Reliability-based citation impact factor and the manipulation of impact factor

Guang Yu · Dong-Hui Yang · Wang Liang

Abstract According to the definition of reliability-based citation impact factor (R-impact factor) proposed by KUO & RUPE and the cumulative citation age distribution model, a mathematical expression of the relationship between R-impact factor and impact factor is established in this paper. By simulation of the change processes of the R-impact factor and impact factor in the manipulation process of the impact factor, it is found that the effect of manipulation can be partly corrected by the R-impact factor in some cases. Based on the Journal Citation Report database, impact factors of 4 normal journals and 4 manipulated journals were collected. The journals' R-impact factors and self-cited rates in the previous two years were calculated for each year during the period 2000 to 2007, and various characteristics influenced by the manipulation were analyzed. We find that the R-impact factor has greater fairness than the impact factor for journals with relatively short cited half-lives. Finally, some issues about using the R-impact factor as a measure for evaluating scientific journals are discussed.

Keywords Impact factor · Manipulation · R-impact factor · Self-cited rate

Introduction

Impact factor is a quantitative tool for evaluating journals in the Journal Citation Report (JCR) database. Librarians rely on the journal impact factor as a tool for selecting periodicals, primarily in scientific disciplines (Cameron 2005). However, impact factor can be manipulated (Hemmingsson 2002), leading to incorrect evaluation of some journals. One method of manipulation is for journal editors to ask authors to cite articles published in their journals when accepting articles for publication (Smith 1997). We define this as self-cited manipulation of the impact factor in this paper (Smith 1997). Since this manipulation was announced by Smith in 1997, similar announcements have emerged in succession

G. Yu (✉) · D.-H. Yang · W. Liang
Library, School of Management, Harbin Institute of Technology, No. 17, Silin Street, Nan Gang, Harbin, Heilongjiang 150001, People's Republic of China
e-mail: yug@hit.edu.cn

(Fassoulaki et al. 2000; Whitehouse 2001; Miller 2002; Hemmingsson 2002; Sevinc 2004; Caldwell 2006; Procianoy 2007; Krauss 2007; Falagas and Alexiou 2008). Manipulation which attempts to increase the nominator would cause misleading evaluation results of scientific journals and cause scientists and researchers doubt whether impact factor is a valid tool to measure the quality of journals (Schreiber 2007). Therefore, manipulation of the impact factor should be banned in the scientific field (Wallner 2009). Jones (2003) proposed that a "clean" impact factor is needed, which should be developed by journal editors (Wallner 2009). It is easy to understand that journal self-citations of the previous two years can result in abnormally increasing *IF* (Falagas and Alexiou 2008), giving the impression a journal is of higher quality. Reliability-based citation impact factor, abbreviated as "R-impact factor", proposed by Kuo and Rupe (2007) considers the influence of citation and length of impact of published journals as measured by the cited half-life. The greater the number of self-citations of a journal, the shorter the cited half-life of the journal. Therefore, we should determine how the R-impact factor of a journal changes when the journal's impact factor is manipulated. Yu and Wang (2007) established a mathematical expression of the relation between the journal self-citation rate and its impact factor by the single-factor method and analyzed the possibility that journal editors manipulate impact factors of their journals by raising the self-cited rate. In this paper, we investigate different aspects of the manipulation process between the impact factor and R-impact factor. The paper is organized as follows: in the next section we deduce the calculation formulas of manipulated impact factor and R-impact factor based on the distribution model of citations, and then simulate changes of the manipulated impact factor and R-impact factor with different aging constants. In Sect. "Analyses on actual journals", we calculate the R-impact factors and self-cited rates in the previous two years of eight journals from 2000 to 2007, and then compare the R-impact factors with the impact factors of those journals.

## Simulation of manipulation processes

Definition of R-impact factor

According to the definition of R-impact factor (Kuo and Rupe 2007), its equation can be expressed as follows:

$$RIF \;=\; IF \;\times\; T_{0.5},\tag{1}$$

where *RIF* is R-impact factor, *IF* is impact factor and $T_{0.5}$ is cited half-life. Amin and Mabe (2000) described the accumulation of citations over time after publication using the generalized citation curve in Fig. 1. The impact factor and cited half-life can be clearly seen in the figure.

There are obvious differences between the R-impact factor and impact factor. The impact factor is an index which considers the relation between the number of cited papers a journal has published in the previous two years and the number that has been cited. The R-impact factor comprises two factors: (1) the impact factor: and (2) the long-lasting impact of published journals as measured by the cited half-life. As is commonly known, very long publication delays of scientific papers would badly influence citation distributions and journal impact factors (Garfield 1999; Yu et al. 2005). If the R-impact factor is used as a measure the same as the *IF*, this bad influence might partly be avoided.
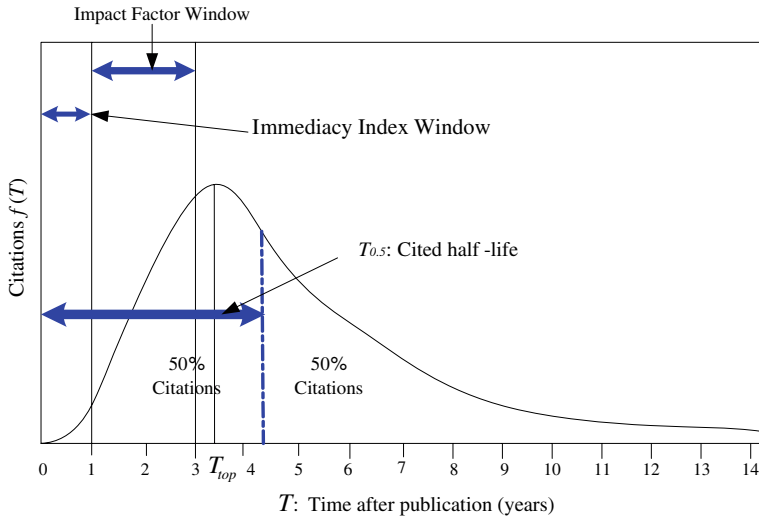
Fig. 1 Generalized citation curve (Amin and Mabe 2000) and ISI indicators

Simulation models

When the number of self-citations published in a journal in the previous two years is manipulated, the self-cited rate in those two years and the impact factor of the journal will be increased in a short time. According to Eq. 1, the cited half-life, one of the two elements of *RIF*, is from the long-term perspective taking into account the reference value of citations, and, therefore, its rate of change is relatively slow. When *IF* is manipulated, the cited half-life correspondly changes in the opposite direction of the change of *IF*. For example, when a journal's self-cited rate is artificially increased, the impact factor rises and the cited half-life becomes shorter. So, the result of the cited half-life multiplied can partly eliminate this increased trend and the negative impact of manipulation can be partly reduced.

In this study, we simulate changes of manipulated *IF* and *RIF* to theoretically prove their values beyond judgment. To achieve this objective, a mathematical model is first chosen to describe the citation distribution of a journal, and then expressions of manipulated *IF* and *RIF* are deduced, which provide the simulation models.

Yu and Li (2009) divide classical citation distribution models into two categories: (1) the Bernal first-order negative exponential model and its improved models; and (2) the Burton and Kebler equation and its improved models. Here, we choose the Bernal first-order negative exponential citation distribution model with non-dimensional parameters. Considering the delay phenomena in the research-citation cycle mainly caused by the publishing process (Egghe and Rousseau 2000; Yu et al. 2005), the model is corrected by pure delay τ. Let $f(T)$ be the cited probability of the journal's published papers and T be citation age, then:

$$f(T) = Ke^{-\alpha(T-\tau)}, \tag{2}$$

where K is the coefficient, α is the aging coefficient, τ is the pure delay for editing and waiting time in the publication process of cited articles after acceptance. Let $C(T)$ be the cumulative probability of citations at age T, then we have

$$C(T) = 1 - e^{-\alpha(T-\tau)} \tag{3}$$

According to the definition of *IF*, its theoretical expression is

$$IF(T) = \lambda(C(3) - C(1)), \tag{4}$$

where $\lambda = M/Y$, M is the number of times cited and Y is the number of papers published in the previous two years of one journal.

When manipulation behavior does not exist in the process of citing, we can obtain $T_{0.5}$ according to the definition of cited half-life, namely, $C(T_{0.5}) = 0.5$ when $T = T_{0.5}$. From Eq. 3, we have

$$T_{0.5} = \tau + \ln 2/\alpha. \tag{5}$$

However, $T_{0.5}$ can not be calculated by Eq. 5 if *IF* is manipulated. We assume that the manipulated citation function $f'(T)$ is k times its theoretical value in the two years. Meanwhile, the number of citations remains unchanged for other ages, so $T \in [1, 3]$, $f'(T) = k \cdot f(T)$. We can use integral calculus for different ages and get the distribution function of cumulative citations which has been manipulated thus:

$$\begin{cases} C(T) = 1 - e^{-\alpha(T-\tau)}, & \tau \le T \le 1 \\ C(T) = 1 + (k-1)e^{-\alpha(1-\tau)} - ke^{-\alpha(T-\tau)}, & 1 < T \le 3 \\ C(T) = 1 + (k-1)e^{-\alpha(1-\tau)} - (k-1)e^{-\alpha(3-\tau)} - e^{-\alpha(T-\tau)}, & T > 3 \end{cases} \tag{6}$$

According to Eqs. 4 and 6, we obtain the formulae of *IF* and *RIF* disturbed by manipulation as follows:

$$IF' = k(e^{-\alpha(1-\tau)} - e^{-\alpha(3-\tau)}) \cdot M/Y$$

$$RIF = T_{0.5} \cdot IF' \tag{7}$$

IF' expresses the manipulated impact factor. According to Eqs. 3 and 6, we have:

$$C(T_{0.5}) = 1 + (k-1)e^{-\alpha(1-\tau)} - (k-1)e^{-\alpha(3-\tau)} - e^{-\alpha(T_{0.5}-\tau)} = 0.5$$

After transforming the expression above, we obtain this formula:

$$T_{0.5} = \tau + \frac{1}{\alpha}(\ln 2 - \ln(1 + 2(k-1)e^{-\alpha(1-\tau)} - 2(k-1)e^{-\alpha(3-\tau)})) \tag{8}$$

Using Eqs. 4–8, the changing processes of *IF'* and *RIF* can be simulated.

It should be explained that the time dynamic characteristic of citedness may differ according to the journal. Therefore, no citation model is a suitable fit for all journals. Here, we study only those journals fitting the corrected first-order negative exponential distribution model. We believe this does not affect the results of this research.

Simulation results

Yu and Wang (2007) found that impact factors are obviously affected by self-cited rates for journals with low impact factors. It is easy to manipulate impact factors by increasing self-cited rates, especially self-citations to papers published in the previous one or two years. From Eq. 5, the smaller the aging coefficient α is, the longer the cited half-life $T_{0.5}$ and the slower the aging. Consequently, when the parameter values of Eq. 8 are chosen, the following assumptions are made: (1) let $\alpha < 0.25$ ($T_{0.5} > 2.97$ years according to Eq. 5)
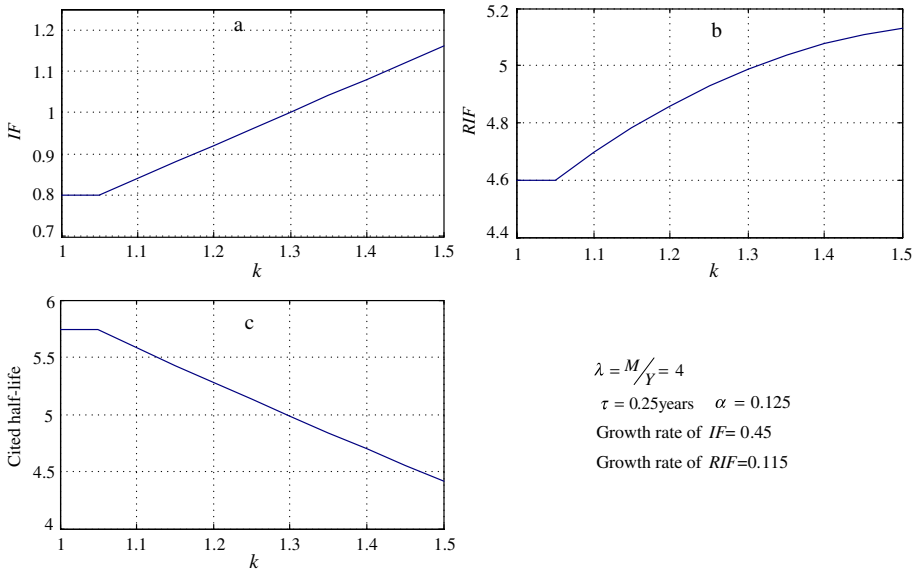
Fig. 2 Simulation results of three parameters of example 1

because a change to a shorter $T_{0.5}$ is more often seen than to a longer $T_{0.5}$; (2) From Eq. 4, the larger $\lambda(\lambda = M/Y)$ is, the higher IF is for the same citation distribution $f(T)$; so we choose small λ, namely the impact factor is low (between 4 and 6); and (3) τ is between 0.2 and 0.3 years because the smallest pure publication delay is commonly about 2–3 months. According to these assumptions, we have four simulation examples of different λ as follows:

(1) $\lambda = 4$, $\tau = 0.25$ years, $\alpha = 0.125$. When $k = 1$, $T_{0.5} = 5.8$ years according to Eq. 5.
(2) $\lambda = 4$, $\tau = 0.25$ years, $\alpha = 0.15$. When $k = 1$, $T_{0.5} = 4.87$ years according to Eq. 5.
(3) $\lambda = 4$, $\tau = 0.25$ years, $\alpha = 0.2$. When $k = 1$, $T_{0.5} = 3.71$ years according to Eq. 5.
(4) $\lambda = 6$, $\tau = 0.25$ years, $\alpha = 0.2$. When $k = 1$, $T_{0.5} = 3.71$ years according to Eq. 5.

Here, k changes in the scope from 1.0 to 1.5, and its mathematical expression is $k \in [1.0, 1.50]$. Four simulation results (Figs. 2, 3, 4, 5) show changes of IF, RIF and $T_{0.5}$ in the four cases.

In Figs. 2 and 3, when IFs are manipulated, the cited half-life is decreased and IF and RIF are increased, but the growth rate of RIF is slower than that of IF. In Figs. 2 and 3, when k is increasing, it is found that the faster the aging is (i.e., the bigger the aging factor α or the shorter the half-life), the greater the decrease of the cited half-life and the smaller the growth rates of RIF. In Fig. 4b and Fig. 5b, we can also see that RIF decreases after reaching a certain level. In example 3 and example 4, manipulation is offset by using R-impact factor to evaluate the journal's impact. It is should be explained that the growth rate of IF or RIF is the change rate between the initial value ($k = 1$) and the final value ($k = 1.5$) of IF or RIF.

Analyses on actual journals

From the simulation results in the previous section, it is shown that RIF can partly offset the effect of manipulation; namely, the shorter the cited half-life is (i.e., the bigger the
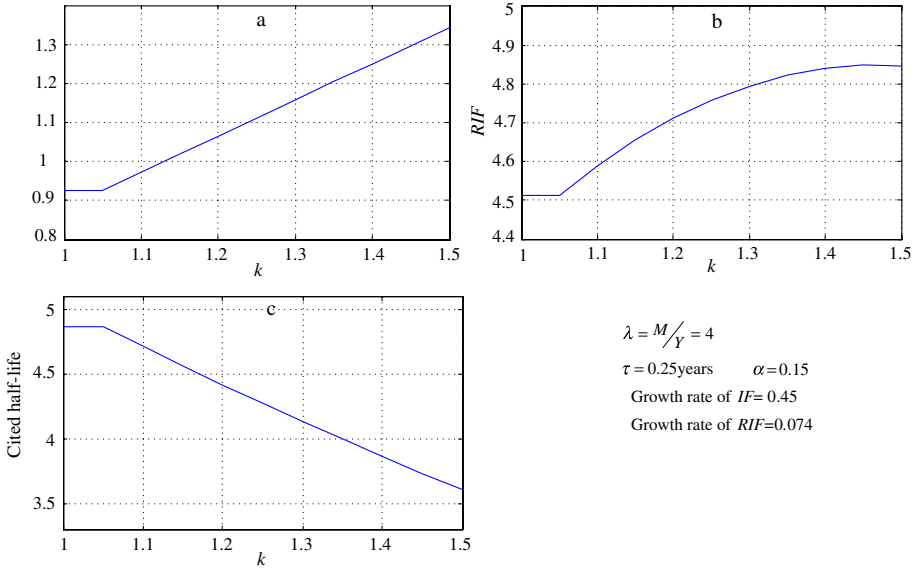
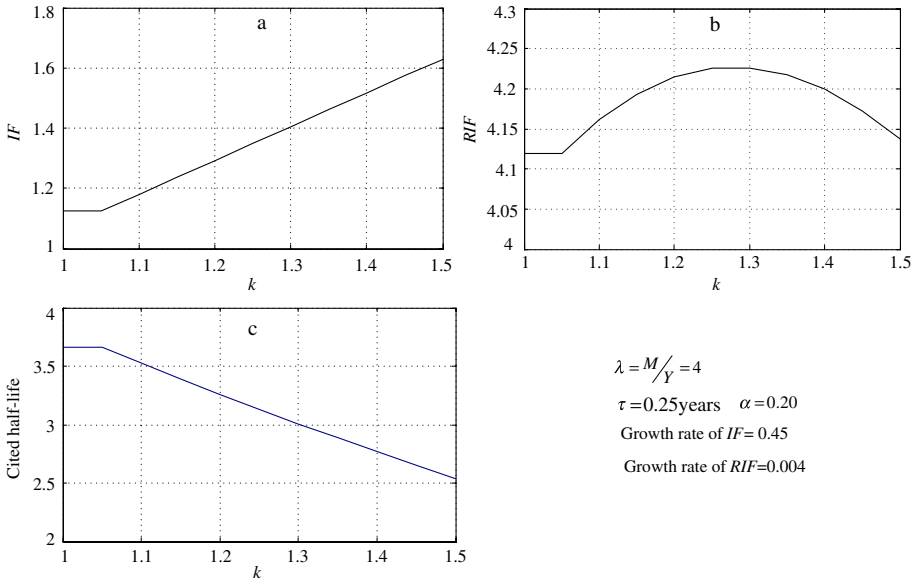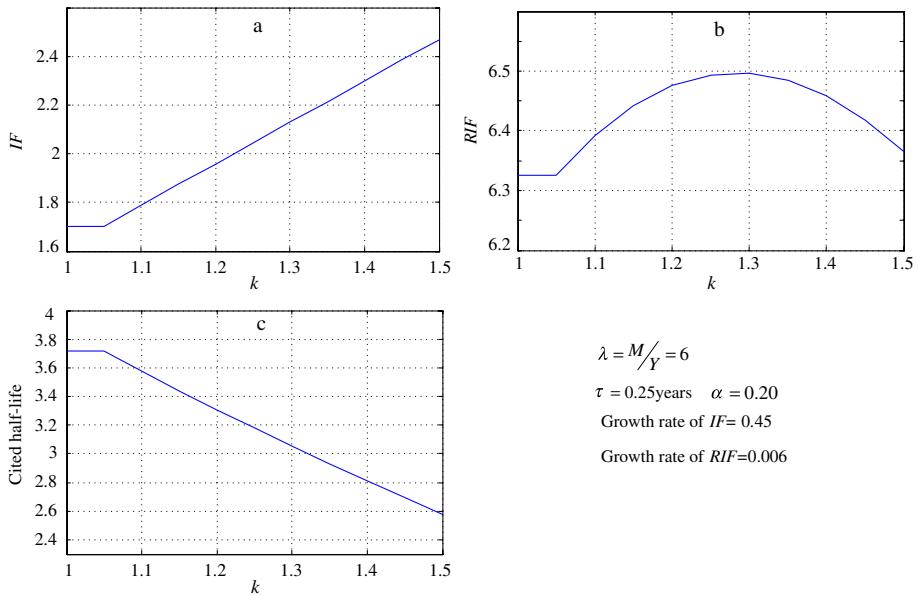Fig. 3 Simulation results of three parameters of example 2

$\lambda = M/Y = 4$

$\tau = 0.25$ years　　$\alpha = 0.15$

Growth rate of IF = 0.45

Growth rate of RIF = 0.074

Fig. 4 Simulation results of three parameters of example 3

$\lambda = M/Y = 4$

$\tau = 0.25$ years　　$\alpha = 0.20$

Growth rate of IF = 0.45

Growth rate of RIF = 0.004

aging coefficient α is), the lower the effect of manipulation on RIF. In this section, we collect total citations, self-citations, impact factors as well as cited half-lives of eight journals over a period of 8 years (2000–2007) from the JCR database, and calculate their self-cited rates in the previous two years and R-impact factors, followed by analysis of

Fig. 5 Simulation results of three parameters of example 4

changes of the self-cited rates in the previous two years, the impact factors and the R-impact factors of these journals.

Journals selection and data collection

In this paper, we selected four "normal" journals from the JCR database: *Scientometrics*, *Journal of Materials Chemistry* (Abbreviated Journal Title: J MATER CHEM), *Transactions of Nonferrous Metals Society of China* (Abbreviated Journal Title: T NONFERR METAL SOC) and *Journal of Materials Processing Technology* (Abbreviated Journal Title: J MATER PROCESS TECH). Here, normal journals are considered as those journals which have not requested authors to cite their journals. According to the explanatory statements and announcements made by some authors submitting their works to the journals, we selected four journals whose editors asked the authors to cite papers published in the journals in the previous two years ("manipulated journals"). In order to protect the identities, we conceal the journal names, instead referring to them as "Journal 1", "Journal 2", "Journal 3" and "Journal 4". We collected the number of times of self-citation, the impact factors and the cited half-lives of the eight journals for each year from 2000 to 2007, and then calculated each journal's self-cited rates in the previous two years and the R-impact factors. See Tables 1 and 2.

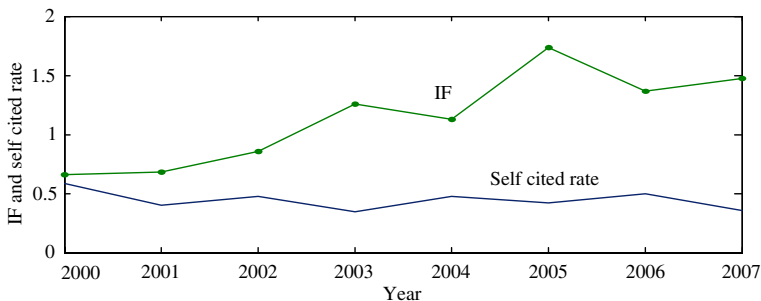Analysis result of self-cited rates and impact factors

The change curves of the impact factors and self-cited rates in the previous two years of the four normal journals from 2000 to 2007 are shown in Figs. 6, 7 and 8. The change curves of the two parameters of four manipulated journals are shown in Figs. 9 and 10 from 2000

Table 1 Self-cited rates in the previous two years of eight journals for 2000–2007

Abbreviated journal title	2000	2001	2002	2003	2004	2005	2006	2007
Scientometrics	0.579	0.401	0.473	0.338	0.471	0.415	0.49	0.356
J Mater Chem	0.129	0.115	0.133	0.08	0.08	0.084	0.072	0.071
T Nonferr Metal Soc	0.538	0.47	0.392	0.516	0.385	0.429	0.418	0.394
J Mater Process Tech	0.248	0.251	0.233	0.177	0.179	0.192	0.116	0.137
Journal 1	0.077	0.157	0.023	0.046	0.119	0.131	0.178	0.651
Journal 2	0.263	0.298	0.304	0.226	0.294	0.345	0.237	0.215
Journal 3	0.207	0.183	0.154	0.076	0.115	0.063	0.124	0.165
Journal 4	0.257	0.328	0.336	0.35	0.451	0.299	0.456	0.265

Table 2 Impact factors of eight journals for 2000–2007

Abbreviated journal title	2000	2001	2002	2003	2004	2005	2006	2007
Scientometrics	0.66	0.676	0.855	1.251	1.12	1.738	1.363	1.472
J Mater Chem	2.379	2.736	2.683	2.659	2.721	3.688	4.287	4.339
T Nonferr Metal Soc	0.294	0.34	0.293	0.322	0.277	0.302	0.264	0.264
J Mater Process Tech	0.237	0.255	0.362	0.45	0.578	0.592	0.615	0.816
Journal 1	0.322	0.264	0.361	0.447	0.54	0.723	1.003	2.147
Journal 2	0.393	0.47	0.55	0.577	0.734	1.039	1.181	1.373
Journal 3	0.49	0.494	0.553	0.605	0.828	0.824	0.918	1.096
Journal 4	0.269	0.418	0.374	0.493	0.517	0.788	0.674	0.611

Fig. 6 Curves of self-cited rate and impact factor of Scientometrics

to 2007. Comparing the change curves of the two parameters of the four normal journals, we find that the change trends of the two parameters of these normal journals are inconsistent, three journals (*SCIENTOMETRICS*, *J MATER CHE* and *J MATER PROCESS TECH*) very obviously (see Figs. 6, 7, 8, respectively), namely, their impact factors are increasing and the self-cited rates are nearly unchanged or decreasing. This indicates that the increases of their impact factors do not rely on their self-cited times in the previous two years. Figures 9 and 10 show the change curves of the self-cited rates in the previous two years and impact factors of the four manipulated journals. It is obvious that the change trends of the two parameters of Journal 1 are the same (see Fig. 9). There are similar
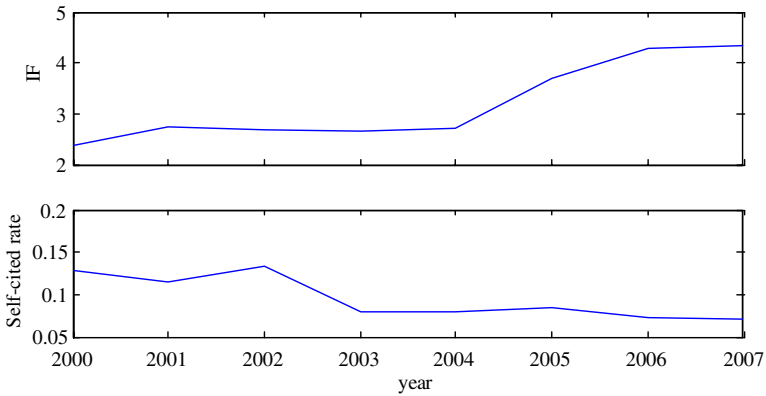
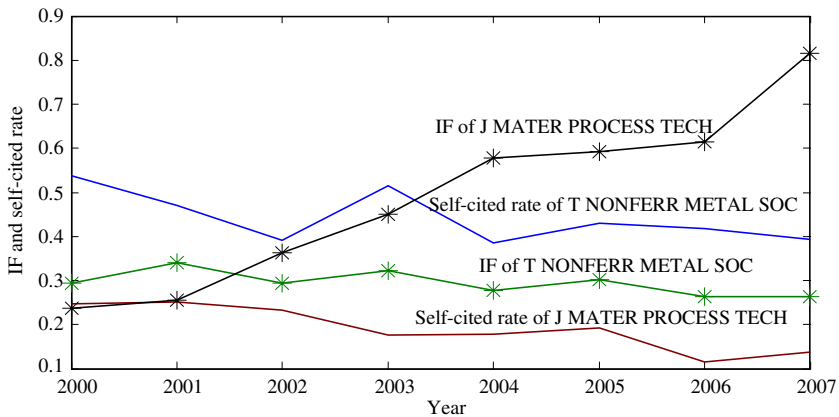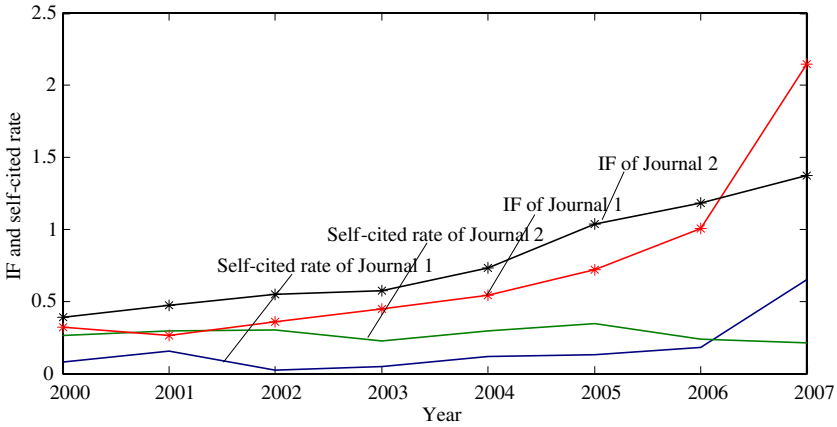**Fig. 7** Curves of self-cited rate and impact factor of J Mater Chem

Fig. 8 Curves of self-cited rates and impact factors of T Nonferr Metal Soc and J Mater Process Tech

change trends of the two parameters for Journal 2 between 2000 and 2005, Journals 3 between 2003 and 2007, and Journal 4 between 2004 and 2007. This means that the growth of these journals' impact factors partially relies on increasing self-citations in the previous two years.

Analysis on *RIF*

Abnormal change of the self-cited rate can influence *IF* and *RIF* to a certain extent. The results of the eight journals' R-impact factors are shown in Table 3. In order to clearly observe the changes of the impact factors and R-impact factors, the change rates of the *IF*s and *RIF*s of the four manipulated journals from 2000 to 2007 are calculated and shown in Fig. 11. Because there is no increase in the self-cited rates of the four normal journals, their R-impact factors' changes are caused by the impact factors and cited half-lives.

   The curves of the rates of change of the four manipulated journals' *IF*s and *RIF*s are shown in Fig. 11. By comparing the four graphs, we see that the rate of change of the *IF* of

Fig. 9 Curves of self-cited rates and impact factors of Journals 1 and 2
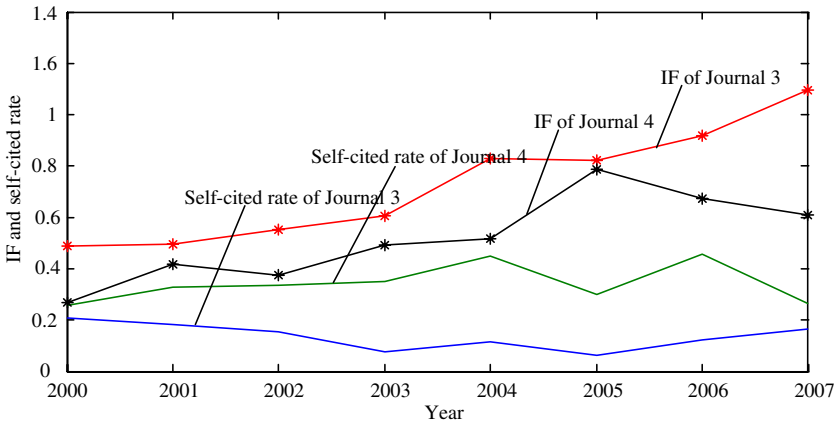

Fig. 10 Curves of self-cited rates and impact factors of Journal 3 and Journal 4

Journal 1 is bigger than that of the *RIF* because there is a large increase in the journal's self-citation which was obviously manipulated. In Table 3 and Fig. 11, it can be seen that the range of the rate of change of the *IF* of Journal 1 is from −0.18 to 1.14 and that of the *RIF* is from −0.18 to 0.574. So, when a journal's IF is obviously manipulated, it is possible that the R-impact factor partially offsets the manipulation effect. Take Journal 1, for example, its impact factor drastically increased in 2007, but its R-impact factor increased much less. However, this is not seen for Journals 2, 3 or 4 because there were no increases in self-citation by the authors. Since there are many factors influencing the impact factor and the cited half-life, the R-impact factor will have a large increase in practice. According to the study results in Sect. "Simulation of manipulation processes", the growth rate of the R-impact factor is much slower than that of the impact factor theoretically and the aging coefficient is particularly large (for example, $\alpha = 0.2$ in Sect. "Simulation of manipulation processes"). However, in the case of manipulation of self-citations where there are small aging coefficients (for example, $\alpha \leq 0.15$), there is no convincing evidence for or against suggested advantage of *RIF* over *IF*.

Table 3 Results of R-impact factors of eight journals for 2000–2007

Journal name	2000	2001	2002	2003	2004	2005	2006	2007
Scientometrics	4.158	4.462	5.387	8.132	6.384	10.602	8.042	8.390
J Mater Chem	8.564	10.39	10.46	11.168	11.97	16.60	20.15	21.70
T Nonferr Metal Soc	1.000	1.054	0.996	1.159	1.053	1.087	0.977	1.003
J Mater Process Tech	1.067	1.224	1.846	2.295	2.890	3.019	3.321	4.814
Journal 1	3.220	2.640	3.610	4.470	5.130	5.639	8.325	13.10
Journal 2	3.262	4.089	4.950	5.308	6.753	9.559	10.98	13.59
Journal 3	3.675	3.606	4.148	4.598	6.400	6.750	7.711	9.097
Journal 4	1.614	2.717	2.356	3.451	3.309	4.964	4.853	4.521
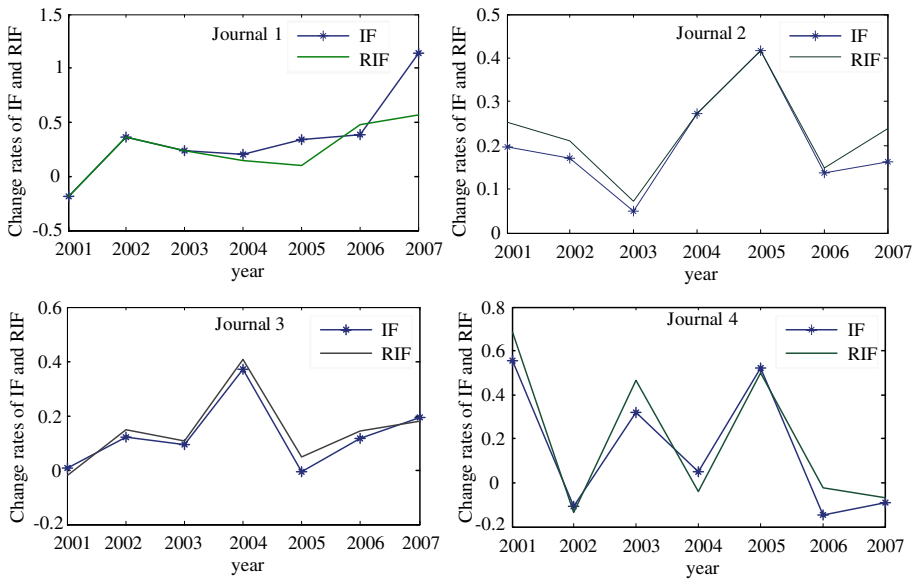
Fig. 11 Change rate curves of the *IFs* and *RIFs* of four manipulated journals

Conclusions

According to the above simulation results and analyses on actual data on impact factor manipulation, the following conclusions are obtained:

(1) As an evaluation tool of scientific journals, the impact factor is easily manipulated according to its definition. So a scientific method needs to be established to identify which journals have been manipulated by increased self-citation.

(2) Through the simulation study on the change process of the R-impact factor caused by the manipulation process in Sect. "Simulation of manipulation processes", it can be seen that the R-impact factor partially offsets the manipulation because the cited half-life changes in the opposite direction against manipulation, especially when the aging coefficient of a journal is large (or the cited half-life is short). So the R-impact factor has a greater fairness than the impact factor when we evaluate journals whose cited half-lives are relatively short.

(3) As the impact factor and the cited half-life are influenced by many factors, similar plot results of the two parameters for journal evaluation may not be obtained in practice. So, whether the R-impact factor can substitute for the impact factor for evaluating scientific journals' impact needs more concrete demonstration.

(4) The R-impact factor is superior to the impact factor for evaluation of a journal's quality in the long term. It has an apparent ability to offset a significant manipulation according to the simulation results and Journal 1's analysis results. However, in this study, we believe there is no convincing evidence of suggested advantage of *RIF* over *IF* in the case of a small aging coefficient (or a long cited half-life).

(5) The R-impact factor can be used as a supplementary tool to monitor the development of journals. However, further research needs to be done before the R-impact factor can be recognized as a stand alone tool for evaluation of the quality of journals.

References

Amin, M., Mabe, M. (2000). Impact factor: Use and abuse. *Perspectives in Publishing*, *1*, 1–6.

Caldwell, T. (2006). Impact factors rocked by manipulation charge. http://www.wame.org/wame-listserve-discussions/manipulating-the-impact-factor.

Cameron, B.D. (2005). Trends in the usage of ISI bibliometric data: Uses, abuses, and implications. *Libraries and the Academy, 5*, 105–125.

Egghe, L., & Rousseau, R. (2000). The influence of publication delays on the observed aging distribution of scientific literature. *Journal of the American Society for Information Science, 51*(2), 158–165.

Falagas, M.E., & Alexiou, V.G. (2008). The top-ten in journal impact factor manipulation. *Archivum Immunologiae et therapiae Experimentalis, 56*(4), 223–226.

Fassoulaki, A., Paraskeva, A., Papilas, K., & Karabinis, G. (2000). Self-citations in six anaesthesia journals and their significance in determining the impact factor. *British Journal of Anaesthesia, 84*, 266–269.

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

Hemmingsson, A. (2002). Manipulation of impact factors by editors of scientific journals. *American Journal of Roentgenology, 178*(3), 767.

Jones, A.W. (2003). Impact factors of forensic science and toxicology journals: What do the numbers really mean? *Forensic Science International, 133*, 1–8.

Krauss, J. (2007). Journal self-citation rates in ecological sciences. *Scientometrics, 70*(1), 79–89.

Kuo, W., & Rupe, J. (2007). R-impact factor: Reliability-based citation impact factor. *IEEE Transaction on Reliability, 56*(3), 366–367.

Miller, J.B. (2002). Impact factors and publishing research. *The Scientist, 16*(18), 11.

Procianoy, R.S. (2007). The journal impact factor today. *Journal De Pediatria, 83*(6), 487.

Schreiber, M. (2007). Self-citation corrections for the Hirsch index. *Europhysics Letters*, *78*(3), 30002-1–30002-6.

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, 314*, 461.

Wallner, C. (2009). Ban impact factor manipulation. *Science, 323*(5913), 461.

Whitehouse, G.H. (2001). Citation rates and impact factors: Should they matter? *British Journal of Radiology, 74*, 1–3.

Yu, G., & Li, Y.J. (2009). Identification of referencing and citation processes of scientific journals based on the citation distribution model, *Scientometrics, 81*(2).

Yu, G., & Wang, L. (2007). The self-cited rate of scientific journals and the manipulation of their impact factors. *Scientometrics, 73*(3), 356–366.

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