Citations as a Measure of the Research Outputs of New Zealand's Economics Departments: The Problem of 'Long and Variable Lags'

John Tressler and David L. Anderson[1]

Abstract

The paper explores the merits of utilising citation counts to measure research output in economics in the context of a nationwide research evaluation scheme: the New Zealand Performance-Based Research Fund (PBRF). Citations were collected for all refereed papers produced by New Zealand's academic economists over the period 2000 to 2008, and used to estimate the time-lags in between publication and the flow of citations; to demonstrate the impact of alternative definitions of 'economics-relevant' journals on citation counts; and to assess the impact of citation measures on departmental and individual performance. We conclude that under certain scenarios around 60 per cent of papers received no citations over the period. Our findings suggest that the time-lags between publication and citation make it difficult to rely on citation counts to produce a meaningful measure of output in a PBRF-like research-evaluation framework, especially one based on the assessment of individual academics.

Introduction

The external evaluation of the research of university faculty has become an important part of academic life in many countries. The processes used to assess research differ significantly in the way in which evaluations are undertaken.[2] Peer-review panels or expert panels are used by a number of countries, but they are costly, both for the assessors and those being assessed. Advances in information technology and the quality of research databases have made citation analysis a more attractive alternative or complement to other assessment

1 University of Waikato; tressler@waikato.ac.nz (corresponding author); Queen's University.

2 For an in-depth overview of processes used in the evaluation of research generally as part performance-based research funding for universities, see OECD (2010).

processes. The primary purpose of this paper is to explore the merits of utilising citations to measure research output in economics in the context of a nationwide research-evaluation scheme.

To provide a context for the evaluation of citation analysis we discuss its potential use in relation to the New Zealand government's Performance-Based Research Fund (PBRF).[3] To assess the quality of research under this program, individual academic staff must submit research portfolios providing a review of their research output, of their contribution to the research environment and of peer esteem over the previous six years. These portfolios are assessed and scored by peer-review panels formed for each subject area. Individual scores are aggregated to provide assessments of subject areas in an institution and the tertiary institution overall. Individual scores are reported to the academics concerned, and subject and institution results are made public. PBRF evaluations of the quality of research were undertaken in 2003 and 2006, and will be undertaken in 2012.[4] The evaluations of research quality determine 60 per cent of the PBRF funding allocations to tertiary institutions, with the remaining 40 per cent based on research degree completions and external research income. In 2010 the results were used to distribute NZ$268 million of research funding; this is equal to 18 per cent of government funding to universities and roughly 9 per cent of total system-wide revenue. Furthermore, and equally important, the PBRF results are aggressively used by the winners in their formal and informal promotional material.

Unlike the Australian ERA process, PBRF peer-review assessments are not supplemented by citation analysis. Over time it is likely that counting citations will play a part. It is our view that after the upcoming 2012 PBRF round, pressure may mount for the government to consider a shift, at least in part, to a metric-based system. If this were to occur, citation counting is likely to be at the heart of any such scheme, given its widespread acceptance as a reliable measure of performance in the physical and biological sciences. This view is based, in part, on the tendency of New Zealand tertiary education policy to follow that of the UK and, to a slightly lesser extent, Australia. In both of these countries, citation counting is now partially incorporated into their nationwide research-evaluation schemes.[5]

In this paper we will explore the merits of the concerns of many in the social science community over the use of citations as a measure of research output.

3 A discussion of the key elements of the scheme can be found in Goldfinch (2003) and Hodder and Hodder (2010). Additional information can be found on the official website: www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF-/Resources/

4 The 2006 round was a partial round in which academics could carry forward their evaluation from 2003 or submit a new portfolio.

5 For details, see Research Excellence Framework (REF) (www.hefce.ac.uk/reserch/ref/) and Excellence in Research for Australia (ERA) (www.arc.gov.au/era).

In order to restrict the task to manageable proportions, we will focus on the discipline of economics. We have collected citation counts to all refereed papers produced by New Zealand's academic economists over the period 2000 to 2008 using the databases of the ISI/Web of Science (henceforth, the ISI) and, in a more limited fashion, Google Scholar. The data collected allows us to assess, among other things, the time-lags in economics between publication of an article and the flow of citations; the impact of alternative definitions of 'economics- relevant' journals; and a comparison of departmental and individual performance using both direct-citation measures and alternative schemes based on journal-specific weights. With respect to the latter, we utilise journal weights based indirectly on citation counts and on reputational surveys.

In general, our findings suggest that the conventional assessment period of six years may be acceptable from a departmental perspective due to averaging effects, but that this is too short a time period for individual assessment. Our work suggests that the output measures generated by citation counts are not highly correlated with traditional output measures based on journal impact factors.

Some doubts explored

Although citation counts have long been used, and generally accepted, in the physical and biological sciences (henceforth denoted as 'the sciences') as a proxy measure of research output, the applicability of this metric for estimating social-science research output is problematic.[6] Concerns have been expressed over purported differences in citation practices across the above-mentioned disciplines. This argument has at least two dimensions: major differences between the sciences and social sciences in the time-lag between the publication of an article and the commencement of a meaningful flow of citations to the article; and differences in the publication frequency and citing habits between the science and social-science disciplines that work to the disadvantage of the latter.

There is also a data-collection issue at play here. Historically, ISI focused on the sciences; it is only in the past few years that this organisation has started to aggressively expand the range of social-science journals for which it collects citations. This means that citation-based performance measures in the social sciences only capture a portion of all citations generated by researchers, especially those publishing in languages other than English and, of greater relevance to the New Zealand scene, to those publishing in regional journals

6 For example, see Centre for Science and Technology Studies (2007).

on regional issues. The latter is a major issue in small countries: governmental funding agencies generally wish to see a substantial degree of research performed on matters deemed to be of relevance to the nation-state. In the social sciences, this often results in articles that have greater interest to national or regional journals than to international journals. Therefore, if only the latter journals are in the database, researchers performing work with a regional focus will appear to be low or even non-publishers. Furthermore, even if the work is published in a journal included in the ISI list of recognised journals, papers discussing local issues are less likely to be cited than those addressing similar issues in a large-country setting.

Although the above issues are important, the primary problem with respect to the social sciences is likely to be the lengthy lag between the typical article's publication date and the commencement of a meaningful flow of citations. In order to demonstrate the importance of this matter, and to illustrate how it may arise, let us refer to the upcoming 2012 PBRF round. For all academic staff employed at the 2012 census date, the PBRF scheme will attempt to assess all research generated by these individuals over the period 1 January 2006 to 31 December 2011. If one were to introduce a measure designed to capture the number of citations generated by papers published over this six-year time period, it is quite apparent that the time-lag issue will be of great importance. If the lags are, say, on average two to three years, it means that much of the research performed over the six-year assessment period will be ignored by the PBRF scoring system; it also means that work published in the early years of the cycle will be deemed to be of greater value than work published at the end of the evaluation period (everything else being equal).

The lag issue creates a special problem for newly hired and newly minted PhDs. In addition to the time required to develop a research program, obtain necessary funding, prepare papers for submission to journals, and go through the review and publication process, we now have to add additional time to reflect the period between publication and a meaningful flow of citations. Even without the citation-lag issue, the PBRF scheme has been modified to treat new entrants (with limited prior experience) differently. In practice, it is widely recognised that institutions are shifting their hiring practices away from the inexperienced to those with a 'good' next-round PBRF-relevant publication record.[7]

All of the above is based on conjecture. We have not been able to find an empirical study of the citation practices in economics (or social science) that addresses the issues raised above. In this study we will attempt to shed some light on these matters, especially the time-lag issue. We employ data from New Zealand-

7 This statement is largely based on anecdotal evidence, but supporting evidence can be found in Cinlar and Dowse (2008).

based economists to generate estimates of the time pattern of citations based on alternative definitions of 'economics-relevant' papers, and, to a limited degree, on alternative citation-capturing schemes. We will also compare the output performance of economics departments and individual economists using direct citation counts and widely employed alternative measures. However, we will not attempt to compare citation practices in economics to those in the biological and physical sciences.

At this point we should mention that the economics literature on research-output measurement is dominated by the journal-based weighting approach (Macri and Sinha 2006).[8] The most common method for generating the desired journal weights is to count citations to each journal in the dataset, over a given time period, and then to divide the total by the number of articles contained in each journal over the same time period. This procedure yields an estimate that is commonly denoted as a journal's impact factor, and frequently assumed to be a measure of a journal's quality. This approach has been modified by a number of economists, through the use of iterative adjustment processes, to yield aggressive journal-weighting schemes that are widely used in the economics literature (Anderson and Tressler 2010). Alternatively, journal-based weighting schemes sometimes rely on 'expert opinion' such as that initially employed in the Australian government's Excellence in Research for Australia (ERA) scheme.[9] Regardless of the underlying approach, the resulting journal weights are applied to all articles in a given journal, and the resulting values are aggregated to arrive at departmental and, sometimes, individual output estimates. In the majority of cases, further adjustments are made to account for the length of the article in terms of the *American Economic Review* (*AER)* page-size equivalents, and to reflect each author's share in multiple-authored papers.

The primary reason for favouring the journal-based approach is a variant of the time-lag argument presented above. Given the desire to generate an estimate of the probable long-term impact of an individual's relatively recent output (say, one to six years), it is necessary to resort to proxy measures. It is generally accepted by economists that the best proxy available is the impact or quality of the journal in which the paper is published. Rephrased, if citations are viewed as the principal indicator of research impact or quality, then the best indicator of the expected number of citations to a paper over the long run is best approximated by the relative importance (however measured) of the host journal.

8 Following convention, we have restricted research output to cover only refereed articles in journals listed in *EconLit*. Rephrased, academic work disseminated in books, conference papers, reports and non-refereed publications are ignored in this study.

9 For details, see ERA's website: www.arc.gov.au/era

However, the merits of journal-based weighting schemes have recently been called into question by Starbuck (2005), Oswald (2007), Wall (2010) and Chang, McAleer and Oxley (2011). For the purposes of this paper, their findings can be summarised as follows: good papers (lots of citations) can be found in lowly ranked journals (relatively few citations), and poor papers (very few citations) can be found in highly ranked journals (many citations). Indeed, Chang, McAleer and Oxley (2011) found that over a 25-year period, approximately 26 per cent of the papers published in the top 40 economics journals (based on ISI's two-year impact factors) did not receive a single citation. Furthermore, they found that 20 per cent of the selected journals had 'zero citation' rates in excess of 40 per cent. All of this work suggests that journal-based impact factors may not yield a good estimate of an individual paper's long-term impact. If this is correct, the search for a better proxy inevitably leads one to explore the use of a direct-citation measure — the counting of citations to a given paper, over a given time period.

Data

We assembled three basic datasets for this study.

- Dataset1: citations — be they in ISI or Google Scholar databases — collected over a 10-year period, for all papers published in 2000 and 2001 by New Zealand's academic economists on staff as at either or both 15 April 2007 or 15 April 2009.[10] More formally, we counted all citations to these papers using both databases; and the citations were collected in early January 2011.[11]

The collection of citations using Google Scholar is relatively straightforward. On the other hand, generating citations from the ISI database requires a number of adjustments and exclusions. First, we restricted our search to citations from ISI-listed journals (that is, we excluded citations from conference papers). Second, one faces an age-old problem in economics: which journals are economics journals? We handled this matter in two ways: we created a 'broad' definition of economics by assuming that all articles published by New Zealand's academic economists in both *EconLit* and ISI-listed journals are relevant to the discipline.[12] We refer to data based on this definition as *ISIB*. The alternative approach was labelled as a 'narrow' definition of economics and is based on the

10 These staff census dates were chosen for pragmatic reasons: we had previously collected publication records for all academic staff employed on these dates. More specifically, we collected data on all permanent staff with the rank of Lecturer, Senior Lecturer, Associate Professor and Professor. We should also note that we used both staff lists to maximise the size of the sample.

11 For papers published in 2000, we collected cites generated over the period 1 January 2000 to 31 December 2009. For 2001 papers, the time period shifted forward by one year.

12 This decision rule was used by Coupe (2003) in deriving his 'Impact' measure. At the time of his study (2000), some 800 journals were listed in *EconLit*; of these, 273 were listed in ISI/JCR. However, only approximately 170 of these journals were listed as economics journals by Journal Citation Reports (JCR).

restrictive practice of recognising only articles published in journals listed as 'economics' in the Journal Citation Reports (JCR).[13] We refer to data based on this definition as *ISIN*. In practice, under the broad definition of economics we include a number of journals in the areas of urban studies and finance that are excluded from the narrow definition list. The third restriction used in our collection exercise was to eliminate self-cites by authors.[14]

- Dataset2: ISI citations attributable to papers published between 2000 and 2008 by the same group of New Zealand economists as noted above. In this case, we counted citations up to the end of 2010. This means that we have a time series of citations ranging from 11 years, for papers published in 2000, to three years for those published in 2008. It should also be noted that in order to restrict the analysis to manageable proportions, we have limited our citation-collection exercise in Dataset2 (and in Dataset3 to follow) to our broad definition of economics (ISIB) rather than the narrow version (ISIN). Our rationale for selecting the broad over the narrow definition of economics for ISI counting purposes is based on our understanding of the current PBRF scheme — that work in boundary areas (such as finance and urban studies) is generally recognised as 'economics-relevant'.[15] Furthermore, our preference for ISIB over Google Scholar is based on the widespread view that ISI (narrowly or broadly defined) is the 'gold standard database' (Chang, McAleer and Oxley 2011).[16]
- Dataset3: constructed to allow us to compare rankings of departments and individual economists using various citation measures with those generated by more traditional measures. For all 135 economists employed by New Zealand's eight university economics departments as at 15 April 2009, we constructed a record of all articles published by these researchers in *EconLit*-recognised journals, over the period 2003–08.[17] In order to restrict the scope of the study, we have arbitrarily selected only two journal-based weighting

13 Most of the well-known journal-based ranking schemes in economics are based on this restrictive definition of economics-relevant journals albeit, in a few instances, with a couple of additions from the finance area. For example, see Liebowitz and Palmer (1984), Laband and Piette (1994), Kalaitzidakis, Mamuneas and Stengos (2003 and 2010), and for their 'economics' journal weights, Kodrzycki and Yu (2006).

14 Self-cites are eliminated to prevent game-playing tactics. Although this is not considered to be a problem in economics, or the social sciences, at this time, it is widely recognised as a problem in the biological and physical sciences. Although we are getting ahead of ourselves, we should note that of the total citations to papers published by New Zealand's academic economists over the period 2000–2008, approximately 15 per cent (441 of 2857) were 'self-cites' (for the citing period ending 31 December 2010). At this time we wish to stress that throughout the remainder of this paper, unless otherwise noted, all references to 'citations' will always be to the 'non-self' variety.

15 For a discussion of these issues, albeit from a Finance perspective, see Cosme and Teixeira (2010).

16 Although Google Scholar is rapidly gaining academic credibility, it has been criticised for lack of transparency in design and scope. For a New Zealand/PBRF-related assessment of Google Scholar, see Smith (2008).

17 Following convention, we allocated shares to individual authors based on the 1/n rule (for example, if a paper has three authors, each is granted a one-third share), and used the size-adjusted page (AER equivalent) as our unit of output (see Macri and Sinha 2006).

schemes for comparison purposes: *KMS2010* to represent an aggressive scheme based indirectly on citation counts;[18] and *ERAB*, the Australian government's journal-weighting scheme based on 'expert opinion' (that is, a perception-based system).[19]

In order to demonstrate the importance of the time-lag issue, we constructed three citation measures.

- *ISIB03–08* is based on a simple count of citations over the period 2003–08 to all papers published over this same time period. This time span corresponds to the timeframe adopted by the PBRF — a six-year period. Therefore, papers published in the first year of the evaluation period are able to generate citations over a six-year period, whereas those published in the last year of the cycle have, at the most, one year to capture citations.
- *ISIB01–06A* is based on papers published over the period 2001–06, with citations collected from 2001 to the end of 2008; therefore, the maximum citation collection period is eight years, and the minimum is two years.
- *ISIB01–06B* is also based on papers published over the 2001–06 period, but the citation-collection period now ends on 31 December 2010. The maximum and minimum period for capturing cites is now 10 and four years, respectively.

Findings

Ten-year citation patterns for articles published in 2000 and 2001

Our first task is to shed light on the time-lag between publication and the generation of a meaningful stream of citations for papers produced by New Zealand's economists. Before doing so, we should note that the nation's 156

18 This weighting scheme was developed by Kalaitzidakis, Mamuneas and Stengos (2010); it is an update of their prior work (Kalaitzidakis, Mamuneas and Stengos 2003). This is an aggressive weighting scheme in that the weights given to the 'top' journals are as large as 1000-times that assigned to lower-end journals. For example, the first-place journal, the *AER*, receives a score of 100.0, whereas the 50th- (*Labour Economics*), 100th- (*Journal of Economic Geography*), and 150th- (*Economic Geography*) placed journals receive scores of 3.06, 0.73, and 0.12, respectively. For a more rigorous discussion of the aggressive nature of this scheme, see Henrekson and Waldenstrom (2009) and Anderson and Tressler (2010).

19 We have adopted a broad version of the ERA scheme (hence the reason for denoting it as ERAB). That is, we recognise all journals listed in both the ERA and *EconLit* regardless of the category that they have been arbitrarily placed in. In practice, this means that a number of papers in finance and urban-studies journals receive a non-zero weighting. Recall that under the narrow definition of economics selection process, these papers would have received a zero weighting. It should also be noted that the ERA officially uses a four-point grading scale: A+, A, B and C. We have arbitrarily converted it to a five-point scale: 4, 3, 2, 1, and 0 (the last score for journals not covered by the ERA scheme but included in *EconLit*).

academics employed by its eight university-based economics departments in 2007 and/or 2009, published 167 papers in *EconLit*-listed journals in 2000 and 2001. As shown in Table 1, over a 10-year period, the average paper received 4.1 and 3.0 ISIB and ISIN citations, respectively; the corresponding number for Google is much larger — 15.9. This difference is not surprising given that Google Scholar collects citations from working papers, public reports, conference papers and books, whereas ISI citations are only collected for, and generated by, JCR-listed journals.

Let us now explore the time pattern of citations. In the first year after publication, very few citations were generated for ISIB and ISIN — less than one per cent of the 10-year total. Indeed, by the end of Year 3, the corresponding estimate is only 6.7 and 7.2 per cent, respectively. It is clear that a relatively steady, but growing, stream of citations does not commence until Year 4. Interestingly, the flow does not abate over the remainder of the collection period as Year 10 citations are higher than those in any preceding year.[20] Google exhibits a somewhat similar year-by-year time pattern; however, the flows in Year 1 and Year 2 are larger, but still below the levels generated in Year 3 and onwards. Not surprisingly, we found the ISI and Google Scholar year-by-year citation patterns to be highly correlated: the Pearson Correlation Coefficients for ISIB/Google and ISIN/Google are 0.951 and 0.906, respectively. However, if we explore the relationship on a per-paper basis, the correlation coefficients are much lower: ISIB/Google 0.262 and ISIN/Google 0.483.

Table 1: Non-self citations for papers published in 2000 and 2001

Citation Scheme	Total	Cites/ Paper	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
ISI, Economics Broadly Defined	686	4.1	4	8	46	65	80	64	96	92	115	117
ISI, Economics Narrowly Defined	497	3.0	3	7	36	54	50	44	60	69	87	88
Google Scholar	2659	15.9	100	129	242	288	330	303	394	362	372	342

Note 1: Based on 167 refereed papers published in EconLit listed journals.

Note 2: The figure reported for Google Scholar does not reflect undated citations of which we found 203.

The above-noted discrepancy between the correlation coefficients generated by our broad and narrow definition of economics is undoubtedly largely attributable to a single paper. Gordon and McCann (2000), published in *Urban Studies,* generated 42.2 per cent of the total ISIB cites captured by all papers published in 2000, and 26.5 per cent of all such cites over the 2000 and 2001 period. For Google, the corresponding figures are 31.4 and 21.1 per cent, respectively. However, *Urban Studies* is not considered to be an 'economics' journal by ISI (it

20 As discussed later in the paper, this result may be attributable to a significant increase in the past few years in the number of journals eligible to generate and receive citations.

is deemed, rather, to be an urban studies publication), and hence generates zero ISIN cites. Although this is an extreme case, the distribution of citations across papers will subsequently be shown to be highly skewed.

As mentioned earlier, it is widely known that many papers in economics fail to receive a single citation over long periods of time (Chang, McAleer and Oxley 2011; Wall 2010; and Oswald 2007). We shall now explore this issue in the New Zealand context. Based on Dataset1, over a 10-year collection period only 40.1 per cent of papers received one or more ISIB citations (for ISIN, the estimate is 37.7 per cent). In contrast, the estimate for Google is almost double — 78.4 per cent. In large part, this discrepancy can be explained by differences in the scope of coverage of the citing and cited journals. Recall that ISIB is based solely on citations to and from JCR-listed journals, whereas Google Scholar sources citations from a much broader range of publications. In 2010, this resulted in only 64.7 per cent of papers in Dataset1 being eligible to receive ISIB citations, as opposed to 100 per cent for Google.[21] Therefore, of ISIB-eligible papers, 62.0 per cent were ultimately cited. (The corresponding numbers for ISIN are 61.1 and 61.7 per cent, respectively).

In the above discussion we have focused on the time pattern of citations per year to all journals in our sample. However, this is only one way of looking at the 'lag' issue. Another way of doing so is to explore the length of time it takes individual papers to receive their first citation. As shown in Table 2, three years after publication, only 16.2 per cent of papers had received one or more ISIB citations; after five years, 32.3 per cent of papers were in this category, and, as discussed earlier, after 10 years the number had increased to 40.1 per cent. On the other hand, the estimates for each time period under Google are dramatically higher: 52.7, 71.9 and 78.4 per cent, respectively.

Table 2: Non-self, non-zero citations per paper, various time periods based on cites to papers published in 2000 and 2001

	Percentage of Papers with Non-Zero Citations		
Citation Scheme	End Year 3	End Year 5	End Year 10
ISI, Economics Broadly Defined	16.2	32.3	40.1
ISI, Economics Narrowly Defined	15.6	31.1	37.7
Google Scholar	52.7	71.9	78.4

21 Economists in New Zealand face the regional-bias problem mentioned earlier in the paper. That is, the nation's only refereed economics journal, the *New Zealand Economics Papers* (*NZEP*), is not included in the ISI/JCR database. For obvious reasons, *NZEP* is the leading publication vehicle for New Zealand economists. If we arbitrarily drop papers in *NZEP* from the dataset, we find that 70.6 per cent of all remaining broadly defined economics papers were published in ISI/JCR-listed journals (the corresponding figure for narrowly defined economics papers is 66.7 per cent).

Despite the evidence of relatively long lags in the citation-generating process, especially for ISIB and ISIN, one can take some comfort from the information displayed in Table 3. For example, for all three measures of output, the correlation coefficients associated with three- and 10-year citation counts (on an individual-paper basis) range from 0.819 to 0.875. As expected, the estimates rise as we increase the collection period: for instance, the correlation coefficients for the five- versus 10-year citation period rise to 0.925 to 0.978 for all three output measures; the corresponding estimates for the seven- versus 10-year citation period range from 0.973 to 0.995. This suggests that if a 10-year collection period is considered to be an ideal time period for generating estimates of citation-based research output, then the use of, say, a five-year collection period could result in acceptable proxy estimates.

Table 3: Correlation coefficients, non-self citations per paper, various time periods (based on cites to papers published in 2000 and 2001)

Cites Per Paper, Various Time Periods	ISI (Broad)	ISI (Narrow)	Google
Year 1-1: Year 1-10	0.523	0.209	0.525
Year 1-2: Year 1-10	0.697	0.691	0.637
Year 1-3: Year 1-10	0.875	0.838	0.819
Year 1-4: Year 1-10	0.900	0.893	0.924
Year 1-5: Year 1-10	0.978	0.925	0.970
Year 1-6: Year 1-10	0.990	0.960	0.987
Year 1-7: Year 1-10	0.994	0.973	0.995
Year 1-8: Year 1-10	0.996	0.981	0.997
Year 1-9: Year 1-10	0.999	0.996	0.999

Citation patterns for all articles published between 2000 and 2008

Let us now move to an analysis based on Dataset2. Recall that the distinguishing feature of this dataset is that we have expanded the publication period from 2000–01 to 2000–08; however, the research group remains the same as in Dataset1. Over this nine-year period, NZ's 156 economists published 871 articles in *EconLit*-listed journals, and by the end of 2010 these publications had received a total of 2470 ISIB citations. The distribution of these by year is shown in Table 4. Note that, with one exception, the citation pattern is similar to that discussed earlier when we explored the 10-year pattern for papers released in 2000 and 2001. Now the collection period ranges from 11 years (for papers published in 2000) to three years (for papers published in 2008). The one exception relates to 2008 publications: it would appear that more citations are generated in Years 2 and 3 than expected. This might be related to the nature of the papers published and their topicality, but it might also be related to the rapid expansion of ISI's journal coverage in economics and, more geneally, the social sciences. This issue will be discussed later in the paper.

Table 4: ISI non-self citations, papers published 2000–2008, broad definition of economics: distribution of citations per year; citations collected up to 31 December 2010

Year Published	# of Papers	Total # of Non- Self Cites	Average # of Non-Self Cites/ Paper	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11
2000	73	478	6.55	2	4	22	35	51	40	61	61	75	74	53
2001	94	262	2.79	2	4	24	30	29	24	35	31	40	43	
2002	101	347	3.44	3	11	23	33	43	55	56	72	51		
2003	95	447	4.71	2	19	50	67	81	68	70	90			
2004	91	144	1.58	1	15	25	24	27	20	32				
2005	101	277	2.74	3	20	48	65	57	84					
2006	115	220	1.91	9	17	40	66	88						
2007	94	171	1.82	6	18	63	84							
2008	107	124	1.16	5	43	76								

Table 5: ISI non-self citations, papers published 2000–2008, broad definition of economics for papers ultimately cited, percentage with non-zero citations at various year-ends

Year Published	# of Papers	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11
2000	73	6.5	16.1	48.4	67.7	80.6	83.9	90.3	96.8	100.0	100.0	100.0
2001	94	5.6	11.1	33.3	55.6	80.6	86.1	94.4	94.4	97.2	100.0	
2002	101	6.7	20.0	22.2	60.0	80.0	91.1	91.1	93.3	100.0		

Although NZ's economists published 871 papers over the period 2000–08, only 41.1 per cent of them received one or more ISIB cites. However, this figure is misleading in two respects. First, only 540 papers were published in currently listed JCR journals. After making this adjustment, we find that 66.3 per cent of eligible papers received one or more ISIB cites. Second, 50 papers in our sample were published in journals that at time of publication were not covered by ISI (that is, these journals were added at a later date). Therefore, if we restrict the sample to papers eligible for citation counting by ISIB, we find that 73.1 per cent of them received one or more citations.

It is interesting to note that of the papers eventually receiving one or more ISIB cites, the vast majority reached that status by the end of Year 5. This can be seen by reference to Table 5. More specifically, for papers with a nine-or-more-year citation-collection period (papers published over the period 2000 to 2002), approximately 80 per cent of papers that were eventually cited had reached that status by the end of Year 5.

Let us now turn to an examination of the distribution of citations across papers. In Table 6 we display the percentage distribution of ISIB cites over various groupings for three different collection periods: 10, eight and five years. As previously noted, approximately 40 per cent of papers receive at least one cite over our catchment period (up to 10 years). However, the number of papers receiving multiple cites drops off rather quickly. For example, across our five-, eight- and 10-year collection periods, only 19.1, 22.6 and 20.4 per cent of papers received five or more cites; the corresponding figures for 10 or more cites are 9.9, 12.7 and 12.0 per cent. It is clear from the data that few papers receive 20 or more cites; even 10 years after publication, only 4.8 per cent of papers published in 2000 and 2001 reached this status.

Citation patterns compared with other measures of research output

We concluded our empirical work by calculating departmental and individual researcher output using various citation measures, and we compared the results with those generated by competing schemes. As noted in the discussion of Dataset3, we constructed three citation measures for academic staff employed as at 15 April 2009: ISIB03–08, ISB01–06A, and ISIB01–06B. It is important to recall that these schemes differ in two ways. First, ISIB03–08 is based on publications over the period 2003–08, our hypothetical PBRF timeframe. By contrast, ISIB01–06A and ISIB01–06B count citations to papers published over the 2001–06 period. Second, each scheme differs with respect to the lag time between the last year of publication and the final year of citation counting. More explicitly, the time-lags are zero, two and four years for ISIB03–08, ISIB01–06A and ISIB01–06B, respectively.

Table 6: ISI non-self citations, broad definition of economics, papers published 2000–2008; cumulative percentage distribution of papers receiving a given number of citations over various time periods

			Percentage of Papers with denoted number of Citations								
Years of Citation Coverage	# of Papers	Total # of Non- Self Cites	Papers with Zero Cites	Papers with > = 1 Cite	Papers with > = 2 Cites	Papers with > = 3 Cites	Papers with > = 4 Cites	Papers with > = 5 Cites	Papers with > = 10 Cites	Papers with > = 15 Cites	Papers with > = 20 Cites
10 Years (Publ. 2000 - 2001)	167	686	59.3	40.7	32.9	28.1	22.8	20.4	12.0	7.8	4.8
8 Years (Publ. 2000- 2003)	363	1480	57.3	42.7	35.8	31.1	25.3	22.6	12.7	7.4	4.4
5 Years (Publ. 2000- 2006)	670	2121	59.3	40.6	32.4	27.3	22.2	19.1	9.9	5.4	3.0

Note: Citations collected from date of publication to 31 December 2010.

For comparison purposes, we derived output estimates for three competing output schemes: KMS2010, ERAB and *EQUAL*. The first two weighting schemes were discussed above, but EQUAL appears for the first time. This metric represents the number of share-adjusted pages of qualifying research (contained in journals listed in *EconLit*); in other words, a 20-page article in the *AER* is deemed to be equivalent to a 20-page article in an obscure regional journal. EQUAL is really a representation of quantity, not quality, but serves as a useful reference point when one is trying to judge the aggressiveness of alternative weighting schemes.

In Table 7 we reveal the relationship between our various measures of departmental output. It is clear that our three citation-based measures are very weakly correlated with KMS2010 (ISIB03–08: 0.10; ISIB01–06A: 0.06; and ISIB01–06B: 0.03).[22] Recall that KMS2010 is an updated version of a widely accepted, aggressive journal-based weighting scheme. On the other hand, the correlation coefficients for ERAB (the Australian government's research-evaluation scheme) and our various citation measures range from 0.88 to 0.91. Perhaps more surprising is the nature of the relationship between EQUAL and ISIB03–08, ISIB01–06A, and ISIB01–06B — they range from 0.93 to 0.94. This result might be explained by the fact that once a journal has been listed by ISI, all citations are deemed to be of equal value; and with respect to New Zealand's economists, papers published in lower-ranked journals appear to be as successful in capturing cites as those published in higher-ranked journals.

Table 7: Pairwise correlation coefficients, departmental output, weighted pages and citations per capita (2003–2008)

	EQUAL	KMS2010	ERAB	ISIB03-08	ISIB01-06A	ISIB01-06B
EQUAL	1.00	0.01	0.93	0.94	0.93	0.93
KMS2010		1.00	0.25	0.10	0.06	0.03
ERAB			1.00	0.88	0.91	0.91
ISIB03-08				1.00	0.96	0.96
ISIB01-06A					1.00	0.99
ISIB01-06B						1.00

Let us now turn our attention to individual economists. Given that the PBRF scheme evaluates individual performance, a movement away from the current peer-evaluation system to a more mechanistic scheme would undoubtedly produce many winners and losers. Although we are not able to generate proxy PBRF results, we are able to capture the nature of the relationship between our various output schemes. As shown in Table 8, we present the pairwise correlation coefficients between our three citation-based schemes (ISIB03–08, ISIB01–06A,

22 These correlations are not significantly different from zero.

and ISIB01–06B), and three alternative schemes (KMS2010, ERAB and EQUAL). For illustration purposes, our sample is restricted to output estimates for the top 30 researchers as ranked by EQUAL. We have done so since highly ranked producers by any measure have more to lose in the adoption of an alternative measure and because many economists in our sample have generated zero output under KMS2010 and all of our citation-based schemes.

Table 8: Pairwise correlation coefficients, individual output, top 30 (ranked by EQUAL), weighted pages and citations per capita, 2003–2008

	EQUAL	ERAB	KMS2010	ISIB03-08	ISIB01-06A	ISIB01-06B
EQUAL	1.00	0.74	0.06	0.43	0.51	0.55
ERAB		1.00	0.47	0.67	0.77	0.78
KMS2010			1.00	0.19	0.34	0.38
ISIB03-08				1.00	0.84	0.80
ISIB01-06A					1.00	0.99
ISIB01-06B						1.00

It is apparent that our three citation schemes are weakly correlated with KMS2010 — with correlation coefficients ranging from 0.19 to 0.38. On the other hand, the perception-based ERAB scheme yields much higher estimates: ISIB03–08 (0.67), ISIB01–06A (0.77) and ISIB01–06B (0.78). Of interest is the fact that, as opposed to the departmental situation, the relationship between EQUAL and the ISI-based measures is only of moderate strength: 0.43 to 0.55. It is clear that as far as individuals are concerned, an evaluation system based on citation counts yields different results from one based on journal weights.

Policy implications

Our findings suggest that the time-lags between publication and citing are such that it would be difficult to rely on citation counts as a meaningful measure of output in a PBRF-like research evaluation framework, especially one based explicitly on individual assessment and a six-year timeframe. Nationwide evaluation schemes such as the UK's Research Excellence Framework (REF),[23] Australia's Excellence in Research for Australia (ERA)[24] and New Zealand's PBRF attempt to provide an indication of recent research productivity. This is evidenced by the fact that they use a stock measure of output; that is, they select a census date that is as close to the portfolio submission date as possible, and then they assess each institution's research activity over the preceding six years.

23 For details, see www.hefce.ac.uk/research/ref The Research Excellence Framework (REF) will carry out its first nationwide evaluation in 2014; it replaces the Research Assessment Exercise (RAE) (www.rae.ac.uk) that, in many ways, served as a model for the PBRF scheme.

24 For details, see www.arc.gov.au/era

Hence, the average paper (assuming a relatively stable publication flow) is only in print for three years prior to the end of the assessment period. As we have shown above, three years after publication the vast majority of papers had not received a single ISI cite; and for those cited, early citation patterns can deviate substantially from those exhibited over a longer time period. This problem is much more severe at the individual level than at the departmental level (due to the effects of averaging). We found numerous cases wherein individual papers did not receive any cites until Year Eight or later, with some as late as 10 years.

On the other hand, an argument can be made that citation counts provide additional information that could be used in a multi-criteria evaluation system. Our work suggests that the output measures generated by citation counts are not highly correlated with traditional output measures based on journal-impact factors. This follows primarily from the fact that some papers in lower-ranked journals generate a relatively large number of citations, and some in highly ranked journals receive few, if any, cites. Therefore, especially if collected over a longer time period than the six-year window currently used by PBRF, citations could provide evaluation committees with useful information. However, if the collection period were extended — say, from six years to eight — it would create even more of an incentive to hire productive, experienced staff rather than young, inexperienced researchers.

Earlier, we drew attention to the fact that the number of JCR-listed economics and social science journals has expanded rapidly over time. For example, when Liebowitz and Palmer (1984) undertook the research that led to their ground-breaking work in constructing adjusted-impact measures, they relied on an ISI/JCR database which, at that time, listed only 107 economics journals. By 1998, the JCR economics list had expanded to 159 journals and by 2003 the number of JCR economics journals had reached 169. However, in recent years the list has expanded dramatically: to 209 in 2008 and 247 in 2009 (the most recent list at the time of writing — March 2011).[25] A similar expansion has undoubtedly taken place in other social science disciplines.

An expanding journal list leads to two effects: first, the percentage of publications eligible for citation collection has increased and will continue to do so; and second, the number of citations per paper should also increase as the number of eligible citation-generating journals has grown (all journals in the ISI database). This has both positive and negative effects on the value of a citation-counting scheme. On the positive side, it will minimise the impact of the regional-journal issue (as more and more are incorporated into the database).

25 The dates 1998 and 2003 were chosen because they represent the journal selection dates adopted by two of the major papers in the journal-based weighting literature: Kalaitzidakis, Mamuneas and Stengos (2003), and Kodrzycki and Yu (2006).

It also helps departments and individuals working in new and emerging areas of the discipline, since journals with a focus on these areas are more likely to be included in the 'eligible list' than in the past.

On the other hand, the less discriminating the 'eligible list' becomes, the more pressure will arise to challenge the assumption that all citations are of equal value. One may find cries to weight citations by, say, the relevant JCR Impact Factor; however, this leads to problems similar to those arising from earlier efforts by economists to apply differential weights to cites in the development of adjusted-citation journal-weighting schemes, of which KMS2010 is a prime example. The primary argument against weighting is that it mixes individual performance (the number of citations to a given paper) with the average performance of others with papers in the same journal, and, indirectly, with the quality of the editorial staff at any point in time (the ability to pick winners!).

Summary and conclusion

In this paper we have attempted to assess the merits of using citation counts per researcher as part of a nationwide research-assessment exercise, with particular reference to the discipline of economics. Two issues gave rise to our interest in this subject. First, the growing interest in using bibliographic techniques in research-assessment exercises driven, in part, by advances in information technology; and second, the concerns expressed by many social scientists over the merits of using citations to measure performance, especially with respect to the nature of the time-lag between publication and the generation of a meaningful flow of citations in their disciplines.

We explored these issues in the context of a single discipline — economics — and a single nation — New Zealand. Our findings, based on a 10-year collection period, suggest that cites are, indeed, initially slow to develop; for example, the proportion of cites collected over a 10-year period that are generated within the first three years of publication is in the order of 10 per cent. This estimate rises to roughly 30 per cent in Year 5. We also found that roughly 40 per cent of papers received one or more citations, 20 per cent five or more citations, and slightly less than 5 per cent received 20 or more citations. However, we must stress that many papers in our sample were not eligible for ISI citation collection. After adjusting for this fact, we found that slightly over 73 per cent of eligible papers were eventually cited within the period of our analysis.

In general, our findings suggest that the conventional assessment period of six years may be acceptable from a departmental perspective due to averaging

effects, but that this is too short a time period for individual assessment.[26] This arises from the fact that the average paper will have only three years to collect citations. Although this problem can be addressed, in part, by expanding the citation collection period, doing so provides an additional incentive for departments to, in effect, buy 'CVs' rather than hire young, inexperienced researchers. Overall, we agree with the view expressed on the REF's website: 'The pilot exercise showed that citation information is not sufficiently robust to be used formulaically or as a primary indicator of quality; but there is considerable scope for it to inform and enhance the process of expert review.'[27]

Although this study is set in a New Zealand context, our findings should be of interest to academics and higher-education policy analysts in other domains who are considering the use of citations as a research-assessment measure in the social sciences. In addition to demonstrating that the lag between publication and the generation of a meaningful flow of citations is problematic in the context of the time span normally employed in a nationwide research-assessment scheme (five or six years), our findings raise other policy-relevant issues. First, our work demonstrated that the choice of a citation collection scheme is not a trivial matter. For instance, we found that Google Scholar not only generates more cites, but does so much sooner after publication than does the ISI. However, the former engine is much less discriminating, and its use in a nationwide program will undoubtedly give rise to cries for 'quality' adjustment — something that is very difficult to implement in a real-world setting. Second, the widespread concerns that young academics are disadvantaged by journal-based evaluation schemes will only be reinforced by citation-based schemes that increase the time between initiation of a research program and the generation of measurable outputs. A third issue relates to the generality of our findings: is the economics discipline representative of the social sciences, and is the research performance of New Zealand's economists similar to that found in many nation-states? We must leave the answer to these questions to others.

26 Note that individuals receive notification of their score, and regardless of confidentiality rules, outcomes are widely known within departments, and perceived to be used in promotion and merit-pay assessments. Hence, the generation of individual scores may have long-term career implications.

27 www.hefce.ac.uk/reserch/ref/Biblio/ (Accessed 25 March 2011).

References

Anderson, D. L. and Tressler, J. 2010, 'The Merits of Using Citation-Based Journal Weighting Schemes to Measure Research Performance in Economics: The Case of New Zealand', Working Paper 10/3, Department of Economics, University of Waikato, Hamilton, New Zealand.

Centre for Science and Technology Studies 2007, *Scoping Study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions*, Leiden University, U.K.

Chang, C. L., McAleer, M. and Oxley, L. 2011, 'What Makes a Great Journal Great in Economics? The Singer not the Song', *Journal of Economic Surveys* 25: 326–61.

Cinlar, N. and Dowse, J. 2008, *Human Resource Trends in the Tertiary Academic Workforce*, Tertiary Education Commission, Wellington, New Zealand.

Cosme, P. and Teixeira, A. 2010, 'Are finance, management and marketing autonomous fields of scientific research? An analysis based on journal citations', *Scientometrics* 85: 627–48.

Coupe, T. 2003, 'Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990–2000', *Journal of the European Economic Association* 1(6): 1309–45.

Goldfinch, S. 2003, 'Investing in Excellence? The Performance-based Research Fund and its Implications for Political Science Departments in New Zealand', *Political Science* 55(1): 39–53.

Gordon, I. R. and McCann, P. 2000, 'Industrial Clusters, Agglomeration and/or social Networks?', *Urban Studies* 37(3): 513–32.

Henrekson, M. and Waldenstrom, D. 2011, 'How Should Research Performance be Measured? A Study of Swedish Economists', *Manchester School* 79(6): 1139–56.

Hodder, A. P. W. and Hodder, C. 2010, 'Research culture and New Zealand's performance-based research fund: some insights from bibliographic compilations of research outputs', *Scientometrics* 84: 887–901.

Kalaitzidakis, P., Mamuneas, T. and Stengos, T. 2003, 'Rankings of Academic Journals and Institutions in Economics', *Journal of the European Economic Association* 1(6): 1346–66.

——— 2010, 'An Updated Ranking of Academic Journals in Economics', Working Paper 9/2010, Economics Department, University of Guelph, Guelph, Canada.

Kodrzycki, Y. K. and Yu, P. 2006, 'New Approaches to Ranking economics Journals', *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy* 5(1), Article 24.

Laband, D. and Piette, M. 1994, 'The Relative Impact of Economics Journals', *Journal of Economic Literature* 32(2): 640–66.

Liebowitz, S. J. and Palmer, J. P. 1984', 'Assessing the Relative Impact of Economics Journals', *Journal of Economic Literature* 22(1): 77–88.

Macri, J. and Sinha, D. 2006, 'Rankings Methodology for International Comparisons of Institutions and Individuals: An Application Economics in Australia and New Zealand', *Journal of Economic Surveys* 20(1): 111–56.

OECD 2010, *Performance-based Funding for Public Research in Tertiary Institutions: Workshop Proceedings*, OECD Publishing.

Oswald, A. J. 2007, 'An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers', *Economica* 74(1): 21–31.

Smith, A. G. 2008, 'Benchmarking Google Scholar with the New Zealand PBRF research assessment exercise', *Scientometrics* 74(2): 309–16.

Starbuck, W. H. 2005, 'How much Better are the Most-Prestigious Journals? The Statistics of Academic Publication', *Organization Science* 16(2): 180–200.

Wall, H. J. 2009, 'Don't Get Skewed Over by Journal Rankings', *B.E. Journal of Economic Analysis and Policy: Topics* 9(1): Article 34.