

**UNIVERSITY OF WAIKATO
Hamilton**

New Zealand

**The Merits of Using Citations to Measure Research Output in
Economics Departments: The New Zealand Case**

David L. Anderson and John Tressler

Department of Economics

May 2011

Working Paper in Economics 11/11

Corresponding Author

John Tressler
Economics Department
University of Waikato
Private Bag 3105
Hamilton
NEW ZEALAND

Tel: +64 (0) 7-838-4045
Email: tressler@waikato.ac.nz

Homepage: <http://wms-soros.mngt.waikato.ac.nz/personal/tressler>

David L. Anderson
School of Business
Queen's University
Kingston, Ontario K7L 3N6
CANADA

Tel: 1-613-533-2362
Email: dla@queensu.ca

Abstract

In this paper we explore the merits of utilizing citation counts to measure research output in economics in the context of a nationwide research evaluation scheme. We selected one such system for study: the New Zealand government's Programme-Based Research Fund (PBRF). Citations were collected for 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 and, to a limited extent, Google Scholar. These data allowed us to estimate the time lags in economics between publication of an article 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 direct citation measures and alternative schemes on departmental and individual performance. Our findings suggest that the time-lags between publication and citing are such that it would be difficult to rely on citations counts to produce a meaningful measure of output in a PBRF-like research evaluation framework, especially one based explicitly on individual assessment.

Keywords

citations
economics departments
journal weighting schemes
PBRF
research output

JEL Codes

A19, C81, J24

1. Introduction

The primary purpose of this paper is to explore the merits of utilizing citations to measure research output in economics in the context of a nationwide research evaluation scheme. We shall focus on one such performance measurement system: the New Zealand government's Programme Based Research Fund (PBRF).¹ Under this programme, the performance of all academics is assessed, and the assigned grades are aggregated across academic units to generate a university-wide score. In 2010 the results were utilized to distribute \$NZ268 million of research funding; this is equal to 18 percent of government funding to universities and roughly 9 percent 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. The current system relies on a labour-intensive, peer-review process; however, it is our view that after the up-coming 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 nation-wide research evaluation schemes.²

In this paper we will explore the merits of the concerns of many in the social science community to the use of direct citations as a measure of research output. 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 utilize journal weights based indirectly on citation counts and on reputational surveys.

2. Critical Issues 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.³ Concerns have been expressed over purported differences in citation practices across the above mentioned disciplines. This argument has at least two dimensions: major differences in the time lag between the publication of an article and the commencement of a

¹ 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/

² For details, see Research Excellence Framework (REF) (www.hefce.ac.uk/research/ref/) and Excellence in Research for Australia (ERA) (www.arc.gov.au/era).

³ For example, see Centre for Science and Technology Studies (2007).

meaningful flow of citations to the article; and differences in the publication frequency and citing habits between the disciplines that work to the disadvantage of social scientists.

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 organization 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 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 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 recognized 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, for the PBRF scheme 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 recognized that institutions are shifting their hiring practices away from the inexperienced to those with a 'good' next round PBRF- relevant publication record.⁴

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-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

⁴ This statement is largely based on anecdotal evidence, but supporting evidence can be found in Cinlar and Dowse (2008).

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).⁵ 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 employed in the Australian government's Excellence in Research for Australia (ERA) scheme.⁶ 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 authors' share in multiple-authored papers.

The primary reason for favouring this 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 individuals' 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.

However, this approach has recently been called in question by Starbuck (2005), Oswald (2007), Wall (2010) and Chang, McAleer and Oxley (2010). For purposes of this paper, their findings can be summarized as follows: good papers (lots of cites) can be found in lowly ranked journals (relatively few cites), and poor papers (very few cites) can be found in highly ranked journals (many cites). Indeed, Chang, McAleer and Oxley (2010) found that over a twenty-five year period, approximately 40 percent of the papers published in *Econometrica* and *Econometric Theory* failed to generate a single citation – even from the authors themselves. 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.

⁵ 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.

⁶ For details, see ERA's website at www.arc.gov.au/era.

3. Data

We assembled three basic datasets for this study. First, we created a file, denoted as Dataset1, containing all citations collected over a ten 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 2009.⁷ More formally, we counted all citations to these papers using both the ISI and Google Scholar databases; the citations were collected in early January 2011. For papers published in 2000, we collected cites generated over the period 1 January 2000 to 31 December 2009, and for 2001 papers, the time period shifted forward by one year.

The collection of cites using Google Scholar is relatively straight forward, albeit time consuming since citing papers are listed according to the number of times they themselves have been cited, rather than by publication date. On the other hand, generating cites 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 cites 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.⁸ We refer to data based on this definition as **ISIB**. The alternative approach was labeled as a 'narrow' definition of economics – it is based on the restrictive practice of recognizing only articles published in journals listed as 'economics' in the Journal Citation Reports (JCR).⁹ 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 utilized in our collection exercise was to eliminate self-cites by authors.¹⁰

⁷ 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 maximize the size of the sample.

⁸ 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)*.

⁹ 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).

¹⁰ 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 recognized as a problem in the biological and physical sciences. Although we are getting ahead of ourselves, we should note that of the total cites to papers published by New Zealand's academic economists over the period 2000-2008, approximately 15 percent (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.

Dataset 2 is based on 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 cites ranging from eleven 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 Dataset 2 (and in Dataset 3 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 recognized as ‘economics-relevant’.¹¹ 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, 2010).¹²

Our third dataset (Dataset3) was 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* recognized journals, over the period 2003-2008. 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 third share), and utilized the size adjusted page (*AER* equivalent) as our unit of output (see Macri and Sinha, 2006). In order to restrict the scope of the study, we have arbitrarily selected only two journal-based weighting schemes for comparison purposes: **KMS2010** to represent an aggressive scheme based indirectly on citation counts;¹³ and **ERAB**, the Australian government’s journal weighting scheme based on ‘expert opinion’ (that is, a perception- based system).¹⁴

In order to demonstrate the importance of the time-lag issue, we constructed three citation measures. Our first scheme, **ISIB03-08**, is based on a simple count of citations over

¹¹ For a discussion of these issues, albeit from a Finance perspective, see Cosme and Teixeira (2010).

¹² Although Google Scholar is rapidly gaining academic credibility, it has been criticized for lack of transparency in design and scope. For a New Zealand/PBRF related assessment of Google Scholar, see Smith (2008).

¹³ 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).

¹⁴ We have adopted a broad version of the ERA scheme (hence the reason for denoting it as ERAB). That is, we recognize 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 latter score for journals not covered by the ERA scheme but included in *EconLit*).

the period 2003-2008 to all papers published over this very same time period. This time span corresponds to the time frame utilized by the PBRF – a six year time 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. In order to address the obvious timing issues associated with **ISIB03-08**, we constructed two additional citation measures based on a two and four year lag, labeled **ISIB01-06A** and **ISIB01-06B**, respectively. The former scheme, **ISIB01-06A**, is based on papers published over the 2001-2006 period, 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. The latter measure, **ISIB01-06B**, is also based on papers published over the 2001-2006 period, but the citation collection period now ends on 31 December 2010. The maximum and minimum period for capturing cites is now ten and four years, respectively.

4. 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. As shown in Table 1, the nation’s 156 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. Also note that over a ten 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.

Table 1: Non-Self Citations for Papers Published in 2000 and 2001

Citation Scheme	Total	Cites/										
		Paper	Year1	Year2	Year3	Year4	Year5	Year6	Year7	Year8	Year9	Year10
ISI, Broad Defn. of Economics	686	4.1	4	8	46	65	80	64	96	92	115	117
ISI, Narrow Defn. of Economics	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.

Let us now explore the time pattern of citations. In the first year after publication, very few cites were generated for **ISIB** and **ISIN** – less than one percent of the ten year total. Indeed, by the end of year three, the corresponding estimate is only 6.7 and 7.2 percent, 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 Year10 cites are higher than those in any preceding year.¹⁵ **Google**

¹⁵ 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.

exhibits a somewhat similar year-by-year time pattern; however the flows in Year1 and Year2 are larger, but still below the levels generated in Year3 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 paper-by-paper basis, the correlation coefficients are much lower: **ISIB/Google** (0.262) and **ISIN/Google** (0.483).

Although a digression, 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 percent of the total **ISIB** cites captured by all papers published in 2000, and 26.5 percent of all such cites over the 2000 and 2001 period. For **Google** the corresponding figures are 31.4 and 21.1 percent, respectively. However, *Urban Studies* is not considered to be an 'Economics' journal by ISI (it is deemed 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 cite over long periods of time (see, for example, Chang, McAleer and Oxley, 2010, 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 percent of papers received one or more **ISIB** cites (for **ISIN**, the estimate is 37.7 percent). In contrast, the estimate for **Google** is almost double – 78.4 percent. In large part, this discrepancy can be explained by differences in the scope of coverage of the exporting and importing journals. Recall that **ISIB** is based solely on cites to JCR-listed journals. In 2010, this restriction resulted in only 64.7 percent of papers in Dataset1 being eligible to receive **ISIB** citations.¹⁶ Therefore, of eligible papers, 62.0 percent were ultimately cited. (The corresponding numbers for **ISIN** are 61.1 and 61.7 percent, 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 cite. As shown in Table 2, three years after publication, only 16.2 percent of papers had received one or more **ISIB** cites; after five years, 32.3 percent of papers were in this category, and, as discussed earlier, after 10 years the number had increased to 40.1 percent. On the other hand, the estimates for each time period under **Google** are dramatically higher: 52.7, 71.9 and 78.4 percent, respectively.

¹⁶ 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 percent of all remaining papers are eligible for **ISIB** cites (the corresponding figure for **ISIN** is 66.7 percent).

Table 2: Non-Self, Non-Zero Citations Per Paper, Various Time Periods
Based on Cites to Papers Published in 2000 and 2001

Citation Scheme	Percentage of Papers with Non-Zero Citations		
	End Year3	End Year5	End Year10
ISI, Broad Definition of Economics	16.2	32.3	40.1
ISI, Narrow Definition of Economics	15.6	31.1	37.7
Google Scholar	52.7	71.9	78.4

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 ten year citation counts (on a paper by paper basis) range from 0.819 to 0.875. As expected, the estimates rise as we increase the citation collection period: for instance, the correlation coefficients for the five versus ten year citation period rise to 0.925 to 0.978 for all three output measures; the corresponding estimates for the seven versus ten year citation period range from 0.973 to 0.995. This suggests that if a ten 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
Year1-1: Year1-10	0.523	0.209	0.525
Year1-2: Year1-10	0.697	0.691	0.637
Year1-3: Year1-10	0.875	0.838	0.819
Year1-4: Year1-10	0.900	0.893	0.924
Year1-5: Year1-10	0.978	0.925	0.970
Year1-6: Year1-10	0.990	0.960	0.987
Year1-7: Year1-10	0.994	0.973	0.995
Year1-8: Year1-10	0.996	0.981	0.997
Year1-9: Year1-10	0.999	0.996	0.999

5. Citation Patterns for all Articles Published between 2000 and 2008

Let us now move to an analysis based on Dataset 2. Recall that the distinguishing feature of this dataset is that we have expanded the publication period from 2000-2001 to 2000-2008; however, the research group remains the same as in Dataset1. Over this nine year period, New Zealand'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 **ISIB** cites 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 3 years (for papers published in 2008). The one exception relates to 2008 publications: it would appear that more cites 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 generally, 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	No. of Papers	Total No. of Non-Self Cites	Average No. of Non-Self Cites/Paper	Year1	Year2	Year3	Year4	Year5	Year6	Year7	Year8	Year9	Year10	Year11
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								

Although NZ's economists published 871 papers over the 2000- 2008 period, only 41.1 percent 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 percent of eligible papers received one or more **ISIB** cites. Secondly, 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 percent 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 percent of papers that were eventually cited, had reached that status by the end of Year 5.

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	No. of Papers	Year1	Year2	Year3	Year4	Year5	Year6	Year7	Year8	Year9	Year10	Year11
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		

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: ten, eight and five years. As previously noted, approximately 40

percent 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 ten year collection periods, only 19.1, 22.6 and 20.4 percent of papers received five or more cites; the corresponding figures for 10 or more cites are 9.9, 12.7 and 12.0 percent. It is clear from the data that few papers receive 20 or more cites; even 10 years after publication, only 4.8 percent of papers published in 2000 and 2001 reached this status.

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	No. of Papers	Total No. 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

Note1: Citations Collected from Date of Publication to 31 December 2010

6. Citation Patterns and 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 Dataset 3, 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 2003-2008, our hypothetical PBRF time frame. By contrast, **ISIB01-06A** and **ISIB01-06B** count citations to papers published over the 2001-2006 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.

For comparison purposes, we derived output estimates for three competing output schemes: **KMS2010**, **ERAB** and **EQUAL**. The first two weighting schemes were discussed in the Data section of this paper, 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 twenty page article in the *AER* is deemed to be equivalent to a twenty 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.

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

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).¹⁷ 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.

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 pair-wise correlation coefficients between our three citation-based schemes (**ISIB03-08**, **ISIB01-06A**, and **ISIB01-06B**), and three alternative schemes (**KMS2010**, **ERAB** and **EQUAL**). For illustration purposes, our sample is restricted to output estimates for the top thirty 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,
Top30 (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

¹⁷ These correlations are not significantly different from zero.

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 an evaluation system based on citation counts yields different results from one based on journal weights.

7. 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 time frame. Nation-wide evaluation schemes such as the UK's Research Excellence Framework (REF)¹⁸, Australia's Excellence in Research for Australia (ERA)¹⁹ and New Zealand's PBRF attempt to provide an indication of recent research productivity. This is evidenced by the fact they utilize 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 institutions research activity over the preceding six years. 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 in the above section, 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 rather than 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; 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 citation 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 groundbreaking work in constructing adjusted-impact measures, they relied on an *ISI*/JCR database that, 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

¹⁸ For details, see www.hefce.ac.uk/research/ref. The Research Excellence Framework (REF) will carry out its first nation-wide 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.

¹⁹ For details, see www.arc.gov.au/era.

list has expanded dramatically: to 209 in 2008 and 247 in 2009 (the most recent list at the time of writing- March 2011).²⁰ 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 and will increase; 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 minimize the impact of the regional journal issue (as more and more are incorporated into the database). 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 cites are of equal value. One may find cries to weight cites 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 cites 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!).

8. Summary and Conclusions

In this paper we have attempted to assess the merits of utilizing citation counts per researcher as part of a nation-wide 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 ten year collection period, suggest that cites are, indeed, initially slow to develop; for example, the proportion of cites collected over a ten year period that are generated within the first three years of publication is in the order of 10 percent. This estimate rises to roughly 30 percent in year 5. We also found that roughly 40 percent of papers received one or more citations, 20 percent five or more cites, and slightly less than 5 percent 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 percent of eligible papers were eventually cited within the period of our analysis.

²⁰ The dates 1998 and 2003 were chosen because they represent the journal selection dates utilized by two of the major papers in the journal-based weighting literature: Kalaitzidakis, Mamuneas and Stengos (2003) and Kodrzycki and Yu (2006).

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.²¹ 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.'²²

²¹ 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.

²² www.hefce.ac.uk/research/ref/Biblio/ (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., M. McAleer and L. Oxley (2010). What Makes a Great Journal Great in economics? The Singer not the Song. Working Paper No. 43/2010, Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand.
- Cinlar, N. And J. Dowse (2008). Human Resource Trends in the Tertiary Academic Workforce. Tertiary Education Commission, Wellington, New Zealand.
- Cosme, P. and A. Teixeira (2010). Are finance, management, and marketing autonomous fields of scientific research? An analysis based on journal citations. *Scientometrics*. 85: 627-648.
- Coupe, T. (2003). Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990-2000. *Journal of the European Economic Association*. 1(6): 1309-1345.
- 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 P. McCann (2000). Industrial Clusters, Agglomeration and/or social Networks? *Urban Studies*. 37(3): 513-532.
- Henrekson, M. and D. Waldenstrom (2009). How Should Research Performance be Measured? Evidence from Rankings of Academic Economists. Working Paper No. 693, Stockholm School of Economics, Stockholm, Sweden.
- Hodder A.P.W. and C. Hodder (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., T. Mamuneas and T. Stengos (2003). Rankings of Academic Journals and Institutions in Economics. *Journal of the European Economic Association* 1(6): 1346-1366.
- Kalaitzidakis, P., T. Mamuneas and T. Stengos (2010). An Updated Ranking of Academic Journals in Economics. Working Paper 9/2010, Economics Department, University of Guelph, Guelph, Canada.
- Kodrzycki, Y.K. and P. Yu (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 M. Piette (1994). The Relative Impact of Economics Journals. *Journal of Economic Literature* 32(2): 640-666.
- Liebowitz, S.J. and J.P. Palmer (1984). Assessing the Relative Impact of Economics Journals. *Journal of Economic Literature* 22(1): 77-88.
- Macri, J. and D. Sinha (2006). Rankings Methodology for International Comparisons of Institutions and Individuals: An Application Economics in Australia and New Zealand. *Journal of Economic Surveys*. 20(1): 111-156.
- Oswald, A.J. (2007). An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. *Economica* 74(1): 21-31.
- Starbuck, W.H. (2005). How much Better are the Most-Prestigious Journals? The Statistics of Academic Publication. *Organization Science*. 16(2): 180-200.
- Smith, A.G. (2008). Benchmarking Google Scholar with the New Zealand PBRF research assessment exercise. *Scientometrics*. 74(2): 309-316.
- 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.