KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Business and Economics

# Apriori vs. a posteriori normalisation of citation indicators. The case of journal ranking

Wolfgang Glänzel, András Schubert, Bart Thijs and Koenraad Debackere

DEPARTMENT OF MANAGERIAL ECONOMICS, STRATEGY AND INNOVATION (MSI)

OR 1014

# Apriori vs. a posteriori normalisation of citation indicators. The case of journal ranking

WOLFGANG GLÄNZEL[1,2], ANDRÁS SCHUBERT[2], BART THIJS[1], KOENRAAD DEBACKERE[1]

[1]*Katholieke Universiteit Leuven, Centre for R&D Monitoring and Dept. MSI, Leuven (Belgium)*

[2]*Hungarian Academy of Sciences, Institute for Research Policy Studies, Budapest (Hungary)*

## ABSTRACT

Two paradigmatic approaches to the normalisation of citation-impact measures are discussed. The results of the mathematical manipulation of standard indicators such as citation means, notably journal Impact Factors, (called aposteriori normalisation) are compared with citation measures obtained from fractional citation counting (called apriori normalisation). The distributions of two subfields of the life sciences and mathematics are chosen for the analysis. It is shown that both methods provide indicators that are useful tools for the comparative assessment of journal citation impact.

## 1. INTRODUCTION

A common problem in comparative bibliometrics at the meso and micro level is the differentiation and specialisation of subject profiles of the objects of analysis. Typical tasks are the evaluation and ranking of research institutes and journals, which requires the application of more sophisticated techniques than customary at the macro level. The necessity and possibility of proper standardisation or cross-field normalisation of scientometric indicators for evaluative purposes has been studied by Schubert and Braun (1996). Other papers on this issue have shown that subfield-specific normalised citation indicators are useful tools for the comparative assessment of objects with deviating profile (Glänzel et al., 2009).

Most notably, the ranking of scientific journals according to their citation impact proved an everlasting story. It is not only the most commonly used ISI Impact Factor as such that is in dispute; also the general question of how journal performance and significance could best be depicted has provoked broad discussions and encouraged a plethora of methodological approaches to solve controversial issues in the context of journal impact measures (e.g., Pinski and Narin, 1976, Lindsey, 1978, Asai, 1981, Tomer, 1986). Some solutions aimed at improving the impact factor (e.g., Asai) others suggested completely different approaches like the weighted journal influence suggested by Pinski and Narin. However, one of the main points of criticism lies in the Impact Factors' subject bias. The page-rank based approach by Gonzalez-Pereira et al. (2010) is linked to the earlier suggestions by Pinski and Narin (1976). In particular, the *SCImago Journal Rank* indicator (SJR) weights citations received by journals by the eminence of the citing journals. *Eigenfactor Score* and *Article Influence* by Thomson Reuters are similar measures, which estimate the relative influence of cited items (cf., Bergstrom et el, 2008). While the first measure expresses the overall prestige of a given journal, the latter indicator provides the per-article citation influence of the journal. Nevertheless, these alternatives cannot solve the subject bias caused by the specific subject and document type. Review journals and life sciences thus remain still on top of the ranking lists.

The idea of subject normalisation for journals has recently taken up again by Beirlant et al. (2007) and Glänzel (2010). The idea was to transform the distribution of Impact Factors over journals within given disciplines so that standard distributions are obtained. Since these approaches are manipulations of the original measures, we can call these methods *aposteriori normalisation*. Other forms of aposteriori normalisation, applied in a broader context, are journal- or subject-normalised relative indicators such as RCR, NMCR, CPP/JSCm and CPP/FSCm used by (Schubert and Braun, 1986, Braun and Glänzel, 1991, Moed et al., 1995). This type of normalisation is also called cited-side normalisation (cf. Waltman and van Eck, 2010).

More recently, Zitt and Small (2008), Zitt (2010), Moed (2010) and Leydesdorff and Opthof (2010) have chosen a different way: Citations to journals are immediately normalised before the indicators are build. These are referred to as citing-side or source normalisation (e.g., Zitt, 2010, Moed, 2010). We will call this solution *apriori normalisation*. In the following we will compare the two specific normalisation methods for journal impact measure, one each apriori and aposteriori. The pros and cons of the two approaches will be discussed.

## 2. THE APOSTERIORI METHOD – NORMALISED IMPACT-FACTOR DISTRIBUTIONS

In an earlier study, Beirlant et al. (2007) have transformed the distribution of journals over impact factors to the standard-normal distribution. In a recent paper, Glänzel (2010) has chosen another way; threshold values of *Characteristic Scores and Scales* (CSS; cf. Glänzel and Schubert, 1988) were used to re-scale the distribution of impact measures over journals. CSS are obtained from a recursive procedure of iteratively truncating a sample according to mean values from the low-end up to the high-end. In particular, samples are iteratively truncated at their mean value $m$, then the mean of the truncated sample is recalculated until the procedure is stopped or no new scores are obtained. The scores are denoted by $\beta_i$, where $\beta_0 = 0$ and $\beta_1 = m$ by definition. $\beta_3$ is ordinarily used to identify top journals. This procedure can preferably be applied to ISI Subject Categories or sufficiently narrow subfields. The transformation

$$u^* = \frac{x}{\beta_2 - \beta_1} ,$$

where $x$ is the journal impact factor and $\beta_1$ and $\beta_2$ the first and second characteristic score of the underlying citation distribution, results in a robust normalisation of journal impact (cf. Glänzel, 2010). This method aimed at the identification of top. Since their modified impact measures become comparable though normalisation, top journals can be compared across different fields. Table 1 illustrates this effect for three selected fields biochemistry/biophysics/molecular biology (B1) and applied mathematics (H1). Here papers are published in 2006, the citation window comprises the three-year period 2006–2008. The threshold $\beta_3$ usually identifies 3% to 5% of all journals of the discipline as top journals. In the case of the two selected fields, 11 top journals were found in the life-science field B1 ($n$ = 333) and 14 in mathematics H1 ($n$ = 380). Also the mean value of the distributions changes considerably. While 6.90 for B1 and 1.82 for H1 1.59 clearly reflect the usual subject-specific peculiarities, the same distribution means take almost subject-invariant values after transformation ($\beta 1^*$ = 1.07 for B1 and 1.41 for H1). The normalisation effect is even more dramatic if the 3rd score is considered (22.69 and 4.47 for $\beta_3$ vs. 3.53 and 3.46 for $\beta_3^*$). Figure 1 and 2 show, in addition, how this transformation smoothes the large deviations at the high end of the distributions.

TABLE 1. CSS values ($\beta_k$) according to the distribution of journal impact measures and their normalised versions ($\beta_k^*$) for two subfields (citation window: 2006-2008).

| $k$ | $\beta_k$ | | $\beta_k^*$ | |
|---|---|---|---|---|
| | **B1** | **H1** | **B1** | **H1** |
| 0 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1 | 6.90 | 1.82 | 1.07 | 1.41 |
| 2 | 13.33 | 3.11 | 2.07 | 2.41 |
| 3 | 22.69 | 4.47 | 3.53 | 3.46 |

The advantage of this method is it robustness: Although the distributions still differ at the lower end, their modes are almost "synchronised" and the deviation between the probabilities around the mode has distinctly decreased. The transformation to the standard form of the distribution would require the knowledge or estimation of all parameters of the distribution; obtaining this standard form can thus hardly be expected from using some robust parameter-free statistics. However, the parameter-free $u^*$ transformation provides a proxy even for different underlying distribution models.

TABLE 2. 11 top journals (CSS class 3) in the subfield B1 based on 3-year diachronous impact factors.

| $x$ | $u^*$ | **Journal title** |
|---|---|---|
| 61.04 | 9.49 | Cell |
| 57.13 | 8.88 | Annual Review of Biochemistry |
| 42.16 | 6.55 | Annual Review of Biophysics and Biomolecular Structure |
| 41.69 | 6.48 | Nature Medicine |
| 27.51 | 4.28 | Molecular Cell |
| 27.27 | 4.24 | Progress in Lipid Research |
| 25.86 | 4.02 | Nature Chemical Biology |
| 25.04 | 3.89 | Nature Methods |
| 24.50 | 3.81 | PLoS Biology |
| 23.43 | 3.64 | Genome Research |
| 22.82 | 3.55 | Nature Structural & Molecular Biology |

TABLE 3. 14 top journals (CSS class 3) in the subfield H1 based on 3-year diachronous impact factors.

| $x$ | $u^*$ | **Journal title** |
|---|---|---|
| 13.81 | 10.70 | PLoS Computational Biology |
| 9.63 | 7.47 | International Journal of Nonlinear Sciences and Numerical Simulation |
| 8.18 | 6.34 | Bioinformatics |
| 6.86 | 5.32 | MATCH-Communications in Mathematical and in Computer Chemistry |
| 6.85 | 5.31 | IEEE-ACM Transactions on Computational Biology and Bioinformatics |
| 6.45 | 5.00 | CMES-Computer Modeling in Engineering & Sciences |
| 6.14 | 4.76 | Communications on Pure and Applied Mathematics |
| 6.04 | 4.68 | Econometrica |
| 5.13 | 3.97 | Foundations of Computational Mathematics |
| 4.82 | 3.74 | Biostatistics |
| 4.76 | 3.69 | Statistical Science |
| 4.67 | 3.62 | Chaos Solitons & Fractals |
| 4.67 | 3.62 | Production and Operations Management |
| 4.66 | 3.61 | Archive for Rational Mechanics and Analysis |

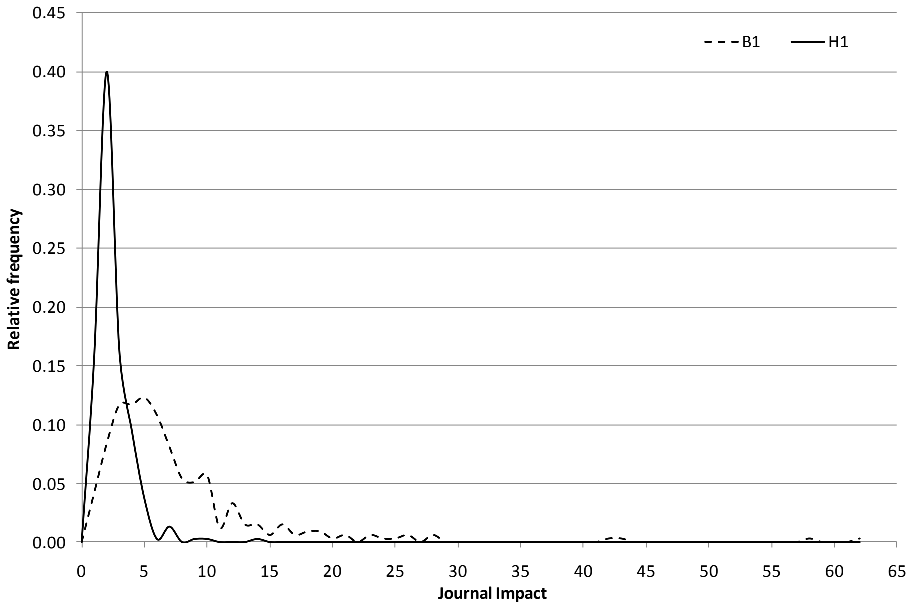


FIGURE 1. Distribution of mean citation rate over journals based on the three-year citation window 2006–2008.

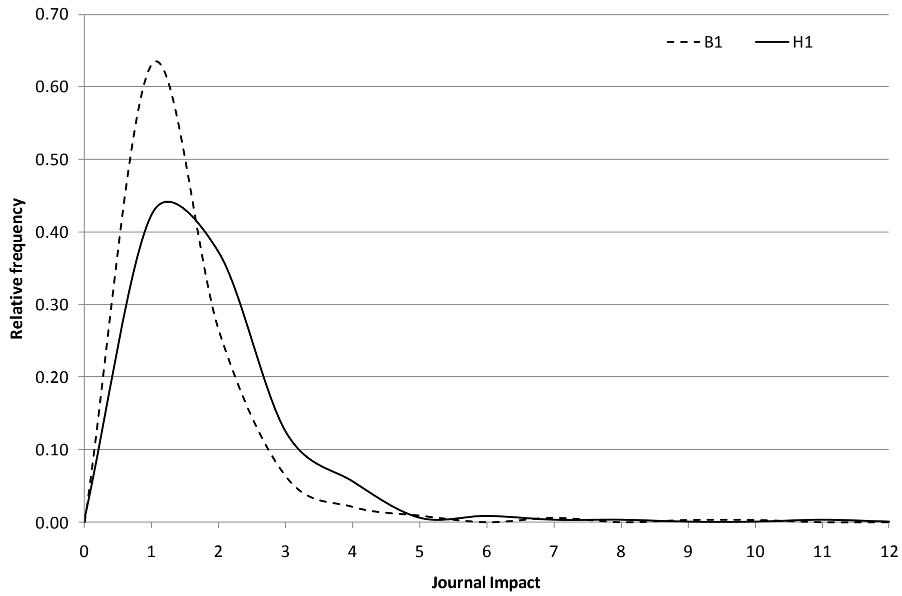


FIGURE 2. Distribution of mean citation rate over journals based on the three-year citation window 2006–2008 after the $u^*$ normalisation.

A further advantage of the aposteriori method is that it might be applied to any journal-based subjects defined according to the corresponding needs. At the same time, it immediately provides sets of top journals as determined, e.g., by the $3^{rd}$ characteristic score.

By contrast, there are two disadvantages. On one hand, this method does not take into account the deviating citation patterns of review journals. On the other hand, the transformation results in different impact measures for one and the same journal if this journal is assigned to different subject categories.

## 3. THE APRIORI METHOD – FRACTIONAL IMPACT-FACTORS

Recently two interesting apriori solutions have been suggested. Henk Moed (2010) introduced his SNIP indicator which is already used as a standard journal measure in Elsevier's SCOPUS database. According to its definitions, SNIP takes into account both the frequency of other papers in the reference lists of citing papers and the coverage of the corresponding subject field in the database. Leydesdorff (2010) suggested a modified journal impact factor which is calculated from fractional citation counts. In the following, we suggest a similar solution, which is, by contrast, based on fractional citation counts using references from *indexed source items* at the level of individual papers, and which thus results in a consistent measure with respect to the total database and to any partition of disjoint subsets of the database. This means if the journal impact measures are summed up over all journals in the database and all papers published in these journals and the result is divided by all papers indexed in the database, the corresponding value of the complete database should be obtained. This certainly holds for the Impact Factor if citation counts are determined on a paper-to-paper basis. For instance, if a journal impact measure is defined on one publication year and a three-year citation window beginning with the publication year, and is calculated from individual citations of papers (as used in the previous section) then the grand total over all papers results in the number of all citations in the three citation years in question and the ratio of the grand total and the number of papers in the first year provide the impact measure of the database. Now we proceed as follows. Again, one publication year and a three-year citation window is chosen. Citations are fractionated for individual reference-source pairs. If a paper A published in year $y$ is cited by paper B say published in year $y+2$ and paper B has $k$ references to papers *indexed in volume y of the same database*, then the corresponding fractional 'citation value' amounts to $1/k$. The case $k=0$, of course, cannot occur once A is cited by B. The resulting fractional journal impact measures are consistent metrics in the above-mentioned sense: The grand total does not provide the total of citations but the total of *citing papers*. Although fractional citation counts are rational numbers, the grand total is an integer again. This value divided by the number of papers indexed in the volume $y$ of the database gives the ratio of citing paper in the period $y$–$y$+2 and the publications in $y$. This is the fractional citation mean of the complete database. This also implies that this notion of consistency also holds for all subsets of the database defined on journal assignment, provided journals are fully covered by the subset. Based on the definition we expect that this metrics compensates for the "surplus" of citations received from both review articles which have by nature long reference lists and the "hard sciences" where most of the references are most recently published journal articles (cf. Price, 1970).

The results for the twelve major science fields according to the Leuven classification scheme (Glänzel et al., 2003) are presented in Table 4. These fields include A = Agriculture & Environment, Z = Biology, B = Biosciences, R = Biomedical research, I = Clinical & Experimental Medicine I (General & Internal Medicine), M = Clinical & Experimental Medicine II (Non-Internal Medicine Specialties), N = Neuroscience & Behaviour, C = Chemistry, P = Physics, G = Geosciences & Space Sciences, E = Engineering and H = Mathematics. The following notations are used: MOCR denotes the Observed Mean Citation Rates based on integer counts, $\mathrm{MOCR}|_{+}$ is the same ratio but calculated for cited papers only (a conditional mean value), $\mathrm{MOCR}|_{F}$ and $\mathrm{MOCR}|_{F+}$ are the corresponding metrics based on fractional counts and $f_0$ is the percentage of uncited papers. We have obviously $\mathrm{MOCR}|_{+} = \mathrm{MOCR}/(1-f_0)$ and $\mathrm{MOCR}|_{F+} = \mathrm{MOCR}|_{F}/(1-f_0)$, respectively.

TABLE 4. Mean citation impact of science fields based on integer and fractional citation counts.

| Field | MOCR | $\mathrm{MOCR}\|_{+}$ | $\mathrm{MOCR}\|_{F}$ | $\mathrm{MOCR}\|_{F+}$ | $f_0$ |
|---|---|---|---|---|---|
| **A** | 3.18 | 4.24 | 1.26 | 1.68 | 25.0% |
| **Z** | 4.60 | 5.84 | 1.51 | 1.92 | 21.3% |
| **B** | 7.93 | 8.91 | 2.05 | 2.31 | 11.0% |
| **R** | 5.55 | 6.66 | 1.61 | 1.93 | 16.7% |
| **I** | 7.18 | 8.93 | 2.03 | 2.52 | 19.6% |
| **M** | 4.28 | 5.69 | 1.47 | 1.95 | 24.9% |
| **N** | 5.68 | 6.88 | 1.74 | 2.10 | 17.5% |
| **C** | 4.35 | 5.82 | 1.42 | 1.89 | 25.3% |
| **P** | 3.90 | 5.30 | 1.47 | 2.00 | 26.4% |
| **G** | 4.57 | 6.09 | 1.53 | 2.04 | 25.0% |
| **E** | 1.71 | 3.40 | 0.81 | 1.60 | 49.7% |
| **H** | 1.85 | 3.21 | 0.98 | 1.69 | 42.4% |

The gap between the MOCR values of biosciences, on one hand, and mathematics and engineering, on the other hand, is well known. The same applies to the share of uncited papers, which is high in mathematics and applied sciences and considerably low in the life sciences (cf. Table 4). Fractional counting somewhat decreases subject-specific differences but the results fall short of the expectation. There is one straightforward explanation, namely fractionation affects only citation counts of *cited items*, i.e., $c/k$ is independent of the number of references $k$ if $c$=0. Therefore we have to split up the fractional indicator into two statistics based on *disjoint* subsets, the conditional mean and the relative frequency of uncited papers. This is also in line with earlier observations concerning the *multi-dimensionality* of journal impact (Glänzel, 2009). Splitting citation impact into an indicator pair does not only reduce subject-specific biases of one of its components as the $\mathrm{MOCR}|_{F+}$ values are almost subject-independent (cf. Table 4), but also paves the way for the extension of $\mathrm{MOCR}|_{F+}$ to non-source items while $f_0$ has to be skipped in these cases. At CWTS such procedure is called extension of citation analysis to 'non-source' items (cf. Butler and Visser, 2006).

Figure 3 presents the distribution of the conditional mean citation rate ($\mathrm{MOCR}|_{F+}$) over journals. Although the modes of the two distributions are almost "synchronised" and the mean value of the two distributions (2.04 for B1 and 1.57 for H1, respectively) do not differ dramatically, the distribution plot in Figure 3 pronouncedly deviates from the corresponding chart in Figure 2. The considerably higher impact of the top journals in the life-science field with respect to applied mathematics is quite striking.

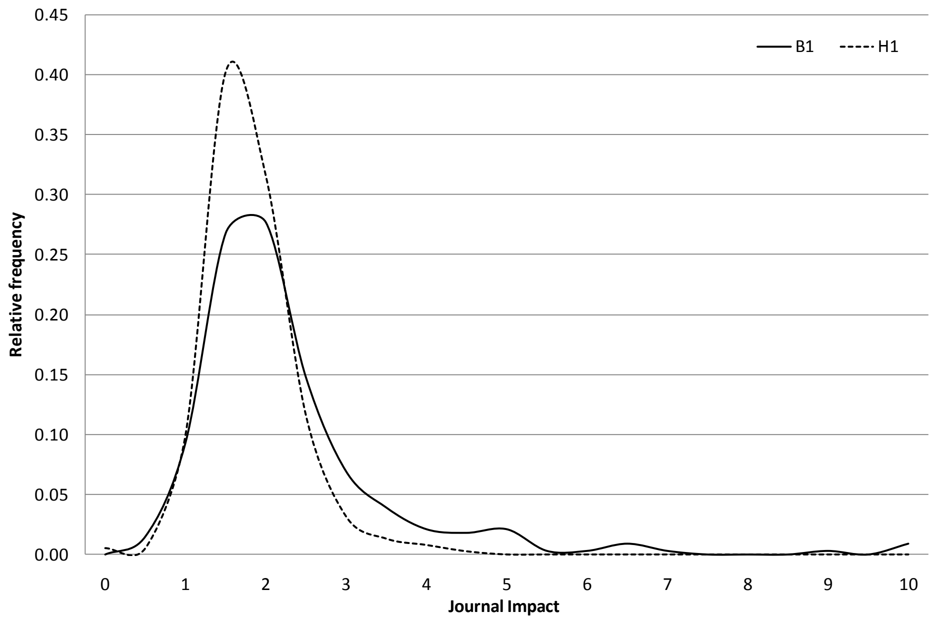


FIGURE 3. Distribution of the conditional mean citation rate ($\mathrm{MOCR}|_{F+}$) over journals based on the three-year citation window 2006–2008 and fractional citation counts.

TABLE 5. 11 top journals in the subfield B1 based on 3-year fractional impact factors according to $MOCR|_{F+}$.

| $MOCR\|_{F+}$ | $f_o$ | Journal title |
|---|---|---|
| 13.04 | 3.3% | Annual Review of Biochemistry |
| 12.48 | 2.4% | Cell |
| 10.03 | 8.5% | Nature Medicine |
| 8.92 | 0.0% | Annual Review of Biophysics and Biomolecular Structure |
| 6.63 | 0.0% | Progress in Lipid Research |
| 6.32 | 1.4% | Nature Chemical Biology |
| 6.25 | 8.8% | Nature Methods |
| 6.04 | 0.6% | PLOS Biology |
| 5.92 | 0.0% | Molecular Cell |
| 5.04 | 0.0% | Natural Product Reports |
| 4.97 | 0.0% | Trends in Biochemical Sciences |

TABLE 6. 14 top journals in the subfield H1 based on 3-year fractional impact factors according to $MOCR|_{F+}$.

| $MOCR\|_{F+}$ | $f_o$ | Journal title |
|---|---|---|
| 4.05 | 5.7% | Econometrica |
| 3.93 | 25.0% | Journal of Educational and Behavioral Statistics |
| 3.75 | 3.7% | PLOS Computational Biology |
| 3.58 | 4.9% | IEEE-ACM Transactions on Computational Biology and Bioinformatics |
| 3.32 | 19.2% | MATCH-Communications in Mathematical and in Computer Chemistry |
| 3.21 | 35.4% | Journal of Applied Econometrics |
| 3.21 | 40.9% | SIAM Review |
| 3.18 | 11.6% | Communications on Pure and Applied Mathematics |
| 3.09 | 12.5% | Foundations of Computational Mathematics |
| 2.97 | 16.3% | Annals of Statistics |
| 2.92 | 5.4% | Archive for Rational Mechanics and Analysis |
| 2.90 | 14.6% | Journal of the Royal Statistical Society Series B-Statistical |
| 2.86 | 20.7% | Review of Economics and Statistics |
| 2.75 | 14.3% | Statistical Science |

The deviation of top journals shown in Figure 3 is reflected by Tables 5 and 6 as well. In both tables, where journals are ranked according to $MOCR|_{F+}$, the same top 11 and 14 journals are displayed as in the corresponding tables based on aposteriori normalisation in the previous section. While the changes in the top of B1 journals are not dramatic, the comparison of Tables 3 and 6 reveals structural changes. The journals *International Journal of Nonlinear Sciences and Numerical Simulation (*rank 2 vs. 17), *Bioinformatics* (rank 3 vs. 26) and *Biostatistics (*rank 10 vs. 77) disappear from the top list when apriori normalisation is applied. The fact that $\mathrm{MOCR}|_{F+}$ is only one component of the citation metrics is at least in part responsible for this effect ($f_0$ = 6.7% for *Bioinformatics* and $f_0$ = 9.8% for *Biostatistics*). Both shares lie much below the standard in mathematics which amounts to about 42%. This provides one more argument fort the application of multi-dimensional impact measures.

## 4. CONCLUSIONS

Citation-impact metrics can be normalised based on two paradigmatic models, prior to building the indicator (apriori) and afterwards (aposteriori). Two examples – one each for a field with high and low citation standard, respectively – have illustrated that, if normalisation is done in a consistent way, the difference between the two methods is not large. Nevertheless, apriori normalisation is versatile and flexible as individual links can be weighted according to many aspects. The method used in this study was based on a simple frequency weight. The method reduces the weight of a citation if the cited work is just "one of many", e.g., a citation received from a review or bibliography. Another advantage is that the fractional citation indicator becomes less sensitive to the citation window since citations are losing weight as the overall number of citation grows in time and the decreasing share of uncited paper is absorbed by the complementary indicator. On the other hand, aposteriori proved robust and can provide standard distributions for cross-disciplinary assessment, but are inferior to apriori methods if complex aspects of normalisation are concerned. The specific task and the availability of data can help decide which type of normalisation should be preferred.

## REFERENCES

Asai, I. (1981), Adjusted Age Distribution and Its Application to Impact Factor and Immediacy Index. *Journal of the American Society for Information Science*, 32, 172–174.

Beirlant, J., Glänzel, W., Carbonez, A., Leemans, H. (2007), Scoring research output using statistical quantile plotting. *Journal of Informetrics*, 1 (3), 185–192.

Bergstrom, C.T., West, J.D., Wiseman, M.A. (2008), The Eigenfactor (TM) Metrics. Journal of Neuroscience, 28 (45), 11433–11434.

Butler, L., Visser, M.S. (2006), Extending citation analysis to non-source items. *Scientometrics*, 66 (2), 327–343.

Glänzel, W., Schubert, A. (1988), Characteristic scores and scales in assessing citation impact. *Journal of Information Science,* 14 (2), 123–127.

Glänzel, W., Schubert A. (2003), A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56 (3), 357–367.

Glänzel, W., Schubert, A., Thijs, B., Debackere, K. (2009), Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics*, 78 (1), 165–188.

Glänzel, W. (2009), The Multi-Dimensionality of Journal Impact. *Scientometrics*, 78 (2), 355–374.

Glänzel, W. (2010), The application of Characteristic Scores and Scales to the evaluation and ranking of scientific journals. *Journal of Information Science*, to be published.

Gonzalez-Pereira, B., Guerrero-Bote, V.P., Moya-Anegon F. (2010), A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics,* 4 (3), 379–391.

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,* (forthcoming; DOI: 10.1002/asi.21371)

Lindsey, D. (1978), Corrected Quality Ratio: A Composite Index of Scientific Contribution to Knowledge. *Social Studies of Science*, 8, 349–354.

Moed, H.F. (2010), Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4 (3), 265–277.

Pinski, G., Narin, F. (1976), Citation Influence for Journal Aggregates of Scientific Publications. *Information Processing and Management*, 12, 297–312.

Price, D.J. de Solla (1970), Citation measures of hard science, soft science, technology, and non-science. In: C.E. Nelson, D.K. Pollak, Eds.: *Communication among Scientists and Engineers* (Heat, Lexington, Mass. 1970) 1-12.

Schubert, A., Braun, T. (1996), Cross-field normalization of scientometric indicators. *Scientometrics,* 36 (3), 311–324.

Tomer, C. (1986), A Statistical Assessment of Two Measures of Citation: the Impact Factor and the Immediacy Index. *Information Processing and Management*, 22, 251–258.