Journal Diffusion Factors – a measure of diffusion?

Tove Faber Frandsen

Journal Diffusion Factors – a measure of diffusion?

Tove Faber Frandsen
Royal School of Library and Information Science, Birketinget 6, DK 2300 Copenhagen S, Denmark
kk02tofa@db.dk

Keywords

Abstract

In this paper we show that the measure of diffusion introduced by Ian Rowlands called the Journal Diffusion Factor (JDF) is highly negatively correlated with the number of citations, leading highly cited journals to get a low JDF, whereas less cited journals gets a high JDF. This property reduces the utility of the JDF as a tool for evaluation of research influence. We present a new definition of the JDF in order to attempt to improve the JDF. This new JDF corrects the strong correlation with the number cites, but has a strongly statistically positive correlation with Journal Impact Factors (JIF). However, the new JDF may still be used as an evaluation tool since, for journals with similar JIF values, the new JDF can be used to differentiate between them. Thereby, journal evaluation will be based on more than one aspect of journal influence when assessing journal influence with similar Journal Impact Factor values.

Introduction

The Journal Impact Factor (JIF) has, according to Amin and Mabe (2000), developed from being an obscure bibliometric indicator to being the central quantitative indicator in measuring journal quality. JIF is often used as the only indicator of journal quality in evaluations of journals, articles or researchers. But JIF is much disputed and some critics recommend that it should not be used in evaluations (e.g. Seglen, 1997); others suggest improving or supplementing it. Rousseau (2002) stresses that journal quality is a multifaceted notion and points out that a whole battery of indicators seems preferable. Bonnevie (2003) provides such a multifaceted portrait of a journal. Ingwersen *et al.*, (2000) suggest applying diachronic citation analysis instead of the synchronous method applied by ISI® as they are immediately understandable and informative also to those being evaluated. A supplement to JIF has been suggested by Rowlands (2002) as he introduces the Journal Diffusion Factor (JDF) as a new approach to measuring research influence. The JDF is the average number of citing journals per 100 source citations within a given time window. To describe the relevance of the JDF Rowlands uses a metaphor of the ripple effect. The size of the splash when pebbles (articles) are thrown into water (the scientific community) is measured by the JIF. But JIF does not capture the extent of the ripples that flow from a particular journal. This requires a description of breadth rather than impact of the journal's citations and JDF is an attempt to describe breadth in the form of a simple indicator like the JIF. JIF describes one

facet of journal influence being the amount of citation activity and JDF another facet being the degree of specialisation of the journal.

Rowlands finds no connection between impact and breadth, but perhaps it is possible to further develop the JDF? In the different modes of influence Rowlands shows that both impact and diffusion can be low as well as high and he argues that they represent different vectors of the citation reception process. But isn't there a connection between impact and breadth? Isn't it more likely that a highly cited journal reaches further out into the scientific community than a less cited journal? A hypothetical example can illustrate this connection between impact and breadth[1]. Consider Journal A from the economics subject category receiving 1000 citations from 100 different journals within a given time window. The JDF of Journal A will then be 10. Consider then Journal B from the same subject category receiving 10,000 citations within the same time window. To achieve the same JDF as Journal A, Journal B has to be cited by 1000 different journals.

The economics subject category in Social Sciences Citation Index (SSCI) includes only 166 journals. Therefore we have to assume that it is easier to be cited by 100 different economics journals than by 1000. In this example the first 166 different citing journals are easier to achieve than the following citing journals that would be outside the cited journal's home discipline. An economics journal cited by 1000 different journals is extremely rare. So it appears that we can expect journals receiving many citations to have a low JDF. Although varying in size this is also the case in other subject categories.

The purpose of this investigation is to confirm the correlation between the JDF and number of citations empirically. We define the JDF suggested by Rowlands and, furthermore, we attempt an improvement of the JDF by defining a slightly altered JDF measure.

In the next section we will present the empirical data for an investigation of the correlation. In addition we suggest another measure of breadth defined as the number of different citing journals per article. We show how the new definition is statistically independent of the number of received citations. Finally we conclude.

Data

Citation analysis is an applicable tool for the sciences dominated by serial publications according to Garfield (1998). Pierce (1992) and Whitley (1991) analyse the publication activity in economics and find it to be dominated by journals. Twenty-eight journals are selected from the subject category 'economics' in SSCI and included in this investigation. This selection of journals consists of the top 25 journals chosen among economists in a survey about their views on economics publications (Bräuninger and Haucap, 2001). The selection is supplemented by three journals of special interest to this analysis. Two health economics journals are included as they have high impact but are ranked low in the survey. A Scandinavian economics journal is included as it has low impact but a tradition of a high

[1] The example here is hypothetical but not unrealistic as the data computed in this investigation showed huge differences in the number of citations between journals ranging from a little over 400 to more than 16,000 citations in a given time window.

degree of breadth. Journals that are core journals in other related disciplines are excluded according to Kalaitzidakis *et al.*, (2001). The list consists of journals with high as well as low impact. A note on the data has to be made as the data set constitutes journals from one discipline only and consists of 16.9 per cent of the journals in the present subject category in question. Therefore, we must take into consideration that the conclusions possibly cannot be generalised. However, it is our belief that these analyses, in spite of the data size, contribute to the investigation of journal breadth.

For these analyses the three Dialog Classic implementations of Arts & Humanities Citation Index (A&HCI), Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) have been used. All three databases have been used, as citations received from journals outside the home discipline are just as interesting as those from within the home discipline. When we look at a measure of breath, we compute three different JDFs as we choose three different time windows. The first JDF is determined according to Rowlands' definition and is a synchronous JDF with two years of publication window and one-year citation window. The next is a JDF with five years publication and citation window and the last is a JDF without publication window and a one-year citation window[2]. The first two JDFs include only the document types review and article, as suggested by Rowlands. In the latter calculation of JDF we include the document types note and letter. Note is included in the ISI calculation of JIF, and Christensen *et al.*, (1997) recommend including letters to assure a more detailed picture. This search limit is not shown in the examples.

The synchronous JDF with two years of publication window and one-year citation window is determined as follows.

S s1(s)cy=1998:1999/2000

A cited journal title may be represented by many different abbreviations across a set of journals, in articles and even within a single article. Therefore we EXPAND on the Cited Work field to identify the different versions of the journal title. Then we remove the duplicates with the *RD* command and get the following search result by using the command *set postings on*.

Set	Items	Postings
S3	210	890

When determining the JDF we use a correction (Christensen *et al.*, 1997) and the total number of citations is determined as (890-210)/2=340.

The number of different citing journals is determined by ranked output of the search set as follows.

Rank fields found in 210 record - - 112 unique terms

[2] We choose three different time windows as Rowlands argues that the size of the time window can vary as well as both synchronous and diachronous methods can be employed. We show that the main point of this paper is not limited to just one computation.

The JDF is therefore determined as 112*100/340 = 32.94.

The JDF with five years publication and citation window is determined as the synchronous JDF with two years of publication window and one-year citation window except that the first search string is replaced with the following

S cw=economica(s)cy=1996:2000/1996:2000

The JDF without publication window and a one-year citation window is determined slightly differently. The first search string is replaced with the following

S cw=economica/2000

When correcting the outcome we do not divide by two as only the *CW* data occurs in the same citation. Normally we divide with two as both the *CW* and *CY* data occur in the same citation.

The Rowlands JDF

Here we define the JDF suggested by Rowlands as follows. Please note that this definition is general and can therefore be applied in both synchronous as well as diachronic analysis. Diachronic analyses consisting of several publication periods each with citation windows of identical length but of different starting points are also possible within this definition if calculate a JDF of each publication period and take the average.

Rowlands JDF =
$$JDF\ (n_p, n_c, y_p, y_c, j) = \frac{R(k,i,j) * 100}{\sum_{i=y_p}^{y_p+n_p-1} \sum_{k=y_c}^{y_c+n_c-1} Cy(k,i,j)}$$

n_p	equals the length of the publication period measured in years,
n_c	equals the length of the citation window measured in years,
y_p	is the beginning year of the publication period,
y_c	is the beginning year of the citation window,
i	is the publication year(s),
k	is the citation year(s),
j	is the cited journal under investigation,
Cy(k,i,j)	is the number of citations that the documents published in year(s) *k* of the journal *j* receives in the year(s) *i.*
R(k,i,j)	is the number of different journals that cites the documents published in year(s) *k* of the journal *j* in the year(s) *i.*

Table I shows that the choice of time window does not determine the outcome when calculating the JDF. The three different JDFs are strongly statistically dependent and thus measure the same phenomenon.

Take in Table I Different Journal Diffusion Factors correlates (Pearson coefficients)

Rowlands finds no statistically significant correlation between JDF and other well-known measures of research influence and neither do we. Table II shows that JDF is independent of JIF, Immediacy Index and cited Half-life. But Table II also shows that JDF and number of citations are statistically dependent. Hence the point of the hypothetical example with Journal A and Journal B in the introduction can be empirically supported.

Take in Table II Journal Diffusion Factor correlates (Pearson coefficients)

However, all three JDF measures show a strong negative correlation with the number of citations received by the journal. No matter the choice of time window the highly cited journals will have a low JDF. Roughly speaking the number of citations determines the JDF of a given journal due to the definition of the JDF.

The numerator (the number of different citing journals) of the fraction cannot vary as much as the denominator (the number of citations) and therefore the highly cited journals will be in a situation of low JDF, as any extra citations from an already citing journal will reduce the JDF. As the denominator of the fraction in the definition of JIF is the number of citations, one can expect a hyperbola and not a linear correlation. The Pearson coefficient can only measure the degree of a linear correlation; therefore we could expect the statistical correlation to be even stronger than suggested by the Pearson coefficients if we allowed for a non-linear relationship.

The correlation between the JDF and the number of citations is also evident in Figure 1, which illustrates the correlation between a synchronous JDF with two years of publication window and one-year citation window and the number citations. This shows that the JDF of a journal is strongly influenced by the number of citations the journal receives.

Take in Figure 1. Two year JDF and number of citations

Introducing a new measure of diffusion

As we saw above the JDF introduced by Rowlands has a built-in injustice to the highly cited journals. The question now is if we can rectify the imbalance and define a new but similar measure? Rowlands leans against the JIF when defining the JDF. But why not lean even more? If we replace the number of citations with the number of publications in the JIF

fraction we get a measure which can be verbally expressed as the average number of different journals an average article is cited by within the given time window[3]. The JDF suggested here is to be calculated as the JIF only replacing the number of citations with the number of different citing journals and we define it as follows:

New JDF = $$JDF\ (n_p, n_c, y_p, y_c, j) = \frac{R(k,i,j)}{\sum_{i=y_p}^{y_p+n_p-1} Py(i,j)}$$

n_p	equals the length of the publication period measured in years,
n_c	equals the length of the citation window measured in years,
y_p	is the beginning year of the publication period,
y_c	is the beginning year of the citation window,
i	is the publication year(s),
k	is the citation year(s),
j	is the cited journal under investigation,
$Py(k,j)$	is the number of citable units published in the year(s) *k* of the journal *j*.
$R(k,i,j)$	is the number of different journals that cites the documents published in year(s) *k* of the journal *j* in the year(s) *i*.

Again we can vary the time window and both synchronous and diachronic methods can be used. By dividing with the number of publications instead of the number of citations we can avoid the injustice and still keep the principal idea of the measure. Another attractive feature is that extra citations from an already citing source do not alter this measure.

As above we employ a synchronous JDF with two years of publication window and one-year citation window and a JDF with five years publication and citation window. The JDF without publication window and a one-year citation window cannot be computed as the lack of publication window makes it impossible to determine the denominator of the fraction.

As we alter the denominator of the fraction we need to replace some of the search strings. The new search strings to determine the number of publishable units are as follows. The first is the search string to compute the synchronous JDF with two years of publication window and one-year citation window and the latter the search string to compute the JDF with five years publication and citation window.

S jn=economica and dt=(review or letter)/1998:1999

S jn=economica and dt=(review or letter)/1996:2000

[3] Notice the strong resemblance to the verbal expression of JIF which is the average number of citations an average article receives within the given time window.

This new definition of the JDF is not determined by the choice of time window. The two new JDFs employed here are strongly correlated as the Pearson coefficient is 0.99 and highly significant at the 0.01 level when we investigate the degree of linear correlation between the two measures.

This new definition of the JDF is statistically independent of the number of citations. As Table III shows the number of citations does not determine this measure.

Take in Table III New Journal Diffusion Factors correlates (Pearson coefficients)

When introducing this new definition of the JDF the strong correlation with the number of cites is no longer problematic. Another correlation exists though between the new JDF and JIF. This may be caused by resemblance in calculation method or resemblance in the phenomenon analysed. One can argue that the definitions of the two measures are alike to a degree that makes the new JDF unable to give us new information on the breadth of journals. However, it is our belief that the focus of the new measure is different from the one of JIF, as JIF focuses on impact rather than breadth.

The dependence could also be caused by the connection between impact and breadth. The two measures may describe the simple phenomenon that highly cited journals also reach further out in the scientific community. This will require further investigations. However, it is our belief that the new measure can provide supplementing information on journal influence, as we will show in Table IV.

Although the new JDF is strongly correlated with the JIF, the measure may still be valuable. Journals with almost the same JIF can be differentiated further with the JDF. In Table IV we see three journals that are differentiated further by adding information about diffusion.

Take in Table IV New Journal Diffusion Factor and JIF

The three journals have almost the same JIF but can be further differentiated by the new JDF, as there are marked differences in the new JDFs. The second journal has a JIF of 1,05 but a new JDF of 1,23 indicates another facet of influence in relation to the first and last journal that have similar new JDFs but the first with a slightly higher JIF.

Conclusion

This paper has demonstrated that the JDF measure introduced by Rowlands is highly correlated with the number of citations, leading highly cited journals getting a low JDF whereas less cited journals get a high JDF. The number of received citations is decisive for the JDF, which can then hardly be used as an adequate measure of breadth. This reduces the utility as a tool for evaluation of research influence.

We presented a new definition of the JDF in order to attempt to correct the dependence of number of cites. This new JDF corrects the strong correlation with the number cites. But then we also showed that this new JDF has a strong statistically positive correlation with JIF. Still the new JDF can be used as an evaluation tool. For journals with similar JIF values, the new JDF can be used to differentiate them further. When evaluating journals, the new JDF can be employed to assure an evaluation based on more than one aspect of journal influence when assessing journal influence of journals with similar JIF values.

References

Amin, M. and Mabe, M. (2000), "Impact factors: use and abuse" *Perspectives in Publishing*, Vol. 1.

Bonnevie, E. (2003), "A multifaceted portrait of a library and information science journal: the case of the Journal of Information Science", *Journal of Information Science*, Vol. 29 No. 1, pp. 11-23.

Bräuninger, M. and Haucap, J. (2001), "Was Ökonomen lesen und schätzen: Ergebnisse einer Umfrage" *Perspektiven der Wirtschaftspolitik*, Vol. 2, pp. 185-210.

Christensen, F.H. et. al. (1997), "Online determination of the journal impact factor and its international properties", *Scientometrics*, Vol. 40 No. 3, pp. 529-540.

Garfield, E. (1998), "From citation indexes to informetrics: is the tail now wagging the dog? *Libri*, Vol. 48, pp. 67-80.

Ingwersen, P. and Christensen, F.H. (1997), "Data set isolation for bibliometric online analysis of research publications: fundamental methodological issues", *JASIS*, Vol. 1, pp. 205-217.

Ingwersen, P., Larsen, B. and Wormell, I. (2000), "Applying diachronic citation analysis to research program evaluations", *Web of Knowledge – A Festschrift in Honor of Eugene Garfield*, pp. 373-387.

Pierce, S.J. (1992), "On the origin and meaning of bibliometric indicators: journals in the social sciences", *JASIS*, Vol. 43 No. 7, pp. 477-487.

Rousseau, R. (2002), Journal evaluation: technical and practical issues", *Library Trends*, Vol. 50 No. 3, pp. 418-439.

Rowlands, I. (2002), "Journal diffusion factor: a new approach to measuring research influence", *Aslib Proceedings*, Vol. 54 No. 2, pp. 77-84.

Seglen, P.O. (1997), "Why the impact factor of journals should not be used for evaluating research", *British Medical Journal*, pp. 314-497.

Whitley, R. (1991), *The organisation and role of journals in economics and other scientific fields. Working Paper 204*, Manchester Business School, Manchester.

Acknowledgements

It is a pleasure to acknowledge the helpful advice and support of Ellen Bonnevie and Rasmus Højbjerg Jacobsen to whom the author is greatly indebted. The author would also like to thank Peter Ingwersen and Birger Larsen for inspirational comments and suggestions.

Table I Different Journal Diffusion Factors correlates (Pearson coefficients)

	A JDF with 2 years of publication window and 1-year citation window	A JDF with 5 years publication and citation window
A JDF with 2 years of publication window and 1-year citation window	---	0,83*
A JDF without publication window and a 1-year citation window	0,65*	0,73*

Note: * Correlation is significant at the 0.01 level.

Table II Journal Diffusion Factor correlates (Pearson coefficients)

	JIF	Immediacy Index	Cited Half-life	Citations
A JDF with 2 years of publication window and 1-year citation window	-0.03	-0.27	-0.14	-0.57*
A JDF with 5 years publication and citation window	-0.02	-0.26	-0.30	-0.65*
A JDF without publication window and a 1-year citation window	-0.18	-0.40**	-0.64*	-0.69*

Notes: * Correlation is significant at the 0.01 level; ** Correlation is significant at the 0.05 level.

Table 3 New Journal Diffusion Factors correlates (Pearson coefficients)

	Cites	JIF
A synchronous new JDF with 2 years of publication window and 1-year citation window	0.36	0.86*
A new JDF with 5 years publication and citation window	0.25	0.86*

Note: * Correlation is significant at the 0.01 level.

Table IV New Journal Diffusion Factor and JIF

	JIF	New JDF
Rand Journal of Economics	1,107	0,69
Journal of Law and Economics	1,05	1,23
Review of Economics and Statistics	1,031	0,63

Figure 1 2 year JDF and number of citations