Joel A.C. Baum

The excess-tail ratio: correcting journal impact factors for citation distributions

M@n@gement, 16(5), 697-706.

Special Issue

© 2013 M@n@gement and the author(s).
The excess-tail ratio: correcting journal impact factors for citation distributions

Joel A.C. BAUM
University of Toronto
jbaum@rotman.utoronto.ca

Abstract
Despite their widespread adoption, journal Impact Factors suffer well-known drawbacks that limit their usefulness in accurately and fairly assessing scientific quality. Among these is the extreme variance and skewness in the citations to articles published by a given journal, which results in their sensitivity to a few highly cited articles, and enables many infrequently cited articles to “free-ride” on citations to these “skewed few.” To address this problem, I adjust journal Impact Factors according to the relative citedness of the few highly cited articles in a journal’s h-core (i.e., the h articles that receive at least h citations) and the many infrequently cited articles in its h-tail (i.e., those that receive fewer than h citations). I gauge the skew of a journal’s citation distribution by $e^2/t^2$, the excess-tail ratio where $e^2$ captures excess citations above the $h^2$ citations received by a few highly cited h-core articles and $t^2$ captures surplus citations received by the many infrequently cited h-tail articles that fall below the h-core. I employ $e^2/t^2$ to adjust raw Impact Factors for 25 selected management and economics journals. The adjusted scores, $IF$, discriminate Impact Factors based on the shapes of journal citation distributions, leading to more accurate evaluation. I find $e^2/t^2 < 1$ (often $<< 1$) for 23 of these journals to be consistent with an overstatement of their quality resulting from the sensitivity of Impact Factors to a few highly cited articles. Adjusted Impact Factors also yield distinctive and more consistent journal rankings over standard two-year and five-year time horizons. I conclude that the “excess-tail” ratio and $IF$ are a useful complement to journal Impact Factors, particularly given their increasing use in the evaluation of individual scholarly output.

Keywords: h-Index, e-Index, h-core, h-tail, excess-tail ratio, journal Impact Factor, IF
INTRODUCTION

Thompson Scientific is a database company that owns and publishes the Institute for Scientific Information (ISI) Web of Knowledge, which includes the Science Citation Index, the Social Science Citation Index, and the Journal Citation Reports. Central to the Journal Citation Reports are journal Impact Factors, which Thompson/ISI (1994) describes as “a systematic and objective means to critically evaluate the world’s leading journals.” The Impact Factor was devised in the 1960s by Eugene Garfield as a way to measure journal usage based on the mean number of citations per article within a specific period of time. A journal’s Impact Factor is calculated by counting the number of current-year citations to articles published by the journal during the preceding two years and dividing this count by the number of articles the journal published in those two years. More recently, the ISI introduced a five-year Impact Factor (i.e., current-year citations to articles published by the journal during the preceding five years divided by the number of articles published in those years) to account for differences in the diffusion and obsolescence of ideas across fields.

Garfield’s original idea was to sort journals by citation rates to aid in determining which to include in library collections (or indexes). Over the last decade, however, increasing electronic availability, along with aggressive marketing by Thomson Scientific, which acquired the ISI in 1992, has transformed the Impact Factor from a sorting device into a definitive quantitative rating of the quality of journals, of particular articles appearing in them and, by corollary, of the academics writing those articles. The journal Impact Factor is now in widespread use to evaluate researchers, serving central functions in academic hiring, peer review, and grant decisions—uses for which it was never intended and which Garfield (2006) himself has called misleading and inappropriate.

As a result, the Impact Factors of journals in which a researcher tends to publish are increasingly central to evaluations of his/her scholarly achievements. Indeed, the tendency has increasingly been to ascribe the Impact Factor of a journal to each article published within it. The veracity of such attributions rests on the assumption that a journal’s Impact Factor is representative of its articles. For this to be true, the citedness of a journal’s articles must follow a Gaussian distribution, with a narrow variance around the mean—that is, around its Impact Factor. It is well-known, however, that the distribution of citations to a journal’s articles is highly skewed, with few articles near the mean. The skewness of article citedness is also problematic for Impact Factors as an index of journal quality because mean journal citedness is disproportionately influenced by a small number of highly cited articles. A small minority of articles, unrepresentative of the journal’s publications, may thus be decisive in determining journal Impact Factors and resultant journal quality rankings (Baum, 2011).

To address this problem, I propose an adjustment for journal Impact Factors to account for the distribution of citations a journal receives. The adjustment is derived from two recent extensions to the $h$-index (Hirsch, 2005): $e^2$, which captures excess citations beyond the $h^2$ citations received by $h$-core articles (Zhang, 2009), and $t^2$, which captures surplus citations received by $h$-tail articles that fall below the $h$-core (Ye and Rousseau, 2010). A journal’s citation distribution is gauged by the “excess-tail” ratio, $e^2 / t^2$, which indicates...
the relative citedness of the few highly cited articles comprising the journal’s h-core and the many infrequently cited articles falling in its h-tail. The excess-tail ratio is used to compute $\hat{IF}$ as $IF \times \frac{e^2}{t^2}$, where $IF$ is the raw journal Impact Factor and $\hat{IF}$ is the adjusted journal Impact Factor. Using the excess-tail ratio to adjust journal Impact Factors maintains the advantage of having a single index with which to evaluate journals, while incorporating important information on journal citation distributions.

I compute excess-tail ratios and $\hat{IF}$ for a sample of 25 management and economics journals. I find $\frac{e^2}{t^2} < 1$ (often $<< 1$) for 23 of the sample journals. These results are consistent with an overstatement of raw Impact Factors, attributable to their sensitivity to a few highly cited articles. I also find that $\hat{IF}$ is more stable than raw Impact Factors across standard two- and five-year time horizons.

**GAUGING JOURNAL CITATION DISTRIBUTION WITH THE EXCESS-TAIL RATIO**

I gauge a journal’s citation distribution using the ratio of, on the one hand, excess citations to the few highly cited articles in the journal’s h-core to, on the other, surplus citations to the many infrequently-cited articles in the journal’s h-tail. The h-index divides a journal’s articles into two groups: the first group is the h-core, each having at least h citations during the period under study, and the second is the h-tail, each having at most h-1 citations.

The h-index, h-core and h-tail can be applied to many source–citation relations over many time windows (Ye and Rousseau, 2010). If there are S source articles and C citations, by definition the h-core consists of h articles and the h-tail consists of S – h articles. The number of citations in the h-core, $C_h$, is a minimum of $h^2$ but has no upper limit. Zhang (2009) recently defined $e^2$, comprised of $C_h - h^2$ citations, to distinguish “excess” citations ignored by the h-index. The number of citations in the h-tail, $t^2$, ranges from 0 to $(S - h)(h - 1)$.

The relation between h-core and h-tail citations is illustrated in Figure 1, which represents the citedness of a journal’s articles assuming a continuous citation function. In the figure, citations to the journal’s h-core articles, $C_h$, are the sum of citations in the $h^2$ and $e^2$ areas. The $t^2$ area represents the surplus citations received by the journal’s h-tail articles. The excess-tail ratio, $\frac{e^2}{t^2}$, gauges a journal’s citation distribution based on the ratio of excess h-core to surplus h-tail citations. When $\frac{e^2}{t^2} > 1$, citations tend to be excess citations to the few articles comprising a journal’s h-core articles. When $\frac{e^2}{t^2} < 1$, citations tend to be surplus citations to the many articles comprising a journal’s h-tail. Thus, the larger the ratio, the greater the extent to which a journal’s citations reflect excess citations to its few most highly cited articles relative to surplus citations to its many infrequently cited articles.\(^3\)

3. Substituting $\frac{e^2}{t^2}$ for $\frac{e^2}{t^2}$ gives identical results; the correlation between the two is $r = .996$ for the sample of journals examined below.
Figure 1. Geometric representation of $e^2$, $t^2$, $h$, and $h^2$

Symmetric distribution

Skewed distribution

Adapted from Zhang (2009); Ye and Rousseau (2010)
The excess-tail ratio: correcting journal impact factors for citation distributions

M@n@gement vol. 16 no. 5, 2013, 697-706

4. Using publication data for the period 2006–2010 and citation data for the period 2006–2011 results in lower values of $e^2/t^2$ for all journals, but does not otherwise alter the main findings or their implications.

5. Articles in a journal’s $h$-tail may receive no citations, and so contribute nothing to $t^2$. When some $h$-tail articles go uncited, the excess-tail ratio may therefore overstate the “quality” of a journal’s citations. To correct for this, $t^2$ can be substituted with the number of “reverse tail” citations $r^2$; that is, the difference in the actual number of $h$-tail article citations and the number of $h$-tail citations if all $h$-tail articles received $h$ citations. More formally, $r^2 = t^2 - (S - h)(h - 1)$, which ranges from 0 to $(S - h)(h - 1)$. Because, for the 25 sample journals, the proportion of uncited $h$-tail articles is relatively small (mean = .058; min = .006; max = .171), the correlation between $e^2/t^2$ and $e^2/r^2$ is 0.949, and the impact of this substitution on adjusted Impact Factors is negligible. Nevertheless, if the proportion of uncited $h$-tail articles is large, this substitution may be material.

Figure 2. Quarterly Journal of Economics citations, 2000-2010

Note: Ninety of the 457 articles published in the Quarterly Journal of Economics during the period 2000-2010 received at least 90 citations, giving $h = 90$ and $h^2 = 8,100$. The total number of citations received by these 90 “$h$-core” articles is, however, 16,021, yielding $e^2 = 16,021 - 8,100 = 7,921$. The total number of citations to the remaining 367 “$h$-tail” articles (each receiving < 90 citations) is 11,025, which gives $t^2$. Accordingly, $e^2/t^2 = 0.718$, indicating an inflated Impact Factor.
Table 1. Citation statistics and ratios for selected management and economics journals

| Journal                                      | h   | h²  | e²  | t²  | e²/t² |
|----------------------------------------------|-----|-----|-----|-----|-------|
| Academy of Management Journal                | 106 | 11236 | 7225 | 20736 | 0.348 |
| Academy of Management Review                 | 85  | 7225 | 9025 | 9604 | 0.940 |
| Administrative Science Quarterly             | 62  | 3844 | 4900 | 3364 | 1.457 |
| American Economic Review                     | 100 | 10000 | 8649 | 33489 | 0.258 |
| Econometrica                                 | 74  | 5476 | 5929 | 11664 | 0.508 |
| International Journal of Industrial Organization | 29  | 841  | 625  | 4489  | 0.139 |
| Journal of Economics & Management Strategy   | 27  | 729  | 576  | 1764  | 0.327 |
| Journal of Economic Behavior & Organization  | 39  | 1521 | 900  | 6889  | 0.131 |
| Journal of Economic Literature               | 58  | 3364 | 6084 | 2209  | 2.754 |
| Journal of Financial Economics               | 81  | 6561 | 5625 | 16384 | 0.343 |
| Journal of Finance                           | 98  | 9604 | 6400 | 22801 | 0.281 |
| Journal of Industrial Economics              | 32  | 1024 | 1156 | 1681  | 0.688 |
| Journal of International Business Studies    | 61  | 3721 | 2401 | 9216  | 0.261 |
| Journal of Law, Economics, & Organization    | 27  | 729  | 576  | 1521  | 0.379 |
| Journal of Management                         | 69  | 4761 | 4624 | 8100  | 0.571 |
| Journal of Management Studies                 | 55  | 3025 | 1849 | 9604  | 0.193 |
| Journal of Political Economics                | 69  | 4761 | 3600 | 7921  | 0.454 |
| Management Science                           | 88  | 7744 | 7225 | 23409 | 0.309 |
| Organization Science                          | 76  | 5776 | 7225 | 10816 | 0.668 |
| Organization Studies                          | 48  | 2304 | 1600 | 6889  | 0.232 |
| Quarterly Journal of Economics                | 90  | 8100 | 7921 | 11025 | 0.718 |
| Review of Economics & Statistics              | 57  | 3249 | 2916 | 10000 | 0.292 |
| Review of Economic Studies                    | 53  | 2809 | 2116 | 6084  | 0.348 |
| Review of Financial Studies                   | 54  | 2916 | 2704 | 8464  | 0.319 |
| Strategic Management Journal                  | 96  | 9216 | 11025| 17956 | 0.614 |

With two exceptions (Administrative Science Quarterly and Journal of Economic Literature), surplus h-tail citations exceed excess h-core citations, resulting in an excess-tail below 1 and for many journals << 1. Among the sample journals, modal citations thus tend to be citations to infrequently cited h-tail articles. As a result, a journal’s $\overline{h}$ is generally smaller (and often significantly so) than its raw Impact Factor (Table 2). This result is consistent with raw journal Impact Factors being inflated by a few highly cited articles in a journal’s h-core, when citations are more typically to one of a larger number of infrequently cited articles in the journal’s h-tail.  

6. The excess-tail ratio, $e^2/t^2$, can also be used to adjust the h-index itself, improving its ability to discriminate the shapes of citation distributions similarly (Zhang 2013).
Table 2. Raw and adjusted impact factors and rankings for selected management and economics journals

| Journal                                      | Two-year IF | Rank | Five-year IF | Rank | Adj. two-year $\bar{IF}$ | Rank | Adj. five-year $\bar{IF}$ | Rank |
|----------------------------------------------|-------------|------|--------------|------|--------------------------|------|----------------------------|------|
| Academy of Management Journal                | 5.250       | 4    | 10.779       | 2    | 1.829                    | 9    | 3.756                      | 7    |
| Academy of Management Review                 | 6.720       | 2    | 11.657       | 1    | 6.315                    | 2    | 10.954                     | 2    |
| Administrative Science Quarterly             | 3.683       | 13   | 7.359        | 5    | 5.365                    | 3    | 10.719                     | 3    |
| American Economic Review                     | 3.150       | 16   | 4.278        | 17   | 0.814                    | 17   | 1.105                      | 19   |
| Econometrica                                 | 3.185       | 15   | 5.330        | 13   | 1.619                    | 10   | 2.709                      | 10   |
| International Journal of Industrial Organization | 0.731     | 25   | 1.247        | 25   | 0.102                    | 25   | 0.174                      | 25   |
| Journal of Economics & Management Strategy   | 1.123       | 22   | 1.656        | 23   | 0.367                    | 23   | 0.541                      | 23   |
| Journal of Economic Behavior & Organization | 0.924       | 23   | 1.355        | 24   | 0.121                    | 24   | 0.177                      | 24   |
| Journal of Economic Literature               | 7.432       | 1    | 8.076        | 3    | 20.469                   | 1    | 22.243                     | 1    |
| Journal of Financial Economics               | 3.810       | 10   | 5.631        | 11   | 1.308                    | 12   | 1.933                      | 11   |
| Journal of Finance                          | 4.141       | 7    | 6.529        | 8    | 1.162                    | 13   | 1.833                      | 12   |
| Journal of Industrial Economics              | 0.796       | 24   | 1.678        | 22   | 0.547                    | 21   | 1.154                      | 18   |
| Journal of International Business Studies    | 4.148       | 6    | 5.539        | 12   | 1.081                    | 14   | 1.443                      | 15   |
| Journal of Law, Economics, & Organization   | 1.595       | 21   | 2.172        | 21   | 0.604                    | 20   | 0.823                      | 22   |
| Journal of Management                        | 3.758       | 12   | 6.210        | 9    | 2.145                    | 7    | 3.545                      | 8    |
| Journal of Management Studies                | 3.817       | 9    | 4.684        | 15   | 0.735                    | 18   | 0.902                      | 20   |
| Journal of Political Economics               | 4.065       | 8    | 6.896        | 6    | 1.847                    | 8    | 3.134                      | 9    |
| Management Science                          | 2.221       | 20   | 3.966        | 19   | 0.685                    | 19   | 1.224                      | 17   |
| Organization Science                        | 3.800       | 11   | 5.838        | 10   | 2.538                    | 5    | 3.900                      | 6    |
| Organization Studies                        | 2.339       | 19   | 3.590        | 20   | 0.543                    | 22   | 0.834                      | 21   |
| Quarterly Journal of Economics               | 5.940       | 3    | 8.053        | 4    | 4.268                    | 4    | 5.786                      | 4    |
| Review of Economics & Statistics             | 3.113       | 17   | 4.300        | 16   | 0.908                    | 16   | 1.254                      | 16   |
| Review of Economic Studies                   | 2.883       | 18   | 4.163        | 18   | 1.003                    | 15   | 1.448                      | 14   |
| Review of Financial Studies                  | 4.602       | 5    | 5.016        | 14   | 1.470                    | 11   | 1.602                      | 13   |
| Strategic Management Journal                 | 3.583       | 14   | 6.818        | 7    | 2.200                    | 6    | 4.186                      | 5    |

Note: $\bar{IF} = IF \times e^{\gamma/t^2}$

After making these adjustments the relative ranks of the journals shift, and sometimes substantially, as illustrated in Table 2. The lower a journal is ranked based on $\bar{IF}$ relative to its raw Impact Factor, the greater the initial overstatement of its impact due to the sensitivity of its raw Impact Factor to a few highly cited $h$-core articles relative to the mass of infrequently cited $h$-tail articles. As illustrated in Figure 3, the correlation between two- and five-year rankings is substantially larger for $\bar{IF}$ ($r = 0.98$) than for raw Impact Factors ($r = 0.88$). $\bar{IF}$ is thus more consistent over different citation time horizons, particularly among higher-ranking journals.
Figure 3. Correlation between two- and five-year raw and adjusted journal Impact Factors
CONCLUSION

As a measure of research quality, journal Impact Factors are problematic. The tendency to attach the same value to each article published in a given journal masks extreme variability in article citedness, and permits a journal’s many infrequently cited articles—and the journal itself—to free-ride on the journal’s few highly cited articles, which are principal in determining the journal’s Impact Factor (Baum, 2011). I propose a correction for this problem whereby a journal’s raw Impact Factor is adjusted to account for its citation distribution, which is gauged by the ratio of its excess \(h\)-core to surplus \(h\)-tail citations. This excess-tail ratio captures the extent to which the journal’s citations are centered on the more or less frequently cited articles, and thus the more or less influential articles it publishes.

I employ the excess-tail ratio to recalibrate Impact Factors for 25 selected journals in management and economics. The excess-tail ratio is less than 1 (and often \(< 1\)) for all but two sample journals. This is consistent with an overstatement of raw journal Impact Factors resulting from their sensitivity to small numbers of highly cited articles, and an inability to discriminate the shapes of the underlying journal citation distributions. Thus, while journal Impact Factors in management and economics are driven by citations to the journals’ small number of influential \(h\)-core articles, more typically their citations are to one of the large number of infrequently cited \(h\)-tail articles they publish. Moreover, adjusted Impact Factors (\(\overline{IF}\)) produce rankings that differ (often substantially) from the raw rankings, and are more consistent across two- and five-year time horizons.

The excess-tail ratio and \(\overline{IF}\) thus appear to provide useful complements to journal Impact Factors in assessing journal impact and quality, particularly given the increasing use of journal Impact Factors in the evaluation of individual scholarly output. Journal Impact Factors adjusted by these ratios carry additional information derived from journal citation distributions. As a result, \(\overline{IF}\) would appear to afford a more accurate single-number metric for the evaluation of journals and the authors who publish in them.

Joel A.C. Baum is Associate Dean, Faculty and George E. Connell Chair in Organizations and Society at the Rotman School of Management, University of Toronto, where he also received his PhD late in the last millennium.
REFERENCES

- Baum, J. A. C. (2011). Free-riding on power laws: Questioning the validity of the Impact Factor as a measure of research quality in organization studies. *Organization, 18*(4), 449-466.

- Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association, 295*(1), 90-93.

- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569-16572.

- Institute for Scientific Information (1994). *The ISI Impact Factor*. Thomson Scientific. Retrieved from: http://thomsonreuters.com/products_services/science/free/essays/impact_factor/

- Monastersky, R. (2005, October 14). The number that is devouring science. *Chronicle of Higher Education*.

- Reedijk J., & Moed, H. F. (2008). Is the impact of journal impact factors decreasing. *Journal of Documentation, 64*(2), 183-192.

- Rousseau, R. (2006). New developments related to the Hirsch index. *Science Focus, 1*, 23-25 (Chinese). English translation retrieved from: http://eprints.rclis.org/6376/

- Wilhite, A. W., & Fong, E. A. (2012). Coercive citation in academic publishing. *Science, 335*(6068), 542-543.

- Ye, F. Y., & Rousseau, R. (2010). Probing the h-core: An investigation of the tail-core ratio for rank distributions. *Scientometrics, 84*, 431-439. doi: 10.1007/s11192-009-0099-6

- Zhang, C. T. (2009). The e-index, complementing the h-index for excess citations. *PLoS ONE, 4*(5), e5429.

- Zhang, C. T. (2013). The h'-Index, effectively improving the h-Index based on the citation distribution. *PLoS ONE, 8*(4), e59912.