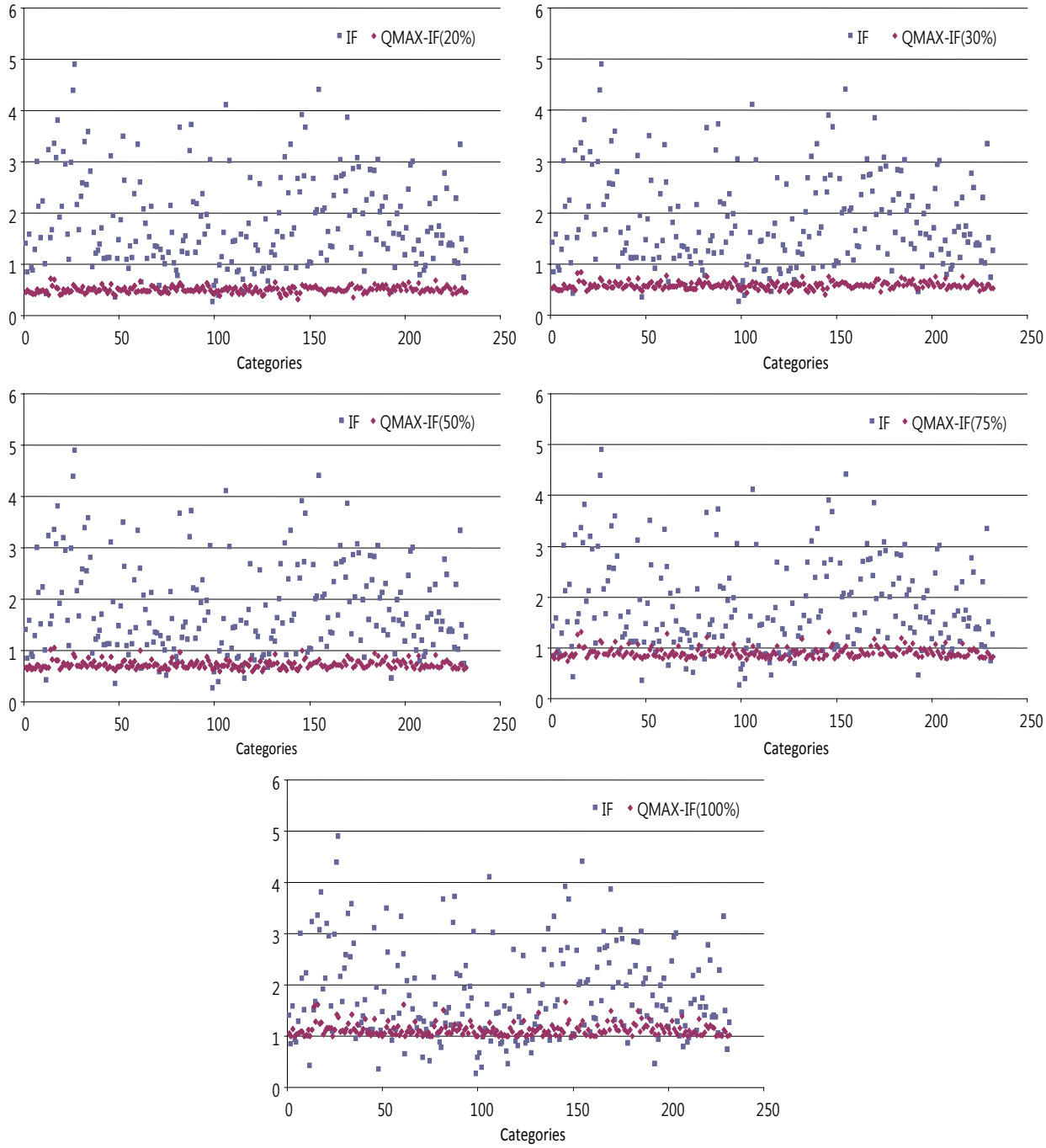

Figure 6. Distributions of impact factor and QMAX-IFs.

cles is 277,386,462. If the order is reversed between two journals, then the number of articles in both journals is counted as a counting error. Based on the n(n-1)/2 cases in each subject category, the articles of both journals are reflected.

In Figure 8, the results are similar to the unweighted case. The worst and the best metrics are not changed. However, the error counting ratios increase slightly in the overall metrics. This increase means that more ordering errors exist between journals with articles more than the average. In general, a few significant journals in a subject category in an upper quantile tend to have many articles than the journals with lower quantiles. By definition of QAVG-IF and QMAX-IF, the significant journals give the critical effect to WOECs for both metrics, especially for QMAX-IF. As shown in Figure 9, the WOEC ratio of QMAX-IFs increases significantly with a maximum of 16.81% for $\mathrm{QMAX}_{20} - \mathrm{IF}$.

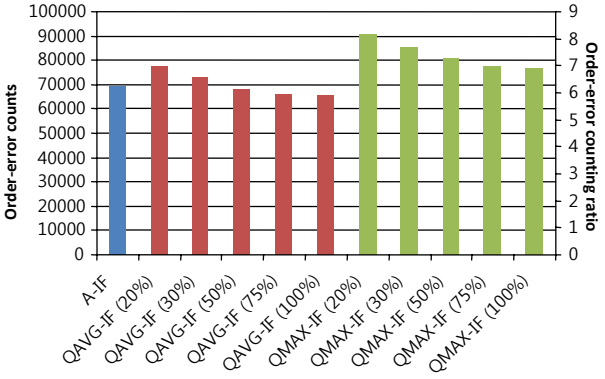

Figure 7. Unweighted ordering error counts and counting ratio of 11 metrics.

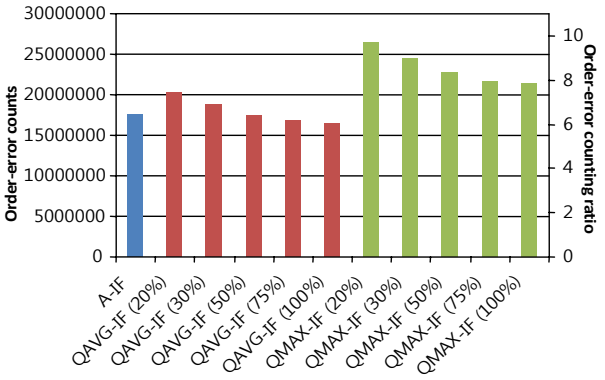

Figure 8. Article-weighted ordering error counting and counting ratio of 11 metrics.

In summary, QMAX-IFs do not seem to be proper evaluation metrics. This metric exhibit an inconsistency problem compared with other metrics. Among 11 metrics, $QAVG_{100}-IF$ shows the best performance in ordering error counting and in increasing the effect of normalization in its descriptive statistics. Journals can be compared with other journals more reliably across subject categories by using $QAVG_{100}-IF$ as an evaluation metric; this metric provides the minimum ordering error counting in the same category.

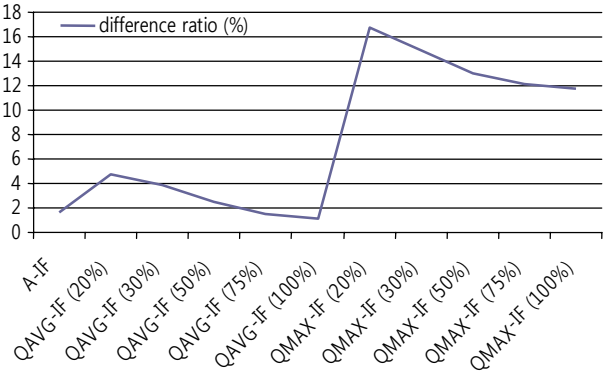

Figure 9. Difference ratio (%) between without and with articles.

5. CONCLUSION

An impact factor has been widely used as a journal evaluation metric for the last decades. The higher the impact factor of a journal is, the more articles in the journal are cited. The impact factor provides one simple numerical value for journal evaluation criteria, but this metric does not reflect the characteristics of various academic fields. Although comparing a journal with other journals in the same field in terms of impact factors may be reasonable, comparing a journal in one category with a journal in another category without any adjustment is neither reasonable nor desirable. The latter issue triggered many problems and controversies in the quantitative evaluation of the academic performances of researchers.

Several A-IF that use 11 metrics are suggested to resolve these problems related to the comparison of journals across academic fields. An adjustment can be conducted by dividing the impact factor into various adjusting values, which can affect the subject categories. An AIF and the average value of journals within quantile ratio are used as the adjusting values.

The normalization effect and ordering error counting are used as criteria to evaluate the performance. The 11 metrics show more normalization effects than the impact factor. As the quantile ratio decreases, the normalization effect tends to be stronger in both QAVG-IFs and QMAX-IFs. QAVG-IFs show better performance in or-dering the error counting than QMAX-IFs in the same quantile does. $QAVG_{100}-IF$ exhibits the best performance, whereas $QMAX_{20}-IF$ shows the worst performance among all metrics. Therefore, $QAVG_{100}-I$ can be used as a reasonable journal evaluation metric to compare journals across subject categories instead of the impact factor that does not reflect the characteristics of a subject category. The proposed metric is quite general and can be readily applied to any base index other than an impact factor. This aspect should be further investigated.

ACKNOWLEDGEMENTS

Part of the results of this paper have been presented at the 16th Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS 2015).

This work was supported by the National Research Foundation of Korea (NRF) grant funded the Korean government (MEST) (No. 2011-0017657).

REFERENCES

Althouse, B. M., West, J. D., Bergstrom, C. T., and Bergstrom, T. (2009), Differences in impact factor across fields and over time, *Journal of the American Society for Information Science and Technology*, **60**(1), 27-34.

Bornmann, L. and Daniel, H. D. (2008), What do citation counts measure? A review of studies on citing behavior, *Journal of Documentation*, **64(1)**, 45-80.

Brody, S. (2013), Impact factor: Imperfect but not yet replaceable, *Scientometrics*, **96**, 255-257.

Chen, J. and Konstan, J. A. (2010), Conference paper selectivity and impact, *Communications of the ACM*, **53**(6), 79-83.

Dorta-Gonzalez, P. and Dorta-Gonzalez, M. I. (2012), Comparing journals from different fields of science and social science through a JCR subject categories normalized impact factor, *Scientometrics*, **95**(2), 645-672.

Garfield, E. (1979), *Citation indexing-Its theory and application in Science*, Technology and Humanities, New York: Wiley and Sons.

Leydesdorff, L. (2012), Alternatives of the journal impact factor I3 and the top-10% of the most highly cited papers, *Scientometrics*, **92**, 355-365.

Pudovkin, A. I. and Garfield, E. (2012), Rank normalization of impact factors will resolve Vanclay's dilemma with TRIF, *Scientometrics*, **92**, 409-412.

Simons, K. (2008), The misused impact factor, *Science*, **322**(5899), 165-165.

Vanclay, J. K. (2011), Impact Factor: outdated artefact or stepping-stone to journal certification?, *Scientometrics*, **92**, 211-238.

Zhang, Z., Cheng, Y., and Liu, N. C. (2014), Comparison of the effect of mean-based method and z-score for field normalization of citations at the level of Web of Science subject categories, *Scientometrics*, **101**, 1679-1693.

Zitt, M., Ramanana, S., and Bassecoulard, E. (2005), Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation, *Scientometrics*, **63**(2), 373-401.

Appendix A. Error counts of 11 metrics

Metrics	Unweighted		Article-weighted		Difference ratio (%)
	Error Countings	Error Counting Ratio (%)	Error Countings	Error Counting Ratio (%)	
A-IF	67,020	6.24	17,581,628	6.34	1.58
QAVG-IF (20%)	74,933	6.98	20,259,227	7.30	4.69
QAVG-IF (30%)	70,430	6.56	18,882,475	6.81	3.81
QAVG-IF (50%)	65,940	6.14	17,445,735	6.29	2.45
QAVG-IF (75%)	63,776	5.94	16,709,844	6.02	1.45
QAVG-IF (100%)	63,152	5.88	16,485,515	5.94	1.08
QMAX-IF (20%)	87,771	8.17	26,476,846	9.55	16.81
QMAX-IF (30%)	82,617	7.69	24,522,175	8.84	14.93
QMAX-IF (50%)	77,720	7.24	22,676,832	8.18	12.98
QMAX-IF (75%)	74,778	6.96	21,662,229	7.81	12.17
QMAX-IF (100%)	74,292	6.92	21,446,822	7.73	11.78