Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 73, No. 3 (2007) 321–330
DOI: 10.1007/s11192-007-1779-8

The self-cited rate of scientific journals and the manipulation of their impact factors

GUANG YU, LIANG WANG

School of Management, Harbin Institute of Technology, Heilongjiang (P. R. China)

Owing to some discussions about manipulating impact factor by requesting authors to increase their citations to the publication journal, we theoretically establish a mathematical expression of a relation between the journal self-citation rate and its impact factor by the single-factor method in this paper. Based on self-citation data of some journals in JCR and the observed relation between journal impact factor and the self-cited rate, we analyze the possibility that journal editors manipulate impact factors of their journals by raising the self-cited rate. Finally, we make some suggestions for supervising this crude way of active manipulating the impact factor.

Introduction

Self-citation is a universal citation form which refers to a citation form by which scientific body with relative stability and literature-production continuity cites its former literatures in its later ones (TAGLIACOZZO, 1977; GLÄNZEL et al., 2004). Self-citations are of two main kinds, author self-citation and journal self-citation; we concentrate on journal self-citation in this paper. The self-cited rate relates a journal's self-citations to the number of times it is cited by all journals, including itself

Received March 7, 2007

Address for correspondence:
GUANG YU
Library, Harbin Institute of Technology,
No. 17, Silin Street, NanGang, Harbin, 150001, Heilongjiang, P. R. China
E-mail: yug@hit.edu.cn

0138–9130/US $ 20.00

(EGGHE & ROUSSEAU, 1990). According to the definition, the expression of self-cited rate in current year is:

$$The\ self\text{-}cited\ rate = \frac{the\ number\ of\ self\text{-}citations}{total\ cites} \tag{1}$$

The expression of Impact Factor (*IF*) in two–year period in *Y* year is:

$$IF_2 = \frac{CIT_{Y-1}(Y) + CIT_{Y-2}(Y)}{PUB(Y-1) + PUB(Y-2)} \tag{2}$$

According to Eq. (1) and Eq. (2), it is obvious that a journal's impact factor has a close relation with its self-cited rate and the number of self-citations to articles of the journal is important to the calculation result of its impact factor. It has been observed by some researchers (FASSOULAKI et al., 2000; SEVINC, 2004) that self-citations of a journal may affect its impact factor and a high self-citing rate of a journal may positively affect its impact factor.

Impact factor can be affected by many factors such as the number of articles published (ROUSSEAU & HOOYDONK, 1996), the difference between disciplines (AMIN & MABE, 2000), the number of cooperation author, the number of cooperation institutions, publication delay (GARFIELD, 1999; WILLIAM, 2000; YU et al., 2005), the journal self-cited rate (FASSOULAKI et al., 2000; MOTAMED et al., 2002) and so on. If these factors were changed naturally or artificially, the impact factor would also change. Therefore, changing the number of self-citations becomes an instrument of the editors who attempt to raise journal impact factors. DONG et al. (2005) pointed one quite crude way of active manipulating *IF* by requesting author to increase references to the papers published in his journal and consequently editors may artificially raise a journal *IF*. SMITH (1997) pointed it out in a News of BMJ that the journal *Leukemia* has been accused of trying to manipulate its impact factor by authors who had submitted a paper to *Leukemia* asking them to increase the number of references to papers published in *Leukemia*. NEUBERGER & COUNSELL (2002) reported another similar case: they described how one journal editor suggested the inclusion of more references to that journal. HEMMINSSON et al. (2002) pointed in the letter to editor-in-chief of AJR that editors of some journals were sending copies of articles previously published in their journals together with the review copy of another article to the referees and were asking them whether it is possible to include those published articles in the reference list. SEVINC (2004) reported that the influence of authors' choice of references distorts the perception of the journal within the scholarly community and is considered as highly unethical behavior. We have gone through such same experiences when we contributed to Chinese core journals for several times.

The citing of literature is a spontaneous behavior when authors write articles; the intervention to the citing process is a manipulation which disobeys the order of nature.

Because the journal self-cited rate has a close relationship with its impact factor, there should be a quantitative calculation method to measure the relationship. The impact factor can be affected by many factors as mentioned above. However, in this paper we only consider the journal self-cited rate, therefore we can establish a mathematical expression between journal self-cited rate and its impact factor by the *single-factor method*, and quantitatively study the relation between the journal impact factor and its self-cited rate using data in JCR; then analyze the possibility that editors raise impact factors of their journals by manipulating the self-cited rate.

Methods

The relation expression between the self-cited rate and IF

Let $C(T)$ as the cumulative citation distribution function and k_s as the self-cited rate. According to the definition of self-cited rate (Eq.(1)), k_s is affected mainly by the number of self-citations and total citations. For studying the question in theory, we firstly suppose that the citation distribution $C(T)$ of one journal is a continuous function and it is the same as that of the distribution by cited year of total citations to articles published in the journal; secondly, the single-factor method is used, namely the self-cited rate is regarded as a main factor influencing the impact factor and other factors are neglected; for example, journal production, total citation quantity, etc...; thirdly, initial values of the cumulative citation distribution $C(T)$ and journal production in current year are regarded as $C(0)=0$ and $PUB(0)=0$, respectively.

The journal impact factor has a close relationship with its cumulative citation distribution probability $C(T)$, and the equation of impact factor in two years is (YU et al., 2006)

$$IF = K \cdot (C(3) - C(1)) \tag{3}$$

We assume K as a constant which has a close relationship with the numbers of total citations and the journal production which were defined in that article (ROUSSEAU & HOOYDONK, 1996), M as the number of total citations, Y as journal production in a period of 2 years and does not change. The relationship between them can be described as $K=M/Y$.

We let k_s be the self-cited rate of a journal in current year, k_0 be the self-cited rate and M_0 be the number of total citations in previous year. The constant K's initial value $K_0=M_0/Y$. Because the single-factor method is used and the function $C(T)$ is a continuous function and dose not change, the number $(1-k_0)M_0$ of other-citations should not change, either.

When the self-cited rate changes from k_0 to k_s, the number of total citations changes from M_0 to $M = (1-k_0)M_0 + k_s M$, so we have

$$M = \frac{1-k_0}{1-k_s} \cdot M_0 \tag{4}$$

Because $K = \frac{M}{Y} = \frac{1-k_0}{1-k_s} \cdot \frac{M_0}{Y} = \frac{1-k_0}{1-k_s} \cdot K_0$, we have

$$K = \frac{1-k_0}{1-k_s} \cdot K_0 \tag{5}$$

By combining Eq. (5) and Eq. (3), we have

$$IF = K_0 \frac{1-k_0}{1-k_s}(C(3) - C(1)) \tag{6}$$

Due to $IF_0 = K_0(C(3) - C(1))$, we have

$$IF = \frac{1-k_0}{1-k_s} \cdot IF_0 \tag{7}$$

According to Eq. (7), we get $(1-k_s) < (1-k_0)$ and $IF > IF_0$ when $k_s > k_0$. That is to say, if a journal self-cited rate rises, its impact factor rises in theory, and it is a natural phenomena. However, for manipulating *IF*, editors usually ask authors to increase the number of references to papers published in their journals in recent two years and *IF* dose not change according to Eq. (7). Eq. (7) theoretically describes the relation between the self-cited rate of a journal and its impact factor.

Methods of collecting and processing data

We collect the data of self-citations and *IF*s of journals in JCR of ISI and observe the real relation between actual *IF* and k_s of any journal. We choose 16 journals in four disciplines (Information Science, Biology Science, Multidisciplinary Science and Engineering Science) from JCR and collect the data of impact factor (*IF*), the total citations (*M*) and the self-citations (*SC*) of every journal from 1998 to 2004 (some of them from 2000 to 2004). Based on these data, the self-cited rate and the theoretic impact factor can be solved by Eq. (1) and Eq. (7), respectively:

- the journal self-cited rate: $k_s = \frac{SC}{M}$;

- the theoretical impact factor after the self-cited rate changes: $IF = \frac{1-k_0}{1-k_s} \cdot IF_0$, here IF_0 is the impact factor of previous year, k_0 is journal self-cited rate of previous year; for example, when the theoretical impact factor of a journal in 2000 is calculated, k_0 should be self-cited rate of the journal in 1999 and IF_0 should be the actual impact factor in 1999.

Results and discussion

Journals with high impact factors

Generally, a journal whose impact factor is high, such as *Science* or *Nature*, has a higher academic status than other journals. Total citations of the two journals are thousands but their self-cited rates are quite low. Their cited half-lives have been steady. Table 1 shows that self-cited rates (k_s) of both journals are lower than 1.5%, the theoretical impact factor (shorten form *IF*) is different from the actual impact factor (shorten form IF_0). Because the self-cited rate is quite low, the change of k_s merely has imperceptible influence on the impact factor. Therefore, the higher the impact factor and the lower the self-cited rate are, the weaker the influence of the self-cited rate on the impact factor is.

Table 1. Impact factors and self-cited rates of *Nature* and *Science* 2000–2004

Journal		2000	2001	2002	2003	2004
Nature	IF_0	25.814	27.955	30.432	30.979	32.182
	IF	25.814	25.793	27.936	30.391	30.967
	$k_{s=SC/M}$	4406/306184	4292/315640	4206/325646	3977/343528	4064/363347
	k_s	0.0144	0.0136	0.0129	0.0116	0.0112
Science	IF_0	23.872	23.329	26.682	29.781	31.853
	IF	23.872	23.892	23.327	26.639	29.760
	$k_{s=SC/M}$	3233/274443	3557/282431	3700/296080	3397/311593	3399/332803
	k_s	0.0118	0.0126	0.0125	0.0109	0.0102

Figure 1 shows change curves of IF_0 and IF and k_s of two journals whose impact factors and the self-cited rate are high enough in their respective discipline, but their impact factors are lower and self-cited rates are higher than *Science* or *Nature*, the change curves of actual and theoretical impact factors of *Theoretical and Applied Genetics* are little similar.

According to Table 1, Figure 1 and Eq. (7), for journals whose impact factors are high, it is difficult that editors manipulate impact factors by adding the number of self-citations whether self-cited rates of these journals are low or high.

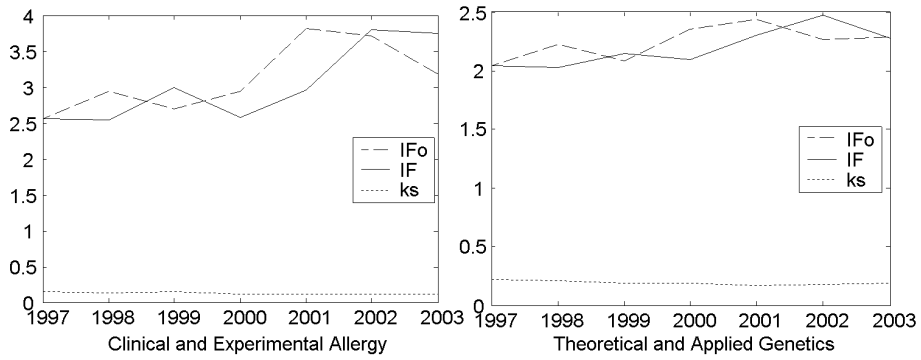

Figure 1. Two journals with high impact and higher self-cited rate

Journals with medium impact factors

Impact factors of most journals in SCI are medium. It is difficult to confirm these journals because the medium of impact factor is different from different discipline. We choose journals of information science and medicine science.

(1) Journals with high self-cited rates

In Table 2, we can see that although self-cited rates of the three journals are higher, but there are big differences between actual IF_0 and theoretical IF. According to the relationship expression, namely Eq. (7), we know that the changing process of the impact factor should be an unartificial process and is influenced by many factors, and the self-cited rate is merely one of these factors. For example, the self-cited rate of *Journal of the American Society for Information Science* got lower in 2004 but its impact factor rose; it is shown that other factors influenced the impact factor and the influence of the self-cited rate was weaker.

Table 2. Three journals' impact factors and self-cited rates of information science

Journal name		2000	2001	2002	2003	2004
Journal of the American Society for Information Science	IF_0	1.226	1.641	1.773	1.473	2.086
	IF	1.226	1.1808	1.6026	1.7209	1.3865
	k_s	0.2550	0.2265	0.2080	0.1840	0.1331
Scientometrics	IF_0	0.660	0.676	0.855	1.251	1.120
	IF	0.660	0.5289	0.7866	0.9441	1.0696
	k_s	0.4944	0.3690	0.4577	0.5089	0.4256
Journal of Information Science	IF_0	0.473	0.707	1.080	1.067	0.899
	IF	0.473	0.4772	0.7448	1.0879	1.0139
	k_s	0.1050	0.1128	0.1579	0.1640	0.1202

Table 3 includes the data of impact factors and the self-cited rates of four medicine journals; we can see that the self-cited rates of these journals are higher than in Tables 1 and 2 and have stronger influences on impact factors. For example, changes IF_0 and IF of *Theoretical and Applied Genetics* are close, similarly, *Clinical and Experimental Allergy*; but changes of IF_0 and IF of *Journal of Pineal Research* and *Clinical* or *Experimental Allergy* are very different and sometimes inverse, such as changes from 1998 to 1999; so there should be no possibility of active manipulating in these four journals. Theoretically, it would be efficient that editors of these journals manipulate their impact factors by increasing self-cited rates, but we could not identify this behavior by data in Tables 2 and 3 according to current means.

Table 3. Impact factors and self-cited rates of four medicine journals

Journal name		1997	1998	1999	2000	2001	2002	2003
Theoretical and Applied Genetics	IF_0	2.040	2.224	2.082	2.358	2.438	2.264	2.287
	IF	2.040	2.0336	2.1449	2.095	2.3043	2.4714	2.2754
	k_s	0.2161	0.2136	0.1846	0.1898	0.1709	0.1821	0.1862
Journal of Pineal Research	IF_0	3.711	2.544	2.571	3.799	4.040	3.913	3.426
	IF_t	3.711	3.6198	2.6225	2.5348	3.8661	3.8865	4.2382
	k_s	0.1959	0.1756	0.2003	0.1889	0.2029	0.1715	0.2350
Clinical and Experimental Allergy	IF_0	2.559	2.939	2.702	2.947	3.826	3.721	3.176
	IF	2.559	2.5357	2.9896	2.5837	2.9563	3.7946	3.7473
	k_s	0.1409	0.1330	0.1477	0.1087	0.1114	0.1041	0.1103
Journal of Allergy and Clinical Immunology	IF_0	3.769	4.509	4.637	4.179	5.506	6.282	6.831
	IF	3.769	3.8306	4.5644	4.5087	4.3266	5.2284	6.3771
	k_s	0.1007	0.1152	0.1259	0.1011	0.1317	0.0856	0.0992

(2) Journals with low self-cited rates

Table 4 shows the data of impact factors, the self-cited rates of three journals and the theoretical results of their impact factors calculated by Eq. (7). These self-cited rates are low, and changes of self-cited rates of these journals weakly influence their impact factors; so behaviors of manipulating impact factor are unmeasured according to data in Table 4.

Table 4. Three journals' impact factors and self-cited rates

Journal name		1997	1998	1999	2000	2001	2002	2003
Neuroscience and Biobehavioral Reviews	IF_0	2.786	3.316	3.595	3.382	5.212	5.504	5.482
	IF	2.786	2.8515	3.2561	3.6394	3.3225	5.2551	5.4853
	k_s	0.0217	0.0441	0.0265	0.0384	0.0212	0.0292	0.0259
Astrophysical Journal Supplement Series	IF_0	1.946	2.075	3.749	3.475	5.214	4.749	6.247
	IF	1.946	1.9299	2.0831	3.7671	3.4502	5.2162	4.7776
	k_s	0.0450	0.0370	0.0407	0.0454	0.0385	0.0389	0.0447
Microscopy Research and Technique	IF_0	1.198	0.765	1.183	1.746	2.165	2.524	2.307
	IF	1.198	1.1602	0.7720	1.162	1.7776	2.1672	2.5232
	k_s	0.0786	0.0486	0.0507	0.0372	0.0198	0.0208	0.0205

Journals with low impact factors

For journals with low impact factors, theoretically, their impact factors are likely influenced by various factors. We choose four engineering journals with low impact factors and calculate their self-cited rates and theoretical impact factors by Eq. (7) from 1998 to 2003. The results are shown in Table 5.

Table 5. Impact factors and self-cited rates of four engineering journals

Journal name		1997	1998	1999	2000	2001	2002	2003
Journal of Engineering Mathematics	IF_0	0.342	0.5170	0.3470	0.4430	0.3580	0.4380	0.6420
	IF	0.342	0.3568	0.5088	0.3410	0.4460	0.3550	0.4461
	$k_{s=SC/M}$	11/203	27/289	19/241	19/304	22/314	21/338	30/379
	k_s	0.0542	0.0934	0.0788	0.0625	0.0701	0.0621	0.0792
Science in China Series E – Technological Sciences	IF_0	0.169	0.1160	0.3330	0.3090	0.3760	0.4120	0.3550
	IF	0.169	0.2682	0.0648	0.3101	0.2396	0.4690	0.3242
	$k_{s=SC/M}$	6/14	16/25	27/76	28/91	15/140	56/197	19/210
	k_s	0.4286	0.6400	0.3553	0.3077	0.1071	0.2843	0.0905
Computer Applications in Engineering Education	IF_0			0.138	0.1040	0.1670	0.1400	0.3950
	IF			0.138	0.1004	0.0794	0.1900	0.1727
	$k_{s=SC/M}$			11/19	8/19	7/29	15/45	17/37
	k_s			0.5789	0.4211	0.2414	0.3333	0.4595
International Journal of Production Economics	IF_0	0.172	0.1660	0.2270	0.2580	0.2880	0.4810	0.4100
	IF	0.172	0.1821	0.1667	0.222	0.2611	0.2915	0.4257
	$k_{s=SC/M}$	45/273	69/327	74/345	98/498	132/569	193/800	118/827
	k_s	0.1648	0.2110	0.2145	0.1968	0.2320	0.2413	0.1427

We can obtain two conclusions as following:

- Because impact factors and self-cited rates are very low, impact factors of these journals are feebly affected by the self-cited rate, such as *Journal of Engineering Mathematics* and *Science in China Series E-Technological Sciences*, IF_0 is different from IF and sometimes inverse. Theoretically, there would be a small effect on the impact factor that editors of these journals manipulate their impact factors by increasing self-citations but the effect is very low.
- Impact factors of journals with high self-cited rates are obviously affected by the self-cited rate, such as last three journals in Table 5, change trends of impact factors (IF) are similar to that of the theoretical impact factor (IF_0). Theoretically, there would be a bigger effect on the impact factor if editors manipulated their impact factors by increasing self-citations.

Suggestions

Based on theoretical and actual analyses and the relationship between the journal impact factor and its self-cited rate, it is obvious that the journal impact factor should be affected by its self-cited rate, but change trends of actual impact factors of most journals do not match with theoretical impact factors; the reason is that the self-cited rate is only one of factors which affect the impact factor and is not a main factor. According to theoretical analysis and empirical research hereinbefore, there are some suggestions:

- For journals with the high impact factor, the large number of total citations and the low self-cited rate, the change of self-cited rate has a low influence on its impact factor. It is impossible to manipulate the impact factor by change the self-cited rate and is not necessary.
- For journals with medium impact factors, their impact factors were affected by their self-cited rates a little if their self-cited rates were low; some of these journals with high self-cited rates might be affected more; but it is difficult to manipulate impact factors of these journals.
- For a journal with low impact factor, the large number of total citations and the low self-cited rate, the change of the self-cited rate has some influence on its impact factor. It is likely to be effectual that a editor manipulate impact factor of his journals by requesting authors to increase references to papers published in their journal of recent two or three years.
- For journals with low impact factors and high self-cited rates, impact factors are affected by self-cited rates obviously, it will be very effectual and easy to manipulate impact factors by increasing self-cited rates, especially self-citations to papers published in recent two or three years.
- The smaller the number of total citations and the impact factor are, the closer the correlativity between the impact factor and the self-cited rate is. Impact factors of those journals which belong to applied subjects or less popular disciplines usually are low and are easily manipulated.
- Because only self-citations of a journal to such papers published in the previous two years can influence *IF* of the journal, the age distribution of self-citations of a journal is an important factor which affects its *IF*, especially a change of self-citations to papers published in recent two or three years. So it is an index for identifying the manipulation behavior of the impact factor that whether the self-citation distribution model is normal.

It is significant that those journals should be supervised, which are easy to be manipulated by asking authors to increase the number of journal self-references; but there is not a proper measure to directly identify whether some journals are manipulated by editors at present. That research should be based on some specific circumstances.

Identifying the manipulating behavior of the journal impact factor is valuable and Pattern Identification theory might be a good choice.

*

The authors thank Dr. Shi-Peng Li for his assistance in language revision. This work is supported by grant G70571020 from the National Nature Science Foundation of China and the TPM center of Harbin Institute of Technology.

References

AMIN, M., MABE, M. (2000), Impact factors: Use and abuse, *Perspectives in Publishing, October* (1) 1–6, Elsevier Science.

DONG, P., LOH, M., MONDRY, A. (2005), The "impact factor" revisited, *Biomedical Digital Libraries*, 2 : 7.

EGGHE, L., ROUSSEAU, R. (1990), *Introduction to Information.* Elsevier Science Publishers.

FASSOULAKI, A., PARASKEVA, K., PAPILAS, K., KARABINIS, G. (2000), Self-citations in six anaesthesia journals and their significance in determining the impact factor, *British Journal of Anaesthesia,* 84 (2) : 266–269.

GARFIELD, E. (1999), Journal impact factor: a brief review, *Canadian Medical Association Journal*, 161 (8) : 979–980.

GLÄNZEL, W., THIJS, B., SCHLEMMER, B. (2004), A bibliometric approach to the role of author self-citations in scientific communication, *Scientometrics,* 59 (1) : 63–77.

HEMMINGSSON, A., MYGIND, T., SKJENNALD, A., EDGREN, J. (2002), Manipulation of impact factors by editors of scientific journals, *American Journal of Roentgenology,* 178 (3) : 767.

MOTAMED, M., MEHTA, D., BASAVARAJ S., FUAD, F. (2002), Title self citations and impact factors in otolaryngology journals, *Clinical Otolaryngology* 27 (5) : 318–320.

NEUBERGER, J., COUNSELL, C. (2002), Impact factors: uses and abuses, *European Journal of Gastroenterol Hepatol,* 14 : 209–211.

ROUSSEAU, R., HOOYDONK, G. V. (1996), Journal production and journal impact factors. *Journal of the American Society for Information Science,* 47 (10) : 775–780.

SEVINC, A. (2004), Manipulating impact factor: an unethical issue or an Editor's choice? *Swiss Medical Weekly,* 134 : 410.

SMITH, R. (1997), Journal accused of manipulating impact factor, *British Medical Journal,* (15) : 314–315.

TAGLIACOZZO, R. (1977), Self-citation in scientific literature, *Journal of Documentation*, 33 : 251–265.

WILLIAM, G. J. (2000), Notes from the Editor: Concerns and innovations, *Journal of Politics,* 65 (2) : 1–4.

YU, G., WANG, X. H., YU, D. R. (2005), The influence of publication delays on impact factors, *Scientometrics*, 64 (2) : 235–246.

YU, G., GUO, R., LI, Y. J. (2006), The influence of publication delays on three ISI indicators, *Scientometrics,* 69 (3) : 511–527.