A General Framework for Relative Impact Indicators / Cadre général pour les facteurs d’impact relatifs

Non Peer-reviewed author version

EGGHE, Leo & ROUSSEAU, Ronald (2003) A General Framework for Relative Impact Indicators / Cadre général pour les facteurs d’impact relatifs. In: Canadian Journal of Information and Library Science, 27(1). p. 29-48.

Handle: http://hdl.handle.net/1942/767

A General Framework for Relative Impact Indicators

LEO EGGHE
LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
and UIA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
E-mail: leo.egghe@luc.ac.be

and

RONALD ROUSSEAU
KHBO, Industrial Sciences and Technology,
Zeedijk 101, B-8400 Oostende, Belgium
& UIA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
& LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
E-mail:ronald.rousseau@kh.khbo.be

Abstract

This article brings the underlying structure of different relative indicators to the forefront. Special attention is given to the relative impact of a journal within a set of journals, a so-called meta-journal. Examples of relative impact factors are calculated for a group of information science, and for a group of management journals. Advantages of relative impact indicators are highlighted. These indicators are further studied in the context of regression analysis. Finally, it is shown that, compared to the Ramírez-García-Del Río renormalized impact factor, the relative impact factor is more sensitive to changes of relative contributions of journals within a journal set.

Keywords

Relative impact, global impact, average impact, journal impact factor, regression lines, Pearson correlation coefficient, information sciences journals, management journals, activity index, attractivity index

Introduction

Consider a set of N journals, let C_j denote *the number of citations received by journal* j over a fixed period of time, t_C, and let P_j denote the number of articles published in this journal over a certain period of time, t_P. Usually (but not necessarily) t_C and t_P refer to different periods, with different lengths. The (t_P : t_C) impact factor of journal j, $j \in \{1,...,N\}$, is then defined as:

$$IF_j(t_P, t_C) = \frac{C_j}{P_j} \qquad (1)$$

Usually we will not specify the period, and write simply:

$$IF_j = \frac{C_j}{P_j} \qquad (2)$$

The Garfield or ISI impact factor [1] is of this type: citations are normalized with respect to the number of publications. There is, however, no normalization with respect to the total number of citations or the total number of publications in the set of journals under study. Normalizing, hence obtaining relative values, with respect to the total number of publications and the total number of citations (in this pool of journals, over fixed periods) yields a relative impact factor, denoted as RIF, and defined as:

$$RIF_j = \frac{\gamma_j}{\pi_j} \qquad (3)$$

where

$$\gamma_j = \frac{C_j}{\sum_{k=1}^{N} C_k} \qquad (4)$$

and

$$\pi_j = \frac{P_j}{\sum_{k=1}^{N} P_k} \tag{5}$$

It is obvious that the relative impact factor is of more importance if one wants to compare the (relative) impact of journals belonging to different pools of journals, e.g. a biomedical journal versus an economic one. Indeed, an impact factor of two may refer to a top journal in one field, and to a mediocre one in another domain. The RIF, on the other hand, refers to the relative impact of a journal within its field.

By a field we mean any group of related journals, but the prototype of such a field is an ISI-JCR subject category. Sometimes we will refer to such a group of related journals, considered as a whole, as a meta-journal. The global impact factor (GIF) [2] of a meta-journal is defined as:

$$GIF = \frac{\sum_{k=1}^{N} C_k}{\sum_{k=1}^{N} P_k} \tag{6}$$

From formulae (2), (3) and (6) we derive:

$$RIF_j = \frac{IF_j}{GIF} \tag{7}$$

Equation (7) shows that ranking journals <u>in a fixed field</u> according to RIF or according to IF yields the same result. This simple observation implies that the ISI-JCR category rankings by impact factor are the same as those for relative impact factors (this is with respect to the category under consideration). We, and probably most scientometricians with us, would appreciate it if ISI would publish besides journal IFs also journal category RIFs. As said, this can be done in the same table. The advantage of doing so is explained above.

If a journal is classified in two categories (denoted as A and B) it has two relative impact factors: RIF(A) and RIF(B). If the global impact factor of category A is denoted as GIF(A) and the global impact factor of category B as GIF(B) then

$$\frac{RIF_j(A)}{RIF_j(B)}=\frac{GIF(B)}{GIF(A)} \tag{8}$$

Hence the RIF of a journal is the larger in that category for which the global impact factor is the smaller. Further, if journals j and k both belong to categories A and B, then

$$\frac{RIF_j(A)}{RIF_k(A)}=\frac{RIF_j(B)}{RIF_k(B)}=\frac{IF_j}{IF_k} \tag{9}$$

This shows that ratios between impact factors are always the same.

The RIF is globally scale invariant. By this we mean that if for every j P_j becomes $k.P_j$ and C_j becomes $k.C_j$ ($k > 0$) then, for every j, the RIF (as well as the IF and the GIF) remains unaltered.

Example

As an example we have calculated the RIFs of all 'information science & library science' and all 'management' journals in the JCR, edition 2000. Table 1 gives the results for the information science journals, Table 2 for the management journals. Abbreviations of journal names and IFs are taken from the JCR. Clearly, rankings for IFs and RIFs are the same. In Table 3 we have combined the two lists of RIFs. As the GIF for the information science journals is 0.553, while the GIF for management journals is 0.808, journals that belong to both categories (indicated in bold) have a higher RIF considered as an information journal than as a management journal. Among the top 10 in the combined list we find 7 management journals, while among the last ten, we find only two. This indicates that the group of management journals contains more stronger and less weaker journals (as considered from a relative citation point of view).

We further note that the GIF for the information and library journals is larger than the average impact factor for this group. This is the case for most meta-journals [3-4] . For the management journals, however, the opposite is true. In this sense, the group of management journals is similar to mathematics and chemistry journals, where the same phenomenon has been observed. Egghe and Rousseau 2,5 studied theoretical foundations for the relation between the global and the average impact of a meta-journal.

Table 1. RIFs for all 'information science and library science' journals in the JCR, edition 2000

Rank	information science	IF	RIF
1	J AM MED INFORM ASSN	3.089	5.590
2	MIS QUART	2.064	3.735
3	J DOC	1.640	2.968
4	J AM SOC INFORM SCI	1.226	2.219
5	INTERNET WORLD	1.167	2.112
	LIBR INFORM SCI	1.167	2.112
7	INFORM SYST RES	1.093	1.978
8	INT J GEOGR INF SCI	0.988	1.788
9	COLL RES LIBR	0.905	1.638
10	KNOWL ORGAN	0.778	1.408
11	RESTAURATOR	0.759	1.374
12	TELECOMMUN POLICY	0.731	1.323
13	INFORM PROCESS MANAG	0.719	1.301
14	INFORM MANAGE	0.683	1.236
15	SCIENTOMETRICS	0.660	1.194
16	INFORM TECHNOL LIBR	0.481	0.870
17	J INFORM SCI	0.473	0.856

18	J HEALTH COMMUN	0.463	0.838
19	ONLINE	0.456	0.825
20	SOC SCI COMPUT REV	0.429	0.776
21	INT J INFORM MANAGE	0.424	0.767
22	LIBR QUART	0.407	0.737
23	INFORM SOC	0.404	0.731
24	INTERLEND DOC SUPPLY	0.400	0.724
25	ASLIB PROC	0.397	0.718
26	DATABASE	0.393	0.711
27	J INFORM TECHNOL	0.382	0.691
28	INFORM SYST J	0.375	0.679
29	PROGRAM-ELECTRON LIB	0.364	0.659
30	LIBR RESOUR TECH SER	0.361	0.653
31	SCIENTIST	0.347	0.628
32	B MED LIBR ASSOC	0.343	0.621
33	J GOV INFORM	0.328	0.594
34	LIBR TRENDS	0.316	0.572
35	LIBR INFORM SCI RES	0.297	0.537
36	J ACAD LIBR	0.296	0.536
37	LIBR J	0.265	0.480
38	SOC SCI INFORM	0.264	0.478
39	J LIBR INF SCI	0.263	0.476
40	ONLINE CDROM REV	0.206	0.373
41	ELECTRON LIBR	0.190	0.344
	GOV INFORM Q	0.190	0.344
43	LIBRI	0.188	0.340
44	REF USER SERV Q	0.185	0.335
45	CAN J INFORM LIB SCI	0.167	0.302
46	J INFORM ETHICS	0.148	0.268
47	LAW LIBR J	0.105	0.190
	Z BIBL BIBL	0.105	0.190
49	J SCHOLARLY PUBL	0.086	0.156
50	LIBR COLLECT ACQUIS	0.071	0.128
51	ECONTENT	0.065	0.118
	LIBR ACQUIS PRACT TH	0.065	0.118
53	NFD INFORM-WISS PRAX	0.029	0.052
54	BEHAV SOC SCI LIBR	0.000	0.000
	average	0.526	
	global average (GIF)	0.553	

Table 2 RIFs for all 'management' journals in the JCR, edition 2000

Rank	management	IF	RIF
1	ACAD MANAGE REV	3.912	4.842
2	ADMIN SCI QUART	3.333	4.126
3	CALIF MANAGE REV	2.877	3.561
4	HARVARD BUS REV	2.561	3.170
5	STRATEGIC MANAGE J	2.531	3.133
6	ACAD MANAGE J	2.375	2.940
7	MIS QUART	2.064	2.555
8	SLOAN MANAGE REV	1.794	2.221
9	HUM RESOURCE MANAGE	1.268	1.570
10	J MANAGE	1.235	1.529
11	ORGAN BEHAV HUM DEC	1.200	1.485
12	MANAGE LEARN	1.186	1.468
13	RES POLICY	1.078	1.334
14	ORGAN SCI	1.052	1.302
15	J INT BUS STUD	1.012	1.253
16	MANAGE SCI	1.011	1.251
17	J PROD INNOVAT MANAG	1.000	1.238
18	ORGANIZATION	0.963	1.192
19	HUM RELAT	0.832	1.030
20	LEADERSHIP QUART	0.830	1.027
21	ORGAN STUD	0.818	1.013
22	ORGAN DYN	0.800	0.990
23	ADV STRATEG MANAGE	0.739	0.915
24	R&D MANAGE	0.737	0.912
25	INFORM MANAGE	0.683	0.845
26	INT J FORCASTING	0.677	0.838
27	J MANAGE INQUIRY	0.649	0.803
28	J OPER RES SOC	0.648	0.802
29	J MANAGE STUD	0.646	0.800
30	INTERFACES	0.629	0.779
31	GROUP ORGAN MANAGE	0.590	0.730
32	NEW TECH WORK EMPLOY	0.524	0.649
33	DECISION SCI	0.473	0.585
34	SYST DYNAM REV	0.455	0.563
35	OMEGA-INT J MANAGE S	0.453	0.561
36	IND MARKET MANAG	0.420	0.520
37	INT J OPER PROD MAN	0.412	0.510
38	INT J SELECT ASSESS	0.395	0.489
39	J INFORM TECHNOL	0.382	0.473
40	J ECON MANAGE STRAT	0.378	0.468
41	J FORECASTING	0.377	0.467
42	TOURISM MANAGE	0.328	0.406
43	IEEE T ENG MANAGE	0.325	0.402
44	NEGOTIATION J	0.317	0.392
45	RES TECHNOL MANAGE	0.299	0.370
46	GROUP DECIS NEGOT	0.294	0.364
47	TOTAL QUAL MANAGE	0.269	0.333
48	CAN J ADM SCI	0.264	0.327

49	J ORGAN CHANGE MANAG	0.250	0.309
50	SYST RES BEHAV SCI	0.242	0.300
51	REV IND ORGAN	0.241	0.298
52	SERV IND J	0.213	0.264
53	LONG RANGE PLANN	0.207	0.256
54	J ORGAN BEHAV MANAGE	0.206	0.255
55	INT J SERV IND MANAGE	0.200	0.248
56	INT J TECHNOL MANAGE	0.198	0.245
57	SYST PRACT ACT RES	0.179	0.222
58	J SMALL BUS MANAG	0.167	0.207
59	INT J MANPOWER	0.116	0.144
60	WORKFORCE	0.075	0.093
	average	0.840	
	global average (GIF)	0.808	

Table 3 Combined ranking of information and management journals, according to their RIFs.

rank	journal	RIF
1	J AM MED INFORM ASSN (inform)	5.590
2	ACAD MANAGE REV (manag)	4.842
3	ADMIN SCI QUART (manag)	4.126
4	**MIS QUART (inform)**	**3.735**
5	CALIF MANAGE REV (manag)	3.561
6	HARVARD BUS REV (manag)	3.170
7	STRATEGIC MANAGE J (manag)	3.133
8	J DOC (inform)	2.968
9	ACAD MANAGE J (manag)	2.940
10	**MIS QUART (manag)**	**2.555**
11	SLOAN MANAGE REV (manag)	2.221
12	J AM SOC INFORM SCI (inform)	2.219
13	INTERNET WORLD (inform)	2.112
	LIBR INFORM SCI (inform)	2.112
15	INFORM SYST RES (inform)	1.978
16	INT J GEOGR INF SCI (inform)	1.788
17	COLL RES LIBR (inform)	1.638
18	HUM RESOURCE MANAGE (manag)	1.570
19	J MANAGE (manag)	1.529
20	ORGAN BEHAV HUM DEC (manag)	1.485
21	MANAGE LEARN (manag)	1.468
22	KNOWL ORGAN (inform)	1.408
23	RESTAURATOR (inform)	1.374
24	RES POLICY (manag)	1.334
25	TELECOMMUN POLICY (inform)	1.323
26	ORGAN SCI (manag)	1.302
27	INFORM PROCESS MANAG (inform)	1.301
28	J INT BUS STUD (manag)	1.253

29 MANAGE SCI (manag)	1.251
30 J PROD INNOVAT MANAG (manag)	1.238
31 INFORM MANAGE (inform)	**1.236**
32 SCIENTOMETRICS (inform)	1.194
33 ORGANIZATION (manag)	1.192
34 HUM RELAT (manag)	1.030
35 LEADERSHIP QUART (manag)	1.027
36 ORGAN STUD (manag)	1.013
37 ORGAN DYN (manag)	0.990
38 ADV STRATEG MANAGE (manag)	0.915
39 R&D MANAGE (manag)	0.912
40 INFORM TECHNOL LIBR (inform)	0.870
41 J INFORM SCI (inform)	0.856
42 INFORM MANAGE (manag)	**0.845**
43 INT J FORCASTING (manag)	0.838
44 J HEALTH COMMUN (inform)	0.838
45 ONLINE (inform)	0.825
46 J MANAGE INQUIRY (manag)	0.803
47 J OPER RES SOC (manag)	0.802
48 J MANAGE STUD (manag)	0.800
49 INTERFACES (manag)	0.779
50 SOC SCI COMPUT REV (inform)	0.776
51 INT J INFORM MANAGE (inform)	0.767
52 LIBR QUART (inform)	0.737
53 INFORM SOC (inform)	0.731
54 GROUP ORGAN MANAGE (manag)	0.730
55 INTERLEND DOC SUPPLY (inform)	0.724
56 ASLIB PROC (inform)	0.718
57 DATABASE (inform)	0.711
58 J INFORM TECHNOL (inform)	**0.691**
59 INFORM SYST J (inform)	0.679
60 PROGRAM-ELECTRON LIB (inform)	0.659
61 LIBR RESOUR TECH SER (inform)	0.653
62 NEW TECH WORK EMPLOY (manag)	0.649
63 SCIENTIST (inform)	0.628
64 B MED LIBR ASSOC (inform)	0.621
65 J GOV INFORM (inform)	0.594
66 DECISION SCI (manag)	0.585
67 LIBR TRENDS (inform)	0.572
68 SYST DYNAM REV (manag)	0.563
69 OMEGA-INT J MANAGE S (manag)	0.561
70 LIBR INFORM SCI RES (inform)	0.537
71 J ACAD LIBR (inform)	0.536
72 IND MARKET MANAG (manag)	0.520
73 INT J OPER PROD MAN (manag)	0.510
74 INT J SELECT ASSESS (manag)	0.489
75 LIBR J (inform)	0.480
76 SOC SCI INFORM (inform)	0.478
77 J LIBR INF SCI (inform)	0.476
78 J INFORM TECHNOL (manag)	**0.473**
79 J ECON MANAGE STRAT (manag)	0.468

80	J FORECASTING (manag)	0.467
81	TOURISM MANAGE (manag)	0.406
82	IEEE T ENG MANAGE (manag)	0.402
83	NEGOTIATION J (manag)	0.392
84	ONLINE CDROM REV (inform)	0.373
85	RES TECHNOL MANAGE (manag)	0.370
86	GROUP DECIS NEGOT (manag)	0.364
87	ELECTRON LIBR (inform)	0.344
	GOV INFORM Q (inform)	0.344
89	LIBRI (inform)	0.340
90	REF USER SERV Q (inform)	0.335
91	TOTAL QUAL MANAGE (manag)	0.333
92	CAN J ADM SCI (manag)	0.327
93	J ORGAN CHANGE MANAG (manag)	0.309
94	CAN J INFORM LIB SCI (inform)	0.302
95	SYST RES BEHAV SCI (manag)	0.300
96	REV IND ORGAN	0.298
97	J INFORM ETHICS (inform)	0.268
98	SERV IND J (manag)	0.264
99	LONG RANGE PLANN (manag)	0.256
100	J ORGAN BEHAV MANAGE (manag)	0.255
101	INT J SERV IND MANAGE (manag)	0.248
102	INT J TECHNOL MANAGE (manag)	0.245
103	SYST PRACT ACT RES (manag)	0.222
104	J SMALL BUS MANAG (manag)	0.207
105	LAW LIBR J (inform)	0.190
	Z BIBL BIBL (inform)	0.190
107	J SCHOLARLY PUBL (inform)	0.156
108	INT J MANPOWER (manag)	0.144
109	LIBR COLLECT ACQUIS (inform)	0.128
110	ECONTENT (inform)	0.118
	LIBR ACQUIS PRACT TH (inform)	0.118
112	WORKFORCE (manag)	0.093
113	NFD INFORM-WISS PRAX (inform)	0.052
114	BEHAV SOC SCI LIBR (inform)	0.000

A generalisation

In this section we will place the observations made in the introduction in a general framework. This framework, leading to a general theory of relative indicators, is built up as follows.

Consider a set of items, denoted as S, partitioned into N disjoint subsets, denoted as $T_{j, j=1...N}$. There are, moreover, given two functions f and g : S $\to \mathbb{R}$. Then for every j, j = 1,...,N, we define numbers P_j and C_j by

$$P_j = \sum_{a \in T_j} f(a) \qquad and \qquad C_j = \sum_{a \in T_j} g(a) \tag{10}$$

An absolute indicator (AI) is of the form

$$AI_j = \frac{C_j}{P_j} \tag{11}$$

Relative values for P_j and C_j are given by:

$$p_j = \frac{P_j}{\sum_{k=1}^{N} P_k} \qquad and \qquad c_j = \frac{C_j}{\sum_{k=1}^{N} C_k} \tag{12}$$

Finally, relative indicators (RI) are then of the form:

$$RI_j = \frac{c_j}{p_j} \tag{13}$$

Relative indicators are based on the empirical distributions (in the statistical sense) of P and C.

Examples of this general framework

1°) The standard journal impact factor

The set S consists here of all articles (in journals covered by ISI) published in the years Y-1 and Y-2. Partitioning occurs according to journals. The function f is the constant function mapping each article to the value 1, while the function g maps each article to the number of citations it receives in the year Y. Clearly, C_j/P_j is the standard journal impact factor of journal j, while c_j/p_j is an example of the relative impact factor introduced above.

2°) The activity index (AI)

The activity index (AI) was introduced by Frame [6]. It characterizes the relative research effort a country devotes to a given field F. Its definition is:

$$\mathrm{AI} = \frac{\text{the country's share in the world's publication output in the given field F}}{\text{the country's share in the world's publication output in all science fields}} \qquad (14)$$

The AI fits into our general framework as follows. S is the set of all publications in a certain period (e.g. one year). This set is partitioned per country. The function f is again the constant function 1, while the function g is the binary function mapping an article to the value 1 if it belongs to the particular sub field under study, and to zero otherwise. For country j, P_j denotes the number of publications of this country and p_j denotes the relative share of this country in the world's publication. Similarly, C_j denotes the number of articles published by country j in sub field F, while c_j denotes this country's share in the world output in sub field F. This shows that the activity index fits into our general framework.

Applying formula (7) to the activity index yields:

$$\mathrm{AI} = \frac{\text{the given field's share in the country's publication output}}{\text{the given field's share in the world's publication output}} \qquad (15)$$

which is a result mentioned e.g. by Schubert and Braun [7].

3°) The attractivity index (AAI) [7]

The attractivity index (AAI) characterizes the relative impact of a country's publications in a given field. Its definition is:

$$\mathrm{AAI} = \frac{\text{the country's share in citations attracted by publications in the given field F}}{\text{the country's share in citations attracted by publications in all science fields}} \tag{16}$$

The AAI fits into our general framework as follows. S is the set of all publications in a certain period (e.g. one year). This set is partitioned per country. The function f associates to each publication the number of citations it receives over the period under study, while the function g is the function mapping an article to its number of received citations if it belongs to the sub field under study, and to zero otherwise. For country j, P_j denotes country j's received number of citations and p_j denotes the relative share of this country's citations in the world's total. Similarly, C_j denotes the number of citations received by country j in sub field F, while c_j denotes this country's share in the world's citation total in sub field F. This shows how the attractivity index fits into the frame.

Applying formula (7) to the attractivity index yields:

$$\mathrm{AAI} = \frac{\text{the given field's share in citations attracted by the country's publications}}{\text{the given field's share in citations attracted by all the publications of the world}} \tag{17}$$

which is also a result mentioned by Schubert and Braun [7].

4°) The relative citation rate [7,8]

A country's (K) relative citation rate (RCR) is defined as the sum of the observed citation rates of all articles published by scientists of this country, divided by sum of the expected citation rates. Here the expected citation rate of an article is the impact factor of the journal in which the article was published. Clearly, the RCR can be calculated for all publications of a country, or for a sub field. Note also that expected and observed citation rates must be calculated over the same period.

Put in our framework, one starts off with the set of all publications (over a certain period). This set is then partitioned according to countries. The function g associates with each article the number of citations it received (during the period under consideration). The other function f associates with each article the impact factor (calculated over the same period) of the journal in which the article was published. In this way, C/P is just the country's RCR. The corresponding relative value, c/p, has not been defined in the literature (to the best of our knowledge), but it easily can.

More examples will be given in another publication.

Mathematical explorations of relative indicators

In this section we derive a number of mathematical relations concerning relative indicators. Here we will use the terminology introduced in the first section, but we stress the fact that, as shown in the previous section, these results can be interpreted in many other contexts.

Notation. If RIF_1, RIF_2, ... , RIF_N are the relative impact factors of a group of N journals (a so-called meta journal), then we denote their arithmetic, geometric and harmonic mean as follows:

$$\overline{RIF} = \frac{1}{N}\sum_{j=1}^{N} RIF_j = \frac{1}{N}\sum_{j=1}^{N}\frac{\gamma_j}{\pi_j} \tag{18}$$

$$g(RIF) = \left(RIF_1 \cdots RIF_N\right)^{\frac{1}{N}} = \left(\frac{\gamma_1}{\pi_1}\bullet\frac{\gamma_2}{\pi_2}\cdots\frac{\gamma_N}{\pi_N}\right)^{\frac{1}{N}} \tag{19}$$

$$h(RIF) = \frac{N}{\sum_{j=1}^{N}\frac{1}{RIF_j}} = \frac{N}{\sum_{j=1}^{N}\frac{\pi_j}{\gamma_j}} \tag{20}$$

Similar notations will be used for the arithmetic, geometric and harmonic average of the impact factors.

Theorem 1

$$GIF = \frac{\overline{IF}}{\overline{RIF}} = \frac{g(IF)}{g(RIF)} = \frac{h(IF)}{h(RIF)} \tag{21}$$

Proof.

a) the arithmetic mean

$$\overline{RIF} = \frac{1}{N}\sum_{j=1}^{N}\frac{C_j / \sum_{k=1}^{N} C_k}{P_j / \sum_{k=1}^{N} P_k} = \frac{\overline{IF}}{GIF}$$

This shows the first result.

b) the geometric mean

$$g(RIF) = \left(\frac{\gamma_1}{\pi_1}\bullet\frac{\gamma_2}{\pi_2}\cdots\frac{\gamma_N}{\pi_N}\right)^{\frac{1}{N}}$$

$$= \left(\prod_{j=1}^{N}\frac{C_j}{\sum_{k=1}^{N} C_k}\frac{\sum_{k=1}^{N} P_k}{P_j}\right)^{\frac{1}{N}} = \left(\prod_{j=1}^{N}\frac{C_j}{P_j}\right)^{\frac{1}{N}}\frac{\sum_{k=1}^{N} P_k}{\sum_{k=1}^{N} C_k} = \frac{g(IF)}{GIF}$$

c) the harmonic mean

$$h(RIF) = \frac{N}{\sum_{j=1}^{N} \frac{\frac{P_j}{\sum_{k=1}^{N} P_k}}{\frac{C_j}{\sum_{k=1}^{N} C_k}}} = \frac{1}{GIF} \frac{N}{\sum_{j=1}^{N} \frac{P_j}{C_j}} = \frac{h(IF)}{GIF}$$

In earlier publications [2,5] we proved and discussed the following result.

Theorem S

Consider the scatterplot of a meta journal $\left\{\left(P_1, \frac{C_1}{P_1}\right), \left(P_2, \frac{C_2}{P_2}\right), \dots, \left(P_N, \frac{C_N}{P_N}\right)\right\}$ and let r

be the Pearson correlation coefficient of the regression line through this scatterplot.

Then

$$\begin{aligned} r > 0 &\Leftrightarrow GIF > \overline{IF} \\ r = 0 &\Leftrightarrow GIF = \overline{IF} \\ r < 0 &\Leftrightarrow GIF < \overline{IF} \end{aligned}$$

Applying this theorem to the first part of Theorem 1 yields the following corollary:

Corollary 1

$$\begin{aligned} r > 0 &\Leftrightarrow 1 > \overline{RIF} \\ r = 0 &\Leftrightarrow 1 = \overline{RIF} \\ r < 0 &\Leftrightarrow 1 < \overline{RIF} \end{aligned}$$

Similarly, we may consider the scatterplot $\left\{\left(\pi_1, \frac{\gamma_1}{\pi_1}\right), \left(\pi_2, \frac{\gamma_2}{\pi_2}\right), \dots, \left(\pi_N, \frac{\gamma_N}{\pi_N}\right)\right\}$.

Denoting the Pearson correlation coefficient of the regression line through this scatterplot as ρ leads to the following result.

Proposition 1

$$r = \rho$$

Proof.

This follows trivially from the fact that Pearson correlation coefficient is co-ordinate wise invariant under scaling. By this we mean that the correlation coefficients of the pairs $(x_i,y_i)_{i=1,\ldots,N}$, $(c_1x_i,y_i)_{i=1,\ldots,N}$, $(x_i,c_2y_i)_{i=1,\ldots,N}$, , $(c_1x_i,c_2y_i)_{i=1,\ldots,N}$ with c_1, $c_2 \in \mathbb{R}_0$ are the same (see also Ahlgren et al. [9]). As

$$\pi_i = \frac{P_i}{\sum_{k=1}^{N} P_k} \quad and \quad \frac{\gamma_i}{\pi_i} = \frac{C_i}{P_i} \cdot \frac{\sum_{k=1}^{N} P_k}{\sum_{k=1}^{N} C_k}$$

the result follows immediately.

A scatterplot $\{(x_j, y_j)_{j=1,\ldots,N}\}$ is said to be increasing if $x_i < x_j$ if and only if $y_i < y_j$. Similarly, a scatterplot $\{(x_j, y_j)_{j=1,\ldots,N}\}$ is said to be decreasing if $x_i < x_j$ if and only if $y_i > y_j$. These notions are used in the following result.

Proposition 2

The scatterplot $\left\{\left(P_j, \frac{C_j}{P_j}\right)_{j=1,\ldots,N}\right\}$ is increasing if and only if the scatterplot $\left\{\left(\pi_j, \frac{\gamma_j}{\pi_j}\right)_{j=1,\ldots,N}\right\}$ is increasing. A similar result holds for decreasing scatterplots.

Proof (for the increasing case).

$$\pi_i < \pi_j \quad \Leftrightarrow \frac{P_i}{\sum_{k=1}^{N} P_k} < \frac{P_j}{\sum_{k=1}^{N} P_k} \Leftrightarrow P_i < P_j$$

and

$$\frac{\gamma_i}{\pi_i} < \frac{\gamma_j}{\pi_j} \Leftrightarrow \frac{C_i}{\sum_{k=1}^{N} C_k} \cdot \frac{\sum_{k=1}^{N} P_k}{P_i} < \frac{C_j}{\sum_{k=1}^{N} C_k} \cdot \frac{\sum_{k=1}^{N} P_k}{P_j}$$
$$\Leftrightarrow \frac{C_i}{P_i} < \frac{C_j}{P_j}$$

Corollary 2

If either $\left\{\left(P_j, \frac{C_j}{P_j}\right)_{j=1,\ldots,N}\right\}$ or $\left\{\left(\pi_j, \frac{\gamma_j}{\pi_j}\right)_{j=1,\ldots,N}\right\}$ is increasing, then

1) the other one is also increasing
2) GIF > $\overline{IF}$
3) 1 > $\overline{RIF}$
4) r > 0
5) ρ > 0

Proof. 1) follows from Proposition 2. From this and Corollary 1 in *Egghe & Rousseau*[2] follows 2). Assertion 3) follows then from (21); 4) follows from Corollary 1, and then finally 5) follows from Proposition1.

The relative impact factor versus the renormalized one

A recent proposal for making impact factors comparable over categories or fields is the Ramírez-García-Del Río renormalized impact factor [10]. According to this proposal a journal j's impact factor, denoted as IF_j, is recalculated (renormalized in their terminology) with respect to the group to which this journal belongs, as follows:

$$F_j = \frac{IF_j - IF_{MED}}{IF_{MAX} - IF_{MED}}$$

where IF_{MAX} denotes the largest impact factor of the group of journals under consideration and IF_{MED} is the median impact factor of this group. By definition $F_{MAX} = 1$, $F_{MED} = 0$ and $-\infty < F_j \leq 1$.

We will compare some properties, in particular the sensitivity, of the renormalized impact factor versus the relative one. For this reason we first prove the following lemma, which is of some interest on its own.

Lemma

Consider a meta-journal with $IF_1 \leq \ldots \leq IF_{MAX}$, and let $GIF_- = \frac{\sum_{k=1}^{MAX-1} C_k}{\sum_{k=1}^{MAX-1} P_k}$ (we removed the journal with the highest impact factor). Then the following relation holds:

$$GIF_- \leq GIF \leq IF_{MAX} \tag{22}$$

with equality if and only if all impact factors are equal.

Proof. We know that for every j = 1, ..., MAX : $\frac{C_j}{P_j} \leq \frac{C_{MAX}}{P_{MAX}}$ or

$$C_j \cdot P_{MAX} \leq P_j \cdot C_{MAX} \tag{23}$$

We first prove the inequality: $GIF_- \leq GIF$. We have to show that:

$$\frac{\sum_{k=1}^{MAX-1} C_k}{\sum_{k=1}^{MAX-1} P_k} \leq \frac{\sum_{k=1}^{MAX-1} C_k + C_{MAX}}{\sum_{k=1}^{MAX-1} P_k + P_{MAX}}$$

$$\Leftrightarrow \left(\sum_{k=1}^{MAX-1} C_k\right)\cdot\left(\sum_{k=1}^{MAX-1} P_k\right)+\left(\sum_{k=1}^{MAX-1} C_k\right)\cdot P_{MAX} \leq \left(\sum_{k=1}^{MAX-1} C_k\right)\cdot\left(\sum_{k=1}^{MAX-1} P_k\right)+\left(\sum_{k=1}^{MAX-1} P_k\right)\cdot C_{MAX}$$

$$\Leftrightarrow \qquad (C_1 + \ldots + C_{MAX-1})\cdot P_{MAX} \leq (P_1 + \ldots + P_{MAX-1})\cdot C_{MAX}$$

This follows immediately from inequality (23). Equality only occurs if all impact factors are equal.

Next we show that GIF ≤ IF_{MAX}. This is: we show that:

$$\frac{\sum_{k=1}^{MAX-1} C_k + C_{MAX}}{\sum_{k=1}^{MAX-1} P_k + P_{MAX}} \leq \frac{C_{MAX}}{P_{MAX}}$$

$$\Leftrightarrow \left(\sum_{k=1}^{MAX-1} C_k\right)\cdot P_{MAX} + C_{MAX}\cdot P_{MAX} \leq \left(\sum_{k=1}^{MAX-1} P_k\right)\cdot C_{MAX} + C_{MAX}\cdot P_{MAX}$$

$$\Leftrightarrow \qquad \left(C_1 + \ldots + C_{MAX-1}\right)\cdot P_{MAX} \leq \left(P_1 + \ldots + P_{MAX-1}\right)\cdot C_{MAX}$$

The last inequality again follows from inequality (23). Also here equality only occurs if all impact factors are equal.

We are now ready to compare the sensitivity of the renormalized impact factor with that of the relative one.

1. In each category F_{MAX} is always 1, and hence it has no sensitivity at all with respect to the global pattern of publication and citation rates of the category. RIF_{MAX} does depend on this global pattern. Note that we make no assertions about whether this is a desirable or a non-desirable feature. We just observe the fact.
2. For any j, F_j only depends on the value IF_{MAX} (and of course on IF_{MED}, but this is not the point we want to make), and not on the weight of the journal with

maximum impact factor in the whole meta-journal. RIF_j does depend on this. We show this assertion. Let $IF_{MAX} = \frac{k\,C_{MAX}}{k\,P_{MAX}}$ $(k > 0)$, $k \neq 1$, and assume that not all impact factors are equal, and all other Cs and Ps remain unchanged. Then

$$RIF_j(k) = \frac{C_j}{P_j} \cdot \frac{\sum_{i=1}^{MAX-1} P_i + k\,P_{MAX}}{\sum_{i=1}^{MAX-1} C_i + k\,C_{MAX}}$$

If $k < 1$ then $RIF_j(k) > RIF_j$, while, if $k > 1$, $RIF_j(k) < RIF_j$. Indeed, denoting – as a shorthand - any summation from i=1 to i = MAX –1 simply as Σ, we have:

$$RIF_j(k) - RIF_j = \frac{C_j}{P_j}\left(\frac{\Sigma P + k\,P_{MAX}}{\Sigma C + k\,C_{MAX}} - \frac{\Sigma P + P_{MAX}}{\Sigma C + C_{MAX}}\right)$$

$$= \frac{C_j}{P_j}\left(\frac{\Sigma P \cdot \Sigma C + \Sigma P \cdot C_{MAX} + kP_{MAX} \cdot \Sigma C + kP_{MAX}C_{MAX} - \Sigma P \cdot \Sigma C - P_{MAX} \cdot \Sigma C - kC_{MAX}\Sigma P - kP_{MAX}C_{MAX}}{(\Sigma C + k\,C_{MAX}) \cdot (\Sigma C + C_{MAX})}\right)$$

$$= \frac{C_j}{P_j} \frac{(k-1) \cdot (GIF\_ - IF_{MAX})\,P_{MAX} \cdot \Sigma P}{(\Sigma C + k\,C_{MAX}) \cdot (\Sigma C + C_{MAX})}$$

By the previous lemma GIF. - $IF_{MAX} < 0$ (unless all impact factors are equal), and consequently: $RIF_j(k) > RIF_j$ if $k < 1$, and $RIF_j(k) < RIF_j$ if $k > 1$.

3. The renormalized impact factor behaves (sometimes) in a counterintuitive way. Consider Table 4 as an example.

Table 4 Hypothetical set of journals, their impact factor, relative impact factor and renormalized impact factor.

P	C	IF	F	RIF
10	40	4	1	1.818
10	30	3	0.5	1.364
10	20	2	0	0.909
10	10	1	-0.5	0.455
10	10	1	-0.5	0.455
50	110			

Here P stands for number of publications, C stands for number of citations and the last row yields the total number of publications, and the total number of citations. Next we make some small changes to this table (Table 5).

Table 5. Journal set of Table 4 with some small changes

P	C	IF	F	RIF
10	46	4.6	1	2.091
10	34	3.4	0.538	1.545
10	20	2	0	0.909
10	5	0.5	-0.577	0.227
10	5	0.5	-0.577	0.227
50	110			

If we concentrate on the second journal then we see that it has a citation increase of 11.8% with respect to Table 4. The first journal shows an increase of 15%. Hence, the second journal has a smaller increase than the first one (in absolute as well as proportional numbers). What is the influence of this increase on journals' impact scores? The F-ratio between the second and the first increases from 50% to 53.8%, while the RIF-ratio decreases from 75% to 73.9%. In view of the absolute as well as relative higher citation increase of the first with respect to the second the change in RIF looks more natural than that in F.

These three observations show that the RIF is more sensitive than the renormalized impact factor F. In our opinion it is the better of the two.

Discussion

Relative impact indicators have been around for quite some while. In this article we have shown the underlying structure of all these indicators. We have studied these indicators in the context of regression analysis. Finally, we have clearly shown that the relative impact factor is more sensitive than the renormalized impact factor to changes of relative contributions of journals within a journal set. More refined weighting methods exist that can change IF rankings within the same group of journals. Examples are the Pinski-Narin influence measure [11] and the Tijssen-Van Raan measure [12] based on net citation balances. Finally we note that the indicators CPP/JCSm and CPP/FCSm, often used in research evaluation [13-16] are of the same form as the relative citation rate [7,8] mentioned before. They are not relative indicators in our sense.

References

1. E. GARFIELD, I.H. SHER, New factors in the evaluation of scientific literature through citation indexing, *American Documentation,* 14 (1963) 195-201.
2. L. EGGHE, R. ROUSSEAU, Average and global impact of a set of journals, *Scientometrics*, 36 (1996) 97-107.
3. G. VAN HOOYDONK, R. GEVAERT, G. MILIS-PROOST, H. VAN DE SOMPEL, K. DEBACKERE, A bibliotheconomic analysis of the impact factors of scientific disciplines, *Scientometrics*, 30 (1994) 65-81.

4. R. ROUSSEAU, G. VAN HOOYDONK, Journal production and journal impact factors, *Journal of the American Society for Information Science*, 47 (1996) 775-780.

5. L. EGGHE, R. ROUSSEAU, Averaging and globalising quotients of informetric and scientometric data, *Journal of Information Science*, 22 (1996) 165-170.

6. J.D. FRAME, Mainstream research in Latin America and the Caribbean, *Interciencia*, 2 (1977) 143-148.

7. A. SCHUBERT, T. BRAUN, Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9 (1986) 281-291.

8. T. BRAUN, W. GLÄNZEL, A. SCHUBERT, *Scientometric Indicators. A 32 country comparison of publication productivity and citation impact*, World Scientific Publishing Co., Singapore, 1985.

9. P. AHLGREN, B. JARNEVING, R. ROUSSEAU, Why Pearson's correlation coefficient should not be used in author cocitation analyses. Preprint 2002.

10. A.M. RAMÍREZ, E.O. GARCÍA, J.A. DEL RÍO, Renormalized impact factor, *Scientometrics*, 47 (2000) 3-9.

11. G. PINSKI, F. NARIN, Citation influence for journal aggregates of scientific publications: theory, with an application to the literature of physics, *Information Processing and Management,* 12 (1976) 297-312.

12. R.J.W. TIJSSEN, A.F.J. VAN RAAN, Net citation balances: a measure of influence between scientific journals, *Journal of the American Society for Information Science,* 41 (1990) 296-304.

13. R.E. DE BRUIN, A. KINT, M. LUWEL, H.F. MOED, A study of research evaluation and planning: the University of Ghent, *Research Evaluation*, 3 (1993) 25-41.

14. H.F. MOED, W.J.M. BURGER, J.G. FRANKFORT, A.F.J. VAN RAAN, The use of bibliometric data for the measurement of university research performance, *Research Policy*, 14 (1985) 131-149.

15. A.F.J. VAN RAAN, Evaluation of performance and trends in basic and applied research by advanced bibliometric methods: a science policy instrument for nations with an economy in transition. In: PAČES, PIVEC, TEICH (eds.) *Science evaluation and its management,* IOS Press, Amsterdam, 1999, 227-245.

16. Th. N. VAN LEEUWEN, H. F. MOED, R.J.W. TIJSSEN, M.S. VISSER, A. F.J. VAN RAAN, Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance, *Scientometrics*, 51 (2001) 335-346.