



# Author-weighted impact factor and reference return ratio: can we attain more equality among fields?

Tolga Yuret<sup>1</sup> 

Received: 21 March 2018 / Published online: 14 June 2018  
© Akadémiai Kiadó, Budapest, Hungary 2018

## Abstracts

Despite its problems, journal impact factor (JIF) is the most popular journal quality metric. In this paper, two simple adjustments of JIF are tested to see whether more equality among fields can be attained. In author-weighted impact factor (AWIF), the number of citations that a journal receive is divided by the number of authors in that journal. In reference return ratio (RRR), the number of citations that a journal receive is divided by the number of references in that journal. We compute JIF, AWIF and RRR of all 10,848 journals included in journal citation report 2012. Science journals outperform social science journals at JIF but social science journals outperform science journals at both AWIF and RRR. Highest level of equality between science and social science journals is attained when AWIF is used. These findings cannot be generalized when narrower subject categories are considered.

**Keywords** Field-neutrality · Impact factor · Cited-references · Number of authors

## Introduction

Journal impact factor (JIF) is the most popular journal quality metric. Despite its popularity, the measure is criticized to have many deficiencies. The problems of JIF include inconsistency between its numerator and denominator, its arbitrary choice of time window and its treatment of self citations (Archambault and Lariviere 2009; Waltman 2016).

One of the major problems of JIF is its lack of field-neutrality. There are vast inequalities among fields when JIF is used. These inequalities matter when research evaluation is done for distinct academic fields. For example, jury members for a research reward of a university should evaluate publication performance of researchers from many different fields. Therefore, they need a journal metric that is field-neutral.

In this paper, we test whether two other metrics that are very similar to JIF in terms of their formulas can achieve more equality among fields. We think that similarity is

---

✉ Tolga Yuret  
tyuret@gmail.com

<sup>1</sup> Department of Economics, Faculty of Management, Istanbul Technical University, 34367 Macka, Istanbul, Turkey

important because the administrators who are already familiar with JIF can easily understand similar metrics. Consequently, there is more chance for similar metrics to be used in practice. Moreover, JIF is a popular metric because of its desirable properties such as its simplicity. By using similar metrics, we retain some of JIF's desirable properties.

JIF is the number of citations received by the journal divided by the number of articles in the journal. Instead of using the number of articles in the journal as the denominator, the author-weighted impact factor (AWIF) uses the number of authors and reference return ratio (RRR) uses the number of cited-references in the articles of the journal. The numerators of both of these metrics are the same as that of JIF.

The idea of normalizing the number of citations by the number of authors is not new. For example, Harzing et al. (2014) show that H-index is more equalitarian among academic fields when per-author number of citations is used in its computation. However, to our best knowledge, there are no previous studies that analyze a metric that simply adjusts JIF by the number of authors. Therefore, we label this new metric as AWIF.

In economics, productivity is defined as production per input used. For example, worker productivity is measured by dividing total production by total number of workers. Number of citations that a journal receives can be considered as a measure for production of the journal. Then, AWIF can be considered as a measure for productivity of authors in a journal.

The idea of normalizing citations by the number publications in the references has also been considered before (Yanovski 1981). Our second metric closely follows from the formulae developed from Nicolaisen and Frandsen (2008) who have labelled the metrics as RRR. The authors present many versions of RRR in their study. We pick the RRR formula that resembles to JIF most as the second metric.

Success indicator compares the number of citations that an article receives to the number of publications in the references (Kosmulski 2011, Franceschini et al. 2012). If the former is greater than the latter, then the article is considered to be successful. RRR builds on a similar idea but constructs a measure for a journal's quality rather than an article's quality.

The number of publications that a journal cite can be considered as an input to the journal. The number of citations that a journal receive can be considered as a measure for production. Then, RRR can be considered as a measure for productivity per reference input.

We test whether AWIF and RRR achieve equality among social science and science fields. We simply compare percentile rank distribution of science journals to social science journals for each of the two metrics.

We take all 10,848 journals in journal citation report (JCR) 2012. The aim of using a large-scale data is to see whether the distribution of all social science journals is equal to the distribution of all science journals. It is possible to claim that a specific field within science has higher quality journals than a specific field within social science. However, it is harder to claim that science field as a whole has higher quality journals than social science field. Therefore, we believe that equality between science and social science fields is a desirable property for a journal quality metric.

JIF differs among fields partly because average number of authors for an article differs among fields. Larger teams of authors have better chance to disseminate their work and attract more citations (Wutchy et al. 2007; Abramo and D'Angelo 2015). Social science researchers are known to write papers in smaller teams (Yuret 2014, 2015). Therefore, AWIF is expected to decrease inequality among science and social fields by considering per-author citations.

The previous studies note that there is positive correlation between the number of cited-references and number of citations received (Ahlgren et al. 2018; Didegah and Thelwall 2013). This is natural because number of articles that a journal cites is a good proxy for the number of citations that are received in its field. In other words, journals in fields that receive more citations would be more likely to cite more articles. Therefore, we expect that RRR be more equalitarian among fields because it controls for the differences in the number of citations between science and social science fields.

Another reasons why we expect RRR to be more equalitarian because it addresses the coverage problem. There are many publications that cite social science journals which are uncovered because Web of Science typically does not cover books or regional publications. Therefore, social science journals have lower JIF because of the coverage problem. (Althouse et al. 2009; Marx and Bornmann 2015).

Coverage problem only affects the numerator of JIF. In contrast, denominator of RRR is also affected because cited-references from journals that are not covered by Web of Science are not counted in the denominator. If the effect on numerator and denominator have comparable magnitude, they may cancel out and lead to a more equalitarian measure.

## Related literature

There are many studies that use subject categories for normalization. One study normalizes impact factor of a journal by average impact factor of its subject category (Sombatsompop and Markpin 2005). A variant of this method uses median instead of averages (Ramirez et al. 2000). Another normalization divides the number of citations that a journal receive to all citations that are received in its subject category (Vinkler 2009). An alternative technique assigns a base subject category and normalize other fields according to the base subject category (Podlubny 2005).

More sophisticated statistical techniques are used for normalizations that use subject categories. For example, *z*-score of a journal is computed by subtracting mean JIF in the subject category from journal's JIF and dividing this by the standard deviation of JIF in the subject category (Lundberg 2007; Zhang et al. 2014).

There is an obvious advantage of using subject categories to normalize JIF. By construction, normalization attains equality among fields. For example, one of the standard ways to normalize JIF of a journal is to use its percentile rank within its subject category. Suppose we rank all science journals and social science journals separately according to their JIFs. Then percentile rank distribution of science and social science journals would be the same. Therefore such a normalization would achieve the best result in terms of the test we apply to our metrics.

However, the subject-based categorization is not without flaws. Waltman and Van Eck (2013) state that there are three main problems with this type of normalization. First, the idea of dividing scientific works into distinct subject categories is artificial because the boundaries between scientific fields are not clear-cut. Second, there are differences in citation behavior of subfields within fields. For example, empirical journals in economics get typically higher citations than theoretical papers (Johnston et al. 2013). Third, broad-scope journals such as Nature do not fit into any of the subject categories.

Another problem of using subject categories is to choose the appropriate subject category classification method. Web of Science subject categories are commonly used for normalization. However, this classification method is criticized and alternative subject categories are constructed (Ruiz-Castillo and Waltman 2015). Moreover, it is shown that

evaluation of a journal depends heavily on the choice of the subject categories (Zitt et al. 2005).

The technical issues about subject-based normalizations are important. However, the benefit of using subject-based normalizations is also huge. Perfect equality can be achieved in any way it is desired. The reason why we choose two metrics that are not subject-based is not technical but policy related. The subject-based normalizations evaluate science and social science journals in separate categories. In other words, science journals do not compete with social science journals under subject-based normalizations. Consequently, social scientists may feel that they are handicapped when they have favorable treatment. In a similar vein, scientists may find the subject-based categorization unfair because highly cited science journals are ranked below social science journals.

The metrics that are not subject based are less likely to attract these negative emotions. Suppose a university uses AWIF to achieve more equality among fields. Then, social scientist would not feel as if they are handicapped because the same criteria is used for all journals. In other words, social science and science journals are competing with each other by the same criteria. They are not put in separate categories and compete within that category. The scientists would not likely to complain that their highly cited journals are put into a disadvantage. If a science journal receives more citations per-author, it would also have a favorable treatment under AWIF as well.

There are many metrics other than AWIF and RRR that are not subject based. For example, Source-Normalized Impact Factor (SNIP) is a widely used metric that does not use subject categories in its computation. SNIP considers the source journal -the journal which cites- for normalization. Each citation that a journal receives is weighted by the inverse of number of publications in the references of the source journal (Moed 2010; Leydesdorff and Bornmann 2011).

The number of citations in the source journal is a proxy for the number of citations in the field. SNIP is expected to be more equalitarian among fields than JIF because citations from citation-rich fields are valued less. Empirical tests support this hypothesis (Zitt and Small 2008; Leydesdorff and Opthof 2010).

AWIF and RRR are different from SNIP. Unlike SNIP, these two metrics use journal's own information for normalization. For example, RRR uses the number of publications in the journal's own references for normalization whereas SNIP uses the number of publications in the source journal's references. Yet, the reason of why RRR and SNIP achieve equality among the fields is similar. Both metrics penalize fields that have long list of cited-references in their articles.

AWIF and RRR are chosen for this study because of their similarity to JIF. The aim of this paper is not to show that AWIF and RRR are best in terms of achieving equality among fields. For example, SNIP may achieve more equality than AWIF and RRR but this is not the focus of this paper. We simply check percentile rank distributions to show how AWIF and RRR metrics perform for equality. Since showing the technical superiority of metrics is not our primary concern, we do not use more sophisticated tests for field-equality as suggested in some previous studies (Radicchi and Castellano 2012).

In a similar vein, we do not use some suggested technical improvements because we want to preserve similarity. For example, we could use harmonic mean instead of arithmetic mean as suggested in previous studies (Leydesdorff and Opthof 2011; Waltman et al. 2011). In other words, we could compute per-author or per-cited reference citations at the article level and then take the average of these normalized citations. However, this method would take us away from JIF formula.

## Formulas for JIF, AWIF and RRR

The formula for JIF is given as follows:

$$JIF = \frac{\# \text{ of citations that publications of the journal in } t - 1 \text{ and } t - 2 \text{ receives at } t.}{\# \text{ articles of the journal in } t - 1 \text{ and } t - 2}$$

The numerator for all three metrics are the same. It is the number of citations that articles in years  $t - 1$  and  $t - 2$  receives in year  $t$ . Suppose that there are 100 citations that 2010 and 2011 publications of Journal X receives in year 2012. Then, the numerator is equal to 100.

The denominator of JIF is simply the number of articles in the journal in years  $t - 1$  and  $t - 2$ . Suppose that there are 50 articles that are published in years 2010 and 2011 in Journal X. Then, the denominator of the JIF is equal to 50.

In the numerator, the number of citations that are received by all publications are considered. In contrast, in the denominator, the number of publications that are articles only are considered. In other words, there is inconsistency between numerator and denominator which is mentioned in the introduction. It would be ideal if the citations that articles of the journal receive can be computed in the numerator. Unfortunately, the document type of cited-references is not available. Another possible solution is to consider all publications in the denominator. However, JCR does not favor this possibility as the citation potential of articles are much different from other type of documents. Since we follow the same formula as JCR, JIF indicator that we compute also stem from this inconsistency.

The formula for AWIF is given as follows:

$$AWIF = \frac{\# \text{ of citations that publications of the journal in } t - 1 \text{ and } t - 2 \text{ receives at } t.}{\# \text{ of authors of articles of the journal in } t - 1 \text{ and } t - 2}$$

The formula for JIF and AWIF are very similar. Numerator of JIF is the same as numerator of AWIF. Denominator of JIF is the number of articles in years  $t - 1$  and  $t - 2$  whereas denominator of AWIF is the number of authors who write these articles. In the previous example, we mention that there are 50 articles in years 2010 and 2011 in Journal X. If there are 200 authors in these 50 articles, then the denominator of AWIF is equal to 200.

The formula for RRR is given as follows.

RRR

$$= \frac{\# \text{ of citations that publications of the journal in } t - 1 \text{ and } t - 2 \text{ receives at } t.}{(\text{References dated } t - 3 \text{ and } t - 2 \text{ in } t - 1 \text{ articles and references dated } t - 4 \text{ and } t - 3 \text{ in } t - 2 \text{ articles})/2}$$

The formula for JIF and RRR are also very similar. Numerator of JIF is the same as numerator of RRR. Instead of using the number of articles in the denominator as in JIF, RRR uses the number of publications that these articles cite. For each of the articles in years  $t - 1$  and  $t - 2$ , a 2-year window is considered. This time frame is chosen to make denominator and numerator consistent because numerator also uses a 2-year window.

We count the number of publications in years  $t - 2$  and  $t - 3$  that articles in year  $t - 1$  cite. Moreover, we count the number of publications in years  $t - 3$  and  $t - 4$  that articles in year  $t - 2$  cite. As a consequence, there are 4 years of references in denominator as

opposed to 2 years of citations in numerator. Hence, we divide denominator by 2 to overcome this inconsistency.

In the previous example, we considered Journal X that have 50 articles that are published in years 2010 and 2011. Suppose that there are 75 publications that are published in 2008 and 2009 that are cited by articles of Journal X in 2010. In addition to this, suppose that there are 125 publications that are published in 2009 and 2010 that are cited by articles of Journal X in year 2011. Then the denominator of RRR would be equal to  $100 \left( \frac{125+75}{2} \right)$

## Data

JIF, AWIF and RRR values for year 2012 are computed. We compute all three measures for all 10,848 journals that are included in journal citation report (JCR) in year 2012. There are 8405 journals in the science field and 3016 journals in the social science field. 573 journals belong to both fields.

Since metrics for many journals are computed, the data requirement is massive. The number of publications that are required to compute three measures are given in Table 1. All 2012 publications in Web of Science are necessary for computation of numerator for all three measures. We need the number of articles from all 10,848 journals in 2010 and 2011 to compute the denominator for JIF. We need the number of authors who write these articles for the denominator AWIF. Lastly, we need the cited-reference list of these articles for the denominator RRR. Therefore, total number of publications necessary to do all computations is 4,392,092 (1,950,839 + 2,441,253).

We download all publications manually from Web of Science web site. We download information about publications from each of the 10,848 journals by using their ISSN numbers for years 2010–2012. Journals in JCR are not all journals covered by Web of Science. For example, Arts & Humanities journals are not in JCR but covered by Web of Science. Therefore, we also download publications that are not in any of the 10,848 journals but included in Web of Science for the year 2012.

We merge downloaded files into a single text file. Then, we compute three measures for each of 10,848 journals. Lastly, we compute percentile distributions for three measures. We write computer programs in Perl to achieve each of these tasks.

The numerators of three measures are the same. The number of citations that 2010 and 2011 publications receive in 2012 should be computed. To do this, we use cited-references of all publications covered by Web of Science in 2012. Web of Science gives the cited-reference information in the CR variable. Each cited-reference is separated by a semi-colon in CR. For each cited-reference, publication year and abbreviated journal title is given. Consequently, it is possible to choose cited-references that are published in 2010 and 2011. Moreover, it is possible to match the abbreviated journal titles in the cited-references to the abbreviated journal titles of 10,848 journals. Therefore, we are able to compute the number of citations that the articles from these journals in 2010 and 2011 receive in 2012.

We should note that matching abbreviated titles in the cited-references to the abbreviated titles of journals is far from perfect. The abbreviated journal titles of 10,848 journals

**Table 1** Data requirement for three measures

Numerator for all three measures	All 2012 publications in Web of Science	1,950,839
Denominator for JIF, AWIF, RRR	All 2010 and 2011 articles in journals from JCR	2,441,253

are gathered from variable that is called ‘Z9’ by Web of Science. By using all publications from 2010 to 2012, we are able to gather Z9s of all 10,848 journals. We notice that there are multiple Z9s for the same journal and we consider them all. However, it is possible that Web of Science may have used abbreviated journal titles in CR that are never used in Z9. In that case, we may not be able to compute all citations received by the journal.

The denominator of JIF is number of articles that are published in 2010 and 2011. The document type is labelled as “DT” in Web of Science. We only take publications that are given as “article” or “review article” for the denominator of JIF.

The denominator of AWIF is the number of authors who write articles that are published in 2010 and 2011. We do not aim to have number of distinct authors. For example, if the same author published three articles in the same journal, then the author is counted three times. Author names are given in the “AU” colon in Web of Science. Each author is separated by a semi-colon. Consequently it is easy to compute the number of authors.<sup>1</sup>

The denominator of RRR is the number of cited-references in the articles that are published in 2010 and 2011.<sup>2</sup> However, we do not include all cited-references in 2010 and 2011 to ensure that denominator and numerator are consistent. Numerator only counts the citations received by journals that are included in Web of Science. Ideally, denominator should also include the cited-references to the journals that are included in Web of Science. However, matching is done with abbreviated journal titles and we only have reliable titles for 10,848 journals in JCR. Therefore, we only include cited-references to the journals in JCR.

JIF that we compute is different from JIF that is computed in JCR. Specifically, JIF computed from our data-set is  $\pm 20\%$  of JIF in JCR for 89% of the journals. For 77% of the journals the difference is less than 10%. The problem of replication of JIF has been reported by Rossner et al. (2007). They note that JCR uses a data-set finer than the data-set that is available for researchers.

A finer data-set helps to get more accurate JIF. For example, we mention that it is hard to match cited-references to the journals because we can only use abbreviated journal titles for the matchings. If the cited-references also contained article titles, then matching would be much more successful.

The failure to replicate JIF is not a very big problem for this study because we compare distribution of journals rather than focusing on individual journals. Table 2 shows the basic statistics for the measures. We see that JIF that we compute is similar to JIF that is computed in JCR in terms of basic statistical indicators.

Figure 1 shows the quantile distributions of all journals. We see that both measures of JIF have very close distributions. JIF that we compute will be reported in the analysis for consistency because AWIF and RRR are also computed with the same data-set. However, none of the conclusions would change if we reported JIF that is computed in JCR instead.

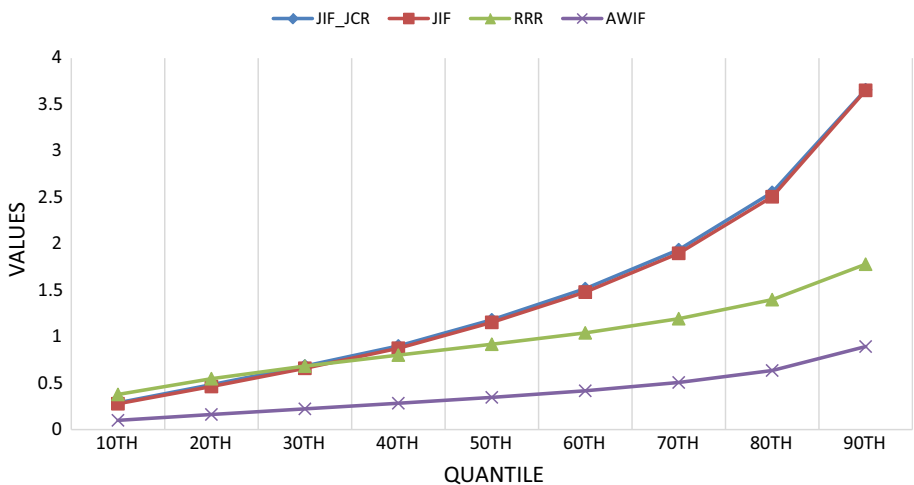
---

<sup>1</sup> The names of the group authors are also given in this column. For example, there are 1067 articles that have more than 100 authors in 2010 and 2011. Consequently, we are able to include articles with group authors in our calculations.

<sup>2</sup> We also do not take distinct number of articles in the cited-references. For example if the same article is cited in Journal X in three different articles, then that cited-article will be counted three times.

**Table 2** Basic statistics for JIF, AWIF and RRR

	Median	Average	SD
JIF_JCR	1.18	1.86	3.04
JIF	1.16	1.84	3.07
AWIF	0.35	0.50	0.82
RRR	0.92	1.08	1.30

**Fig. 1** Distribution of JIF, AWIF, RRR

## Analysis: social science versus science

Figure 2 shows JIF distribution of journals that belong to science and social science subject categories. We exclude journals that belong to both of these subject categories. We see that there is a vast difference between JIF distributions of science and social science journals. Science journals outperform social science journals by a large margin. 80th percentile social science journal is below 60th percentile science journal. Likewise, 70th percentile social science journal has JIF less than median science journal.

Figure 3 shows the distribution of journals with respect to AWIF and RRR. At each quantile, social science journals outperform science journals in both measures. In other words, dividing the number of citations simply by number of authors gives social science journals an edge. Moreover, dividing citations by number of references also gives an edge to social science journals.

Figure 3 also reveals that highest level of equality is attained by AWIF. For instance, 80th percentile social science journal is slightly above 80th percentile science journal but considerably below 90th percentile science journal. In contrast, RRR is less equitable than AWIF. 80th percentile science journal have RRR less than 70th percentile social science journal.



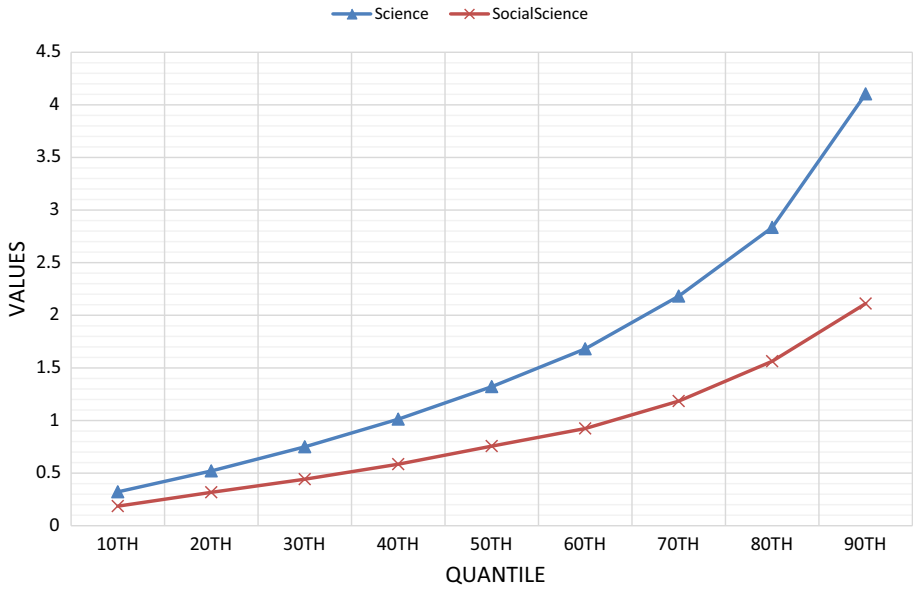


Fig. 2 Distribution of science and social science journals with respect to JIF

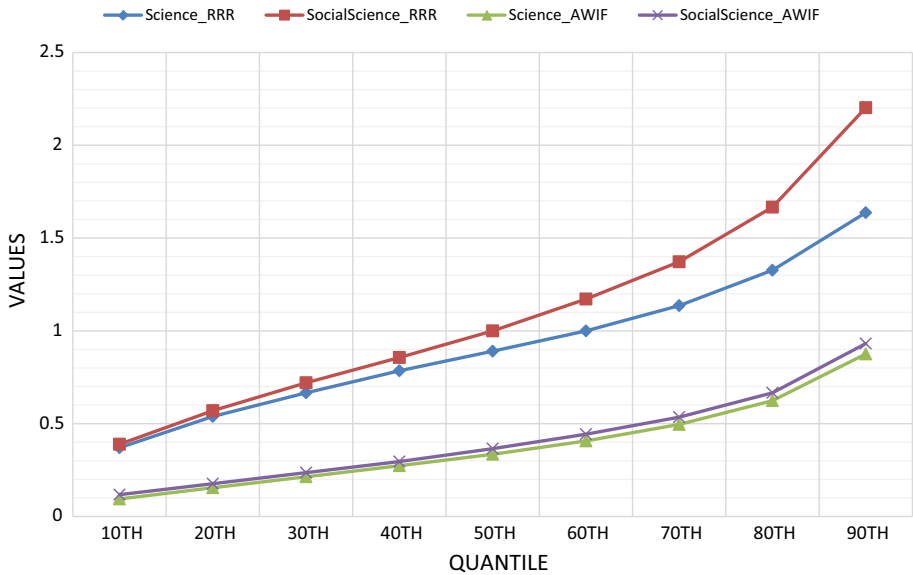


Fig. 3 Distribution of science and social science journals with respect to AWIF and RRR

## Analysis: narrower subject categories

In the previous section, we find two main results. First, AWIF is an equitable measure. Second, social science journals outperform science journals at both AWIF and RRR. In this section, we see whether these results can be generalized when we consider narrower subject categories. To do this, we choose a social science and a science subject category and make a comparison between two.

Figure 4 shows the JIF distributions of history (social science) and engineering-mechanical (science) subject categories. Engineering-mechanical has a considerably better distribution than history. Median engineering-mechanical journal has a comparable JIF to 90th percentile history journal.

Figure 5 shows the AWIF and RRR distributions of history and engineering-mechanical subject categories. Two subject categories have almost indistinguishable quantile distributions under AWIF. The disadvantage of history under JIF reverses under RRR.

To sum, AWIF is more equitable than JIF and RRR. Moreover, the advantage in JIF reverses when we consider RRR. Therefore, this example has similar findings that we attain for science and social science categories in the previous section.

Unfortunately, findings that we draw from previous section cannot be generalized to all subject categories. Next, we provide two counter-examples to show that generalization fails.

Figure 6 shows the JIF and AWIF distributions of law (social science) and biology (science). We see that JIF is highly unequatable between biology and law journals. Biology journals clearly outperform law journals. AWIF reverses this advantage. Law journals outperform biology journals under AWIF. Moreover, AWIF has a level of inequality comparable to the level of inequality under JIF. Therefore, the result that AWIF is more equitable than JIF cannot be generalized to all subject categories.

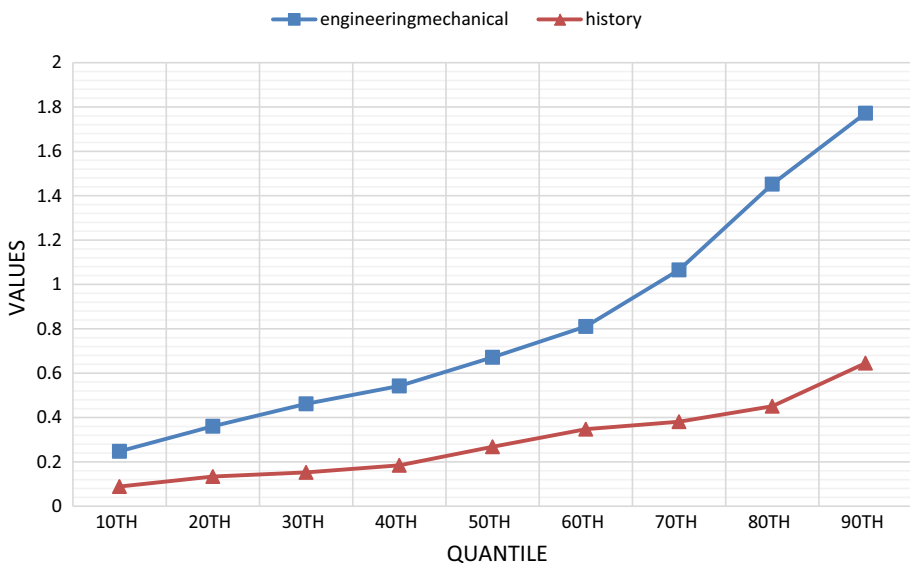


Fig. 4 JIF of engineering-mechanical and history

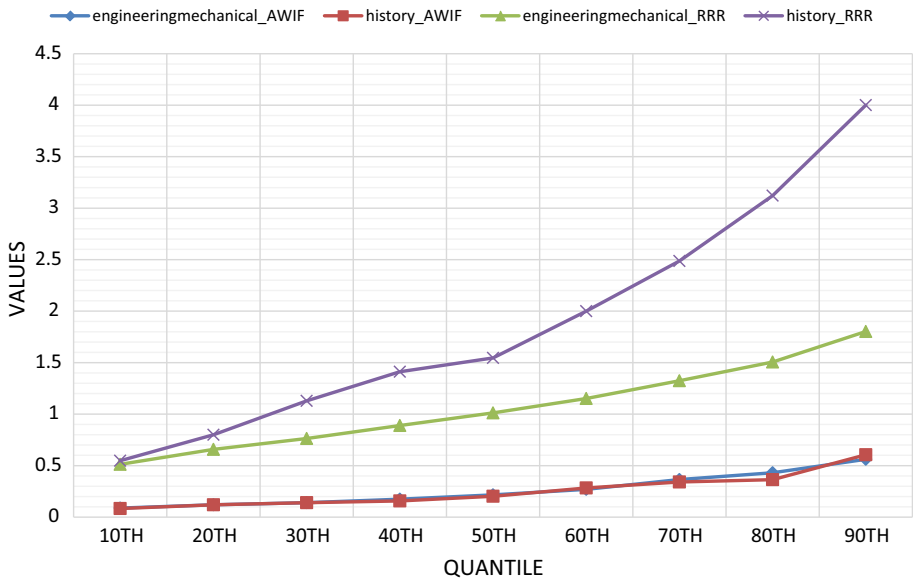


Fig. 5 AWIF and RRR of engineering- mechanical and history

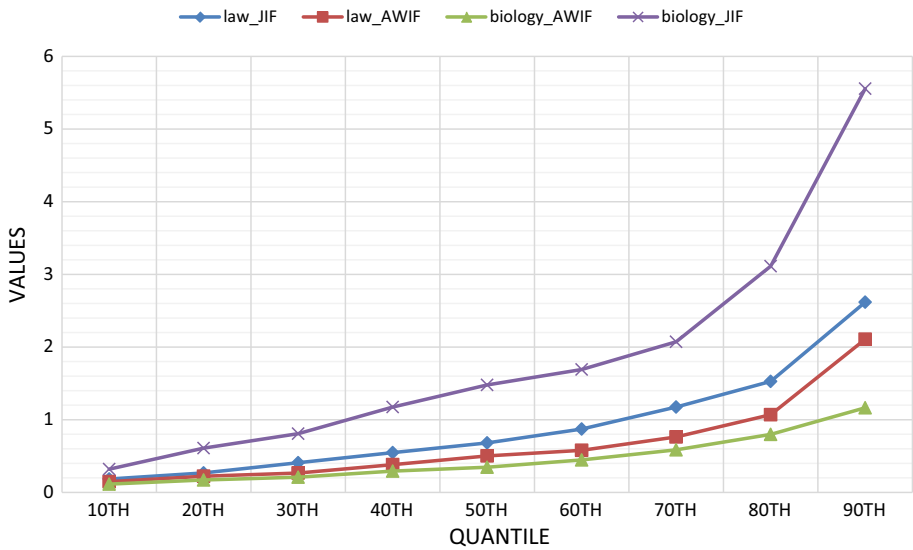


Fig. 6 JIF and AWIF of law and biology

Figure 7 shows JIF and RRR distribution of statistics and probability (science) and public administration (social science). Statistics and probability journals outperform public administration journals at both JIF and RRR. Therefore, the result that the advantage of JIF reverses when RRR is considered cannot be generalized to all subject categories.

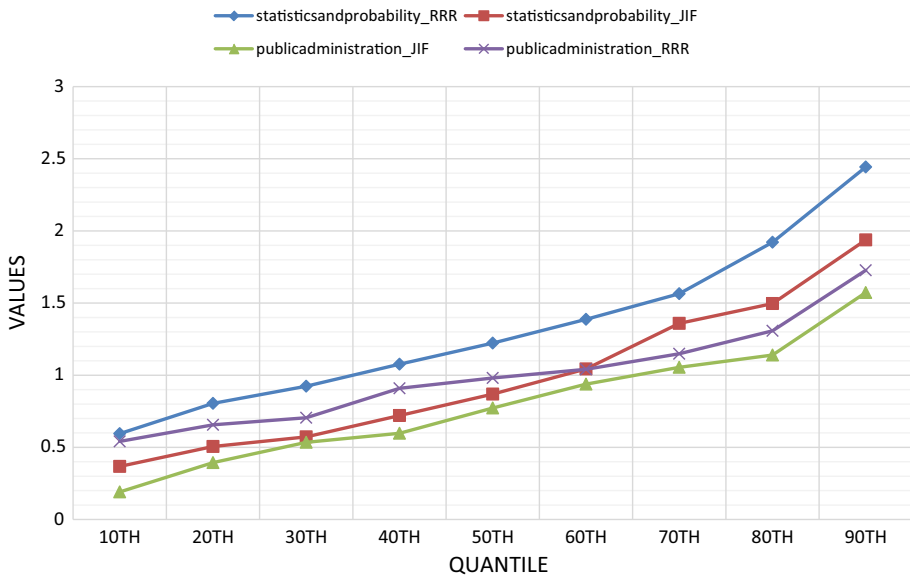


Fig. 7 JIF and RRR for statistics and probability and public administration

## AWIF and RRR for publications with a large number of authors

Some publications that are written by groups such as ATLAS have a large number of authors. There is an upward trend for these publications (King 2012). The phenomenon is widespread in fields such as particle physics. For example, one publication from Physical Review Letters broke the record by having 5000 authors (Castelvecchi 2015). However, the problem is not endemic to particle physics, but can be seen in all physical sciences. For example, one biology paper exceeded the threshold of 1000 authors (Woolston 2015).

Table 3 gives some examples of physics journals that contain publications written by a large number of authors. JIF of these journals are high partially because of their field. RRR corrects for field differences, and two out of six journals become below average journals. Another reason for high JIF is that the metrics does not take number of authors into account at all. When AWIF corrects for the number of authors, all six journals fall below average.

**Table 3** Physics journals that have a large number of authors

Journal	JIF	AWIF	RRR	Field
European Physical Journal C	4.81	0.04	1.36	Physics, particle
Journal of High Energy Physics	4.71	0.15	0.93	Physics, particle
Physical Review C	3.27	0.23	1.26	Physics, nuclear
Physical Review D	4.49	0.21	1.00	Astronomy and astrophysics
Physical Review Letters	7.61	0.28	2.72	Physics, multidisciplinary
Physics Letters B	4.28	0.06	1.60	Physics, multidisciplinary
Average of all JCR journals	1.84	0.50	1.08	

The quality of journals probably lie somewhere inbetween their JIF and AWIF values. If the problem stems from properties of AWIF, then, there are easy fixes. AWIF can be computed by harmonic means. That is, the average of per-author citations of each publication can be computed instead of per-author citations of the whole journal. Per-author citations of group-authored papers is practically zero. A further adjustment can be done to overcome this. Instead of computing per-author citations, number of citations can be divided by a concave function of the number of authors.

The reason for the divergence between AWIF and JIF values may be because authors in publications with group authors do not comply with the standard definition of authorship. Therefore, the number of authors in these publications does not reflect the correct number of authors. In this case, AWIF correctly makes this problem more visible, and no adjustments to the formula is necessary.

## Conclusion

JIF is adjusted for number of authors and references in AWIF and RRR respectively. Two main conclusions from the analysis of these metrics are drawn in this study. First, social science journals outperform science journals in both of these measures. Second, AWIF is highly equalitarian between science and social science journals.

These results are not generalizable when narrow subject categories are considered. Counter-examples are found when one science and one social science subject category are picked and compared. For example, AWIF is not equalitarian among all subject categories that we consider. However, this result is inevitable as we do not normalize by considering subject categories in the first place. It is hard to come up with a metric that is constructed without taking subject categories into account and is able to equate hundreds of subject categories.

AWIF and RRR has two main advantages. First, AWIF and RRR has similar formulas to that of JIF. The similarity would make administrators understand two metrics more easily. Moreover, these metrics preserve some desirable properties of JIF because of the similarity.

Second, AWIF and RRR are not subject-based metrics. We do not use any subject categorization when we compute these metrics. We believe that a metric that is not subject-based has some advantages in practice. Social scientists would not be treated as if they are handicapped because social science journals are not put in a separate category. Scientist have less objections for the metrics because science journals compete with social science journals on equal footing.

AWIF and RRR are better than JIF in terms of equality between social science and science journals. However, it may be the case that they are worse proxies for quality. It may be the case that better researchers have a better knowledge about the related literature and add more publications in their reference list. Ahlgren et al. (2018) suggest that articles with higher number of cited-references may have higher quality because they receive citations even after articles are controlled for their subject categories. However, this is not the ultimate proof that articles with more cited-references is of a higher quality. For instance, review articles which have many cited-references attract more citations. That does not mean that review articles are better than other articles.

The same argument can be extended to the number of authors. The articles with more authors may be of better quality. There is a clear advantage in collaboration. Articles that

have many authors utilize the expertise of many researchers. However, there is no clear empirical evidence that larger teams of authors write better articles than smaller teams.

More research should be done to understand the relationship between the quality of an article and number of cited-references and authors of the article. If more cited-references and authors are shown to produce better quality journals, then the credibility of AWIF and RRR as a journal quality metric would be hurt. Otherwise, administrators may find AWIF and RRR as useful metrics because these metrics treat scientists and social scientists fairly.

## References

- Abramo, G., & D'Angelo, C. A. (2015). The relationship between the number of authors of a publication, its citations and the impact factor of the publishing journal: Evidence from Italy. *Journal of Informetrics*, 9(4), 746–761.
- Ahlgren, P., Colliander, C., & Sjogarde, P. (2018). Exploring the relation between referencing practices and citation impact: A large-scale study based on Web of Science data. *Journal of the Association for Information Science and Technology*, 69(5), 728–743.
- Althouse, B. M., West, J. D., Bergstrom, C., & 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.
- Archambault, E., & Lariviere, V. (2009). History of the journal impact factor: Contingencies and consequences. *Scientometrics*, 79(3), 635–649.
- Castelvecchi, D. (2015) Physics paper sets record with more than 5,000 authors. *Nature News* (May 15).
- Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, 7(4), 861–873.
- Franceschini, F., Galetto, M., Maisano, D., & Mastrogiacomo, L. (2012). The success-index: An alternative approach to the h-index for evaluating an individual's research output. *Scientometrics*, 92(3), 621–641.
- Harzing, A., Alakangas, S., & Adams, D. (2014). hIa: An individual annual h-index to accommodate disciplinary and career length differences. *Scientometrics*, 99(3), 811–821.
- Johnston, D. M., Piatti, M., & Torgler, B. (2013). Citation success over time: Theory or empirics? *Scientometrics*, 95(3), 1023–1029.
- King, C. (2012). Multi-author papers: onward and upward. *Sciencewatch newsletter*.
- Kosmulski, M. (2011). Successful papers: A new idea in evaluation of scientific output. *Journal of Informetrics*, 5(3), 481–485.
- Leydesdorff, L., & Bornmann, L. (2011). How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology*, 62(2), 217–229.
- Leydesdorff, L., & Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology*, 61(11), 2365–2369.
- Leydesdorff, L., & Opthof, T. (2011). Remaining problems with the "New Crown Indicator" (MNCS) of the CWTs. *Journal of Informetrics*, 5(1), 224–225.
- Lundberg, J. (2007). Lifting the crown—Citation z-score. *Journal of Informetrics*, 1(2), 145–154.
- Marx, W., & Bornmann, L. (2015). On the causes of subject-specific citation rates in Web of Science. *Scientometrics*, 102(2), 1823–1827.
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265–277.
- Nicolaisen, J., & Frandsen, T. F. (2008). The reference return ratio. *Journal of Informetrics*, 2(2), 128–135.
- Podlubny, I. (2005). Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*, 64(1), 95–99.
- Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics*, 6(1), 121–130.
- Ramirez, A. M., Garcia, A. O., & Del Rio, J. A. (2000). Renormalized impact factor. *Scientometrics*, 47(1), 3–9.
- Rosner, M., Van Epps, H., & Hill, E. (2007). Show me the data. *The Journal of Cell Biology*, 179(6), 1091–1092.

- Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, 9(1), 102–117.
- 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.
- Vinkler, P. (2009). Introducing the Current Contribution Index for characterizing the recent, relevant impact of journals. *Scientometrics*, 79(2), 409–420.
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391.
- Waltman, L., & van Eck, N. J. (2013). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, 96(3), 699–716.
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37–47.
- Woolston, C. (2015) Fruit-fly paper has 1,000 authors. *Nature News* (May 13).
- Wutchy, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316, 1036–1038.
- Yanovski, V. I. (1981). Citation analysis significance of scientific journals. *Scientometrics*, 3(3), 223–233.
- Yuret, T. (2014). Why do economists publish less? *Applied Economics Letters*, 21(11), 760–762.
- Yuret, T. (2015). Interfield comparison of academic output by using department level data. *Scientometrics*, 105(3), 1653–1664.
- Zhang, Z., Cheng, Y., & 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(3), 1679–1693.
- Zitt, M., Ramanana-Rahary, S., & 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.
- 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.