

Dynamic Impact Factors and Escalating Journal Self Citations

C.-L. Chang^{ab}, E. Maasoumi^c and M. McAleer^{def}

^a*Department of Applied Economics, National Chung Hsing University, Taiwan.*

^b*Department of Finance, National Chung Hsing University, Taiwan.*

^c*Department of Economics, Emory University*

^d*Department of Quantitative Finance, National Tsing Hua University, Hsinchu, Taiwan*

^e*Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands.*

^f*Tinbergen Institute, The Netherlands*

Email: changchialin@nchu.edu.tw

Abstract: The paper focuses on the robustness of rankings of academic journal quality and research impact in general, and in Economics, in particular, based on the widely-used Thomson Reuters ISI Web of Science citations database (ISI). The paper analyses journals using quantifiable Research Assessment Measures (RAMs), and highlights the similarities and differences in various RAMs which are based on alternative transformations of citations and influence. All existing RAMs to date have been static, so two new dynamic RAMs are developed to capture changes in impact factor over time and escalating journal self citations. Alternative RAMs may be calculated annually or updated daily to determine When, Where and How (frequently) published papers are cited. The RAMs are grouped in four distinct classes that include impact factor, mean citations and non-citations, journal policy, number of high quality papers, and journal influence and article influence. These classes include the most widely used RAMs, namely the classic 2-year impact factor including journal self citations (2YIF), 2-year impact factor excluding journal self citations (2YIF*), 5-year impact factor including journal self citations (5YIF), Eigenfactor (or Journal Influence), Article Influence, h-index, and PI-BETA (Papers Ignored - By Even The Authors). As all existing RAMs to date have been static, two new dynamic RAMs are developed to capture changes in impact factor over time ($5YD2 = 5YIF/2YIF$) and Escalating Self Citations. We highlight robust rankings based on the harmonic mean of the ranks of RAMs across the 4 classes. It is argued that emphasizing the 2-year impact factor of a journal, which partly answers the question as to When published papers are cited, to the exclusion of other informative RAMs, which answer Where and How (frequently) published papers are cited, can lead to a distorted evaluation of journal quality, impact and influence relative to the harmonic mean of the ranks.

Keywords: *Research assessment measures, Impact factor, IFI, C3PO, PI-BETA, STAR, Eigenfactor, Article Influence, h-index, 5YD2, ESC, harmonic mean of the ranks, economics, journal rankings*

1. INTRODUCTION

The perceived quality of academic journals is routinely based on untested expert assessments of journal impact and influence, the number of high quality papers, journal policy, and quantitative or qualitative information about a journal, as well as quantifiable bibliometric Research Assessment Measures (RAMs).

In this context, the leading database for generating RAMs to evaluate the research performance of individual researchers and the quality of academic journals is the Thomson Reuters ISI Web of Science (2011) database (hereafter ISI), where most RAMs are based on alternative transformations of citations and influence data. All existing RAMs to date have been static, so two new dynamic RAMs are developed to capture changes in impact factor over time and escalating journal self citations.

Although there are important caveats regarding the methodology and data collection methods underlying any database (see, for example, Seglen (1997) and Chang *et al.* (2011a, b, c, d) for caveats regarding ISI), the ISI citations database is the oldest and most prestigious source of RAMs, and undoubtedly the benchmark against which other general databases, such as SciVerse Scopus, Google Scholar and Microsoft Academic Search, social science open access repositories, such as the Social Science Research Network (SSRN), and discipline-specific databases, such as Research Papers in Economics (RePEc), are compared.

Journal publishers promote the ISI impact factor (see below) of their journals and, if their journals do not yet have an impact factor, publicize the fact that their journals have either been selected for coverage in ISI or have applied for inclusion in ISI.

Various RAMs have been used to compare journals in a wide range of ISI disciplines, such as the 40 leading journals in Economics and the leading 10 journals in each of Management, Finance and Marketing (Chang *et al.* (2011a)), the leading 6 journals in each of 20 disciplines in the Sciences (Chang *et al.* (2011b)), the leading journals in a sub-discipline of Economics, namely Econometrics, and Statistics (Chang *et al.* (2011c)), and the leading 26 journals in Neuroscience (Chang *et al.* (2011d)).

Not all of the leading journals in the various ISI disciplines have yet been analysed in terms of citations, quality and impact, so this would be a useful area of research in the future.

When impact factors and other RAMs-based citations data are used without appropriate care, misleading, unintended inferences may be drawn. Seglen (1997) cautioned against using impact factors of journals to evaluate scientific research. Nevertheless, as quantified metrics, citations are necessary for evaluating the impact and visibility of high quality and significant scientific research output.

Embracing journal citations as a valid measure of scientific research output, Hirsch (2005) suggested a widely-used measure, the h-index, for quantifying an individual researcher's scientific research output. Although citations data are used more widely as a measure of research productivity in the sciences than in the social sciences, the h-index is now widely used to evaluate both the research output of individual researchers and to quantify the number of highly-cited publications in academic journals in both the sciences and social sciences.

The perceived research performance of individual researchers is a key issue in hiring, tenure and promotion decisions. The perceived quality of academic journals has long been used as a suitable proxy for quality, especially for less established scholars, and especially in the social sciences, as leading journals tend to publish significant scientific research output. Evaluations of individual researchers, institutions and journals have been undertaken over an extended period (see, for example, Neary *et al.* (2003), and the other papers in the same issue of the *Journal of the European Economic Association*).

Emphasis on the appropriate weights attached to journals has been analysed, with the suggestions that it can be important, especially in ranking individuals, though not necessarily institutions. Kalaitzidakis *et al.* (2003) determine weights for the thirty leading research journals, while Axaroglou and Theoharakis (2003) analyse the survey results of journal quality perceptions of 2,103 AEA economists worldwide. The overall impression given by these interesting papers is that consensus is difficult to reach in terms of ranking the leading journals in economics. A similar comment would apply to any discipline that is included in ISI.

The convention in the sciences and social sciences is such that the acceptance of a paper for journal publication is based on the expertise of a few editors and referees. Although the number varies considerably across disciplines, acceptance of a paper for journal publication undeniably relies on a handful of decision makers, who determine the explicit rejection rate of a journal before publication. As editors and referees are not immune from making type 1 and type 2 errors regarding the latent quality and likely future impact of submitted papers, the rejection of a paper by a journal is not necessarily a correct decision, just as acceptance of a paper for publication is not a guarantee that it will have future impact and influence.

In comparison with the rejection rate of a journal before publication, there is an equally important implicit rejection rate after publication. Rather than relying on a small number of editors and reviewers, the rejection rate after publication relies on the worldwide scientific community. As argued in Chang *et al.* (2011c), the proportion of published papers that is ignored by the profession, and possibly by the authors themselves, is an important impact performance measure after publication.

The paper is also concerned with highlighting the upsurge in journal self citations in recent years. It would seem useful to present RAMs that capture such an escalation of journal self citations over time, and also to mitigate such an effect. One new dynamic RAM addresses the different speeds at which citations are accrued in the sciences and social sciences, and a second new dynamic RAM captures the escalation of journal self citations over time.

The RAMs may be classified according to four distinct classes, namely Class 1: “impact factor, mean citations and non-citations”, Class 2: “journal policy”, Class 3: “number of high quality papers”, and Class 4: “journal influence and article influence”. It is argued that emphasizing the 2-year impact factor of a journal to the exclusion of other informative RAMs can lead to a distorted evaluation of journal quality, impact and influence relative to the harmonic mean of the ranks of 13 existing and 2 new dynamic RAMs across the 4 classes. Together with the arithmetic and geometric means, the harmonic mean is one of the three Pythagorean means, and is defined as the reciprocal of the arithmetic mean of the reciprocals.

This paper examines the importance of RAMs as viable rankings criteria, and attempts to answer some important questions raised in Chang *et al.* (2011a, b, c), namely When, Where and How (frequently) are published papers cited in leading journals in a discipline. In this paper, we suggest a robust rankings method of alternative RAMs using the harmonic mean of the ranks. The rankings based on any single RAM, such as the h-index or the 2 year impact factors are placed in context, and may be seen as extremes since they are clearly subsumed by the harmonic mean of the ranks when all other RAMs are given zero weights, except the RAM in question.

The plan of the remainder of the paper is as follows. Section 2 presents some key RAMs using ISI data that may be calculated annually or updated daily, including the most widely used RAM, namely the classic 2-year impact factor including journal self citations (2YIF), 2-year impact factor excluding journal self citations (2YIF*), 5-year impact factor including journal self citations (5YIF), Immediacy (or zero-year impact factor (0YIF)), Eigenfactor (or Journal Influence), Article Influence, C3PO (Citation Performance Per Paper Online), h-index, PI-BETA (Papers Ignored - By Even The Authors), 2-year Self-citation Threshold Approval Ratings (2Y-STAR), Historical Self-citation Threshold Approval Ratings (H-STAR), Impact Factor Inflation (IFI), and Cited Article Influence (CAI). Two new dynamic RAMs are developed, namely 5YD2 (5YIF Divided by 2YIF) and ESC (Escalating Self Citations). Section 3 gives some practical suggestions as to how to rank journal quality and impact.

2. RESEARCH ASSESSMENT MEASURES (RAM)

A widely-used RAM database for evaluating journal impact and quality is the Thomson Reuters ISI Web of Science (2011). As discussed in a number of papers (for example, Chang *et al.* (2011a, b, c)), the RAMs are intended as descriptive statistics to capture journal impact and performance, and are not based on a mathematical model. Hence, in what follows, no optimization or estimation is required in calculating the alternative RAMs.

As the alternative RAMs that are provided in ISI and in several recent publications may not be widely known, this section provides a brief description and definition of 13 RAMs that may be calculated annually or updated daily to answer the questions as to When, and Where and How (frequently), published papers are cited (for further details, see Chang *et al.* (2011a, b, c)).

Two new dynamic RAMs that are calculated annually, namely 5YD2 and ESC, are also suggested. The answers to When published papers are cited are based on the set {2YIF, 2YIF*, 5YIF, Immediacy}, and the answers to Where and How (frequently) published papers are cited are based on the set {Eigenfactor, Article Influence, IFI, 5YD2, H-STAR, 2Y-STAR, ESC, C3PO, h-index, PI-BETA, CAI}, as will be discussed below.

2.1 Annual RAM

With three exceptions, namely Eigenfactor, Article Influence and Cited Article Influence, existing RAMs are based on citations data and are reported separately for the sciences and social sciences. RAMs may be computed annually or updated daily. The annual RAMs given below are calculated for a Journal Citations Reports (JCR) calendar year, which is the year before the annual RAM are released. For example, the RAMs were released in late-June 2011 for the JCR calendar year 2010.

(1) 2-year impact factor including journal self citations (2YIF):

The classic 2-year impact factor including journal self citations (2YIF) of a journal is typically referred to as “the impact factor”, is calculated annually, and is defined as “Total citations in a year to papers published in a journal in the previous 2 years / Total papers published in a journal in the previous 2 years”. The choice of 2 years by ISI is arbitrary. It is widely held in the academic community, and certainly by the editors and publishers of journals, that a higher 2YIF is better than lower.

(2) 2-year impact factor excluding journal self citations (2YIF*):

ISI also reports a 2-year impact factor without journal self citations (that is, citations to a journal in which a citing paper is published), which is calculated annually. As this impact factor is not widely known or used, Chang *et al.* (2011c) refer to this RAM as 2YIF*. Although 2YIF* is rarely reported, a higher value would be preferred to lower.

(3) 5-year impact factor including journal self citations (5YIF):

The 5-year impact factor including journal self citations (5YIF) of a journal is calculated annually, and is defined as “Total citations in a year to papers published in a journal in the previous 5 years / Total papers published in a journal in the previous 5 years.” The choice of 5 years by ISI is arbitrary. 5YIF excluding journal self citations is not presently available. Although 5YIF is not widely reported, a higher value would be preferred to lower.

(4) Immediacy, or zero-year impact factor including journal self citations (0YIF):

Immediacy is a zero-year impact factor including journal self citations (0YIF) of a journal, is calculated annually, and is defined as “Total citations to papers published in a journal in the same year / Total papers published in a journal in the same year.” The choice of the same year by ISI is arbitrary, but the nature of Immediacy makes it clear that a very short run outcome is under consideration. Immediacy excluding journal self citations is not presently available. Although Immediacy is rarely reported, a higher value would be preferred to lower.

(5) 5YIF Divided by 2YIF (5YD2):

As both 2YIF and 5YIF include journal self citations, if it is assumed that journal self citations are uniformly distributed over the 5-year period for calculating 5YIF, their ratio will eliminate the effect of journal self citations and capture the increase in the citation rate over time. In any event, the impact of journal self citations should be mitigated with the ratio of 5YIF to 2YIF. We define a new dynamic RAM as:

5YD2 as “ $5YD2 = 5YIF / 2YIF$ ”.

In the natural, physical and medical sciences, where citations are observed with a frequency of weeks and months rather than years, it is typically the case that $5YIF < 2YIF$ (see Chang *et al.* (2011b, d)), whereas the reverse, $5YIF > 2YIF$, seems to hold generally in the social sciences, where citations tend to increase gradually over time (see Chang *et al.* (2011a, c)). Thus, emphasizing the different speeds at which citations

are accrued over time, a lower 5YD2 would be preferred to higher in the sciences, while a higher 5YD2 would be preferred to lower in the social sciences.

(6) Eigenfactor (or Journal Influence):

The Eigenfactor score (see Bergstrom (2007), Bergstrom and West (2008), Bergstrom, West and Wiseman (2008)) is calculated annually (see www.eigenfactor.org), and is defined as: “The Eigenfactor Score calculation is based on the number of times articles from the journal published in the past five years have been cited in the JCR year, but it also considers which journals have contributed these citations so that highly cited journals will influence the network more than lesser cited journals. References from one article in a journal to another article from the same journal are removed, so that Eigenfactor Scores are not influenced by journal self-citation.” The value of the threshold that separates ‘highly cited’ from ‘lesser cited’ journals, as well as how the former might ‘influence the network more’ than the latter, are based on the Eigenfactor score of the citing journal. Thus, Eigenfactor might usefully be interpreted as a weighted total citations score, or a “Journal Influence” measure. A higher Eigenfactor score would be preferred to lower.

(7) Article Influence (or Journal Influence per Article):

Article Influence (see Bergstrom (2007), Bergstrom and West (2008), Bergstrom, West and Wiseman (2008)) measures the relative importance of a journal’s citation influence on a per-article basis. Despite the misleading suggestion of measuring “Article Influence”, as each journal has only a single “Article Influence” score, this RAM is actually a “Journal Influence per Article” score. Article Influence is a scaled Eigenfactor score, is calculated annually, is standardized to have a mean of one across all journals in the Thomson Reuters ISI database, and is defined as “Eigenfactor score divided by the fraction of all articles published by a journal.” A higher Article Influence would be preferred to lower.

(8) IFI:

The ratio of 2YIF to 2YIF* is intended to capture how journal self citations can inflate the impact factor of a journal, whether this is an unconscious self-promotion decision made independently by publishing authors or as an administrative decision undertaken by a journal’s editors and/or publishers. Chang *et al.* (2011a) define Impact Factor Inflation (IFI) as “ $IFI = 2YIF / 2YIF^*$ ”. The minimum value for IFI is 1, with any value above the minimum capturing the effect of journal self citations on the 2-year impact factor. A lower IFI would be preferred to higher.

(9) H-STAR:

ISI has implicitly recognized the inflation in journal self citations by calculating an impact factor that excludes self citations, and provides data on journal self citations, both historically (for the life of the journal) and for the preceding two years, in calculating 2YIF. Chang *et al.* (2011b) define the Self-citation Threshold Approval Rating (STAR) as the percentage difference between citations in other journals and journal self citations. If HS = historical journal self citations, then Historical STAR (H-STAR) is defined as “ $H-STAR = [(100-HS) - HS] = (100-2HS)$ ”. If HS = 0 (minimum), 50 or 100 (maximum) percent, for example, H-STAR = 100, 0 and -100, respectively. A higher H-STAR would be preferred to lower.

(10) 2Y-STAR:

If 2YS = journal self citations over the preceding 2-year period, then the 2-Year STAR is defined as “ $2Y-STAR = [(100-2YS) - 2YS] = (100-2(2YS))$ ”. If 2YS = 0 (minimum), 50 or 100 (maximum) percent, for example, 2Y-STAR = 100, 0 and -100, respectively. A higher 2Y-STAR would be preferred to lower.

(11) Escalating Self Citations (ESC):

As self citations for many journals in the sciences and social sciences have been increasing over time, it would seem useful to present a dynamic RAM that captures such an escalation over time. The difference 2YS – HS measures Escalating Self Citations in journals over the most recent 2 years relative to the historical period for calculating citations, which will differ across journals. We define a new dynamic RAM as:

$$\text{“ESC} = 2YS - HS = (H-STAR - 2Y-STAR) / 2\text{”}.$$

Given the range of each of H-STAR and 2Y-STAR is (-100, 100), the range of ESC is also (-100, 100), with -100 denoting minimum, and 100 denoting maximum, escalation. A lower ESC would be preferred to higher.

2.2 Daily Updated RAM

Some RAMs are updated daily, and are reported for a given day in a calendar year rather than for a JCR year.

(12) C3PO:

ISI reports the mean number of citations for a journal, namely total citations up to a given day divided by the number of papers published in a journal up to the same day, as the “average” number of citations. In order to distinguish the mean from the median and mode, the C3PO of an ISI journal on any given day is defined by Chang *et al.* (2011a) as “C3PO (Citation Performance Per Paper Online) = Total citations to a journal / Total papers published in a journal.” A higher C3PO would be preferred to lower. [Note: C3PO should not be confused with C-3PO, the Star Wars android.]

(13) h-index:

The h-index (Hirsch, 2005) was originally proposed to assess the scientific research productivity and citations impact of individual researchers. However, the h-index can also be calculated for journals, and should be interpreted as assessing the impact or influence of highly cited journal publications. The h-index of a journal on any given day is based on historically cited and citing papers, including journal self citations, and is defined as “h-index = number of published papers, where each has at least h citations.” The h-index differs from an impact factor in that the h-index measures the number of highly cited papers historically. A higher h-index would be preferred to lower.

(14) PI-BETA:

This RAM measures the proportion of papers in a journal that has never been cited, As such, PI-BETA is, in effect, a rejection rate of a journal after publication. Chang *et al.* (2011c) argue that lack of citations of a published paper, especially if it is not a recent publication, reflects on the quality of a journal by exposing: (i) what might be considered as incorrect decisions by the members of the editorial board of a journal; and (ii) the lost opportunities of papers that might have been cited had they not been rejected by the journal. Chang *et al.* (2011c) propose that a paper with zero citations in ISI journals can be measured by PI-BETA (= Papers Ignored (PI) - By Even The Authors (BETA)), which is calculated for an ISI journal on any given day as “Number of papers with zero citations in a journal / Total papers published in a journal.” As journals would typically prefer a higher proportion of published papers being cited rather than ignored, a lower PI-BETA would be preferred to higher.

(15) CAI:

Article Influence is intended to measure the average influence of an article across the sciences and social sciences. As an article with zero citations typically does not have any (academic) influence, a more suitable measure of the influence of cited articles would seem to be Cited Article Influence (CAI). Chang *et al.* (2011b) define CAI as “CAI = (1 - PI-BETA)(Article Influence)”. If PI-BETA = 0, then CAI is equivalent to Article Influence; if PI-BETA = 1, then CAI = 0. As Article Influence is calculated annually and PI-BETA is updated daily, CAI may be updated daily. A higher CAI would be preferred to lower.

3. A ROBUST RANKING OF JOURNALS

As no single RAM captures adequately the quality, impact and influence of a journal, any general measure of journal quality and impact, such as a harmonic mean of the ranks as a robust rankings method of alternative RAMs (see, for example, Chang and McAleer (2013)), should depend on the following four distinct classes:

(i) **Class 1:** “impact factor, mean citations and non-citations” (2YIF, 2YIF*, 5YIF, Immediacy, C3PO, PI-BETA);

(ii) **Class 2:** “journal policy” (IFI, H-STAR, 2Y-STAR, 5YD2, ESC);

(iii) **Class 3:** “number of high quality papers” (h-index);

(iv) **Class 4:** “journal influence and article influence” (Eigenfactor, Article Influence, CAI).

As each of the four classes has equal weight in the calculation of the harmonic mean of the ranks, the h-index has the single highest weight of the 15 RAMs. For journals that have been included in ISI for less than five years, Class 1 does not include 5YIF, Class 2 does not include 5YD2, and Class 4 does not include Article Influence and CAI, in calculating the harmonic mean of the ranks of the RAMs. Class 3 includes only the h-index. When RAM data for only Eigenfactor are available, Class 4 would be a “journal influence” rather than “journal influence and article influence” class.

ACKNOWLEDGMENTS

For financial support, the first author wishes to thank the National Science Council, Taiwan, and the third author wishes to acknowledge the Australian Research Council and National Science Council, Taiwan.

REFERENCES

- Axaroglou, K. and V. Theoharakis (2003), Diversity in economics: An analysis of journal quality perceptions, *Journal of the European Economic Association*, 1(6), 1402-1423.
- Bergstrom C. (2007), Eigenfactor: Measuring the value and prestige of scholarly journals, *C&RL News*, 68, 314-316.
- Bergstrom, C.T. and J.D. West (2008), Assessing citations with the Eigenfactor™ metrics, *Neurology*, 71, 1850–1851.
- Bergstrom, C.T., J.D. West and M.A. Wiseman (2008), The Eigenfactor™ metrics, *Journal of Neuroscience*, 28(45), 11433–11434 (November 5, 2008).
- Chang, C.-L. and M. McAleer (2013), Ranking journal quality by harmonic mean of ranks: An application to ISI Statistics & Probability, *Statistica Neerlandica*, 67(1), 27-53.
- Chang, C.-L., M. McAleer and L. Oxley (2011a), What makes a great journal great in economics? The singer not the song, *Journal of Economic Surveys*, 25(2), 326-361.
- Chang, C.-L., M. McAleer and L. Oxley (2011b), What makes a great journal great in the sciences? Which came first, the chicken or the egg?, *Scientometrics*, 87(1), 17-40.
- Chang, C.-L., M. McAleer and L. Oxley (2011c), Great expectatrics: Great papers, great journals, great econometrics, *Econometric Reviews*, 30(6), 583-619.
- Chang, C.-L., M. McAleer and L. Oxley (2011d), How are journal impact, prestige and article influence related? An application to neuroscience, *Journal of Applied Statistics*, 38(11), 2563-2573.
- 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), 16569-15572 (November 15, 2005).
- ISI Web of Science (2011), *Journal Citation Reports, Essential Science Indicators*, Thomson Reuters ISI.
- Kalaitzidakis, P., T.P. Mamuneas and T. Stengos (2003), Rankings of academic journals and institutions in economics, *Journal of the European Economic Association*, 1(6), 1346-1366.
- Neary, J.P., J.A. Mirrlees and J. Tirole (2003), Evaluating economics research in Europe: An introduction, *Journal of the European Economic Association*, 1(6), 1239-1249.
- Seglen, P.O. (1997), Why the impact factor of journals should not be used for evaluating research, *BMJ: British Medical Journal*, 314(7079), 498-502.