Jointly published by Akadémiai Kiadó, Budapest
and Springer, Dordrecht

Scientometrics,
Vol. 64, No. 2 (2005) 235–246

The influence of publication delays on impact factors

GUANG YU,[a] XIAO-HONG WANG,[a] DA-REN YU[b]

[a] *School of Management, Harbin Institute of Technology, Harbin (P. R. China)*
[b] *School of Energy, Harbin Institute of Technology, Harbin (P. R. China)*

Based on the convolution formula of the disturbed aging distribution (EGGHE & ROUSSEAU, 2000) and the transfer function model of the publishing delay process, we establish the transfer function model of the disturbed citing process. Using the model, we make simulative investigations of disturbed citation distributions and impact factors according to different average publication delays. These simulative results show that the bigger increment the average publication delays in a scientific field, the larger shift backwards of the citation distribution curves and the more fall the impact factors of journals in the field. Based on some theoretical hypotheses, it is shown that there exists theoretically an approximate inverse linear relation between the field (or discipline) average publication delay and the journal impact factor.

Introduction

The impact factor (*IF*) reflects the number of citations of a journal' materials in the preceding several-year (two-year being generally used) period divided by the number of papers published by that same journal within the same period. The Garfield impact factor in the year Y was defined as:

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

The impact factor over different periods was defined as (EGGHE & ROUSSEAU, 1990):

$$IF_n(Y) = \frac{\sum_{i=1}^{n} CIT_{y-i}(Y)}{\sum_{i=1}^{n} PUB(Y-i)} \tag{2}$$

From Eq. 1 and Eq. 2, a journal impact factor is based on 2 elements: the numerator, which is the number of citations in the current year to any items published in a journal

Received March 3, 2005

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

0138–9130/US $ 20.00

in the previous *n* years, and the denominator, which is the number of substantive articles (source items) published in the same *n* years (GARFIELD, 1999). Actually, *IF* may be influenced by some factors, including a number of articles published in this journal, the scientific field to which the journal belongs, the journal's self-citation rate, the language of published articles, the number of articles published by the journal in a given period of time, publication delays of a scientific field and a journal, etc. In recent years, it has caught some scientist's attention that too long publication delays would badly influence the citation distributions and a journal impact factor. LUWEL & MOED (1998) studied preliminarily the influence of publication delays on the age distribution of the references and regarded that the ratio of three to two year old references is the best predictor for the publication delay. GARFIELD (1999) proposed that the time required to review manuscripts may also affect *IF* and references to articles that are no longer current may not be included in the impact calculation if reviewing and publication are delay. RAY et al. (2000) referred that time to publication had a weak negative correlation with the journal impact factor. MARCHI & ROCCHI (2001) presumed that there were an inverse relationship between the publication delay and the impact factor. CHEN & XU (2002) regarded that longer publication delay would reduce *IF* and have a bad influence on the innovation of S&T productions. ORIVE (2003) referred that lag times are a measure of the editorial agility of a journal that aspires to be at the forefront of one research field. EGGHE & ROUSSEAU (2000) found that the variation in publication delays of papers citing a given paper may affect evaluation of aging, and then proposed that the observed aging distribution was a result of convolution of the "undisturbed" aging distribution and the publication delay distribution. In this paper, we should quantificationally research the effects of publication delays on citation distributions and impact factors. The "notion" publication delay is defined here as the average publication delay of articles published in one scientific discipline (or field), the citation distribution is regarded as the age distribution of references in the discipline (or field), the impact factor is regarded as the impact factor of one journal in the discipline (or field) or the average discipline (or field) impact factor.

The delay effect of literature citation

An actual citing behavior occurs at the same time of creating manuscripts not publishing. Though peoples are accustomed to calculate citation ages from publication time, publication delays consequentially exist in citation distributions. EGGHE & ROUSSEAU (2000) pictured the research-citation cycle by Figure 1 and the role of publication delays on the citation of an article (Article A) are shown in Figure 1.

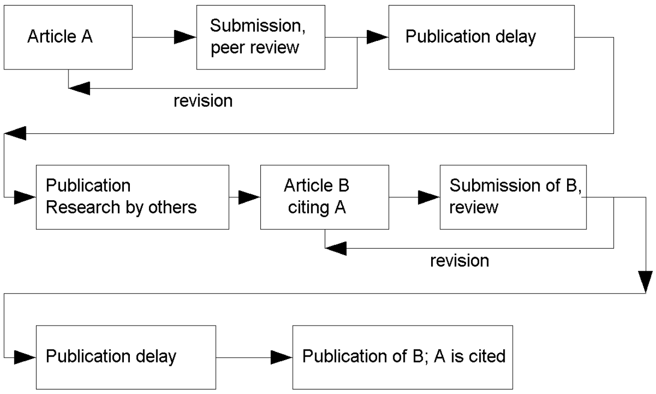

Figure 1. The research-citation cycle (EGGHE & ROUSSEAU, 2000)

In this paper, we call it as *the delay effect of literature citation* that the disturbed citation distribution results from the interplay of publication delay function and the undisturbed citation distribution. From the concept of impact factor (Eq. 1 or Eq. 2), it is composed by the number of citations in period of years, and relates with the number of cumulative citations in *n*–year period (usually two-year). So the delay effect of literature citation undoubtedly is transferred to impact factors and can be reflected by the disturbed citation distribution. Hence, according to the convolution expression of the observed aging distribution (EGGHE & ROUSSEAU, 2000) and the steady state solution of the mathematic model of publication delay process (YU et al., 2000), we can study the influence of publication delays on impact factors in the scientific field using citation analysis.

The transfer function model of the disturbed citing process

The actual citing and publishing processes of scientific literature are shown as Figure 2. In this article, we use the transfer function method to make the investigation for simplifying.

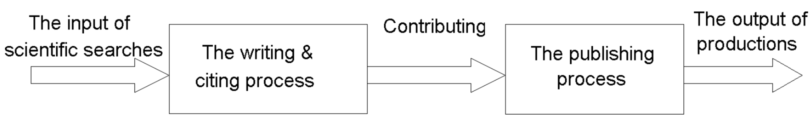

Figure 2. The citing and publishing processes of scientific literature

So an actual citing process can be regarded as a physical system as Figure 2, the output of the system changes along with its input change. The sketch solving the output of the disturbed citing system by Laplace transforms is shown in Figure 3, its input *X(T)* is citing action of authors and is described as: *total citing frequency/the total of citations=I*. It is a step function and dimensionless, its expression is:

$$X(T)=\begin{cases}0, T=0\\1, T>0\end{cases} \tag{3}$$

In Eq. 3, *T* is the literature age. In Figure 3, the output function *C(T)* is the disturbed citation distribution function; *C(s)* is the Laplace transform of *C(T)*; *s* is Laplace variable; $W_1(s)$ is a transfer function model of the "undisturbed" citing process; $W_2(s)$ is a transfer function model of the publishing process.

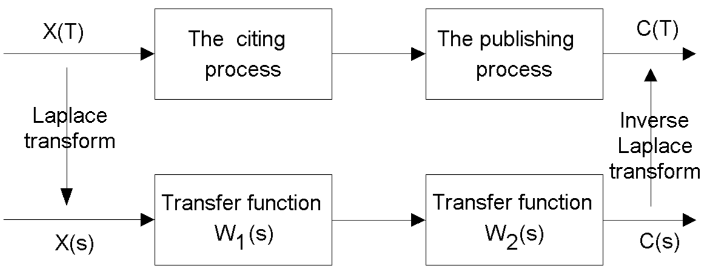

Figure 3. The sketch solving the output of the disturbed citing system by Laplace transforms

Transfer function model of the undisturbed citing process

We use a classical citation distribution model: Bernal negative exponential aging model as the undisturbed age distribution of references, its expression is

$$f(T)=K\cdot e^{-\alpha T} \tag{4}$$

In Eq. 4, *T* is the age of literature, *K* is a constant, α is the aging coefficient. By integrating Eq. 4, we obtain the cumulative citation distribution function (dimensionless):

$$C(T)=1-e^{-\alpha T} \tag{5}$$

Eq. 5 is the output function of *X(T)*, namely a step response. Based on Laplace transform theory, when a step response is Eq. 5, the transfer function model of the process should be the Laplace transform of Eq. 4. Letting $T_1 = 1/\alpha$, so we get the transfer function model of the undisturbed citing process:

$$W_1(s) = \frac{1}{T_1 s + 1} \tag{6}$$

In Eq. 6, T_1 is called as time constant and nearly related to the aging coefficient α.

The transfer function model of the publication delay process

YU & YU (1996) established transfer function models of the publication delay process by identification method, a first (or second) order plus time delay transfer function model, and the parameters of models were identified using the publication delay datum of six Chinese journals. In this paper, we choose one first order transfer function model:

$$W_2(s) = \frac{e^{-\tau s}}{T_s s + 1} \tag{7}$$

In Eq. 7, τ is pure delay and T_s is time constant. It is need to be illuminated that YU et al. (2000) established the partial difference equation model of the process and deduced the steady state solution – the delay (or age) distribution of the relative published flux:

$$y(T) = \frac{Y^2}{N} e^{-(T-\tau)/(N/Y)} \tag{8}$$

By integrating Eq. 8, we obtain the delay (or age) distribution of the cumulative publication flux:

$$Y(T)/Y = 1 - e^{-(T-\tau)/(N/Y)} \tag{9}$$

Eq. 9 is the output of the publishing process. In Eq. 8 and Eq. 9, N is the deposited contribution quantity at steady state, Y is the published contribution flux at steady state, T is the age (or delay) of literature. So the transfer function model of the publishing process can be expressed as:

$$W_2(s) = \frac{e^{-\tau s}}{(N/Y)s + 1} \tag{10}$$

Comparing Eq. 7 and Eq. 10, we get the same result as YU et al. (2000): $T_s = N/Y$. According to the indicator of journal average publication delay (YU et al., 2004): $\overline{T} = \tau + N/Y$, we get the relation between T_s and $\overline{T}$:

$$\overline{T} = \tau + T_s \tag{11}$$

The transfer function model of the disturbed citing process

Based on mathematical considerations about the observed age distribution of references proposed by EGGHE & ROUSSEAU (2000), the disturbed citation distribution function $c(T)$ can be obtained by the convolution combination of the "undisturbed" citation distribution function $f(T)$ and the publication delay function $y(T)$. According to the definition of convolution operation:

$$(f_1 * f_2)(x) = \int_0^x f_1(u) f_2(x-u) du,$$

the expression of $c(T)$ is

$$c(T) = (f * y)(T) = \int_0^T f(u) \cdot y(T-u) du \tag{12}$$

According to integral transform principle, the convolution of two functions in time region equals the inverse Laplace transform of a product of two transfer functions in the frequency domain (FANG & XIAO, 1988), namely $L^{-1}[W(s)] = L^{-1}[W_1(s) \cdot W_2(s)]$. Then we obtain the transfer function model of the disturbed citing process:

$$W(s) = \frac{e^{-\tau s}}{(T_1 s + 1)(T_s s + 1)} \tag{13}$$

In Eq. 13, T_1 is one time constant related with the aging coefficient α, T_s is another time constant related with the average publication delay. According to Eq. 13, we can simulate this operation using MATLAB SIMULINK (ZHONG &WANG, 2004).

Simulation results

In Figure 4, curve 1 and curve 2 display the cumulative citation distribution disturbed by publication delays at $\overline{T} = \tau + T_s = 0.3 + 0.2$ (years), and $\overline{T} = \tau + T_s = 0.3 + 1.2$ (years), respectively, when $T_1 = 2$ (equally the aging coefficient $\alpha = 0.5$). It is proved that the larger the average publication delay $\overline{T}$, the larger backwards shift of the disturbed distribution curve. Because τ is invariable, the initial points of curves don't change. In Figure 5, curve 1 and curve 2 are the disturbed age distribution of citations of curve 1 and curve 2, respectively, of Figure 4; curve 3 is an undisturbed age distribution of references ($\overline{T}=0$). The peak value of citation distribution curve falls and shifts backwards along with increasing average publication delays. From point of view of aging, the larger the average publication delay, the larger fall the cited frequency and the slower the literature-aging rate, and the cited half-life of the scientific discipline would be prolonged.

In Figure 6, curve 1 and curve 2 are the cumulative citation distribution disturbed by different publication pure delays ($\overline{T}=\tau+T_s=0.3+0.5$ (years) and $\overline{T}=\tau+T_s=0.6+0.5$ (years)), respectively, when $T_1=2$ (equally the aging coefficient 0.5), curve 3 shows that $C(2)$ equals zero when $\tau=2$ years and it is predicated that impact factor equals zero in this instance. In Figure 7, curve 1 and curve 2 are the disturbed age distribution of citations of them, and curve 2 shifts backwards 0.3 years along the age axis.

According to simulation results, the change of τ causes the shift of the distribution curve, and the increase of time constant $T_s = N/Y$ causes that the peak value of the distribution curve drops and shifts backwards.

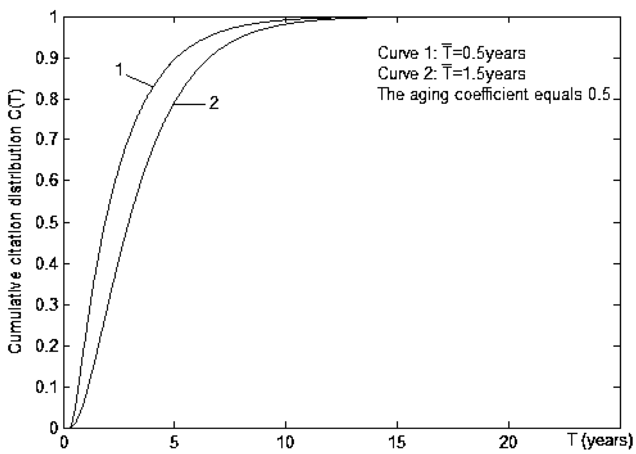

Figure 4. The cumulative citation distribution curves disturbed by different publication delays

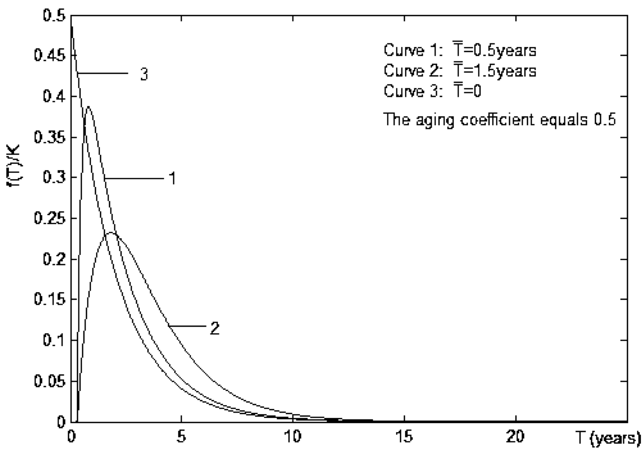

Figure 5. The relative age distribution curves of references disturbed by different publication delays

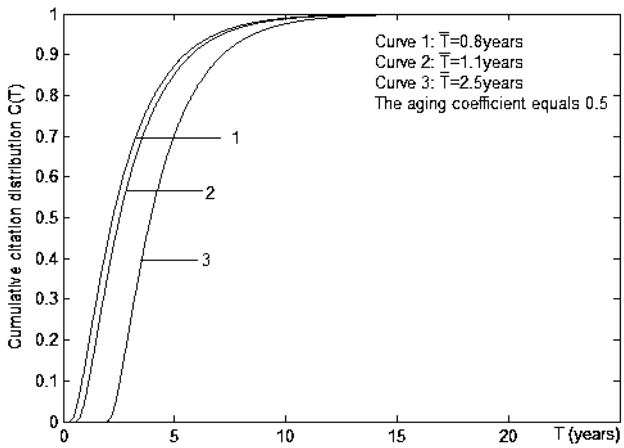

Figure 6. The cumulative citation distribution curves disturbed by different publication delays

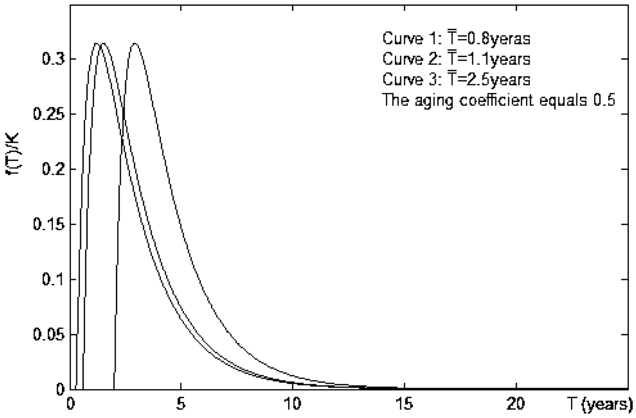

Figure 7. The relative age distribution of references curves disturbed by different publication delays

The simulative calculation of disturbed impact factor

We assume that the impact factor over different periods and these distribution functions are the continuous functions of age *T*. According to the concept of the journal impact factor, its expression is rewritten by combining of Eq. 2 and Eq. 6 as follows:

$$IF(T) = \frac{(C(T) - C(0)) \cdot \lambda}{\int_0^T PUB(T)dT} \tag{14}$$

In Eq. 14, λ is the total cited frequency of one journal, T is literature age which is zero at the statistical year, $C(T)$ is the cumulate citation distribution of the journal, $C(0)$ is the cited frequency rate at the statistical year, $PUB(T)$ is the published paper quantity at age T.

Above all, we should make some *theoretical hypotheses*: firstly, the function $C(T)$ is a continuous function and as same as the citation distribution mode of the discipline (or field) covering the journal; secondly, the single-factor method is used, namely the publication delay is regarded as a main factor influencing the impact factor and other factors are neglected, for example, journal productions, total citation quantity, etc.; thirdly, in terms of the definition of Impact Factor, $C(0)$ is not included in the calculation of *IF*, so the initial conditions are hypothesized: $C(0)=0$, $PUB(0) = 0$.

Letting, $\lambda' = \dfrac{\lambda}{\int_0^T PUB(T)dT}$

Eq. 14 is rewritten:

$$IF(T) = \lambda' \cdot C(T) \tag{15}$$

Therefore λ' is regarded as a constant that related with the total cited frequency and the published literature flux of the journal, *C(T)* can be obtained by above simulation method. In this paper, letting T = 2 years,
Eq. 15 is rewritten:

$$IF(2) = \lambda' \cdot C(2) \tag{15'}$$

In Eq. 15', $\lambda' = \frac{\lambda}{\int_0^2 Y(T)dT}$, *C*(2) can be solved by above simulation method. Because our aim is to find a relation between the publication delay and the disturbed impact factor, we hypothesize $\lambda' = 1$ for predigesting, and the influence of the publication delay on *IF*(2) is reflected by *C*(2).

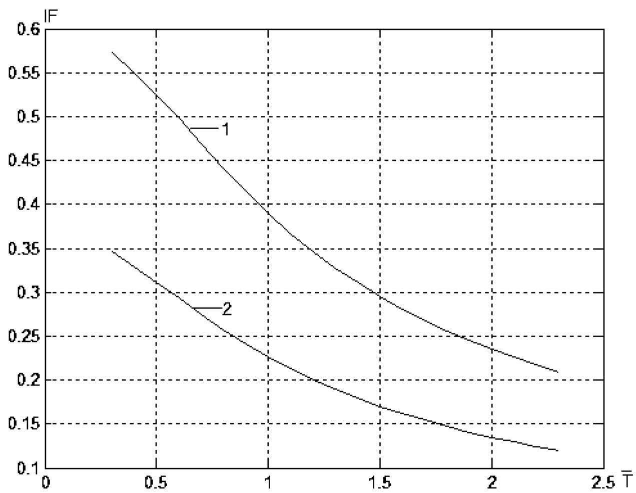

Figure 8 The relation between the average publication delay and *IF*(2)

In Figure 8, when the average publication delay $\overline{T}$ continuously change from 0.3 to 2.3 years ($\tau = 0.3$ years), curve 1 and curve 2 are respectively showed the relation between journal *IF* and the average publication delay at $T_1 = 2$ (the aging coefficient is 0.5) and T_1=4 (the aging coefficient is 0.5). According to curve1 and curve 2, when $\overline{T}$

increases, journal impact factors decrease and approximate to linearity descending. It is shown that the less the average publication delay ($\overline{T} = 0.3$ to 1.0), the rapider change the impact factor. In the other hand, aging coefficient is a important factor to influence on the impact factor, in the instance of equal average publication delay and equal published literature flux, the bigger the aging coefficient (curve 1, $T_1 = 2$), the higher the impact factor; contrariwise the lower the impact factor. When $\tau \geq 2$ years, $C(2) = 0$ and *IF* is zero too.

Conclusion

The larger the average publication delay $\overline{T} = \tau + T_s$ in a scientific field, the larger the shift in the age distribution curve of references, and the peak value of age distribution curve falls and shifts backwards. If the pure delay τ were increased the age distribution curve would be shift parallel backwards. If the time constant $T_s = N/Y$ were increased, the cited half-life of the scientific discipline would be prolonged.

This influence of publication delays on impact factors is serious. The larger the average publication delay in a scientific field (or discipline), the more fall the impact factor of a journal in the field (or discipline). The relation between the impact factor and the average publication delay approximates to linearity descending, and the impact factor is trend to zero along with the delay increasing. An inverse relationship between the publication delay and the impact factor (MARCHI & ROCCHI, 2001) is proved in theory. When the pure delay is bigger than 2 years, namely $\tau \geq 2$ years, the impact factor is zero.

Apparently, if the average publication delay of any journal increases, impact factors of other journals that refer to articles of this journal will fall, and then the delay will transfer to self-citing process and its impact factor will fall too. It is special important to research and control publication delays in the scientific field.

*

Project (HIT.200081) supported by the Scientific Research Foundation of Harbin Institute of Technology.

References

CHEN, F.J., XU, F. (2002), The main impact factors of innovation level of Chinese scientific periodicals, *Journal of the China Society for Science & Technical Information*, 21 (2) : 242–249.

EGGHE, L., ROUSSEAU, R. (1990), *Introduction to Informetrics*, Elsevier Science Publishers, pp. 263-265.

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.

FANG, C. Z., XIAO, D. Y. (1988), *Process Identification*, Beijing: Tsinghua University Press.

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

MARCHI, M., ROCCHI, M. (2001), The editorial policies of scientific journals: Testing an impact factor model. *Scientometrics*, 51 (2) : 395–404.

ORIVE, G. (2003), Reflections on the Impact Factor, *Archivos de Bronconeumología,* 39 : 409–417.

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

RAY, J., BERKWITS, M., DAVIDIFF, F. (2000), The fate of manuscripts rejected by a general medical journal, *The American Journal of Medicine,* 109 : 131–135.

YU, G., YU, D. R. (1996), Identification of the dynamic model of journal literature publishing process. *Journal of the China Society for Science & Technical Information,* 15 (2) : 146–153.

YU, G., YU D. R., RONG, Y. H. (2000), The mathematical models of the periodical literature publishing process, *Information Processing & Management*, 36 (3) : 401–414.

YU, G., YU, D. R., LI, Y. J. (2004), The universal equations of periodical average publication delay at steady state, *Scientometrics,* 60 (2) : 121–129.

ZHONG, L., WANG, F. (2004), *MATLAB Simulation Technology and Application Tutorial*. Beijing: National Defence Industry Press.