UNIVERSITY OF WAIKATO
Hamilton
New Zealand

The Merits of Using Citation-Based Journal Weighting Schemes

to Measure Research Performance in Economics:

The Case of New Zealand

David L. Anderson and John Tressler

Department of Economics

Working Paper in Economics 03/10

May 2010

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 study we test various citation-based journal weighting schemes, especially those based on the Liebowitz and Palmer methodology, as to their suitability for use in a nation-wide research funding model.   Using data generated by New Zealand's academic economists, we compare the performance of departments, and individuals, under each of our selected schemes; and we then proceed to contrast these results with those generated by direct citation counts. Our findings suggest that if all citations are deemed to be of equal value, then schemes based on the Liebowitz and Palmer methodology yield problematic outcomes.   We also demonstrate that even between weighting schemes based on a common methodology, major differences are found to exist in departmental and individual outcomes.

Keywords

Economics Departments
research output
citations
impact factors

JEL Codes

A19, C81, J24

1. Introduction

It is becoming relatively common for nations to allocate research funding, at least in part, on the basis of various measures of research output. It is also apparent that many governmental funding agencies have adopted, or are considering adopting, bibliographic techniques for generating required measures of institutional performance.[1] Therefore, it is important that both administrators and researchers understand the impact of selected indicators on the resulting output estimates and rankings. We have chosen to examine this issue in the context of one discipline only: economics. More specifically, we will analyze the impact of measuring refereed paper output by utilizing citation counts to weight individual journals and hence papers within these journals.[2] Furthermore, we will employ data from New Zealand to demonstrate that the choice of a journal weighting scheme, even between schemes employing a common methodology, can have significant impact on determining winners and losers.

Much of the prior work in measuring the performance of economics departments and individual economists has been based on a relatively small number of highly cited international journals. However, this approach is of limited value when conducting a national performance review with the purpose of allocating research funds to help the nation state achieve its economic and social objectives. First, such an approach is likely to lead to a politically untenable solution since it is probable that a small percentage of institutions will capture a lion's share of the available funds. Secondly, and related to the first point, research output with a local or regional focus will be largely ignored in the calculation process since such work tends to be published in regional journals that are not likely to be included in the 'top' journal list. Thirdly, it is quite possible that a substantial proportion of a nation's research output will be largely ignored in the measurement process. This follows from the fact that there are over 1200 recognized economics journals (as listed in *EconLit*) and it is rare for more than 200 of these publications to be included in evaluation studies (in fact, many older studies used 30 or fewer journals). Therefore, universities must defend the following proposition: that a significant proportion of their total research output is of no value to society.

Not only is this 'bad politics' for institutions that receive a substantial proportion of their total revenue from public sources, it is also a difficult position to support from a theoretical position. As the recent work of Oswald (2007) and Wall (2009) demonstrates, over time, highly cited papers can be found in lower ranked journals and papers that fail to receive a single cite over a period as long as twenty-five years, can be found in the so-called 'top' journals. This suggests that an evaluation system based on a limited number of highly cited international journals is unlikely to generate a socially and politically acceptable solution.

In this paper we shall examine a number of conventional schemes for measuring research output in the economics literature, and then apply these to economics departments and individual economists in a small nation state – New Zealand. In the next section we shall discuss the evolution of research measurement schemes in economics, with particular emphasis on the post-1984 period – the start of the prevailing system for adjusting citations to account for various factors. Section III contains information on the New Zealand university system, a discussion of our underlying dataset, and a rationale for the various assumptions employed in the analysis. Our results are contained in the Section IV, and this is followed by our concluding comments.

2. Literature Review

2.1. The Early Years

The first formal study of the ranking of economics departments is generally said to be that of Fusfeld (1956), despite the fact that this was not the author's intention. Fusfeld was trying to demonstrate the narrowness of the selection process for papers presented at the annual meetings of the American Economics Association. Nevertheless, his ranking of institutions on the basis of the number of papers presented over the 1950-1954 period was interpreted by many as a quality ranking of US economics departments.[3] In a 1974 paper, Bush, Hamelman and Staaf published arguably the first rigorous citation-based ranking of economics journals.[4] Fourteen leading economics journals were selected[5] and the number of citations between these journals was counted over a five year period. The authors presented a ranking of the total number of cites to each journal and another ranking after adjusting for self-citations.

More generally, over the 1956 to 1982 period, most ranking studies used either reputational surveys or the number of publications in a pre-determined list of 'top' journals to rank departments and individuals in US-based schools.[6] By the end of this period, the then state-of-the-art approach to research output measurement can be found in the work of Graves, Marchand and Thompson (1982). They selected 240 schools and 24 'top' journals [7] for study, and derived rankings on the basis of both total and per capita output. They also utilized the size-adjusted page as the basic unit of output, rather than the number of articles published over a given time period. More specifically, pages in each of the 24 journals in the study were expressed in terms of *American Economic Review (AER)* equivalents. The relevant time period for study was five years, and output shares on multi-authored papers were allocated on the basis of the 1/n rule (where n is the number of authors). Furthermore, only tenured and tenure-track staff were included in the analysis. With two exceptions, one minor, the other major, the various procedures and adjustments described above remain in widespread use today.

The minor issue is as follows: Graves, Marchand and Thompson (1982) assigned articles (and thus pages) to institutions based on the institutional affiliation listed on the article at

the time of publication (the Flow method). Since that time most research output has been allocated via the Stock method wherein refereed papers are assigned to an author's current employer.[8] The major issue is of fundamental importance to this paper: the implicit assumption that all journals are of equal value; and, hence, that all articles and size-adjusted pages are of equal value.[9]

2.2.   Adjusted Impact Citation Factors

Arguably the single most influential paper in the research output measurement literature was published in 1984.[10] In this work, Liebowitz and Palmer (1984) attempted to address the major weakness of prior approaches: the implicit assumption that all journals (and the articles therein) are of equal value. To redress this situation, Liebowitz and Palmer utilized 1980 citation data from the Social Science Citation Index (SSCI) to generate a ranking of 108 economics journals.[11] The procedure they adopted continues to serve as the underpinnings of a number of schemes employed in the rankings literature to the present time. The details of the calculation are as follows.

In step 1, Liebowitz and Palmer calculated the number of citations from all journal articles published in 1980, in all journals in the SSCI database (approximately 4200 in 1980), to each of 108 economics journals. The countable cites were restricted to those generated by articles published over the 1975-1979 period (denoted as the age adjustment procedure). The resulting estimates can be labelled as the 'unadjusted impact citation' values. In Step 2, Liebowitz and Palmer adopted an iterative process to yield a new set of values that we will denote as the 'adjusted impact citation' weights. This involved using the weights generated in Step 1 ('unadjusted impact values') as the initial set of journal weights. Citations are then weighted by the relevant journal value, save for cites from non-economic journals that are given a zero value. The process was then repeated a number of times, each time using the journal values generated in the prior iteration to weight the citations to each journal included in the study, to yield a relatively stable solution.   The end result is a set of weights denoted as the 'adjusted impact citation' factors.

Liebowitz and Palmer made an additional adjustment in order to take journal size into account. That is, a journal publishing twice as many papers per year as another journal, will, everything else being equal, have a higher 'adjusted impact citation' weight. Thus, the need to adjust for size differences between journals. The authors suggested two approaches: first, to simply divide the 'adjusted impact citation' value by the number of articles published by each journal over the survey period; or second, to generate an estimate of the number of 'adjusted impact citations' per character. Although Liebowitz and Palmer favoured the latter size-adjustment procedure, both approaches have been used in subsequent studies, and roughly in equal proportions.[12]

The effect of Liebowitz and Palmer's iteration adjustment process (denoted above as Step 2) is rather dramatic. For example, after adjusting the output scores so that the most

highly ranked journal has a value of 100, Liebowitz and Palmer found the *Journal of Economic Literature* to be ranked first under both the 'unadjusted impact citation' approach (Step 1) and the 'adjusted impact citation' approach (Step 2). However, as one moves down the list, the aggressive nature of the LP84 scheme (as the Liebowitz and Palmer findings will henceforth be denoted) becomes very clear. For example, the 30th placed journal under the 'unadjusted impact citation' scheme receives a score of 21.25, whereas under the 'adjusted impact citation' approach the corresponding value is 8.95.[13] In other words, it takes slightly fewer than five articles in the 30th ranked journal to equal one in the *Journal of Economic Literature* if one uses unadjusted citation scores to rank journals, but slightly more than 11 articles if one uses the 'adjusted impact citation' approach.

The difference between the schemes becomes greater as one works down the rankings. For example, the 50th and 90th 'unadjusted' ranked journals have scores of 14.46 and 6.05, respectively. The corresponding scores under the 'adjusted impact citation' method are 2.29 and 0.12. The last number listed can be restated as follows: under the LP84 adjusted impact citation weighting scheme it takes almost 1000 articles in *World Development* to generate an output score equivalent to that generated by one article in the *Journal of Economic Literature*.

There are two issues that should be addressed before continuing with our literature review. First, LP84's 'adjusted impact citation' estimates are based solely only on citations from economics journals on the selected list (108 journals) to other journals on the same list. [14] In other words, the end result is a very narrowly defined 'economic' impact measure; not only does it ignore the citations from non-economic journals in the SSCI database to our 108 selected journals, it also ignores citations from all other refereed economics journals (now in excess of 1200 according to *EconLit*) to the LP84 list of recognized economics journals. Second, it should be noted that Liebowitz and Palmer do not provide an extensive rationale for their iterative approach.[15]

Liebowitz and Palmer's work was updated by Laband and Piette (1994). They utilized SSCI data from 1990, and calculated 'adjusted impact citation' weights for 130 economics journals.[16] More specifically, Laband and Piette counted the citations in 1990 from the selected journals to articles published in these same 130 journals over the period 1985 to 1989. They then employed the Liebowitz and Palmer methodology to generate a set of updated weights (henceforth denoted as LP94 weights) that have been widely utilized in ranking studies until relatively recently.

It should be noted that Laband and Piette also adjusted for size differences between journals by providing weights based on a per-article and per-character basis. Although the LP94 ranking of individual journals differs somewhat from that of LP84 due to different time periods (1984-1989 versus 1974-1979) and the expansion of the recognized journal list (130 versus 108), the overall nature of the weighting scheme remains the same. For example, under LP94, the 30th, 50th and 90th journals receive weights of 6.1, 2.1 and 0.1, respectively.

That is, it takes approximately 16 papers in the 30th placed journal to generate the same score as one article in the top journal (*Journal of Financial Economics*), and 1000 articles in the 90th placed journal (*Economic History Review*) to equal one in the most highly ranked journal. We should also mention that under the LP84 and LP94 weighting schemes, only 99 and 102 journals, respectively, receive a non-zero weighting; articles in all other economics journals are, in effect, deemed to be worthless.

In the late 1990s a number of ranking studies were undertaken, especially in Europe and Australia based, in part, on LP94 weights. However, over time this scheme lost some of its lustre for the simple reason that its weights were based on 'old' citation data (from 1990) and its journal list did not include a number of newer publications covering emerging areas within the discipline. This led Kalaitzidakis, Mamuneas and Stengos (2003) (henceforth denoted as KMS) to update the work of Liebowitz and Palmer (1984) and Laband and Piette (1994). They utilized the 1998 SSCI database to count citations in all 159 SSCI listed economics journals to articles published over the 1994-1998 period in these same journals. KMS followed the Liebowitz and Palmer methodology with one exception: they corrected for self-citations (citations from a journal to an article previously published by the same journal).

After adjusting for self-citations and size, KMS generated non-zero weights for 143 of the 159 journals in their database. As mentioned previously, the years covered for citation collection and the list of journals differs somewhat from that of the LP84 and LP94 studies;[17] nevertheless, the general nature of the weighting schemes remains the same. For example, we find that the 30th, 50th and 90th journals possess KMS weights of 7.84, 3.87 and 0.76, respectively. Rephrased, it takes 13 papers in the *National Tax Journal* (the 50th ranked journal) and 132 papers in the *Journal of the Japanese and International Economies* (the 90th ranked journal) to generate approximately the same output as a single paper in the *American Economic Review* (KMS's number one ranked journal). Before continuing with our literature review, we wish to mention that the KMS weighting scheme has come to be seen, in the eyes of many, as the industry standard, and has been extensively utilized in ranking studies. [18]

With the exception of Liebowitz and Palmer's (1984) ground-breaking paper, arguably the most important work in the development of 'adjusted impact citation' measures is that of Kodrzycki and Yu (2006). They utilized 2003 SSCI citation data covering papers published over the period 1996 to 2003 in selected journals. Although Kodrzycki and Yu (henceforth denoted as KandY) adopted the basic LP84 methodology, they made a number of adjustments that address some of the possible weaknesses of the initial iterative model. First, KandY used their own judgement to generate a list of journals widely utilized by economists rather than relying almost exclusively on the JCR classification system. More specifically, KandY selected 146 of the 169 journals listed as 'economics journals' on the JCR (2003) website, and an additional 35 journals from nine other JCR categories for a total of

181 journals. Although subjective, the selection process acknowledges the fact that it is exceedingly difficult to precisely determine the boundary between economics and non-economics articles in areas such as finance, regional science and public policy.

KandY's second adjustment was to address a restrictive assumption employed by LP84, LP94 and KMS, wherein only citations from other journals on the 'list' are to be counted. In other words, the resulting 'adjusted impact citation' factor is really an 'adjusted **economic** impact citation' measure. To address this problem, KandY constructed three sets of journal weights. One scheme is based on the traditional approach of only counting citations from journals on the 'list' to other journals on the 'list'; they labelled the resulting weights as economic impact factors (KYEI). KandY then calculated a new set of weights based on citations from all social science journals in the JCR (2003) dataset (1714 journals); this yields a set of weights they call the 'overall impact' factors (KYOI). This measure gives an indication of the impact of various economics journals on the overall social science literature (including economics).

The final set of weights derived by KandY attempts to measure the impact of economics on public policy. To do so, the authors selected 87 journals that they deemed to be 'policy' journals; however, 44 of these journals were already on the original list of 181. The estimation procedure is the same as that discussed above, save for the fact that the only citations that are counted are those from the 87 policy journals to articles in the chosen 181 economics journals. The resulting estimates are denoted as policy impact factors (KYPI). Not surprisingly, KandY found substantial differences in journal rank between schemes. For example, the Spearman Rank Order Correlation Coefficient between KYEI and KYOI is 0.74, but declines to 0.55 for the relationship between KYEI and KYPI. We shall return to this discrepancy later in the paper, but at this point we wish to indicate that the underlying issue in the weight- selection process is 'what is the nation state attempting to encourage: economic research that has a major impact on the economics profession itself or on public policy?' The question is phrased in this fashion since our focus is on measuring institutional and individual performance for the purpose of allocating public research funds.

KandY's third modification of the LP84 approach was to correct for reference intensity between the various sub-disciplines of economics. It is well known that there are major differences in the citation practices of economists across various fields. For example, papers on finance and econometric topics contain, on average, substantially more citations than those in economic history. Therefore, finance and econometric journals, everything else being equal, will exhibit larger impact factors than those in economic history. To address this problem, KandY employed the procedure developed by Palacios-Huerta and Volij (2004). Let us now look at the nature of the journal weights generated by KandY. Following past practice, we shall focus on the 30th, 50th and 90th journals. Note that the relevant KYEI scores are 12.39, 5.46 and 2.18, respectively. This suggests that the KandY estimates for economic impact do not decline as sharply as those generated by LP84, LP94 and KMS; for

these journals, the 90th ranked journal generated a weight of 0.12, 0.10 and 0.76, respectively. Although all of these studies employ the Liebowitz and Palmer methodology, KandY utilized a broader set of journals, a longer period for collecting citations, and, of course, a different time period for analysis; any one or more of these factors could account for the above noted differences.

KandY's results are of particular interest since they address an issue that is of fundamental importance to this paper: what is the relevant audience of economic research, at least in the eyes of public funding bodies? For this reason, let us look at the rankings of some well known journals under both the KYEI (economic impact) and KYPI (policy impact) schemes. First, it should be noted that several general interest economics journals receive a similar rank and weighting under each scheme; journals in this group include the *Quarterly Journal of Economics* (2nd under KYEI; 1st under KYPI), *Journal of Economic Literature* (3rd under KYEI and 4th under KYPI) and *Journal of Political Economy* (6th under KYEI and 5th under KYPI). However, other journals fare somewhat differently. For example, the *Journal of Finance* ranks 5th under the KYEI scheme but falls to 15th under the policy impact scheme (KYPI). Other finance journals fare less well: the *Review of Financial Studies* is the 7th ranked journal on the basis of economic impact (KYEI) but falls to 61st under the KYPI scheme; the corresponding rankings for the *Journal of Financial Economics* are 4th and 22nd.

As expected, applied journals generally fare better under the policy impact rather than the economic impact approach. For illustration purposes, consider *World Development;* under KYEI it is ranked 113th and under KYPI it has climbed to 57th. However, for output measurement purposes the differences are much greater: an article in *World Development* carries almost 8 times the weighting under KYPI than under KYEI (7.31 versus 0.94). Even more dramatic is the situation facing researchers who publish in *Housing Policy Debate:* this journal is ranked 97th under KYEI but rises to 9th position in the KYPI ranking; the corresponding weights are 1.63 and 28.01. It should be clear that departments and individuals who focus on applied issues are much better off under the policy impact approach than under the economic impact approach.

We consider the modifications made to the LP84 methodology by KandY to be significant. To summarize, they are first: revision and expansion of the relevant journal list to account for boundary issues between economics and near disciplines; second: the decision to allow non-economic journals to generate countable citations back to the selected economics journals in order to provide estimates of overall and policy impact; and third: the adjustment to address relative differences in citation practices between sub-groups within economics. However, it must be stressed that all of these adjustments involve some degree of subjectivity.

2.3.   Basic Citation-based Impact Factors

The impact factor concept can be traced back to Garfield (1972), and gained importance with the computerization of the SSCI database. The definition of an impact factor is exceedingly simple: the total number of citations to a journal in year t to articles published in years t-1 and t-x, divided by the number of articles published by the receiving journal in years t-1 to t-x where x can range from 2 to the number of years the journal has been in publication. However, in practice, a journal's impact factor is frequently calculated as follows: the total number of citations in year t to articles published in years t-1 and t-2, divided by the number of articles published by the receiving journal in years t-1 and t-2. This is the definition used by the JCR for deriving its well known set of journal impact factors.[19] However, the suitability of a two year citation attribution period is questionable for economics, and, more generally, the social sciences. Given the normal time period required for manuscript preparation, refereeing, revision, and publication, a two year cut-off for citations seems extremely tight. However, the increasing use of online working papers (such as that offered by RePEc) may mitigate the issues raised above, but even then citations are likely to be limited to those generated by researchers working in the same narrowly defined research area.

In a widely referenced study, Coupe (2003) employed a variant of the JCR impact factor scheme (denoted henceforth as JCRIF). He calculated the average of JCRIF for the years 1994 to 2000 for 273 journals.[20] His adjustment addresses, in part, the possible variation in JCRIF results due to the relatively short period over which citations are collected; however, it does not capture the effect of articles that generate a large number of citations three or more years after publication. For comparison purposes, we shall use both the JCRIF scheme and Coupe's variant (CoupeIF) in our subsequent analysis. Before moving on, we should mention that these basic citation impact schemes are based on citations from all journals in the SSCI database to the selected economics journals. This is an important distinguishing feature between adjusted impact citation schemes (save for the work of KandY) and basic citation schemes.

2.4. Indirect Citation-Based Weighting Schemes

Although somewhat dated, the weighting scheme developed by Bauwens (1998) exhibits some interesting properties. Bauwens used the 1996 SSCI/JCR database to generate a set of short-term and long-term impact indicators; he then multiplied the two values together to generate a journal-specific score. More specifically, Bauwens used the reported JCR 1996 impact factors as his short-term measure of journal impact; and the total number of citations to each journal in 1996 (regardless of the age of the article cited) as his long-run impact measure. Bauwens then proceeded to assign, somewhat arbitrarily, all journals for which he had scores to one of five categories,[21] with the highest ranked journals receiving a grade of '5'. It should be noted that all journals listed in *EconLit* in 1998 but not in JCR (1998), were given a grade of '1'. The end result is a per-article ranking for 619 journals,

with 1.33%, 7.64%, 7.64%, 7.64%, and 75.75% of journals receiving grades of 5 through 1, respectively.

The Bauwens weighting scheme has been used by a number of researchers, probably because it gives a non-zero weighting to all papers published in *EconLit* listed journals, and because it gives credit to both recent and past articles published in all JCR classified economics journals.   However, it must be stressed that the scheme is based on a large number of subjective procedures.   At this point we wish stress the obvious: under the Bauwens scheme any article in an *EconLit* listed journal will generate, at a minimum, a value of 20% of the value of a paper in the most highly ranked journal.

2.5.   Other Schemes Provided for Comparative Purposes

At the opposite end of the objectivity scale from 'adjusted impact citation' weighting schemes are those generated by surveys of 'experts'. We will discuss two such schemes in order to contrast outcomes between pure and partial citation-based schemes and those based on perceptions of a journal's worth and reputation.[22]   Although the use of reputational surveys to generate a set of journal rankings appears to have fallen out of favour after manipulation friendly versions of the SSCI/JCR database became available, there is one notable exception- the work of Mason, Steagall and Fabritius (1997).

In 1993, the authors surveyed US economics department chairs and asked them to rank journals on the basis of 'four' for the best to 'zero' for the worst (non-integer values were acceptable).   Based on the responses from 216 chairs (out of the 965 surveyed), Mason, Steagall and Fabritius (henceforth denoted as MSF) derived explicit values for 157 journals, with the leading journal, the *AER*, having a weight of 3.83 and the last placed journal, the *Nevada Review of Business and Economics*, exhibiting a weight of 0.65. Keeping with the approach utilized earlier in the paper, note that the 30th, 50th and 90th placed journals have scores of 2.90, 2.67, and 1.77, respectively. This suggests that the MSF weighting scheme is much less aggressive than the schemes discussed above, save for Bauwens (a hybrid scheme). It should be noted that the MSF scheme has been utilized in a number of studies to provide a comparison with results based on alternative approaches.

Recently, as part of the Australian government's effort to evaluate research in all of the nation's universities, the Australian Research Council (2008) developed another reputational weighting scheme.[23]   The overall evaluation framework is more formally known as the Excellence for Research in Australia (ERA) program, and it should be noted that the measurement of research by way of a journal-based ranking scheme is only one of several indicators that will be used to provide an overall evaluation of universities.[24] Under the ERA programme, journals are allocated to one of four categories: A+, A, B and C (and some are not ranked at all). With respect to the discipline of economics, 640 journals were formally recognized.[25] The actual percentage of journals assigned A+ through C is 7.8, 17.5, 27.0 and 47.7%; this compares to the overall programme target of 5, 15, 30 and 50%, respectively.

For purposes of this paper, we have assigned a grade of 4 to an A+, and so on to a 1 for a C. Economics journals not officially recognized by the ERA, but listed in *EconLit*, are given a grade of 0. Therefore, in reality, the ERA ranking scheme is based on a five-point scale. We must stress that the numerical weights are ours, not those of the ERA. We have chosen our weights based on academic convention – a version of the traditional weighting scheme employed by many universities, especially in North America, for converting from letter to numerical grades.[26] At this point we should note one distinguishing feature of the ERA, at least with respect to economics: the explicit attempt to recognize the role of national and regional journals in generating research of importance and relevance to Australia. In total, eleven Australian-produced journals are listed under the economics grouping of the ERA scheme.[27]

The final journal-based weighting scheme to be included in this study is based on the assumption that all papers are of equal value (we label such a scheme as EQUAL). This is really an indicator of quantity, but it is included to provide a reference point – an indication of how much the results generated by the other ten weighting schemes differ from that associated with the extreme assumption that all papers in *EconLit* listed journals are of equal value.

To this point, we have discussed a total of eleven journal-based weighting schemes. In broad terms, they can be categorized as follows: adjusted impact citation (LP84, LP94, KMS, KYEI and KYPI); basic citation (CoupeIF and JCRIF); hybrid (Bauwens); reputational (MSF and ERA); and other (EQUAL). Before ending this discussion, we wish to add another scheme to the 'other' category: a direct citation indicator of performance. More specifically, we will estimate the number of non-self citations generated by each researcher's share-adjusted papers published over the period 2001-2006.[28] This additional measure, to be noted as CITNS, has been added for comparison purposes, and to provide a link between the methodology used to measure research output in this paper (journal-based weighting approach) and an important competitor that is widely used in the natural and engineering sciences (the direct citation approach).[29]

3. Institutional and Database Information

As noted in the introduction, our objective is to assess the impact of journal-based weighting schemes on departmental and individual economist's output measures, with particular emphasis on the 'adjusted impact citation' approach initially developed by Liebowitz and Palmer (1984).[30] More specifically, we indicated that we intended to examine these schemes in the context of a nation state. To do so, we have chosen New Zealand, a small country of four million people, with eight universities and 135 economists (in 2009).[31]

In constructing our database, we adopted the prevailing approach on a number of matters that are themselves somewhat controversial. For example, we have defined

research to be articles published in journals listed in *EconLit* (as at 15 April 2009)*,* over the period 1 January 2003 to 31 December 2008. For multi-authored papers, shares have been allocated on the basis of the 1/n rule where n is the number of authors on the paper. An equally important allocation rule pertains to the process for assigning refereed papers to departments; we have chosen the Stock method whereby all of an individual researchers' output (over the 2003-2008 period) is assigned to her/his employer as at 15 April 2009. Our choice is based on the widespread view that the Stock method provides a reasonable proxy of a department's current and medium term research capability.[32]

A more contentious choice is that related to the appropriate unit of output. Once again, we have adopted the prevailing approach, and have selected the size adjusted page as the relevant measure of output. This decision is based on the premise that long papers are, on average, of greater significance than short papers. The alternative is to select the article as the unit of account since it is said to be the measure used by most academics to measure the volume of their activity. Given that we have selected the page as our output measure, we will again follow industry norms and convert all pages into *AER* equivalents. Fortunately, we have benefited from the generosity of Sinha and Macri who provided us with their page correction factors for over 500 journals.[33]

4.   Results

In Table 1 we present various descriptive statistics that should provide the reader with some appreciation of the degree of aggressiveness of each of our selected schemes. First, let us examine the percentage of output per indicator that is generated by the top thirty journals utilized by New Zealand's economists, as ranked by each weighting scheme.[34] As expected, our five 'adjusted impact citation' schemes are in a class by themselves; in all cases, over 78% of total output is attributable to the Top30 journals. Indeed, if we look at the relevant number generated by 'economic' adjusted impact measures, that is, schemes wherein only citations from economics journals on the 'list' are counted, we find between 87 and 97 percent of output is attributable to the Top30 journals. The two 'basic impact citation' schemes exhibit substantially lower estimates (58 to 50 percent for CoupeIF and JCR08IF, respectively). On the other hand, our solo hybrid scheme (Bauwens) is seen to be a relatively 'soft' weighting scheme as its top thirty journals account for only 30 percent of total output.

Table 1.   Characteristics of Selected Journal-based Weighting Schemes

	LP84	LP94	KMS	KYEI	KYPI	CoupeIF	JCR08IF	Bauwens	MSF	ERA
Year Published or Year of Version Utilized	1984	1994	2003	2006	2006	2003	2008	1998	1997	2010
Journals in Calculation	108	130	143	181	181	273	209	1200+	150	640
% of Total Output from Top 30 Journals	97.0	95.7	87.3	88.0	77.6	58.3	50.0	30.2	46.2	N/A

% of Total Pages given Non-Zero Weighting	33.8	36.5	46.5	55.4	55.1	56.8	57.1	100.0	39.3	94.3
Active Researchers as a % of Total Researchers	49.6	54.1	62.2	64.4	64.4	65.9	67.4	77.8	57.1	77.8
Gini Coefficient	0.92	0.93	0.90	0.87	0.83	0.72	0.71	0.31	0.73	0.28

Another way of looking at the aggressive/passive weighting scheme issue is to ask the following: what percentage of total size-adjusted pages produced by New Zealand's economists, over the period 2003 to 2008, are given a non-zero weighting by each of our eleven schemes? The answers range from 34 percent for LP84 to 94 percent for ERA and 100 percent for Bauwens.[35] It is rather surprising that for MSF only 39 percent of total pages are deemed worthy of inclusion in the output calculation; this estimates falls below that of KMS, generally held to be an aggressive weighting scheme and one based on the LP84 methodology.

Let us now explore the aggressive/passive weighting issue in a more rigorous fashion. In Table 1 we display the relevant Gini coefficient for each weighting scheme.[36] The data closely follows the patterns displayed by the descriptive statistics discussed above. The greatest degree of inequality is associated with the five 'adjusted impact citation' schemes. Even within this class, the coefficients follow prior patterns with schemes based on larger journal coverage, and more recent data, exhibiting less extreme results. Of particular interest is the difference in the size of the Gini Coefficients between KYEI and KYPI. It is also important to note that Bauwens and ERA possess Gini coefficients that are substantially lower than those exhibited by any other scheme in our study (at 0.31 and 0.28, respectively).

4.1. Departments

Let us now examine the impact of our various weighting schemes on departmental output indicators. In Table 2 we display the performance of New Zealand's eight economics departments under all of the previously discussed weighting schemes. Note that over the 2003 to 2008 period, 105 of New Zealand's 135 academic economists published, in whole or in part, at least one article in a journal listed in *EconLit*. At this point we must state that we have standardized our results to set the performance of the leading institution at 100, and all other scores represent the given department's score as a percentage of the leading department's score. We shall follow this practice throughout the paper.

Table 2. Normalized Departmental Scores: Weighted Pages Per Capita (2003-2008)

	EQUAL	KMS	MSF	CoupeIF	Bauwens	ERA	KYEI	KYPI	JCR08IF	LP84	LP94	CITNS
Auckland	39.8	100.0	100.0	81.5	52.2	55.9	100.0	100.0	100.0	96.4	100.0	35.1
AUT	45.8	2.3	40.5	28.2	37.1	51.8	4.6	15.7	25.5	2.2	0.5	22.0
Canterbury	49.3	39.9	89.7	87.4	54.1	59.5	68.7	95.2	82.5	100.0	28.4	40.7
Lincoln	27.9	1.8	16.6	10.5	19.2	17.9	2.7	4.1	14.6	16.3	1.7	12.4

Massey	25.4	3.3	19.6	13.1	20.1	25.1	2.5	9.2	17.9	4.2	0.9	14.7
Otago	46.2	14.8	78.6	54.4	49.7	57.8	22.0	37.2	64.5	54.7	9.2	53.1
Victoria	36.2	47.3	64.2	93.3	47.4	50.1	88.5	98.0	88.5	96.5	54.0	32.0
Waikato	100.0	16.5	85.0	100.0	100.0	100.0	22.5	56.4	99.6	18.6	2.3	100.0
Average	46.2	28.2	61.5	58.4	47.4	52.3	38.9	51.9	61.5	48.6	24.6	38.8

Perhaps the best place to start our performance review is to look at the winners: Auckland leads in six categories, Waikato in five (three of our ten core schemes and both of our reference measures: EQUAL and CITNS) and Canterbury in one. First, let us look at Waikato's performance in more detail; note that it 'wins' when output is measured by the two schemes with the lowest Gini coefficients listed in Table1, and this becomes three since the EQUAL scheme has a Gini coefficient of zero. Waikato also does well when output is measured by basic citation-based impact factors or what we might call raw impact factors. They are ranked first under CoupeIF and are a close second to Auckland under the JCRIF scheme (scoring 99.6 percent of Auckland's output level). This suggests that Waikato fares well under 'low' and 'medium' powered schemes. However, we must mention that Waikato also leads the pack in the number of non-self citations per staff member (CITNS) – an alternative indicator of performance and one that is sometime touted as the preferred measure research output.[37] In summary, Waikato does well under schemes that are relatively egalitarian in nature, but also performs well under a scheme that directly, rather than indirectly, measures a widely accepted indicator of research impact.

The situation for Auckland is somewhat different. They possess the highest share-adjusted page count per capita on four of the five measures based on the Liebowitz and Palmer methodology (LP94, KMS, KYIE and KYPI) and are a close second on the fifth such metric (LP84). In addition, Auckland also is the leader under MSF (a reputational scheme with the highest Gini coefficient of the group not based on the Liebowitz and Palmer methodology). Although it is clear that Auckland does extremely well under so-called 'strong' weighting schemes, it is interesting to note that their academic staff generate only 35 percent as many non-self citations per capita as do staff members at Waikato.

It is clear that if the New Zealand government were to adopt a research funding model based, in whole or in part, on any of our selected journal-weighted schemes, the two schools mentioned above would have a lot to gain or lose. Indeed, all schools, with the possible exception of Canterbury who do relatively well under most measures, would have reason to lobby for either strong or weak weighting schemes if research funding is to be allocated on the basis of any of the schemes utilized in this study.

Table 3. Pairwise Correlation Coefficients, Departmental Output Weighted Pages Per Capita, 2003-2008

	EQUAL	KMS	MSF	CoupeIF	Bauwens	ERA	KYEI	KYPI	JCR08IF	LP84	LP94	CITNS
EQUAL	1.00	-0.04	0.53	0.61	0.94	0.93	-0.04	0.21	0.56	-0.10	-0.18	0.95
KMS		1.00	0.72	0.62	0.25	0.22	0.93	0.85	0.73	0.81	0.98	0.07
MSF			1.00	0.88	0.75	0.76	0.72	0.83	0.93	0.72	0.60	0.64
CoupeIF				1.00	0.82	0.79	0.75	0.89	0.98	0.70	0.53	0.69
Bauwens					1.00	0.98	0.28	0.49	0.80	0.21	0.12	0.96
ERA						1.00	0.24	0.47	0.76	0.18	0.09	0.94
KYEI							1.00	0.96	0.78	0.94	0.92	0.08
KYPI								1.00	0.90	0.92	0.80	0.31
JCR08IF									1.00	0.73	0.64	0.68
LP84										1.00	0.80	0.06
LP94											1.00	-0.07
CITNS												1.00

Let us now look at the general relationship between our chosen weighting schemes. In Table 3 we present the pair-wise correlation coefficients for the previously discussed departmental output data. For discussion purposes, we shall focus on the relationship of our ten core measures with the two 'other' measures we introduced for comparison purposes. First, we will discuss the relationship between EQUAL and the remaining variables. What stands out is the weak relationship between EQUAL and the five 'adjusted impact citation' schemes in this study (LP84, LP94, KMS, KYEI and KYPI). In fact, the relevant correlation coefficients range from -0.18 to +0.21. This further suggests that schemes based on the Liebowitz and Palmer methodology can be considered aggressive weighting vehicles. At the other extreme, the relevant correlation coefficients for ERA/EQUAL and Bauwens/EQUAL are 0.94 and 0.93, respectively. Therefore, it is not surprising that Waikato ranks first under the EQUAL, ERA and Bauwens measures.

The second metric to focus on is CITNS, our measure of non-self citations per capita. Note the strong correlation between CITNS and the variables discussed immediately above. More specifically, correlation coefficients for EQUAL/CITNS, ERA/CITNS and Bauwens/CITNS are 0.95, 0.94 and 0.96, respectively. At the other extreme, we find the corresponding coefficients for LP84/CITNS, LP94/CITNS and KMS/CITNS to be 0.06, -0.07 and 0.07. At this point all we can say is that if all citations are of equal value, and if the preferred method of evaluating a paper's impact is the number of direct citations it receives, then schemes based on the Liebowitz and Palmer methodology yield problematic outcomes.[38]

4.2. Individuals

It can be expected that individual researchers have an interest in how they perform under various ranking schemes, especially those that 'count' in the eyes of their department head and Dean, and in the eyes of potential employers. That is, in addition to bragging rights, a superior performance level on the 'right' indicator can generate economic rents for researchers, especially if there is a direct link between departmental performance and research grant allocations. Let us now see if economists in New Zealand have reason to favour one scheme over another. In Table 4 we present the results for all twelve measures for the top thirty economists as ranked by their performance under the EQUAL weighting scheme. For explanatory reasons, note that Researcher1 has produced more size-adjusted pages of output, over the period 2003-2008, than any another performer, and hence has a score of 100 under EQUAL. However, Researcher1's performance under most other measures is not nearly as strong; in fact, he/she fails to produce more than 10 percent of the leader's score under any other measure save for Bauwens (69%) and CITNS (49%). In fact, under our five 'adjusted impact citation' schemes, this researcher's top score is 3.1 percent of the leader's.

On the other hand, Researcher4 generates strong results across most indicators; he/she holds 4th place under the EQUAL category (51% of the leader's output) but holds first place on seven measures. It should be noted that Researcher4 is the leader under three of our five schemes based on the LP84 methodology, and second on one other measure. However, he/she produces only 25 percent of the leader's output under KMS largely due to the fact that the KMS scheme does not cover some of Researcher4's finance publications. In addition, we wish to note that despite this researcher strong overall record, he/she only has 21 percent of the citations of the leader under the CITNS scheme, resulting in an 11th place ranking.

In general, it can be seen that the weighting scheme selection process matters to active publishers. In fact, no researcher places in the top five positions under all of our ranking schemes (even if we drop EQUAL from the list). However, Researcher4 comes close to doing so, but, as noted above, places 11th under the CITNS scheme. This issue can be addressed more rigorously by looking at the pair-wise correlation coefficients for the scores of the top 30 performers as ranked by EQUAL; the results are presented in Table 5. [39] As in the prior section, we shall focus on the relationship between our ten core measures and our 'other' schemes (EQUAL and CITNS). The results follow the same pattern as for New Zealand's economics departments, although the values are slightly higher. Once again, the relationship between our five 'adjusted impact citation' schemes and EQUAL is very weak with correlation coefficients ranging from -0.02 to 0.25. At the other extreme, we once again find a somewhat stronger relationship between ERA/EQUAL, Bauwens/EQUAL and CITNS/EQUAL at 0.80, 0.76 and 0.53, respectively.

Table 4. Normalized Research Output for Top 30 Staff Ranked by Various Weighting Schemes
New Zealand Academic Economists: 2003-2008[1]

Code#	EQUAL	CoupeIF	LP84	LP94	ERA	JCR08IF	Bauwens	KMS	MSF	KYEI	KYPI	CITNS
1	100.00	5.84	3.06	0.59	76.04	8.42	68.46	1.78	9.31	1.48	2.40	48.99
2	53.51	24.57	34.48	5.43	80.69	30.84	73.43	9.14	79.31	7.38	10.40	74.87
3	51.17	27.76	2.13	0.37	88.55	31.82	76.94	1.52	39.01	2.54	6.11	94.22
4	50.62	100.00	100.00	63.80	100.00	100.00	100.00	25.11	100.00	100.00	100.00	21.36
5	46.61	19.56	2.39	0.76	67.29	34.62	60.74	5.36	45.17	5.90	14.34	100.00
6	45.66	19.47	1.98	0.42	52.13	29.72	61.05	8.10	24.73	4.52	9.59	38.69
7	38.72	6.92	1.95	1.20	43.47	9.23	33.34	2.33	2.96	1.53	4.97	12.56
8	37.29	1.41	0.05	0.52	38.33	9.32	23.70	0.09	4.84	0.38	0.85	28.89
9	33.74	40.67	34.25	100.00	66.84	73.68	69.69	100.00	76.39	48.22	34.36	87.44
10	32.09	10.93	5.22	4.59	45.92	16.83	38.71	12.24	47.61	6.33	8.99	6.28
11	30.86	0.00	0.00	0.00	16.16	0.99	17.73	0.00	2.61	0.00	0.00	0.00
12	26.48	11.91	12.82	5.96	44.57	14.89	39.95	10.84	47.28	7.89	11.31	20.10
13	23.33	6.16	2.99	3.76	33.43	16.83	21.58	7.75	13.90	3.87	5.43	12.81
14	22.11	1.86	0.42	0.06	27.96	4.52	16.62	0.75	12.52	0.29	0.88	0.00
15	21.02	0.00	0.10	0.04	16.82	0.26	12.08	0.00	0.00	0.00	0.00	0.00
16	20.81	6.15	0.00	0.16	28.46	6.04	22.16	0.44	5.88	0.19	0.90	3.34
17	20.71	5.92	0.39	0.13	27.23	8.54	22.53	0.58	25.12	0.83	3.38	33.17
18	20.57	4.94	0.30	0.17	23.43	11.75	19.15	1.21	15.17	0.65	2.50	7.04
19	18.83	7.74	0.23	0.03	25.72	14.78	27.36	2.10	18.38	1.38	6.68	13.82
20	18.71	4.08	0.95	0.80	28.42	12.35	20.04	2.25	29.70	0.70	1.91	0.00
21	16.91	18.06	20.62	8.51	16.76	17.57	18.31	4.40	13.16	15.10	18.29	7.54
22	16.71	0.00	0.09	0.04	17.35	0.00	9.60	0.00	0.00	0.00	0.00	25.13
23	16.24	5.16	0.99	1.27	25.75	10.93	20.69	2.40	17.22	2.11	5.29	2.51
24	16.01	3.17	3.68	1.02	22.38	6.80	15.17	2.09	6.88	1.34	1.99	11.81
25	15.84	16.33	20.48	46.28	28.07	31.93	28.20	43.46	34.89	19.32	12.41	59.05
26	15.76	11.88	0.00	0.43	27.08	17.12	21.38	2.48	19.85	2.26	4.06	12.56
27	15.56	2.50	2.52	0.32	18.96	5.97	16.49	1.92	19.45	1.24	1.87	16.33
28	15.53	4.70	2.27	3.24	23.33	10.98	17.97	2.29	12.57	4.25	4.63	3.27
29	14.76	5.52	0.40	0.92	19.61	11.24	16.84	3.56	4.84	2.55	2.98	13.82
30	14.53	17.94	14.66	48.00	28.04	31.93	31.60	61.81	35.96	21.87	14.20	6.78

[1] Top 30 staff are chosen using EQUAL.

The relationship between our indicator of direct citations (CITNS) and the scores from the other eleven indicators is less extreme than under the departmental analysis, but still follows the same pattern. It can be seen that the relevant coefficients for our five schemes based on the Liebowitz and Palmer methodology lie between 0.18 and 0.36. As expected, ERA, Bauwens and EQUAL display a much stronger relationship with CITNS; the relevant correlation coefficients are 0.70, 0.69 and 0.53, respectively.

Table 5. Pairwise Correlation Coefficients, Normalized Individual Scores of Top 30 Weighted Pages Per Capita, 2003-2008[1]

	EQUAL	CoupeIF	LP84	LP94	ERA	JCR08IF	Bauwens	KMS	MSF	KYEI	KYPI	CITNS
EQUAL	1.00	0.33	0.25	0.04	0.80	0.29	0.76	-0.02	0.31	0.18	0.23	0.53
CoupeIF		1.00	0.94	0.68	0.72	0.95	0.79	0.45	0.83	0.94	0.97	0.35
LP84			1.00	0.70	0.59	0.87	0.66	0.45	0.80	0.95	0.96	0.21
LP94				1.00	0.36	0.82	0.46	0.94	0.66	0.81	0.69	0.33
ERA					1.00	0.71	0.98	0.25	0.74	0.54	0.59	0.70
JCR08IF						1.00	0.79	0.66	0.87	0.92	0.91	0.48
Bauwens							1.00	0.36	0.79	0.62	0.66	0.69
KMS								1.00	0.56	0.58	0.43	0.36
MSF									1.00	0.75	0.75	0.52
KYEI										1.00	0.98	0.19
KYPI											1.00	0.18
CITNS												1.00

[1]Top 30 are chosen using EQUAL.

Perusal of the data displayed in Tables 4 and 5 suggest that the weighting scheme selection process clearly matters to researchers that perform well on at least one of our indicators. Indeed, this is true even between the five weighting systems based on the Liebowitz and Palmer methodology. For example, consider Researcher30; that is, the 30th ranked individual under EQUAL. Under this scheme Researcher30 obtained a score of 14.5 percent of the category's leader. Now let us look at how this individual's scores vary across the members of our 'adjusted impact citation' group. From Table 4 we see that Researcher30 obtains a score of 62 percent under KMS, and 48 percent under LP94, but if any of the other three members of the group were to serve as the official weighting scheme, he/she would not fare nearly as well. More specifically, Researcher30 would see her/his scores decline to 15, 22 and 14 percent under the LP84, KYEI and KYPI schemes, respectively, essentially the same score as he/she recorded under EQUAL (as always, the reported score represent the individual's score as a percent of the category leader's score).

Let us explore the above issue in a somewhat more rigorous fashion, starting with LP84, the earliest of the 'adjusted impact citation' schemes. From Table5 we see that LP84 is highly correlated with KYEI and KYPI (0.95 and 0.96, respectively) but less so with the other two members of the group (LP84/LP94= 0.70 and LP84/KMS= 0.45). A similar pattern is exhibited by LP94 and KMS, KYEI and KYPI; the respective correlation coefficients are 0.94, 0.81 and 0.69. Although the results indicate a strong relationship pattern, in some cases it is far from 1.0. Therefore, the selection process should still matter to individual researchers

even if they believe Liebowitz and Palmer methodology to be the best available technique for measuring research performance.

5. Conclusion

The rationale for our study is the belief that more and more nations will adopt bibliographic techniques as a short-cut for measuring the level of research output in their nation's universities, and that the results of this exercise will be used, at least in part, to allocate funding between the competing institutions. This led us to test the applicability of various citation-based journal weighting schemes as instruments for measuring economic research output in the context of a nation-wide research funding model. To do so, we divided the task into two elements. First, we discussed the evolution of and nature of citation-based journal weighting schemes, with particular attention paid to those based on the 'adjusted impact citation' methodology first developed by Liebowitz and Palmer (1984). Second, we tested our selected weighting schemes by using data generated by academic economists located in economics departments of New Zealand's universities.

If our premise is correct, it is vitally important that all actors involved understand the implications of using one journal-based weighting scheme over another in the prevailing institutional context. From a given university's perspective, it is important for researchers and administrators to understand how the journal selection process affects the output performance and, hence, research funding. From the public granting agency's perspective, it is important that the selected scheme encourage activity in line with government's goals and objectives. For example, is research on local and regional issues of importance to the nation state; and, if so, does the journal weighting scheme reward such activity? Similarly, does the journal weighting system reward economists for their contribution to public policy debates or does it primarily reward economists for making a contribution to the discipline of economics?

Our findings suggest that the quantity/quality issue, as traditionally defined, is an important one for all concerned. More formally, we explored this issue in the context of strong versus weak journal weighting schemes. Recall that all five 'adjusted impact citation' schemes displayed Gini coefficients in the 0.87 to 0.93 range, whereas the ERA and Bauwens measures exhibited Gini coefficients of 0.28 and 0.31, respectively. Clearly some schemes are less egalitarian than others. We also found that under some schemes virtually all research output was deemed to be of value, whereas under the 'adjusted impact citation' schemes only 34 to 55 percent of published pages were given a non-zero weighting.

Our departmental and individual results demonstrate clearly that the weighting scheme selection process does matter to those involved. It is clear from our departmental analysis that two departments in New Zealand dominate the ranking game: Auckland and Waikato lead in six and five of our twelve categories, respectively. Waikato was seen to do very well under our so-called 'weak' schemes, and Auckland performs exceedingly well under the

'strong' schemes. Similarly, the results for individual researchers demonstrate that, with one notable exception, bragging rights are generally metric specific. All of this suggests that if New Zealand were to implement a citation-based journal weighting scheme for measuring research output for the purpose of allocating funding, individual researchers and university administrators should not be indifferent between approaches. They have much to lose or gain depending upon their research strategies.

Also of interest is the extent to which results, at both the departmental and individual level, vary between measures based on the same methodology. More specifically, the five schemes employing the Liebowitz and Palmer (1984) approach do, on many occasions, generate substantially different outcomes. All of this suggests that the range of journals selected for inclusion in the 'list' of relevant journals, the time period over which citations are collected, the number and type of journals from which citations are collected, and how differing citation practices between sub-disciplines are accounted for, are likely to affect departmental and individual results.

In conclusion, we agree with Engemann and Wall (2009. p.138) that 'There is no such thing as the correct ranking of economics journals. Instead, there is a universe of rankings, each the result of a set of subjective decisions by its constructor.' However, it is important for all participants to understand the nature of possible weighting schemes they may face, and to either adjust their research practices accordingly if they wish to participate successfully in the research funding arena, and/or to lobby for measurement systems that reward activities they deem appropriate for both academia and society at large.

Notes

[1] See Oswald (2007) for a brief discussion of this issue.

[2] It must be stressed that in this paper, and throughout the economics-based output measurement literature, countable research consists only of refereed articles published in a given set of journals (usually a subset of journals listed in *EconLit*). Obviously, this definition of research ignores many other forms of scholarly output such as books, monographs, conference papers, working papers, and refereed journals not included in the study.

[3] For a more detailed review of the rankings literature in economics, see Macri and Sinha (2006).

[4] It should be noted that in 1972, Billings and Viksnins published the results of a much more limited citations- based study. They tracked the citations from three journals (*American Economic Review, Econometrica and the Economic Journal*) to other journals and proceeded to generate a set of journal rankings.

[5] It is not clear how the fourteen journals were selected.

[6] Over the period 1956 to the late 1980s, the literature appears to be almost entirely U.S. focussed.

[7] Once again, it is not clear how the authors selected these journals. This is a recurring theme until the pioneering work of Liebowitz and Palmer (1984).

[8] We shall return to a discussion of the relative merits of the stock and flow method later in this paper. It should be noted that there are circumstances under which the flow method is appropriate.

9   Another problem, albeit relatively minor, with the work of Graves, Marchand and Thompson (1982) and, undoubtedly, prior studies employing the same methodology, was pointed out by Hirsch, Austin, Brooks and Moore (1984). They suggest that such studies claim to be reporting on the output and rankings of economics departments, but are in fact reporting on the output of economists within a given institution regardless of place of employment. Therefore, institutions with a large number of economists in business schools, agricultural economics departments, and law schools will bias the results in their favour. It should be noted that this problem continues to plague many studies (as noted by Macri and Sinha (2006)).

10   In this same year, Davis and Papanek also employed citation analysis to address the major deficiency of prior work: the assumption that all journals, within a group selected for study, are of equal value. However, Davis and Papanek (1984) rejected the journal-based approach and moved directly to a ranking of PhD granting economics departments (in the USA) based on the number of citations attributed to their faculty.

11   The journals selected for inclusion in this study are almost exclusively those classified by SSCI as economics journals as at 1980.

12   Although we are getting ahead of ourselves, it should be noted that the per-character adjustment process is essentially the same as the standardized page approach adopted by many researchers. Under the latter approach, all journal pages are adjusted so that they contain the same word count as an average page in the *American Economic Review*. It must also be denoted that the per-article adjustment process (dividing the 'adjusted impact citation' factor by the relevant number of journal articles), appears to be making a comeback.

13   In this section of the paper, we have selected the article, rather than the character (or page), as the appropriate scale adjustment procedure.

14   Although 108 journals are included in the calculation, only 99 journals receive a non-zero weighting under the 'impact adjusted citation' approach.

15   For a general critique of the Liebowitz and Palmer methodology, see Engemann and Wall (2009).

16   Laband and Piette (1994), with a few minor adjustments, adopted the SSCI economics journal list.

17   It should be noted that the *Journal of Finance* is not on the KMS list, but is included in the LP84 and LP94 weighting schemes.

18   For example, Macri and Sinha (2006) refer to the KMS weighting scheme as the 'industry standard'; and, although in disagreement, Henrekson and Waldenstrom (2008) state that the scheme is held in high regard by many influential economists.

19   See Garfield (2003) for a discussion of the origins of the impact factor concept and for the two year citation collection period in the JCR impact factor calculation process.

20   Coupe (2003) included all of the 800 plus journals listed in *EconLit* in 2000. However, only 273 of these journals are included in the JCR database.

21   Bauwens does not provide a rationale for his cut-off point selection process. He also fails to provide details on the number of journals for which he calculated scores based on citation counts. However, his selection process is the same as that of Coupe (2003): he started with all *EconLit* journals and then selected those also listed on JCR without regard to JCR's classification system.

22   Strictly speaking, citation-based schemes attempt to measure the impact of a researcher's work, not its quality. However, this distinction is frequently glossed over, as discussion of departmental and individual rankings are generally presented, either explicitly or implicitly, as quality measures. For an extensive discussion of this important distinction, see Beed and Beed (1996).

23   Although the ERA programme was first announced in 2008, the final journal weights were only released to the public in February 2010. We should note that we have used the final weights in this study, not the draft weights that were released in June 2008. For a user-friendly presentation of the final ERA journal weights see John Lamp's website: www.lamp.infosys.deakin.edu.au/er/.

24 For details on the scheme see the Excellence in Research for Australia programme's official website at www.arc.gov.au/era.

25 Although 640 of the 1200 plus journals listed in *EconLit* were assigned to the 'economics' classification, many other journals recognized by *EconLit* are included in other ERA categories such as 'Banking, Finance and Investment', 'Urban and Regional Planning' and 'Policy and Administration'.

26 For a more detailed assessment of this scheme, see Anderson and Tressler (2009).

27 Since we are using New Zealand data to assess the impact and nature of our various weighting schemes, we face a similar problem in how to threat the *New Zealand Economic Papers*, the leading economics journal in the country. We have granted the *New Zealand Economic Papers* the same grade as the *Economic Record* throughout this study. The rationale for doing so is the same as that used by the ERA designers- the need to overcome the small nation bias in citation-based schemes. Indeed, one might argue that the problem facing New Zealand journals is even more severe than that facing their Australian counterparts given the relative size of the two nations.

28 Given the lag between the publication of a refereed paper and the release of a citing article, it is unrealistic to use the same publication period that we have used for measuring refereed paper output (2003-2008). Instead, we have adopted a two year lag between initial publication date and the start of the citation collection period. More formally, on 19 February 2009, we counted non-self citations from all SSCI listed publications to all papers in our database published over the period 1 January 2001 to 31 December 2006.

29 It is not our intention to get involved in assessing the relative merits of these two competing approaches for measuring research output, other than to repeat what has been discussed previously in this paper. That is, the nature of the publication/citation time lags facing researchers in economics (and more generally, the social sciences) are generally thought to be much longer than those facing researchers in science-based disciplines. It is for this reason that journal weights, based indirectly on citation counts, are used as a proxy for the likelihood of the number of citations that an article will ultimately generate.

30 It should be noted that there are only five pure economics departments in the country. Another institution has a combined department but formally lists the economists within the group. Two institutions have combined departments, with economists officially comingled with finance staff and, in one case, also with accounting staff. In order to separate economists from the broader pool in these two cases, we initially provided our list of economists to the relevant department heads for comment. This was done in mid-2007, and in order to determine staff changes over the remaining period of our study, we utilized departmental and personal web-pages to determine teaching and research interests. Given the small number of individuals involved, this was not a difficult task and we feel confident in our findings.

31 We have restricted our dataset to academics holding 'regular' appointments at New Zealand universities. By 'regular' we mean an appointment akin to a tenure or tenure track position in a North American setting. A regular appointment need not be a full-time appointment. As a result, Peter Phillips, an internationally renowned economist, is not included in the University of Auckland data. We should also note one other adjustment with respect to another prolific publisher: due to his explicit cross-listing between the economics and finance departments at Victoria University, only 50% of Graeme Guthrie's output is included in Victoria's economics department output, but no such adjustment has been made in calculating individual output.

32 The alternative scheme is known as the flow approach. This scheme results in a measure of the volume of output produced by a given department over a given period of time. It provides a measure of past performance, but it may say little about current capability, especially if the institution has lost its best researchers at some point over the period of analysis. In a world of small departments, as is the norm in New Zealand, this can be an important issue.

[33] By using Sinha/Macri page correction factors, we were able to adjust for page size differentials for 178 of the 244 journals utilized by NZ academics during the period 2003-2008. Those journals for which we lacked page correction factors were almost always unranked or ranked in the lowest possible category by all of our selected weighting schemes. For these journals, we used 0.72, the average page correction factor for Gibson's (2000) lowest ranked journal group. For details on the Sinha/Macri page correction procedure, see Sinha and Macri (2002).

[34] We must stress that by Top30 journals, we mean the thirty most important journals for each of the selected weighting schemes. That is, the Top30 for one scheme will often be quite different from the Top30 under an alternative scheme. Furthermore, we should clarify the phrase 'as utilized by New Zealand's economists'. This means that we have selected the top thirty journals (for each weighting scheme) that contain at least one article published over the period 2003 to 2008 that was authored by a staff member (as at 15 April 2009) of a New Zealand economics department.

[35] It should be noted that we have opted for a broad as opposed to a narrow definition of relevant journals with respect to the ERA journal weighting scheme. That is, as long as a journal is listed in *EconLit*, we assign it the relevant ERA weighting regardless of the discipline category to which it was arbitrarily assigned. For example, researchers with publications in the *Journal of Finance* and *Regional Studies* are given the grade assigned to them in their home discipline categories. The 'narrow' approach is to grant recognition only to journals listed under the economics discipline, and to ignore the many journals in 'border' areas such as finance, urban and regional studies and policy.

[36] The Gini coefficients presented herein are calculated over the weights of the journals in which New Zealand economists published during the period of this study. That is, they are not Gini coefficients calculated over the weights of all possible economics journals or over the pages published by New Zealand economists.

[37] It must be stressed that CITNS is based on the assumption that all citations are of equal value; that is, we have not weighted them by any measure of relative worth.

[38] Since our focus is on the five members of the 'adjusted impact citation' group, we should examine the relationship between the departmental scores generated by these weighting schemes. It can be seen from Table 3 that all of the correlation coefficients involving pairs of this group are relatively high, ranging from 0.80 to 0.98.

[39] We must acknowledge that the correlation coefficients would be higher if we were assessing the scores of all active researchers (105) and all researchers (135). We justify focussing on the top30, as opposed to either of the above alternatives, since performance scores are most likely to be of interest to researchers who do well under at least one measures- these are the people who will be in demand in the open market and who will be in a strong bargaining position within their institution. We should also note that as we expand the range of coverage, the number of 'zero' observations increases and this reduces the degree of 'difference' between our schemes.

References

Anderson, D. L. and Tressler, J. (2009) The 'Excellence in Research for Australia' Scheme: a Test Drive of Draft Journal Weights with New Zealand Data. *Agenda* 16(4): 7-24.

Australian Research Council (2008) *New ERA for Research Quality: Announcement of Excellence in Research for Australia Initiative*. Media Release, 28 February.

Bauwens, L. (1998) *A New Method to Rank University Research and Researchers in Economics in Belgium*. Unpublished Paper, CORE, Universite Catholique de Louvain, Louvain, Belgium. (www.core.ucl.ac.be/econometrics/bauwens/rankings/method.doc)

Beed, C. and Beed, C. (1996)   Measuring the Quality of Academic Journals: the Case of Economics. *Journal of Post Keynesian Economics* 18(3): 369-396.

Billings, B. B. and Viksnins, G. J. (1972)   The Relative Quality of Economics Journals: An Alternative Rating System. *Western Economics Journal* 10(4): 467-469.

Bush, W.C., Hamelman, P.W. and Staaf, R.J. (1974)   A Quality Index for Economic Journals. *Review of Economics and Statistics* 51(1): 123-125.

Coupe, T. (2003)   Revealed Performances: Worldwide Rankings of Economists and Economics Departments,. 1990-2000. *Journal of the European Economic Association* 1(6): 1309-1345.

Davis, P. and Papanek, G.F. (1984) Faculty Ratings of Major Economics Departments by Citations. *American Economic Review* 74(1): 225-230.

Engemann, K.M. and Wall, H.J. (2009) A Journal Ranking for the Ambitious Economist. *Federal Reserve Bank of St. Louis Review* May/June: 127-139.

Fusfeld, D.R. (1956) The Program of the American Economic Association Meetings. *American Economic Review* 46(4): 642-644.

Garfield, E. (2003) The Meaning of the Impact Factor. *International Journal of Clinical and Health Psychology* 3(2): 363-369.

Garfield, E. (1972) Citation Analysis as a Tool in Journal Evaluation. *Science* 178: 471-479.

Gibson, J. (2000)   Research Productivity in New Zealand University Economics Departments: Comments and Update. *New Zealand Economic Papers* 34(1): 73-87.

Graves, P.E., Marchand, J.R. and Thompson, R. (1982) Economics Departmental Rankings: Research Incentives, Constraints, and Efficiency. *American Economic Review* 72(5): 1131-1141.

Henrekson, M. and Waldenstrom, D. (2008)   *How Should Research Performance be Measured? Evidence from Rankings of Academic Economists.* Stockholm School of Economics, Working Paper No. 693, Stockholm, Sweden.

Hirsch, B.T., Austin, R., Brooks, J. and Moore, J.B. (1984)   Economics Departmental Rankings: comment. *American Economic Review* 74(4): 822-826.

Kalaitzidakis, P., Mamuneas, T. and Stengos, T. (2003)   Rankings of Academic Journals and Institutions in Economics. *Journal of the European Economic Association* 1(6): 1346-1366.

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): 640-666.

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 to Economics in Australia and New Zealand. *Journal of Economic Surveys* 20(1): 111-156.

Mason, P., Steagall, J. and Fabritius, M. (1997) Economics Journal Rankings by Type of School: Perceptions Versus Citations. *Quarterly Journal of Business and Economics* 36(1): 69-79.

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

Palacios-Huerta, I. and Volij, O. (2004)   The Measurement of Intellectual Influence. *Econometrica* 72(3): 963-977.

Sinha, D. and Macri, J. (2002) Rankings of Australian Economics Departments, 1988-2000. *Economic Record* 78(241): 136-146.

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.