

UNIVERSITY OF WAIKATO
Hamilton
New Zealand

**Are Researcher Rankings Stable
Across Alternative Output Measurement Schemes
in the Context of a Time Limited Research Evaluation?
The New Zealand Case.**

David L. Anderson and John Tressler

Department of Economics

October 2015

Working Paper in Economics 10/15

Corresponding Author

John Tressler

Economics Department
University of Waikato
Private Bag 3105
Hamilton, 3240
NEW ZEALAND

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

David L. Anderson

School of Business
Queen's University
Kingston
Ontario K7L 3N6
CANADA

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

Abstract

This paper focuses on the stability of rankings of academics by research productivity in the context of short-term decision making. In particular, the growing use of national research assessment exercises (NRAE) has increased interest in identifying the contributions of individual researchers to an assessment unit's output and ranking. Our primary finding is that despite statistical evidence of a high degree of rank stability across a number of plausible journal weighting schemes, the journal selection process is of great importance to individual researchers. This applies with particular force to academics working within a NRAE environment based on individual assessment such as New Zealand's PBRF.

Keywords

research measurement
journal ranking schemes
time pattern of citations

JEL Codes

A14, C81, I23, J24

1. Introduction:

The growing use of national research assessment exercises (OECD, 2010) and other resource allocation schemes based on research performance has increased interest in identifying the contributions of individual researchers to an assessment unit's output and ranking. From a management perspective, the identification of highly and lowly ranked performers is a vital component in hiring and retention strategies that are designed to improve a unit's rankings and frequently its share of available funding. Unfortunately, the selection of the 'best' ranking scheme is not an easy task for at least two reasons: first, the selection process is inherently subjective as typified by the numerous research assessment measures (RAMs) in widespread use; and second, virtually all current national research assessment exercises (NRAE) are primarily based on peer review, thus requiring research managers to speculate on the 'true' underlying evaluation scheme. The uncertainty associated with this selection process leads to the obvious question: does it really matter which RAM is selected for the identification of leading researchers? Rephrased, how stable are researcher rankings across a range of plausible measurement schemes?

Let us now identify and characterize the range of ranking schemes referenced above. At a macro-level the debate is between citation and journal- based schemes. If citations are deemed to be the basic building block of the assessment scheme, one has to choose between well-known, multi-disciplinary citation collectors such as Google Scholar (GS), Web of Science (WoS), and Scopus. Thereafter, one has to choose among a large number of metrics or devise an arbitrary weighting scheme to combine various measures for the generation of a unique ranking of researchers (for example, see Chang and McAleer 2013). Furthermore, should one count raw citations, adjust citations to reflect the importance of the citing journal, or employ a citation-based scheme, such as the h-index (Hirsch 2005), designed to reflect differences in citation patterns across a given body of work. The options are numerous.

Advocates of journal weighting schemes (henceforth denoted as JWS) also face a number of operational issues (Macri and Sinha 2006). A choice must be made between measures based on reputational surveys, citation-based impact factors, and schemes that differentiate between citations from low and highly ranked journals. In addition, the basic unit of output must be selected: is it the number of pages produced by the researcher or the number of articles? If pages, should one account for differences in average word count per page across journals? Finally, and an issue that applies to both citation and journal-based schemes, for multi-authored papers, should each author receive full or partial credit for her/his contribution to the final product? For example, in a jointly authored paper, should each author receive credit for all cites to the paper, or half of the total cites, or some other arbitrarily determined share?

In this paper we shall examine the sensitivity of rankings of New Zealand's academic economists to various RAMs. This country was an early entrant into nation-wide research

assessment exercises, and in 2012 completed the third round of its NRAE – the Performance Based Research Fund (PBRF).¹ Although our paper does not formally fit within the structure of the PBRF, it does rely on the scheme for context. The PBRF assesses the performance of almost every academic expected to undertake research in all eight of the nation's universities. For example, in the last round of the PBRF (2012), all researchers received a letter grade for work completed over the 2006-2011 period. Letter grades were subsequently converted to numerical scores and then aggregated across departments and other groupings to generate university-wide scores.

Approximately 20 percent of the government's total funding to the universities is allocated through the PBRF scheme and the winners, at both the departmental and university level, aggressively promote the results on their websites and in promotional material. As an aside, we note that only 70 percent of the possible score for each individual was based on research performance over the six year assessment period,² and that evaluation was by peer review rather than bibliometrics or some combination of the two. Nevertheless, given advances in bibliometrics, the financial cost and administrative burden of operating a peer review assessment, and the inherent 'black-box' nature of peer-review, we suggest that it is highly likely that bibliometrics, in some form, will eventually be incorporated into the PBRF or successor research evaluation regimes.

As noted above, the 2012 PBRF serves as a backdrop to this study. First, the research evaluation period is six years: 2006-2011. This constraint lies at the heart of such exercises; that is, NRAEs typically restrict relevant output to a five to seven year period (we assume six years since it is the period used by the PBRF). Rephrased, it is not life-time citations or published papers that are rewarded, it is research activity over the evaluation period. Secondly, the various measurement schemes used in the study were selected, in part, because they have relevance to the New Zealand academic scene. Thirdly, the desire by universities to have the results of the exercise conveyed as quickly as possible, and to have the associated spoils delivered soon thereafter, precludes or limits the relevance of evaluation tools that require a longer time period than roughly three years for reliable data collection and analysis or require the decision period to be delayed so that additional data can be collected.³

¹ Details of the third round of the PBRF can be found in TEC, 2013. A fourth round is scheduled for 2018. For a discussion of PBRF-specific measurement issues see Anderson, Smart and Tressler (2013), Hodder and Hodder (2010), and various papers by Warren Smart, especially Smart (2009).

² For the 2012 PBRF evaluation, two other assessments of individual research performance were undertaken by a discipline specific review panel: peer esteem (15%) and contributions to the research environment (15%).

³ Although the assessment period is six years, the critical period is three years since this is the average period of time for research outputs to attract citations or any other form of recognition.

As noted above, the objective of this paper is to explore the stability of ranks for individual researchers across various research assessment measures. In order to do so, we proceed as follows: the following section discusses our data sources and our selection of RAMs; this is followed by a presentation of our results; and the paper concludes with a summary of work and a discussion of the implications of our findings.

2. Data and the Selection of Research Assessment Measures

For purposes of this study, we have restricted countable research to refereed papers published in journals listed in EconLit as at 17 July 2012. We acknowledge that this ignores books, conference papers and working papers; however, it is widely acknowledged that the primary currency in the PBRF decision process is refereed articles. Our staff census date was 15 April 2012: at that time there were 134 academics employed in New Zealand's eight economics departments. However, only 113 of these researchers were research active: that is only 84% of academic economists published a paper, or part thereof, in at least one EconLit listed journal over the period 1 January 2006 to 31 December 2011. We deem members of this group to be 'Active Staff'; and given that the focus of this paper is on the stability of researcher rankings across various research assessment measures, this is the obvious study target group.

The primary information source for publications was individual CVs listed on departmental and/or personal websites, supplemented by scrutiny of other websites such as Web-of-Science (WoS), EconLit, RePEc and Google Scholar (GS).⁴ In total, New Zealand economists published 743 refereed papers in EconLit listed journals over the six year period commencing on 1 January 2006. Of relevance to later work is the fact that publications were not evenly distributed over the six year period. Instead the data were somewhat skewed towards the latter part of the period. More specifically, 59.1 percent of papers were published in the last half of the six year cycle, with the largest number published in 2011.⁵ Since, for an individual researcher, the PBRF letter grade stays the same for six years, this 'rush to the finish line' is to be expected, especially for new entrants and those with promotion prospects or desires to relocate.⁶

⁴ Only formally published papers (those with a formal volume, issue and page numbers (or item number if an on-line journal using this format)) have been counted, not 'early bird' versions or similarly named releases. However, It should be noted that the PBRF rules were relaxed late in the process to allow researchers to count accepted papers that were released on-line prior to 31 December 2011.

⁵ The time pattern of publication of the 743 articles produced by New Zealand economists over the most recent PBRF evaluation period is as follows: 110 (2006); 84 (2007); 108 (2008); 133 (2009); 149 (2010) and 159 (2011).

⁶ Although individual results are technically confidential, many high and medium level achievers find ways of disseminating their results. Hence it frequently becomes relatively easy for administrators and others serving on promotions and hiring committees to determine an applicant's PBRF grade.

Before considering specific RAMs, we must address a critical assumption that affects the results generated by all research measurement schemes: the treatment of output shares associated with multi-authored papers. The conventional approach is to allocate output shares on the basis of the $1/n$ rule. For example, on a paper with three authors, each researcher is granted credit for $1/3$ of a paper and $1/3$ of the total pages associated with the paper. To do otherwise rewards game playing activities (Liebowitz 2014). Although we support the share allocation approach (or complete proration in the words of Liebowitz), we also generate results based on the polar assumption of full credit to all authors so as to compare the resulting rankings with those forthcoming from the $1/n$ rule.

As noted in the introduction, there are two basic approaches to measuring individual research output: schemes based directly or indirectly on citation counts to individual papers, and those that rely on the source of publication (the journal) as a proxy measure of the expected impact of a paper. We have opted for the latter for a number of pragmatic reasons. First, as Liebowitz (2014, p.1268) note 'Survey evidence indicates that economics departments prefer to use an ex ante measure of quality, the journal of publication, as the main indicator of success even though more direct ex post measures, such as citations, are available'. Arguably the primary reason for this convention is that the results of individual efforts are available immediately upon acceptance of a paper, and hence useful in time-limited personnel decisions related to hiring, promotion and tenure.⁷ The underlying rationale being that the average number of citations per paper in a given journal serves as a proxy for the expected number of cites to the paper under discussion. However, it must be acknowledged that many papers in good journals are rarely cited and many papers in lower ranked journals are sometimes highly cited (Oswald, 2007 and Chang, McAleer and Oxley 2011).⁸

A second reason for favouring the journal approach over citation counting relates to time-lag issues. Recall that the focus of this paper is rank stability in the context of a NRAE-like environment. Hence, papers have, on average, approximately three years to accumulate citations; in fact, in the context of the 2012 PBRF evaluation round, the median economics paper had slightly less than 2.5 years in which to do so.⁹ It is widely accepted that this is too

⁷ For a discussion of short-term versus longer term indicators of paper success, see Sgroi and Oswald (2013) and Stern (2014).

⁸ Other limitations of journal rankings can be found in Laband, 2013.

⁹ Given that 743 papers were published over the 2006-2011 period, it is reasonable to assume that the median paper (the 372nd) was not published until the second half of 2008 (Year 4). From footnote 3, note that the sum of papers published in Years 1-3 is 302 (110+84+108); hence the 70th paper published in Year 4 is the median paper. Since only 133 papers were published in Year 4, the median paper was released in the latter half of the year (assuming an even flow of publications over the year). Hence the median paper has slightly less than 2.5 years in which to gather citations.

short a period for the production of reliable measures of performance based on citation counts to individual papers in economics.¹⁰ For this reason, and perhaps others, all known NRAEs currently rely on peer review for assessment of economic research. Although we believe that the journal weighting scheme approach is superior in the context of the PBRF and, more generally, economics, we realize that the citation time-lags facing physical and natural scientists are much shorter and may justify the use of direct citation measures in NRAEs (for example, see Abramo, Cicero and D'Angelo 2011). In summary, the citation/journal debate centres on whether one approach fits all disciplines.

Many JWS exist, so we were forced to be somewhat arbitrary in our selection process; nevertheless, we have selected schemes that encompass all three competing methodologies (peer assessment; impact factors, and recursive adjustment/eigenfactors) and also address issues of relevance to the New Zealand scene. We have chosen five JWS based on diverse assumptions and methodologies that should shed light on the stability of rankings. In addition, we have created two additional schemes that are based directly on the five independently generated JWS.

Our first selection is a perception-based ranking scheme that is widely used in New Zealand business schools to assess research quality – the Australian Business Deans Council's Journal Quality List (2010). Although produced in a neighbouring country, the two nations share many values and have university systems that have much in common. We have used the 2010 version since it corresponds most closely to the time period of our study, and we label the scheme as **ABDC**. The **ABDC** covers many disciplines in addition to economics, and the rankings are based on inputs from various discipline groups and ultimately determined by a committee of the Australian Business Deans Council. Journals are ranked on a letter grade system: A*, A, B, C and implicitly X (not ranked). We have arbitrarily assigned numerical values to each grade: A* (4), A (3), B (2), C (1) and X (0).¹¹

¹⁰ For example, Tressler and Anderson (2012) found that WoS cites to economics papers over the 2001-2007 period were insufficient to generate reliable measures of performance New Zealand's academic economists – the lags were too long between publication and commencement of a meaningful flow of citations. On the other hand, Stern (2014) and Burns and Stern (2015) argue that citation flows in the first two to three years after publication can be used to reliably identify economics papers that are ultimately highly ranked. However, the relevance of these findings are limited in the context of this paper: first, between 40 and 60% of the papers in our dataset (see footnote 5) were published short of the time period Stern (2014) identifies as being of sufficient length to generate reliable estimates; second, WoS data only covers 60% of the papers in our database; and third, the Stern study focuses on top- end papers (the top quintile of total citations at the end of year 7).

¹¹ As noted, the numerical scores used to weight pages and papers for measuring ABDC output were arbitrarily assigned by the authors. The ABDC (2010) report is silent on this issue. Therefore, we have adopted the numerical weighting scheme referenced by Abelson (2009) in his discussion of the development of the Economic Society of Australia's journal weighting scheme that served as an important input into the development of the ABDC (2010) list of economics journals.

Although the economics list covers 907 journals, we recognize that boundary issues exist, especially with respect to finance and urban studies. We have, therefore, adopted a broad definition of relevant work: all papers in EconLit listed journals are deemed to be countable papers regardless of the ABDC category under which they appear. As shown in Table 1, this scheme assigns scores to 95 percent of the papers published by New Zealand’s academic economists in EconLit journals over the period 2006 to 2011.

Table 1: Coverage of 2006-2011 Publications Provided by Various Journal Weighting Schemes

	Rrec/Rsif	CLm/CLh	ABDC/NZABDC
Percent of total papers covered by JWS	83.4	95.3	95.0
Percent of share adj papers covered by JWS	82.8	96.6	95.1
Percent of share adj and size adj pages covered by JWS	86.0	96.3	95.0

Our second perception based scheme is based, in part, on the Australian Business Deans Council journal rankings, and a set of journal weights derived by Anderson, Smart and Tressler (2013). These weights are those associated with the highest Pearson correlation coefficient between the evaluations of the research of individuals by a proxy of **ABDC** and the research output scores assigned by the PBRF peer evaluation committee¹². We have labeled this RAM as **NZABDC** in recognition of the fact that the scoring system is based on both New Zealand data and the Australian Business Deans Council’s journal ranking scheme.

Although we characterized the decision process as choosing between citation-based measures and journal weighting schemes, most of the latter are, in fact, based indirectly on citations. However, the focus is on the journal that published the paper, not the individual paper itself. Indirect citation schemes can be divided into two groups: those based on simple impact factors that give equal weight to all cites regardless of citing source, and those assigning different weights to citations depending upon the importance of the citing journal. The former is based on Garfield (1972) and the latter on the pioneering work of Liebowitz and Palmer (1984).

We have chosen two JWS developed by RePEc that are based on differing methodologies: the RePEc Simple Impact Factor (**Rsif**) and the RePEc Recursive Impact Factor (**Rrec**)¹³. The journal weights for Rsif and Rrec used in this study are those listed on

¹² In this instance, the ERA (Excellence for Research in Australia) ranking scheme is used as a proxy for **ABDC**. Both these schemes relied heavily on the same source- a survey of professors by the Economics Society of Australia.

¹³ For a discussion of these impact factors and additional RAMS generated by RePEc see: Zimmerman (2013) and Kim, Min and Zimmermann (2011).

the RePEc website on the 25th and 29th of June 2015, respectively. As shown in Table 1, **Rsif** and **Rrec** assign non-zero weights to over 83 percent of the papers in our database. The RePEc version of a simple impact factor (**Rsif**) is based on the total number of cites to all articles in a given journal up to the date of calculation divided by the number of articles published since the start of the collection period. **Rrec** is based on the same dataset as **Rsif** but employs a recursive adjustment/eigenfactor methodology to assign differential weights to citations based on the source of the citation¹⁴.

Two additional JWS have been selected for study based on a mix of methodologies in order to generate scores for a very broad range of economics journals. Combes and Linnemer (2010) provide scores for 1168 economics and related journals based on a combination of WoS and GS citation data, and various methodologies such as impact factors, field specialization adjustment indexes and h-indexes. They arbitrarily impose restrictions on the data to yield JWS that display varying degrees of convexity. In practical terms, Combes and Linnemer alter the rate at which scores decline as one moves down the list of journals, and they labeled the scheme exhibiting the steepest decline in scores as **CLh** and the one exhibiting smaller declines as **CLm**. Note that both **CLm** and **CLh** generate non-zero scores for over 95% of the papers in our database

In addition to the weighting regimes presented above, we have added a seventh-**AVJWS**. This scheme is simply the arithmetic mean of the rankings of the previously presented JWS (**ABDC**, **NZABDC**, **Rsif**, **Rrec**, **CLm** and **CLh**). Although admittedly arbitrary in design, **AVJWS** is an example of a plausible attempt to reduce alternative rankings to a scalar measure - a common practice in an environment characterized by uncertainty over the 'true' underlying RAM (Chang and McAleer 2013).

We have previously referenced the degree of convexity inherent in our selected weighting schemes, and in Table 2 we address the matter in a somewhat more formal manner. We have selected **CLm** as a reference point, and have arbitrarily selected journals at several places in the rankings. We then display the number of papers required for publications in each of these journals to generate the same score as that assigned to a single paper in the top ranked journal - the *Quarterly Journal of Economics (QJE)*. For example, the 40th ranked

¹⁴ Arguably the most celebrated recursive adjustment journal weighting scheme in economics is that developed by Kalaitzidakis, Mamuneas and Stengos (2003, 2011) and generally labeled as KMS. However, we have not adopted this RAM since it is arguably best suited to evaluating work of leading institutions and leading economists. For example, with respect to our database, KMS assigns non-zero weights to only 46 percent of the papers produced by New Zealand economists over the period 2006-2011. In addition, for a large number of regional journals, the weights are trivial. For example, under the 2011 version of KMS, a researcher must publish 221 papers in the *Economic Record*, 1988 in the *Australian Economic Review* and an infinite number in the *New Zealand Economic Papers* (unranked by KMS) to equal one in the *Quarterly Journal of Economics*. We should note that the *New Zealand Economic Papers* is New Zealand's leading economics journal and the journal that published the most papers produced by New Zealand's academic economists over the 2006-2011 period.

journal under **CLm** is *Health Economics*. The number of publications in this journal required to match one in *QJE* ranges from 1.33 under **ABDC** to 46.0 under **Rrec**. Two examples have been provided that have special relevance to New Zealand economists: the relative impact of publishing in arguably the top Australasian journal, the *Economic Record (ER)*, and in New Zealand's leading economics journal, the *New Zealand Economic Papers*. The *QJE/ER* paper ratio ranges from 1.33 under the **ABDC** to 91.0 under **Rrec**. The corresponding values for *QJE/NZEP* are 2.0 for **ABDC** and 208.3 for **CLh**.

Table 2: Number of Papers Required to Equal a Single Paper in the Quarterly Journal of Economics (Various Weighting Schemes)

CLm Rank	Journal Name	Rrec	Rsif	CLm	CLh	ABDC	NZABDC
9	Ec Journal	5.7	3.1	1.6	2.4	1.0	1.0
20	Games and Ec Behavior	7.4	6.6	1.8	3.4	1.0	1.0
30	J of Env Ecs and Management	10.1	3.8	2.7	7.1	1.3	3.0
40	Health Economics	46.0	13.7	3.2	10.1	1.3	3.0
50	Ecological Ecs	91.0	12.5	3.4	11.5	1.3	3.0
75	Ec Dev & Cultural Change	32.3	11.6	4.3	18.9	1.3	3.0
100	Resource and Energy Ecs	30.1	9.3	5.5	29.8	1.3	3.0
149	Cambridge J of Ecs	70.8	14.3	7.2	52.4	1.3	3.0
203	Mgmt and Decision Ecs	96.9	37.2	11.1	125.0	4.0	192.0
250	Computational Ecs	53.1	20.2	10.3	106.4	2.0	13.9
300	Emerging Mkts Rev	32.8	11.5	11.5	131.6	1.3	3.0
Australasian Journals							
153	Ec Record	91.0	31.2	7.3	53.8	1.3	3.0
275	Aust Ec Papers	159.3	40.0	11.0	122.0	4.0	192.0
302	Aust J of Agr and Res Ecs	159.3	16.1	11.5	131.6	1.3	3.0
314	Aust Ec Rev	202.7	57.3	11.7	137.0	2.0	13.9
435	NZ Ec Papers	193.9	62.0	14.4	208.3	2.0	13.9
439	Aust Ec History Rev	371.6	139.6	14.5	208.3	1.3	3.0
479	Aust J of Labour Ecs	318.5	68.2	15.2	232.6	2.0	13.9
450	Agenda	NR	NR	14.5	212.8	2.0	13.9
777	Austral J of Reg Studies	NR	NR	19.3	1000.0	2.0	13.9
NR	Economic Papers	637.00	171.30	NR	NR	2.0	13.9

Note 1: Rank is that associated CLm.

Note 2: Journals ranked 10th, 150th and 200th did not publish articles by NZ-based academics over the 2006-2011 period;

For each of our seven JWS we calculate a number of plausible measures of research output. Following the prevailing approach in the literature, we share adjust pages and papers using the 1/n rule (where n is the number of authors on the paper), and for page output estimates, we adjust for size differences between journals with the *American*

Economic Review (AER) serving as the standard.¹⁵ In summary, the share and sized adjusted page and the share adjusted paper are our base measures, with a personal preference for the former. However, we also generate researcher rankings based on relaxation of the above assumptions.

3. Results

Our intention is not to select the 'best' journal weighting scheme or unit of output, something that is neigh impossible given the subjectivity involved, but to focus on the stability of the rankings across various plausible RAMs. To assess rank stability we use both Spearman rank correlation coefficients (SP) and the Kendall Tau (b) correlation coefficients (KT). The SP provides a measure of the linear relationship between the ranks resulting from the application of two RAMs. Since differences in ranks are squared in calculating SP, it places greater emphasis on large differences in ranks. The KT is particularly useful for this paper since it provides a direct measure of changes in the ordering of pairs of academics in moving from rankings using one measure to another. The KT is based on the differences between the proportion of pairs that are in the same order in the two measures (concordant) and the proportion that are in the opposite order (discordant), i.e. the probability that a pair selected randomly is concordant less the probability that it is discordant.¹⁶

In Table 3(a) we show SP and KT for various combinations of JWS, with the unit of output defined as share and size adjusted pages.¹⁷ Note the relatively high degree of correlation between all combinations presented in the table: the SP (shown in parenthesis) range from 0.83 for **CLh:ABDC** to 0.99 for **CLm:AVJWS**, with the vast majority of pairs displaying SP greater than 0.90. As expected, the rankings generated by the two schemes based on iterative/ eigenfactor methodology, **CLh** and **Rrec**, are highly correlated (SP= 0.95, and KT= 0.80). On the other hand, **ABDC**, a journal weighting scheme that assigns a relatively narrow range of values between the top and bottom group of journals, exhibit scores that are at the bottom end of the range of all correlation coefficients listed in the table. Also of interest are the correlation coefficients associated with **AVJWS** (the arithmetic mean of the output estimates of all other JWSs): the minimum values for SP and KT are 0.94 and 0.80, respectively. This suggests, at least in the context of our sample, that an arbitrarily constructed scheme that covers the range of plausible journal weighting schemes appears to generate a highly stable ranking of researchers.

¹⁵ Size adjustment factors for over 500 journals were kindly provided by Joseph Macri; for all other journals, we used a conversion factor of 0.72. This estimate is derived from Gibson (2000).

¹⁶ The interpretation of KT in this way is exact if there are no ties in the rankings or the Kendall-Tau (a) measure is used. We have reported the more conventional Kendall-Tau (b) that adjusts for ties differently. In our sample the differences between these two KT measures are very small.

¹⁷ It is important to recall that for pragmatic reasons we have restricted countable research to refereed papers in EconLit listed journals. We acknowledge that this constraint understates total research activity since we ignore conference papers, monographs, text books and working papers. In doing so, we are following the prevailing approach to be found in the economics literature.

**Table 3(a): Kendall's Tau and Spearman (in Parenthesis) Correlation Coefficients
Active staff, Share and Size Adjusted Pages**

	CLm	CLh	ABDC	Rsif	Rrec	NZABDC	AVJWS
CLm	1	0.82 (0.95)	0.80 (0.94)	0.82 (0.95)	0.78 (0.93)	0.80 (0.94)	0.92 (0.99)
CLh	0.82 (0.95)	1	0.65 (0.83)	0.80 (0.94)	0.80 (0.95)	0.74 (0.91)	0.83 (0.96)
ABDC	0.80 (0.94)	0.65 (0.83)	1	0.71 (0.88)	0.67 (0.85)	0.79 (0.94)	0.80 (0.94)
Rsif	0.82 (0.95)	0.80 (0.94)	0.71 (0.88)	1	0.88 (0.97)	0.81 (0.94)	0.88 (0.98)
Rrec	0.78 (0.93)	0.80 (0.95)	0.67 (0.85)	0.88 (0.97)	1	0.77 (0.92)	0.84 (0.96)
NZABDC	0.80 (0.94)	0.74 (0.91)	0.79 (0.94)	0.81 (0.94)	0.77 (0.92)	1	0.86 (0.97)
AVJWS	0.92 (0.99)	0.83 (0.96)	0.80 (0.94)	0.88 (0.98)	0.84 (0.96)	0.86 (0.91)	1

**Table 3(b): Percentage of Discordant Pairs,
Active Staff, share and Size Adjusted Pages, Various JWS**

	CLm	CLh	ABDC	Rsif	Rrec	NZABDC	AVJWS
CLm	0.00	9.20	9.80	9.04	10.73	9.77	3.87
CLh	9.20	0.00	17.73	10.10	9.81	12.83	8.63
ABDC	9.80	17.73	0.00	14.14	16.51	10.30	10.08
Rsif	9.04	10.10	14.14	0.00	6.08	9.50	6.12
Rrec	10.73	9.81	16.51	6.08	0.00	11.19	7.84
NZABDC	9.77	12.83	10.30	9.50	11.19	0.00	7.08
AVJWS	3.87	8.63	10.08	6.12	7.84	7.08	0.00
Average	7.49	9.76	11.22	7.85	8.88	8.67	6.23

Another way of looking at the stability of rankings is also displayed in Table 3(b). Here we show the percentage of discordant pairs associated with the Kendall's Tau (b) calculation. Note that for the pair **CLh** (an aggressive weighting scheme) and **ABDC** (the weakest weighting scheme), 17.73 percent of pairs are discordant; although a relatively high number, the associated rank correlation coefficients are 0.65 (KT) and 0.83 (SP). At the other extreme, only 3.87 percent of the pair **ABDC: AVJWS** are discordant. On average, the percentage of discordant pairs associated with **AVJWS** is 6.23, lower than that for all other JWS. As expected, given the SP and KT presented above, the journal weighting scheme displaying the highest percent of discordant pairs is **ABDC** (11.22 percent). Overall, the SP and KT, for output measured in terms of share adjusted and size adjusted pages, suggest a strong degree of association between all pair combinations of JWS. Nevertheless, the data on discordant pairs suggests that for many individual researchers rankings may differ somewhat across various journal weighting schemes.

In Table 4(a) we display the SP and KT for pairwise combinations of JWS when output is specified in terms of share adjusted papers. Note that the results are almost identical to those presented in Table 3(a) wherein output is measured in terms of share and size adjusted pages. The SP range from 0.81 (**CLh:ABDC**) to 0.98 (**CLm:AVJWS**), and the

corresponding KT are 0.62 and 0.90, respectively. Although the reported SP and KT for output measured in share adjusted papers are frequently lower than for share adjusted and size adjusted pages, the estimates differ by very little – at most, 0.03. Hence, the results displayed in Table 4(b) are to be expected. In all cases, the percent of discordant pairs for all combinations of JWS are everywhere slightly higher than those reported in Table 3(b). For example, when output is measured in terms of share adjusted papers, the minimum and maximum percentages of discordant pairs are 4.96 and 18.66; the corresponding values for share and size adjusted pages (see Table 3(b)) are 3.87 and 17.73, respectively. In summary, rank stability appears to be slightly stronger across all JWS when the unit of output is specified in terms of share and size adjusted pages than share adjusted papers.

Table 4(a): Kendall's Tau and Spearman (in Parenthesis) Correlation Coefficients

	CLm	CLh	ABDC	Rsif	Rrec	NZABDC	AVJWS
CLm	1	0.78 (0.93)	0.80 (0.95)	0.79 (0.94)	0.75 (0.91)	0.78 (0.93)	0.90 (0.98)
CLh	0.78 (0.93)	1	0.62 (0.81)	0.79 (0.94)	0.80 (0.94)	0.74 (0.90)	0.81 (0.95)
ABDC	0.80 (0.95)	0.62 (0.81)	1	0.70 (0.87)	0.64 (0.82)	0.78 (0.92)	0.79 (0.93)
Rsif	0.79 (0.94)	0.79 (0.94)	0.70 (0.87)	1	0.87 (0.97)	0.81 (0.94)	0.87 (0.97)
Rrec	0.75 (0.91)	0.80 (0.95)	0.64 (0.82)	0.87 (0.97)	1	0.76 (0.91)	0.83 (0.96)
NZABDC	0.78 (0.93)	0.74 (0.90)	0.78 (0.92)	0.81 (0.94)	0.76 (0.91)	1	0.85 (0.94)
AVJWS	0.90 (0.98)	0.81 (0.95)	0.79 (0.93)	0.87 (0.97)	0.83 (0.96)	0.85 (0.97)	1

Table 4(b): Percentage of Discordant Pairs, Active Staff, Share Adjusted Papers, Various JWS

	CLm	CLh	ABDC	Rsif	Rrec	NZABDC	AVJWS
CLm	0.00	10.78	9.69	10.19	12.48	11.08	4.96
CLh	10.78	0.00	18.66	10.59	9.73	13.08	9.18
ABDC	9.69	18.66	0.00	14.36	17.56	10.97	10.40
Rsif	10.19	10.59	14.36	0.00	6.51	9.34	6.15
Rrec	12.48	9.73	17.56	6.51	0.00	11.79	8.23
NZABDC	11.08	13.08	10.97	9.34	11.79	0.00	7.41
AVJWS	4.96	9.18	10.40	6.15	8.23	7.41	0.00
Average	8.45	10.29	11.66	8.16	9.47	9.10	6.62

In Tables 3(a/b) and 4(a/b) our focus was on the stability of ranks across various journal weighting schemes. Now we turn our attention to another aspect of rank stability. In Table 5(a) we report SP and KT for various combinations of output measures for each JWS. For example, consider the case of **CLh**. A comparison of the ranks associated with output measured as share and size adjusted pages (our preferred measure) and output measured in size adjusted pages (that is, no adjustment of co-authorship) yields Spearman and Kendall's Tau (b) estimates of 0.98 and 0.88, respectively. This suggests that if the 'true' journal weighting scheme is **CLh**, the rankings obtained by New Zealand economists are relatively insensitive to the assumption made with respect to the allocation of shares for multi-authored papers. The results for all other JWS, as displayed in the first data column of Table 5(a), are consistent with those for **CLh**. In the second data column we present the results of a pairwise comparison of ranks when output is defined as share and size adjusted pages and

unadjusted pages. Note that the SP and KT are identical to those in the first column for six of the seven JWS. The exception is **ABDC**; however, the differences are trivial – 0.01 and 0.02 for SP and KT, respectively. In other words, the common practice of adjusting for page size appears to have negligible impact on rankings, at least with respect to any of the JWS employed in this paper. This result is clearly demonstrated by reviewing the third data column wherein the SP and KT are displayed at three decimal places in order to reveal the impact of page size adjustments.

**Table 5(a): Kendall's Tau and Spearman (in parenthesis) Correlation Coefficients
Alternative Output Measures, All active Staff, Publications 2006-2011**

Journal weighting scheme	Output: Pages			Output: Papers	Output: Pages and Papers	
	Share and size adj: size adj only	Share and size adj: NO share, NO size adj	Size adj only: NO share, NO size adj	Share adj: NO share adj	Share and size adj pages: size adj papers	Share and size adj pages: NO share adj papers
CLm	0.82 (0.95)	0.82 (0.95)	0.943 (0.993)	0.81 (0.95)	0.82 (0.95)	0.78 (0.93)
CLh	0.88 (0.98)	0.88 (0.98)	0.965 (0.997)	0.87 (0.98)	0.89 (0.98)	0.86 (0.97)
ABDC	0.82 (0.95)	0.80 (0.94)	0.939 (0.992)	0.82 (0.95)	0.82 (0.95)	0.77 (0.91)
Rsif	0.87 (0.97)	0.87 (0.97)	0.954 (0.995)	0.86 (0.97)	0.87 (0.97)	0.83 (0.96)
Rrec	0.89 (0.98)	0.89 (0.98)	0.963 (0.996)	0.89 (0.98)	0.89 (0.98)	0.86 (0.97)
NZABDC	0.86 (0.97)	0.86 (0.97)	0.948 (0.995)	0.85 (0.97)	0.86 (0.97)	0.82 (0.95)
AVJWS	0.85 (0.96)	0.85 (0.97)	0.947 (0.994)	0.85 (0.97)	0.85 (0.97)	0.82 (0.95)

Table 5(b): Percentage of Discordant Pairs

Journal weighting scheme	Output: Pages			Output: Papers:	Output: Pages and Papers	
	Share and size adj: size adj only	Share and size adj: NO share, NO size adj	Size adj only: NO share, NO size adj	Share adj: NO share adj	Share and size adj pages: share adj papers	Share and size adj pages: NO share adj papers
CLm	8.83	8.98	2.86	9.40	8.75	10.97
CLh	5.93	5.96	1.77	6.26	5.56	6.94
ABDC	9.20	9.77	3.02	7.93	8.58	10.65
Rsif	6.46	6.65	2.26	6.80	6.38	8.50
Rrec	5.47	5.34	1.82	5.06	5.52	6.40
NZABDC	6.87	6.83	2.58	7.06	6.89	8.74
AVJWS	7.38	7.43	2.61	7.33	7.51	9.04

N= 113

Table 5(a) also displays information on the impact of adjusting for co-authorship when output is defined in terms of papers. As displayed in the fourth data column, SP for all pairs of share adjusted/ no-share adjusted papers range from 0.95 to 0.98, and for KT, from 0.81 to 0.89. Once again, the results suggest that the ranking of individual researchers is relatively insensitive to adjustments co-authorship. Additional support for this statement can be found in the fifth data column of Table 5(a). A comparison of the ranks generated by output measured in terms of share and size adjusted pages and share adjusted papers yields SP that range from 0.95 to 0.98; the corresponding values for KT are 0.81 and 0.89.

Despite the very high SP displayed in Table 5(a), reference to Table 5(b) suggests some caution should be exercised in interpreting the results with respect to rank stability. For example, consider the case of **ABDC**. A comparison of the ranks associated with share and size adjusted pages and size adjusted pages (no adjustment for co-authorship) yields a SP of 0.95; however, 9.20 percent of the pairs are discordant suggesting that for some researchers the specification of the unit of output does matter. In summary, the large values for SP and KT obtained when comparing rankings based on various definitions of output suggests a high degree of rank stability. Nevertheless, reference to the underlying percent of discordant pairs suggests that this is not always the case at the level of the individual researcher. It is this matter that we now address in greater depth.

In Table 6 we present arbitrarily selected rankings for individual researchers for each JWS. More specifically, we selected three groups of researchers based on their **AVJWS** rank when output is measured as share adjusted and size adjusted pages. The groups cover researchers ranked 1-10, 41-50 and 91-100, and for each journal weighting scheme we list each researcher's ranking for output measured in terms of pages (share and size adjusted) and papers (share adjusted). It is apparent that despite the impressively high rank correlation coefficients displayed in Tables 3, 4 and 5, individual performance is frequently sensitive to the RAM selection process. For example, consider the case of the R#69.¹⁸ This researcher displays extremely strong results over all JWS for both page and paper output measures—always first or second. Indeed, the top three performers in terms of **AVJWS**(page) dominate the rankings over all JWS. On the other hand, consider R#55. She/he ranks 6th per **AVJWS**(page), achieves relatively consistent ranks (5th to 8th) on all other JWS save for Rrec; in the latter case, R#55's performance drops sharply to 19th.

For researchers in the second and third tiers of the table, more variation is frequently displayed. Consider R#16. This researcher's ranking is highly variable: 92nd by **AVJWS**(page), 70th under **CLh**(page), and 103rd by **ABDC** (page). Even more extreme are the rankings associated with R#74: 49th (**AVJWS**(page)), 27th (**CLh**(page)), and 75th (**ABDC**(page)). We have admittedly selected individual researchers for whom the journal weighting scheme selection process matters greatly.

¹⁸ The identifiers for all 113 researchers were randomly selected.

Table 6: Rank for Various Journal Weighting Schemes, Output Share Adjusted Pages and Papers, Staff at 15 April 2012, Publications 2006-11

Rank Per AVJWS	ID#	CLm		CLh		ABDC		Rsif		Rrec		NZABDC		AVJWS		Max Diff in Rank	
		Page	Paper	Page	Paper	Page	Paper	Page	Paper	Page	Paper	Page	Paper	Page	Paper	Page	Paper
1	69	1	1	2	1	1	2	2	1	2	2	2	1	1	1	1	
2	37	3	4	4	5	3	5	1	2	1	1	1	2	2	3	3	4
3	9	2	3	1	2	2	3	4	4	4	3	3	3	3	2	3	2
4	28	7	12	5	7	11	20	3	8	3	5	8	15	4	12	8	15
5	86	4	8	6	6	6	9	10	12	7	10	6	8	5	9	6	6
6	55	5	6	8	9	5	4	7	7	19	20	4	4	6	7	15	16
7	66	6	10	3	3	13	13	12	13	9	13	5	6	7	11	10	10
8	12	11	11	15	15	7	8	5	5	5	8	7	7	8	10	10	10
9	48	10	7	7	4	10	10	9	10	8	6	10	10	9	5	3	6
10	63	8	5	12	11	8	6	8	6	11	12	11	11	9	5	4	7
41	38	40	56	38	48	50	66	38	54	32	44	45	63	41	56	18	22
42	49	37	23	35	25	39	25	45	33	50	40	41	30	42	27	15	17
43	41	42	44	34	32	45	47	46	51	39	37	43	43	43	43	12	19
44	51	53	53	52	54	43	40	37	37	41	46	46	52	44	49	16	17
45	93	44	40	32	29	56	58	51	48	45	41	57	54	45	45	25	29
46	43	45	32	31	26	65	47	47	39	42	33	44	36	46	35	34	21
47	99	48	57	51	58	53	62	42	53	37	43	55	60	47	55	18	19
48	75	47	52	47	50	44	44	40	45	67	65	36	37	48	50	31	28
49	74	52	37	27	23	75	62	54	47	40	32	42	31	49	40	48	39
50	59	50	30	61	51	42	18	48	32	54	51	56	39	50	33	19	33
91	35	89	88	71	64	97	96	88	88	94	94	89	84	91	87	26	32
92	16	96	95	70	61	103	105	90	87	76	68	93	95	92	89	33	44
93	68	82	76	92	90	86	76	93	94	89	89	99	99	93	88	17	23
94	103	92	92	88	88	96	93	70	71	93	94	95	93	94	91	26	23
95	5	94	91	98	94	85	82	95	96	90	89	92	88	95	97	13	15
96	97	91	83	93	94	88	85	99	103	96	103	97	102	96	100	11	20
97	100	97	94	94	89	98	88	96	92	95	88	98	96	97	94	4	8
98	24	100	99	89	85	104	99	92	85	97	91	94	85	98	94	15	14
99	45	95	80	99	92	90	81	98	93	100	93	106	106	99	90	16	26
100	4	93	55	103	101	94	55	104	104	105	104	103	91	100	83	12	49

More generally, the extent of rank variation across JWS can be seen in the final column of the table. For explanatory reasons, let us revisit researchers referenced above. At one extreme we see that R#69 is largely indifferent between JWS as the maximum difference between her/his ranking across all JWS is 1. On the other hand, the maximum difference in rank for R#74 is much higher: 48 positions for pages and 39 for papers. For R#16, the variation is equally extreme: 33 positions when output is measured in pages and 44 when measured in papers. Although less extreme, it can also be seen from a perusal of Table 6 that the decision to select the page or the paper as the official unit of account matters to many. For example, consider R#59. This researcher is ranked 42nd for pages and 18th for papers under **ABDC** (as expected, this person publishes many short papers).

In Table 7 we display the impact of alternative JWS on researcher rankings in the context of New Zealand's PBRF. For the 2012 round of the PBRF (covering the period 2006 to 2011), the assignment of letter grades for the economics group was as follows: 17 'As', 55 'Bs' and 'Cs' for all others given a ranking. Using these results, and arbitrarily selecting **CLh** as our 'true' journal weighting scheme (henceforth, the individual journal weighting scheme is noted as 'JWS'), we assign researchers 1 to 17 (per **CLh**) a grade of 'A', researchers ranked 18 to 72 a 'B' and those ranked 73 to 113 a 'C'.

Table 7: Rank Changes Relative to CLh, Share and Size Adjusted Pages

Top 17 per CLh, output: share and size adjusted pages							
PBRF Gr	AVJWS	CLm	CLh	ABDC	Rsif	Rrec	NZABDC
A	14	15	17	12	14	14	12
B	3	2	0	5	3	3	5
C	0	0	0	0	0	0	0
Rank 18 to 72 per CLh, Output: Share and Size Adjusted Pages							
PBRF Gr	AVJWS	CLm	CLh	ABDC	Rsif	Rrec	NZABDC
A	3	2	0	5	3	3	5
B	47	47	55	38	46	44	41
C	5	6	0	12	6	8	9
Rank 73 to 113 per CLh, Output: Share and Size Adjusted Pages							
PBRF Gr	AVJWS	CLm	CLh	ABDC	Rsif	Rrec	NZABDC
A	0	0	0	0	0	0	0
B	5	6	0	12	6	8	9
C	36	35	41	29	35	33	32

We now explore the following questions: what if **CLh** is not selected as the ‘official’ JWS and does it matter? Let us now assume that **AVJWS** is selected as the ‘official’ JWS. We now find that only 14 of our 17 ‘true As’ are ranked in **AVJWS**’s top 17; and only 47 of our 55 ‘true Bs’ and 36 of 41 ‘true Cs’ received their proper ranks. The discrepancy in ‘true’ versus ‘official’ letter grades is more extreme if **NZABDC** is selected as the ‘official’ JWS. We now find that 29.4 percent of ‘true As’ received a ‘B’ (5 out of 17) and only 74.5 percent of ‘true Bs’ received the correct grade (41 of 55). It follows that only 32 of 41 ‘true Cs’ were properly judged, with the remaining 9 being assigned a ‘B’. Given that the SP for the pairs **CLh:AVJWS** and **CLh:NZABDC** are 0.96 and 0.94, respectively, the above results are somewhat surprising. They suggest that failing to select the ‘true’ JWS as the ‘official’ scheme, in a PBRF-like environment, is likely to create serious hardship for many researchers, especially since letter grades remains the same for a six year period. In the context of this study, this statement holds for any combination of ‘true’ and ‘official’ journal weighting schemes.

Before concluding this section of the paper, we explore the impact on letter grade assignment when the unit of output employed in a PBRF-like environment differs from the ‘true’ measure. Our approach is similar to that discussed in the preceding paragraph: we arbitrarily designate share and size adjusted pages as the ‘true’ output measure, and **CLh** as the ‘true’ JWS; and then explore the impact on letter grade assignment if a ‘non-true’ scheme is selected as the ‘official’ unit of account. The results of this exercise are displayed in Table 8. Under our ‘true’ RAM, 17 researchers receive a grade of ‘A’, 55 a ‘B’ and 41 a ‘C’. If, however, officials deem all pages as being of equal value (regardless of the number of authors listed on the paper and the number of characters on the page), one ‘true A’ researcher finds herself/himself relegated to ‘B’ standing.¹⁹ The grade changes for the ‘true Bs’ and ‘true Cs’ are also relatively minor. More specifically, 51 of 55 ‘true Bs’ hold their position with one bumped up to an ‘A’ and three dropping to a ‘C’; and 38 of 41 ‘true Cs’ retain their proper grade with the other three ‘true Cs’ now receiving a ‘B’.

Let us consider the ramifications of another failure to select the ‘true’ output measure. Assume that the authorities decide that share adjusted journal articles are a better measure of output than the number of share and size adjusted pages. We now find that two of the 17 ‘true As’ are ‘Bs’,²⁰ and four of the 55 ‘true Bs’ fail to maintain that grade (two are bumped up to an ‘A’ and two drop to a ‘C’). Also note that 39 of 41 ‘true Cs’ are unaffected by the change in the ‘official’ unit of output. In summary, the impact on individual researchers of failure to select the ‘true’ output measure is much less severe than the failure to select the ‘true’ journal weighting scheme. Nevertheless, it is

¹⁹ Interestingly, this researcher just failed to hold on to an ‘A’ - missing by one rank (finishing in 18th position).

²⁰ These two researchers were ranked 19th and 20th under **CLh,pp,s**.

still important to those individuals that find themselves holding a lower or higher grade than warranted. Although this discussion is based on the implicit assumption that **CLh** is the ‘true’ journal weighting scheme, the above results hold, in general, for all other JWS covered in this paper.

Table 8: Rank Changes Relative to CLh, Share and Size Adjusted Pages

Top 17 per CLh, Share and Size Adjusted Pages					
PBRF Grade	CLh,pg,a,b	CLh,pg,c,b	CLh,pg,c,d	CLh,pp,a	CLh,pp,c
A	17	16	16	15	15
B	0	1	1	2	2
C	0	0	0	0	0

Rank 18 to 72 per CLh, Share and Size Adjusted Pages					
PBRF Grade	CLh,pg,a,b	CLh,pg,c,b	CLh,pg,c,d	CLh,pp,a	CLh,pp,c
A	0	1	1	2	2
B	55	52	51	51	49
C	0	2	3	2	4

Rank 73 to 113 per CLh, Share and Size Adjusted Pages					
PBRF Grade	CLh,pg,a,b	CLh,pg,c,b	CLh,pg,c,d	CLh,pp,a	CLh,pp,c
A	0	0	0	0	0
B	0	2	3	2	4
C	41	39	38	39	37

Notes:

pg: Pages

pp: Papers

a: Share adjusted for multiple authorship.

b: Size adjusted pages.

c: No adjustment for multiple authorship.

d: No adjustment for page size differences.

4. Conclusions

As noted in the introduction, this paper focuses on the stability of rankings of academics by research productivity in the context of short-term academic decision making. For instance, hiring decisions at the junior and intermediate levels rely heavily on publication activity in the recent past. This is also the case for most promotion decisions and merit awards. Renewal of research grants is often predicated, at least in part, on publications forthcoming from the original grant. Furthermore, in academic environments in which research performance is central to the granting of tenure or permanent employment for those hired directly from graduate school, decisions are typically based in large part on research activity over a period not exceeding six years.

Added to this mix of important academic decisions is the recent movement towards the use of nation-wide research assessment schemes that typically focus on research activity over a five to seven year period (assume six). This means that the average publication has approximately three years in which to demonstrate impact. For this reason, short-term academic decisions for economists are almost universally based on a formal or informal ranking of journals.

Recently, with the development of national research assessment exercises, the adjudication process has become much more formal. Most NRAEs employ a five to seven year assessment period and a formal peer review process for evaluating economic research. However, as previously noted, we believe that advances in information science and technology, efforts to reduce or control expenditures, and growing frustration over the ‘black-box’ nature of the peer review process, will lead to the introduction of formal bibliometric techniques into NRAEs in the near future.

In order to investigate the rank stability issue we have considered the research performance of New Zealand academic economists and have conducted a two part exercise. First, we followed the traditional approach of estimating Spearman rank correlation coefficients for various pairs of RAMs. We also estimated Kendall’s Tau (b) correlation coefficients since we judge this to be a superior measure in the context of this paper. More specifically, the latter measure allows us to directly explore rank stability in terms of concordant and discordant pairs, and to then estimate the percentage of discordant pairs associated with various pairs of research assessment measures. Secondly, we explored the rank stability at the individual researcher level. We have done so by illustrating the rank attained by individual researchers across a range of plausible RAMs. This part of the study also included an analysis of the impact of various JWS and output measures on rankings within the context of a PBRF-like decision framework.

Our primary finding is that despite statistical evidence of a high degree of rank stability across a number of plausible journal weighting schemes, the journal selection process is of great importance to the individual researcher. More specifically, high Spearman rank correlation coefficients were found to mask major changes in rank for individual researchers. Aside from a small group of individuals that dominate the top-end of all of our journal-based rankings, the vast majority of individual researchers perform much better on some rankings than others. For example, in the context of a PBRF-like environment, we found that between 12 and 29 percent of ‘true As’ receive a ‘B’ if an error is made in the selection of the ‘official’ journal weighting scheme. If different journal rankings are reflective of differences in opinions of peer review panel members about relative quality, then these ratings would also be similarly affected even ignoring differences of opinion on the merit of particular research.

In contrast, we found a much greater degree of rank stability with respect to the output unit selection process. In particular, attempts to adjust for variation in the number of characters per page (denoted herein as size adjustment) appear to have negligible impact on rank stability. Although there is some impact at the individual researcher level, it is very limited. Decisions with respect to the treatment of co-authorship and the page versus the paper as the unit of account are of greater importance. However, the impact on rank stability is much less severe than for the journal weighting scheme selection process.

We believe our study to be the first to assess rank stability in the context of a time-limited, nation-wide research assessment. Our findings suggest that future movement from a peer-review to a bibliometric-based research evaluation scheme will continue to be plagued by controversy. They also suggest likely instability in the decisions of peer review panel members unless an attempt is made to come to an agreed view on relative research standards. As shown in this paper, even the selection between two competing research assessment schemes that are based on similar methodologies and are highly correlated with respect to rank, can have extremely important career implications for many academics. This applies with particular force to academics working within a NRAE environment based on individual assessment such as New Zealand's PBRF.

References

- ABDC. (2010) *Australian Business Deans Council Journal Quality List*. www.abdc.edu.au, downloaded 12 October 2012.
- Abelson, P. (2009) 'The Ranking of Economics Journals by the Economic Society of Australia', *Economic Papers*, 28/2: 176-180.
- Abramo, G., Cicero, T. and D'Angelo, C.L. (2011) 'Assessing the varying level of impact measurement accuracy as a function of the citation window length', *Journal of Informetrics*, 5/3: 659-667.
- Anderson, D.L., Smart, W. and Tressler, J. (2013) 'Evaluating Research- Peer Review Team Assessment and Journal based Bibliographic Measures: New Zealand PBRF Research Output Scores in 2006', *New Zealand Economic Papers*, 47/2: 140-157.
- Burns, S.B. and Stern, D.I. (2015) 'Research Assessment Using Early Citation Information', Crawford School Research Paper No. 15-01, Australian National University, Canberra, Australia.
- Chang, C-L., McAleer, M. and Oxley, L. (2011) 'What Makes a Great Journal Great in Economics? The Singer Not the Song', *Journal of Economic Surveys*, 25/2: 326-361.
- Chang, C-L and McAleer, M. (2013) 'Ranking journal quality by harmonic mean of ranks: an application to ISI statistics and probability', *Statistica Neerlandica*, 67/1: 27-53.
- Chang, C-L and McAleer, M. (2013) 'Ranking Leading Econometrics Journals Using Citations Data from ISI and RePEc', *Econometrics*, 1/3: 217-235.
- Combes, P.P. and Linnemer, L. (2010) 'Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of all Journals in Economics', *Groupement de Recherche en Economie Quantitative d'Aix Marseille (GREQAM)*, Document de Travail, no 2010-28.

- Garfield, E. (1972) 'Citation Analysis as a Tool in Journal Evaluation', *Science* 178/4060: 471-479.
- Gibson, J. (2000) 'Research Productivity in New Zealand University Economics Departments: Comment and Update', *New Zealand Economic Papers*, 34/1:73-88.
- Hirsch, J. (2005) 'An Index to Quantify an Individual's Scientific Research Output', *Proceedings of the National Academy of Sciences*, 102/45: 16569-16572.
- Hodder, A.P.W. and Hodder, C. (2010) 'Research Culture and New Zealand's Performance-Based Research Fund: Some Insights from Bibliographic Compilations of Research Outputs', *Scientometrics*, 84: 887-901.
- Kalaitzidakis, P., Mamuneas, T. and Stengos, T. (2003) 'Rankings of Academic Journals and Institutions in Economics', *Journal of the European Economic Association*, 1/6: 1346-1366.
- Kalaitzidakis, P., Mamuneas, T. and Stengos, T. (2011) 'An Updated Ranking of Academic Journals in Economics', *Canadian Journal of Economics*, 44/4: 1525-1538.
- Kim, J-Y, Min, I. and Zimmermann, C. (2011) 'The Economics of Citation', *Korean Economic Review*, 27/1: 93-114.
- Laband, D. (2013) 'On the Use and Abuse of Economics Journal Rankings', *Economic Journal*, 123/570: F223-F254.
- Liebowitz, S.J. and Palmer, J.P. (1984) 'Assessing the Relative Impact of Economics Journals', *Journal of Economic Literature*, 22/1: 77-88.
- Liebowitz, S. (2014) 'Willful Blindness: The Inefficient Reward Structure in Academic Research', *Economic Inquiry*, 52/4: 1267-1283.
- 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 Economics Surveys*, 20/1: 111-156.
- OECD. (2010) Performance-based Funding for Public Research in Tertiary Education Institutions. Workshop Proceedings. OECD Published, <http://dx.doi.org/10.1787/9789264094611-en>.
- Oswald, A.J. (2007) 'An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers', *Economica*, 74/296: 21-31.
- Seiler, C.S. and Wohlrabe, K.W. (2012) 'Ranking Economists on the basis of many indicators: An Alternative Approach Using RePEc Data', *Journal of Informetrics*, 6/3: 389-402.
- Sgroi, D. and Oswald, A.J. (2013) 'How Should Peer-Review Panels Behave?' *Economic Journal*, 123/570: 255-278.
- Smart, W. (2009) 'The Impact of the Performance-Based Research Fund on the Research Productivity of New Zealand Universities', *Social Policy Journal of New Zealand*, 34: 1-18.
- Stern, D.I. (2014) 'High-Ranked Social Science Journal Articles Can Be Identified from Early Citation Information', *PLoS ONE*, 9/11: article e112520.
- TEC. (2013) *Performance Based Research Fund: Evaluating Research Excellence- the 2012 Assessment (Final Report)*. Wellington, New Zealand: Tertiary Education Commission.
- Tressler, J. and Anderson, D.L. (2012) 'Using Citations to Measure Research Output in New Zealand Economics Departments: The Problem of 'Long and Variable' Lags', *Agenda: A Journal of Policy Analysis and Reform*, 19/1: 17-37.
- Zimmermann, C. (2013) 'Academic Rankings with RePEc', *Econometrics*, 1/3: 249-280.