

**UNIVERSITY OF WAIKATO
Hamilton
New Zealand**

**Evaluating Research - Peer Review Team Assessment
and Journal-Based Bibliographic Measures:
New Zealand PBRF Research Output Scores in 2006**

David L. Anderson, Warren Smart and John Tressler

Department of Economics

Working Paper in Economics 03/12

April 2012

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

Email: dla@queensu.ca

and

Waikato Management School
University of Waikato
Private Bag 3105
Hamilton, NEW ZEALAND

Email: anderson@waikato.ac.nz

Warren Smart

Principal Research Analyst
Ministry of Education
Wellington
NEW ZEALAND

Email: warren.smart@minedu.govt.nz

Abstract

This paper concerns the relationship between the assessment of the research of individual academics by peer or expert review teams with a variety of bibliometric schemes based on journal quality weights. Specifically, for a common group of economists from New Zealand departments of economics the relationship between Performance-Based Research Fund (PBRF) Research Output measures for those submitting new research portfolios in 2006 are compared with evaluations of journal based research over the 2000-2005 assessment period. This comparison identifies the journal weighting schemes that appear most similar to PBRF peer evaluations. The paper provides an indication of the 'power or aggressiveness' of PBRF evaluations in terms of the weighting given to quality. The implied views of PBRF peer review teams are also useful in assessing common assumptions made in evaluating journal based research.

Keywords

bibliometrics
journal weighting schemes
PBRF
research measurement

JEL Codes

A19, C81, J24

Acknowledgements

We would like to thank the Tertiary Education Commission for access to the PBRF data used in this study. Ric Scarpa assisted with the maximum likelihood estimation.

Introduction

In the last two decades research assessment programmes have had a significant impact on universities in a number of countries. They have resulted in the evaluations of institutions, departments, and in some cases, individual researchers. These results have influenced public funding decisions, the development strategies of universities and departments, and the incentives and employment prospects faced by academics. Typically research assessment exercises are based on peer review or expert panels, supplemented in some cases by bibliographic methods, including the analysis of citation patterns. In New Zealand the Performance-Based Research Fund (PBRF) processes have produced evaluations of tertiary institutions, and subject areas within them, based on the assessment by expert panels of the research portfolios submitted by individual academics. The process and its outcomes have had a significant impact on the incentives faced by academics, research outputs and the staffing of tertiary institutions in New Zealand.¹

In economics and other disciplines there is an existing literature dealing with the evaluation of academic journals, departments and individuals. Much of this research has used journal ranking and weighting schemes to assess research outputs. The journal ranking and weighting schemes are in turn often based on bibliographic citation analysis or the perceptions of academic economists. Surveys of the research and journal evaluation literature in economics are provided by Macri and Sinha (2006), Anderson and Tressler (2011) and Chang, McAleer and Oxley (2011).

Given these two approaches, researchers have considered the relationship between panel based research assessments and journal based bibliographic assessments. For example, Taylor (2011) shows that quantitative measures, including the UK Association of Business Schools (ABS) rankings, are highly correlated with the outcomes of the UK 2008 Research Assessment Exercise (RAE). He concludes that bibliographic measures should be used in research assessment exercises.² Viewing the relationship the other way around, Geary, Marriott and Rowlinson (2004) use the 2001 RAE publication submissions data and outcomes to rate core journals in Business and Management on a 1-7 scale. Mingers, Watson and Scaparra (2009) apply linear programming techniques to 2008 RAE publication submission data and outcomes to provide rankings of 700 journals in Business and Management on a 0-4 scale.

Previous research examining the association between PBRF and bibliometric performance is limited. Smart (2007) compared the PBRF average quality scores in various subject areas (including business and economics) with the number of citations in indexed journals per eligible staff member. The study found there was a degree of correlation between the two measures in some subject areas, but the magnitude of the correlation was not particularly high. The study noted that one possible reason for the limited level of association was because the analysis used university subject area as the unit of assessment rather than the individual researcher.

The purpose of this paper is to consider the relationship between the assessment of the research of individual academics by peer/expert review teams as part of New Zealand's PBRF programme with the evaluation of the research of the same individuals based on a variety of journal weighting

¹ For an international review see OECD (2010). For review of PBRF see Evidence (2008)

² There are also a number of studies of the relationship between citations and RAE outcomes, see e.g. Norris and Oppenheim (2003). Franceschet and Costantini (2011) look at the relationship between various citation based research quality measures and outcomes from the first Italian research assessment exercise.

schemes. To our knowledge, this is the first time these different types of research assessments for individual academics have been compared, at least for research in economics.

Comparing these two approaches to research assessment provides a number of insights. First, we are able to consider how closely bibliographic techniques based on journal rankings match the considerably more costly assessments of individual academics by expert panels. Secondly, we are able to provide quantitative data that indicates which journal ranking schemes most closely resemble the assessments of PBRF expert panels in economics. Thirdly, we are able to provide a quantitative indication of the 'power or aggressiveness' of the research evaluations of expert panels, i.e., the weight placed on research published in the top level relative to low level journals. This factor has been largely ignored by existing studies of research assessment. The comparison also throws some light on common assumptions about research assessment often made in the literature that evaluates research based on journal rankings.

Specifically, for a common group of economists from New Zealand Economics departments, this paper considers the relationship between the evaluations of journal based research output between 2000 and 2005 using journal weighting schemes with PBRF Research Output (RO) assessments. We consider the general nature of the relationship for a large number of journal based weighting schemes commonly used in the literature. We also consider, in more detail, six schemes which capture a range of power or aggressiveness in evaluations, including two influential Australian journal ranking schemes. Finally the paper considers how various factors affect the relationship including: the treatment of co-authored papers, papers versus pages, all journal publications or the 'best four' and the treatment of New Zealand's regional economics journal: the *New Zealand Economics Papers (NZEPP)*.

Background and Data

New Zealand's Performance Based Research fund was established in 2003. In 2010, universities received around \$236 million from the Performance-Based Research Fund. Sixty per cent of the fund is allocated on the basis of research quality assessed using Evidence Portfolios submitted by individual researchers. These reviews are undertaken by expert panels that include overseas researchers. In the 2006 evaluation, researchers were able to list up to 34 research outputs in their Evidence Portfolios, with four self-nominated as most significant and discussed by the submitter³. Research outputs are defined broadly and include: journal publications, working papers, conference papers, books, book chapters and doctoral theses. Evidence Portfolios also include information relating to claims of Peer Esteem and Contributions to the Research Environment. The expert panels that assess the portfolios produce scores on Research Output (RO), Peer Esteem (PE) and Contributions to the Research Environment (CRE), each on an eight point scale (0 to 7).⁴ These scores are combined to provide an overall score with a weight of 70% given to RO and 15% to each

³ The maximum number of journal articles published by an individual in our sample is 32. Thus the limit of 34 research contributions should not have significantly limited the ability of researchers to list relevant journal based research outputs.

⁴ Since only integer scores were recorded it could be argued that the scores only represented ordinal assessments. However, PBRF procedures weight individual scores and manipulate them using standard arithmetic operations implying that they are treated as cardinal assessments.

of PE and CRE scores. Using the resulting aggregated score as a guide each individual is assessed as being an A, B, C or R researcher. Individuals are advised of their assessments and can obtain access to their RO, PE and CRE scores. PBRF assessments of research are said to be '...primarily about quality, not quantity.'⁵

This paper considers results from the 2006 PBRF evaluations. The staff census date for this round was the 14th of June 2006, and research output between the 1st of January 2000 and the 31st of December 2005 could be submitted in Evidence Portfolios. The 2006 round was a partial round where staff who also participated in the 2003 evaluation had the option of submitting a new evidence portfolio (31 percent chose this) or carrying over the results from 2003 (around 35 percent chose this option). The remaining staff submitted portfolios for the first time. In the 2006 PBRF evaluation, two additional categories (C(NE) and R(NE)) were introduced for new and emerging staff. Portfolios for new and emerging staff were assessed against different criteria. This paper concentrates on RO scores for economists submitting new portfolios in 2006 who were not new and emerging scholars and had published journal based research during the period, although some results are provided for all matching economists and those carrying over portfolios from the 2003 evaluations.⁶

The evaluation of the research output of New Zealand economists using journal weighting schemes is based on a dataset that includes the journal publications of all economists in New Zealand economics departments as at the 15th of April 2007. For this group of economists, all publications between the 1st of January 2000 and the 31st of December 2005 in Econlit journals as at the 15th of April 2007 are included. In total, 103 academic staff published at least one Econlit journal paper over the 2006 PBRF period. Of the 103, 98 could be matched to PBRF data. Of the 98, 65 submitted new portfolios in 2006 and 33 carried over evaluation scores from 2003. Of the 65 who submitted new portfolios, 10 were new and emerging scholars.⁷

We follow prevailing practice by allocating author shares on multiple authored papers by utilizing the 1/n rule, where n is the number of authors, although we also consider varying this rule. Once again we follow convention and adopt the 'weighted page' as our unit of output, with adjustments to allow for page size differentials between journals. The average size of an American Economic Review page is used as a reference point and given a value of 1.⁸

There are a large number of journal-based weighting schemes used by economists in research evaluations. These weighting schemes differ in three key ways: the ranking of the journals evaluated, the range or comprehensive of the evaluations as indicated by the number of non-zero weights

⁵ Performance Based Research Fund: Quality Evaluation guidelines, June 2011, Tertiary Education Commission, p. 104.

⁶ For further details on PBRF see www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF-/Resources/. An application was made to the Tertiary Education Commission for access to PBRF data. Only Warren Smart had access to this data.

⁷ If all academic staff is included, there were 139 in the April 2007 census, 122 matched PBRF data, including 81 who submitted a new portfolio in 2006 and 41 who carried over portfolios from 2003.

⁸ This conversion is made using page correction factors for over 500 journals provided by J. Macri and D. Sinha. For journals not covered we use a conversion factor of 0.72 as suggested by Gibson (2000). For details of the page conversion procedure use see Sinha and Macri (2002).

given and the 'aggressiveness' of the non-zero weights allocated, i.e. the differences between the weights given to highly ranked and relatively low ranked journals.

Table 1 describes some of the characteristics of the journal weighting schemes that will be referred to in this paper. To represent the overall characteristics of these schemes as applied to publications of New Zealand economists between 2000 and 2005, we show the percentage of publications that would receive a non-zero weight (range), and the percentage weight received by the 30th ranked publication compared with the highest ranked publication (as an indicator of aggressiveness). To provide an example of the differences between weighting schemes we show the number of standardised pages in the *Economic Record* that would receive the same weight as a page in the *American Economic Review*.⁹ The Gini Coefficient is widely used as a measure of inequality. Here we employ it using the weights applied to journals in which New Zealand economists published over the review period. We use it to provide an index of the 'power' of each weighting scheme, i.e. the degree of inequality in the weights given to different journals. It is important to note that the inequality or 'power' of a weighting scheme can derive from inequality in the weights applied to the journals actually evaluated (aggressiveness) or the number of journals that receive no evaluation or a zero weight (range). Further discussion of the weighting schemes used here can be found in Anderson and Tressler (2008 and 2011).

Across the journal weighting schemes outlined in Table 1 notice that the commonly used journal weighting schemes, based on citation counts adjusted for the impact of the journals in which the citations occur, are high-powered, both in their relatively narrow range and the weight assigned to highly ranked journals relative to lower ranked journals. In contrast, journal weighting schemes based on the perceptions of economists tend to be less aggressive in the weights applied, but have various degrees of coverage. To provide a benchmark we also include Equal as a weighting scheme, i.e. equal weights to all publications, or an index of quantity.

Later in this paper we consider, in more detail, six weighting schemes that are representative of the range of schemes considered in the literature or of particular interest to economists in Australasia. LP94 is the weighting scheme developed by Laband and Piette (1994) which updated the pioneering work on journal weighting schemes of Liebowitz and Palmer (1984). The journal weights are derived from citations weighted by the impact factor of the journal in which the citations occurs using an iterative process. This weighting scheme is sometimes referred to as the 'industry standard' or 'gold standard'. It is a highly aggressive and narrow scheme. For example, a page in the *American Economic Review* receives the same weight as 402 pages in the *Economic Record* and only 33.5% of publications by New Zealand economists between 2000 and 2005 receive a non-zero weight. The Gini Coefficient for LP94 is 0.93, it is the highest of all the weighting schemes considered.

KMS2010 is the weighting scheme published by Kalaitzidakis, Manuneous and Stengos (2010) which updates their earlier 2003 weights. Like the Laband and Piette weights, these are based on impact factor adjusted citations, but cover a much broader and more recent list of journals. Under this scheme 59% of the papers published by New Zealand economists between 2000 and 2005 receive a non-zero weight and 80 *Economic Record* pages equate to an *American Economic Review* page. The Gini Coefficient remains high at 0.86, despite the relatively wide coverage of this scheme.

⁹ We use the *Economic Record* as a number of these schemes did not rank the *New Zealand Economic Papers*.

Table 1 Summary of Journal Weighting Schemes

| Weighting Scheme | Date | Method | Non-Zero Weighted Pubs., NZ Economists 2000-2005 % | Weight of the 30 th Publication to the Top Ranked, NZ Economists 2000-2005 % | Pages in <i>Econ Record</i> Equivalent to a Page in the <i>AER</i> ¹ | GINI Coef, Pubs by NZ Economists 2000-2005 |
|---|------|---------------------------------|--|---|---|--|
| Liebowitz & Palmer (LP84A1) | 1984 | Impact Adj. Citations | 32.5 | 7.1 | 35.2 | 0.93 |
| Laband & Piette (LP94A1) | 1994 | Impact Adj. Citations | 36.9 | 9.8 | 400.2 | 0.93 |
| Kalaitzidakis, Mamuneous & Stengos (KMS1) | 2003 | Impact Adj. Citations | 50.1 | 18.7 | 34.1 | 0.88 |
| Kalaitzidakis, Mamuneous & Stengos (KMS2) | 2010 | Impact Adj. Citations | 59.0 | 10.1 | 80.0 | 0.86 |
| Research Papers in Economics (RePEc) | 2009 | Impact Adj. Citations | 69.1 | 23.3 | 18.0 | 0.77 |
| Kodrzycki and Yu, Economics, (KYEI) | 2006 | Impact Adj. Citations Economics | 55.3 | 9.9 | 45.2 | 0.84 |
| Kodrzycki and Yu, Policy (KYPI) | 2006 | Impact Adj. Citations Policy | 55.3 | 14.3 | 49.6 | 0.79 |
| Thomson Reuters Journal Citation Report (JCR08IF) | 2008 | Impact Adj. Citations | 52.2 | 35.6 | 3.0 | 0.66 |
| Bauwens (Bauwens) | 1998 | Basic Citations (5 groups) | 100 | 80 | 2.5 | 0.30 |
| Mason, Steagall and Fabritius (MSF) | 1997 | Perceptions | 38.7 | 75 | 1.7 | 0.72 |
| Excellence in Research for Australia (ERA) ² | 2008 | Perceptions | 86.3 | 100 | 1.3 | 0.26 |
| Economic Society of Australia (ESA) ² | 2008 | Perceptions | 93.4 | 100 | 1.3 | 0.31 |
| Gibson (Gibson) | 2000 | Labour Market Estimation | 100 | 64 | 2.9 | 0.54 |
| Equal | | | 100 | 100 | 100 | 0.00 |

¹The *Economic Record* is used here as the *New Zealand Economic Papers* is not given a weight by a number of schemes.

²For ERA and ESA weights of 4,3,2 and 1 are assigned to assessed grades of A*, A, B and C respectively.

MSF is the widely used weighting scheme suggested by Mason, Steagall and Fabritius (1997). It is based on the perceptions of chairs of American economics departments obtained from a survey undertaken in 1993. This scheme evaluates only 38.7% of the journals in which New Zealand economists published over the review period, but the weights applied to the journals evaluated are not aggressive. For this scheme 1.7 *Economic Record* pages equate to a page in the *American Economic Review*. The discriminatory power of this weighting scheme lies primarily in the relatively small proportion of journals included, yielding a relatively high Gini Coefficient of 0.72.

ERA is the journal rating scheme that was part of the Excellence in Research for Australia assessment. It was developed by a committee of the Australian Research Council based on perception information from a variety of sources. Journals are assigned ratings of A*, A, B and C, and some journals are not ranked. Here these ratings have been converted to a numerical scale with A* receiving a weight of 4 and C a weight of 1. A large number of journals receive a non-zero rank, thus the scheme is relatively comprehensive, encompassing 86% of the journal publications by New Zealand Economists between 2000 and 2005. Under this scheme 1.3 pages in the *Economic Record* would equate to a page in the *American Economic Review*. This is a low-powered weighting scheme with a Gini Coefficient of 0.26 given the numerical scores assigned to the grades.

ESA is a similar journal rating scheme to ERA, prepared as part of the ERA process by the Economic Society of Australia (2008). It is based on a survey of the perceptions of professors of economics in Australian economics departments. Journals are again rated on an A* to C basis, and numerical weights of 4 to 1 have been applied here. For this scheme 602 economics journals were rated, including 93% of the journals in which New Zealand economists published between 2000 and 2005. The Gini Coefficient is 0.31.

Gibson is the journal weighting scheme developed by Gibson (2000). Unlike the other weighting schemes, it was based on an econometric analysis of the hiring and promotion decisions of New Zealand economics departments. Journals were placed in four classes using the earlier work of Towe and Wright (1997). All publications receive some weight under this scheme, and 20 pages in the *Economic Record* equates to one page in the *American Economic Review*. Although this scheme is comprehensive, as all journals receive a non-zero weight, the Gini Coefficient is relatively high at 0.54.

PBRF Research Output (RO) Scores and Journal Based Measures of Research Output – an Initial Assessment

In this section we compare the assessment of research based on the commonly used journal weighting schemes with PBRF Research Output (RO) scores. In considering this relationship we are not only concerned with whether the rank or ordering of the individuals assessed by a particular journal-based measure is the same as that implied by the PBRF score, but also the relative magnitude of the evaluations. Both the journal-based measures and PBRF RO scores are relevant to the assessment of the relative size of research outcomes.

As noted above, the RO scores were based on portfolios that reported on research undertaken between the 1st of January 2000 and December 31, 2005. We use the journal publications over this

same period for economists submitting new portfolios in 2006 to evaluate research output using the journal weighting schemes discussed above. However, there are two important differences in the scope of the research assessments compared. First, PBRF portfolios allow for the identification of four self-nominated research outputs and descriptions of the research contributions made by these publications.¹⁰ However, the purpose of the PBRF evaluations is to assess the quality of each individual's overall research contribution over the period. The gathering of additional information about the four nominated research outputs is part of the PBRF process for making this judgement. Thus a comparison of all research as measured by both journal weighting schemes and the PBRF assessment is relevant to the research question addressed herein. Secondly, the PBRF assessment covers a much wider range of research than journal publications. Thus, much of the research described in research portfolios is not represented in the bibliographic analysis of journal publications.¹¹

In journal weighting schemes changes to the scale of a measure do not change the relative weights of different journals or the evaluation of journals that are assessed compared with those that are not considered. Correspondingly, the relative research assessments of different academics are not affected by changes in scale of the weights applied. In general terms other transformations of the weights will change the structure of the research evaluation scheme. Here we use Pearson correlation coefficients to assess the linear relationship between the assessment of research output using journal weighting schemes and RO assessment scores. For comparison purposes we provide correlation coefficients for both those submitting new portfolios in 2006 and those carrying over RO scores for portfolios submitted in 2003.

Table 2 shows the correlation coefficients for all journal weighting schemes outlined in Table 1, for both a narrow group and all matching economists. The narrow group includes all those that had at least one journal paper over the review period that could be assessed by the journal based weighting schemes (research active) and that received PBRF evaluations on the same criteria (were not new and emerging (NE)). For the narrow group, and those submitting new portfolios in 2006, the majority of the journal based measures are only moderately positively correlated with RO research assessments. With the exception of the highest powered measures, LP84 and LP94, correlation coefficients are greater than +0.3. What is surprising is that the differences in the correlations between very different measures of journal based research output are quite small. For example, Equal which simply gives the same weight to any journal publication is as highly correlated as five of the 13 other measures. Relatively high-powered and low coverage measures, such as KMS, are as highly correlated as measures that are very low-powered but incorporate more journals, e.g. ESA. However, it must be recognised the correlation coefficients are likely to be influenced by the lack of variation in RO scores. Scores are integers ranging between 2 and 7 with approximately 60% of the scores values of 5 or 6.

Those carrying over RO scores from the 2003 PBRF assessment are assessed on the basis of different research outputs. However, for many of the journal weighting schemes the correlation coefficients are very close, and the majority of carryover correlations are higher (9 of 14). Since the productivity of individual researchers is not likely to have changed significantly over the three year difference

¹⁰ Copies of these nominated research outputs are also available to peer-review teams.

¹¹ Since we have compared research assessment only for those judged to be research active on the basis of journal publications, the impact of this difference is reduced.

between the 2003 and 2006 evaluations, and as some publications could be included in both assessments, it is not surprising that the correlations for those carrying over portfolios are positive and generally significant. However, higher correlations would be expected for those submitting new portfolios since all journal based research assessed is the same for this group. It should also be noted that those carrying over portfolios from 2003 are a self-selected group of researchers who: may have been those rated as A researchers in 2003; may not have had 2003 evaluations that were close to the cut-offs between scores; or those whose research output for the six years prior to 2006 was not better than that prior to 2003.

Table 2 Pearson Correlation Coefficients between 2006 PBRF RO Scores and Journal Based Measures of Research Output of NZ Economists¹

| Weighting Scheme | Research Active, Not NE ² | | All Matching Researchers | |
|---|--|----------------------|--|----------------------|
| | New Submission and Resubmissions n = 55 | Carryovers n = 33 | New Submission and Resubmissions n = 81 | Carryovers n = 41 |
| Liebowitz & Palmer (LP84A1) | 0.16 | 0.18 | 0.26** | 0.22 |
| Laband & Piette (LP94A1) | 0.22 | 0.25 | 0.29* | 0.29*** |
| Kalaitzidakis, Mamuneous & Stengos 2003 (KMS1) | 0.51* | 0.49* | 0.57* | 0.51* |
| Kalaitzidakis, Mamuneous & Stengos 2010 (KMS2) | 0.40* | 0.44* | 0.47* | 0.47* |
| Research Papers in Economics (RePEc) | 0.42* | 0.41** | 0.51* | 0.49* |
| Kodrzycki and Yu, Economics, (KYEI) | 0.30** | 0.23 | 0.40* | 0.27*** |
| Kodrzycki and Yu, Policy (KYPI) | 0.31** | 0.30 | 0.44* | 0.34** |
| Thomson Reuters Journal Citation Report (JCR08IF) | 0.49* | 0.48* | 0.61* | 0.52* |
| Bauwens (Bauwens) | 0.47* | 0.53* | 0.61* | 0.57* |
| Mason, Steagall and Fabritius (MSF) | 0.44* | 0.48* | 0.52* | 0.51* |
| Excellence in Research for Australia (ERA) ² | 0.45* | 0.56* | 0.59* | 0.60* |
| Economic Society of Australia (ESA) ² | 0.47* | 0.56* | 0.59* | 0.60* |
| Gibson (Gibson) | 0.44* | 0.59* | 0.52* | 0.61* |
| Equal | 0.32** | 0.51* | 0.50* | 0.56* |

¹ Correlation coefficients marked with a *, ** and *** are significantly different from zero at the 1%, 5% and 10% level respectively

² Research active are those who published at least one journal paper between 2000 and 2005.

Correlation coefficients increase in all cases when matching researchers who did not publish a journal paper between 2000 and 2005 and new and emerging researchers are included. This reflects the additional zero or generally low scores for both the journal based and PBRF evaluations.

Overall, the simple comparison of bibliographic journal based measures and PBRF RO scores provides little guidance as to the nature of PBRF assessments. Nor does it provide confirmation that they represent high-powered, quality oriented assessments of research output in the sense in which this is measured by internationally accepted journal weighting based assessments of research. While it might be argued that there is some evidence that PBRF RO scores are consistent with journals based measures of research output, it is clear that there are significant differences in the nature of the assessment.

The 'Aggressiveness' of Research Evaluations and PBRF Research Output Scores

In this section we consider six of the journal weighting schemes in more detail in order to investigate the relationship between the aggressiveness of the weights assigned to differently ranked journals and PBRF evaluations using the narrow group of matching economists, i.e. those who submitted a new EP, were research active and not NE. Two questions will be addressed. First, do PBRF Research Output (RO) evaluations provide any guidance on the relative weights that are or should be applied to differently ranked journals in research assessments. Secondly, for journal based assessment schemes such as ERA, ESA or the Australasian Business Deans rankings, do PBRF RO evaluations give any indication of the relative weights that might apply to journals receiving different scores, e.g. journals receiving an A* grade compared with a B? This analysis also enables us to identify the journal based weighting schemes that are most consistent with PBRF evaluations after adjusting for the aggressiveness of the weights applied.

As noted above, journal weighting schemes differ in three essential ways: the range of the journals that are evaluated (given a non-zero weight), the ranking of the journals evaluated and the aggressiveness of the weights assigned. For any set of journal based research subject to evaluation, the range and aggressiveness of the scheme will determine its 'power'. As noted above, the Gini Coefficient can be used as an indicator of the power of the scheme. Here we consider changes to the aggressiveness of the weights applied by the schemes studied using the following simple transformation of weights:

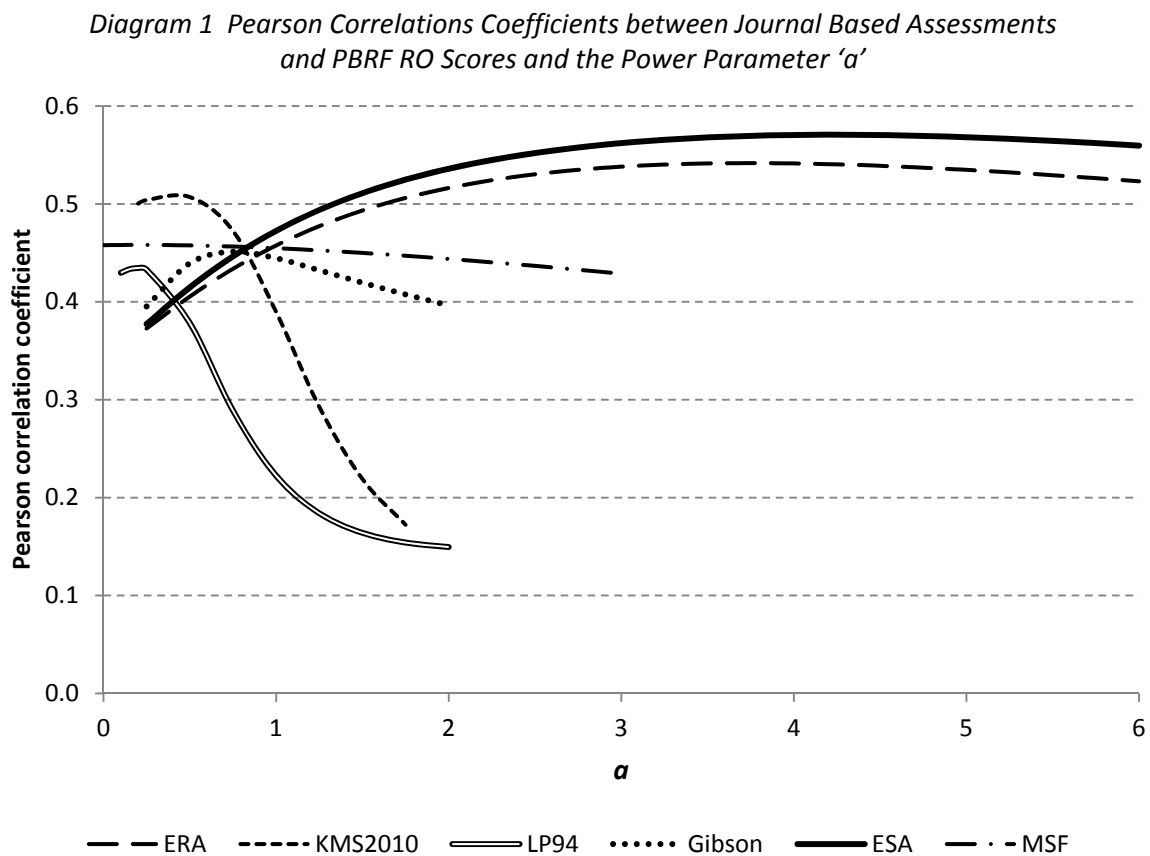
$$nw_i = ow_i^a,$$

where nw_i and ow_i are the new and old weights applied to journal i and a a parameter. The parameter a represents the elasticity of the new weights with respect to the old weights as weights vary across the journals evaluated. Changing a provides a simple way of varying the aggressiveness of journal weighting scheme.¹² Values of $a > 1$ will increase the aggressiveness of a weighting schemes while values of $a < 1$ reduce it. A value of $a = 0$ would imply that all the journals evaluated by the scheme receive the same weight. For example, the Gibson scheme implies that 20 standardised

¹² This transformation of weights is of course arbitrary. Any positive monotone increasing transformation of the weights would maintain the same ranking, and would in general change the aggressiveness of the scheme.

pages in the *Economic Record* would receive the same weight as one page in the *American Economic Review*. If these weights were transformed as above with $a=2$, this trade-off between the two journals would be 400, or with $a=0.5$ it would be 4.5.

We have chosen the following journal weighting schemes for further evaluation: KMS2010, LP94, MSF, ERA, ESA and Gibson. For each of these schemes we used a grid based search to determine the value of a that would result in the highest correlation between the evaluations of the research of individuals by the journal weighting scheme and PBRF RO scores. Diagram 1 describes the variation of the Pearson correlation coefficients with a for each of the six schemes. For KMS2010 and LP94 correlations initially increase significantly as a falls. In contrast, correlations initially rise with a for ERA and ESA. Correlations for the MSF scheme, which has relatively low coverage of journals, do not change much with a . Correlations for the Gibson scheme that is based on labour market assessments reach a maximum at a value of a close to one, i.e. the aggressiveness of the original schemes gives close to the highest association.



For the six journal weighting schemes, the approximate values of a that give the highest correlation coefficients are shown in Table 3. For the Australasian schemes, ERA and ESA, a value of a equal to 3.75 and 4.25 respectively give the highest correlations. This suggests that the PBRF quality standards are significantly more aggressive than the numerical values used above. In contrast, for KMS2010 and LP94, the values of a are 0.4 and 0.2 respectively. This suggests that the quality standards implied by PBRF evaluations are significantly less aggressive than that associated with these two widely used journal based evaluation schemes. In terms of aggressiveness, the Gibson

scheme is the most similar to the PBRF evaluations with a value of $a=0.75$. As shown in Diagram 1, at these values of a , the ESA scheme gives the closest reflection of PBRF RO evaluations. This scheme is comprehensive, covering 93% of the journals in which New Zealand economists published between 2001 and 2005.¹³

Table 3 Approximate Values of a giving the Highest Values of the Pearson Correlation Coefficient

| <i>Journal Weighting Scheme</i> | <i>a Value</i> |
|---------------------------------|-----------------------------|
| KMS2010 | 0.4 |
| LP94 | 0.2 |
| MSF | 0.2 |
| ERA | 3.75 |
| ESA | 4.25 |
| Gibson | 0.75 |

In Table 4 we show, for both adjusted and unadjusted weighting schemes, the number of standardised pages that would equate with a page in the *American Economic Review* in a variety of journals in which New Zealand economists published over the period. Note that the adjusted weights are those transformed using the values of a as shown in Table 3.

Table 4 Journal Page Trade-Offs for Initial Weighting Schemes and Adjusted Weights

| | <i>Adjusted Weights</i> | | | | | |
|------------------------|-----------------------------|-------------|------------|---------------|------------|------------|
| | <i>(Unadjusted Weights)</i> | | | | | |
| | <i>KMS2</i> | <i>LP94</i> | <i>MSF</i> | <i>Gibson</i> | <i>ERA</i> | <i>ESA</i> |
| | 4.60 | 1.73 | 1.04 | 1.40 | 2.94 | 3.40 |
| Economica | (45.45) | (15.46) | (1.21) | (1.56) | (1.33) | (1.33) |
| | 1.97 | 1.07 | 1.02 | 1.40 | 1.00 | 1.00 |
| J. of Econ. Literature | (5.45) | (1.40) | (1.08) | (1.56) | (1.00) | (1.00) |
| | 2.26 | 1.00 | 1.04 | 1.00 | 1.00 | 1.00 |
| Rand J. of Economics. | (7.70) | (1.00) | (1.21) | (1.00) | (1.00) | (1.00) |
| | 2.50 | 2.02 | 1.07 | 1.40 | 1.00 | 1.00 |
| J. of Dev. Econ. | (9.91) | (33.50) | (1.43) | (1.56) | (1.00) | (1.00) |
| | 3.34 | 1.91 | 1.07 | 2.25 | 1.00 | 1.00 |
| J. of Urban Econ. | (20.33) | (25.13) | (1.38) | (2.94) | (1.00) | (1.00) |
| | 3.77 | 1.51 | 1.09 | 2.25 | 2.94 | 3.40 |
| Int. J. of Ind. Org. | (27.62) | (7.73) | (1.56) | (2.94) | (1.33) | (1.33) |
| | 2.48 | 1.68 | 1.10 | 1.40 | 2.94 | 3.40 |
| Economic Letters | (9.65) | (13.40) | (1.62) | (1.56) | (1.33) | (1.33) |
| | 4.93 | 2.51 | 1.11 | 9.46 | 2.94 | 3.40 |
| Applied Economics | (54.05) | (100.50) | (1.67) | (20.00) | (1.33) | (1.33) |
| | 10.65 | 3.32 | 1.12 | 2.25 | 2.94 | 3.40 |
| Economic Record | (370.37) | (402.00) | (1.73) | (2.94) | (1.33) | (1.33) |
| | 11.17 | 1.56 | 1.12 | 9.46 | 13.45 | 19.03 |
| J. of Econ. Education | (416.67) | (9.35) | (1.73) | (20.00) | (2.00) | (2.00) |

¹³ As an alternative to the grid search for the a that maximizes the correlation coefficients, we estimated a using maximum likelihood methods. The values of a obtained were KMS2010 0.25, LP94 0.1, ERA 1.72, ESA 3.95 and Gibson 0.27. For MSF the maximisation did not converge. For KMS2010, LP94 and ESA, the magnitudes of the estimates are similar. We have reported the correlation based estimates as we sought to find the best linear relationship.

The results for the adjusted weighting schemes yield more modest trade-offs between journals than those generated by the unadjusted schemes. For major field journals, two to three standardised pages is equivalent to a page in the *American Economic Review*. However, very well-regarded field journals such as the *Rand Journal of Economics* are treated as equivalent to the *AER* by many schemes. The ERA and ESA schemes treat many internationally recognised field journals as equivalent to the *AER* as all receive the same A* rating. Except for KMS2 and MSF, two to three pages in the *Economic Record* correspond to a page in the *American Economic Review*. Thus by international standards, PBRF evaluations in economics do not appear to be aggressive in the emphasis placed on journal quality.

As noted above, the ERA and ESA evaluation schemes rate journals as A*, A, B or C. In Section 2 numerical values of 4,3,2 and 1 were applied to these scores. While recognising the numbers assigned to the scoring categories and the transformations applied above are arbitrary, the value of α that results in the highest correlations with PBRF evaluations gives some indication of the relative values of these scores. Table 5 shows the implied numerical weights using a scale of 100 for A* journals.

Table 5 Implied Weights for ERA and ESA Journal Evaluation Schemes
A*=100

| | A* | A | B | C |
|-----|-----|-------|------|------|
| ERA | 100 | 34.25 | 7.18 | 0.55 |
| ESA | 100 | 29.56 | 5.25 | 0.28 |

Given these weights, approximately three pages in an A grade journal compares with one page in an A* journal. However, the trade-off increases as we move down the scale: four to five pages in a B journal equates to a level A page, and 10 to 20 pages in a C journal equates to a page in a B rated journal. Since these trade-offs are driven by the transformation function used herein, and the initial arbitrary numerical values, these results should be treated with some caution.

Other Issues in Research Evaluation and Evidence from PBRF Research Output Scores

In evaluating research output on the basis of journal publications and comparing these with PBRF Research Output (RO) scores, a number of other important issues arise. One of the most contentious matters relates to the treatment of research with a domestic focus. More explicitly, it is to be expected that national research priorities would be reflected in government funded research evaluation schemes. If this is the case, one might expect a higher weight to be given to publications in relevant national journals than would be the case if one employed international standards. For example, there is evidence to suggest that the Australian ERA and ESA journal ranking systems give higher scores to some Australian journals than those embodied in most international journal ranking schemes.¹⁴ It could be argued that this reflects a preference for research relevant to the nation state. *New Zealand Economic Papers* plays this role in New Zealand. To study this issue we grant the *New*

¹⁴ For a discussion of this issue, see Anderson and Tressler (2009).

Zealand Economic Papers the same weight as the *Economic Record* in order to see if the correlations with PBRF RO scores increase in value.¹⁵

It is usual in assessing journal based research outputs to use standardised pages as the unit of output. Thus long descriptive papers are judged as representing more research output than short mathematical papers or notes. As an alternative, we explore the possibility that correlations with PBRF scores are higher with weighted papers rather than pages. That is, does the length of a paper matter?

The PBRF evaluation process requires researchers to identify up to four research outputs and gives them the opportunity to briefly explain the significance of the research chosen.¹⁶ Up to 30 other research outputs can be listed. In the analysis above, all of an individual researcher's journal based research output is used in assessing the research undertaken over the period. Here we consider the impact of considering only the four best publications, where best is judged on the basis of the journal based research assessment scheme being used. We do this for both weighted pages and weighted papers.

For co-authored papers, we previously allocated the pages equally amongst all authors of the paper – the '1/n rule'. For some authors this will understate contributions, while for others, it will be overstated.¹⁷ This rule also ignores a possible preference for team based research in evaluations. To provide alternatives we consider the implications of assigning 100% of the weight associated with a paper to all the co-authors; and, as an in between case, for papers with two or more authors, 1/nth of 150% of the normal weight associated with a paper is allocated to each co-author.

The impacts of these alternatives on the correlation coefficients for the six journal based schemes considered in the last section are shown in Table 6. For the base case in this analysis we use the adjusted weighting schemes with the values of the parameter a shown in Table 3, and consider the narrow group of matching economists. In the base case the *New Zealand Economic Papers* receives the weight it would receive in each evaluation scheme; assessments are based on weighted pages; all journal publications are evaluated; and all co-authors are allocated an equal 1/nth share of the measured output of the article. In considering each of the alternative assumptions outlined above, all other aspects reflect the base case.

The results provide no evidence to support the suggestion that the *New Zealand Economic Papers* receives the same weighting as the *Economic Record*. Although correlation differences are very small, for all measures the correlation is lower when the *NZEP* is given the same weight as the *Economic Record*. It is of course possible that while a preference for relevant research is reflected in the assessment of quality, increasing the weight of the *NZEP* to that of the *Economics Record* is too crude a way of reflecting this.

¹⁵ Clearly this test is arbitrary, i.e. the weight given to the *Economic Record* might be more or less than the appropriate weight, but this adjustment has been used in other assessments, see for example Anderson and Tressler (2008).

¹⁶ For co-authored papers, researchers are also expected to explain their contribution to the research.

¹⁷ PBRF panels have information provided by researchers on their contributions to the four research outputs nominated. If accurate this should increase the research contribution for some researchers and decrease it for others.

Across the six measures it is not clear whether pages or papers are the most relevant in assessing research in a way that is consistent with PBRF evaluations. For weighting schemes that encompass most of the journals in which New Zealand economists publish (ERA, ESA and Gibson), weighted pages give slightly higher correlations. For the relatively selective schemes (KMS2010, LP94 and MSF), papers are the most relevant.

In all but one case, research assessments based on the top four papers generate higher correlations with PBRF scores than the base case, independent of our choice of weighted pages or papers as the appropriate output measure.¹⁸ However, for scenarios in which all publications are subject to evaluation, the results are mixed with respect the choice of weighted pages or papers.

When all co-authors receive the full weight for the publication rather than a 1/nth share of the weighted pages, the correlation coefficients for all measures increase. Moreover, when the co-authors receive 1/nth of 150% of the weight of the paper, the correlation coefficients are all lower than when each co-author receives the full weight. While this may indicate preference for team based research in PBRF evaluations, it is hard to avoid the conclusion that there is some double counting of research contributions in PBRF evaluations.

Table 6 Impact of Variations in Research Assessment on Correlations with PBRF RO Evaluations

| | <i>Pearson Correlation Coefficients</i> | | | | | |
|---------------------------------------|---|--------------|--------------|--------------|--------------|---------------|
| | <i>KMS2010</i> | <i>LP94</i> | <i>MSF</i> | <i>ERA</i> | <i>ESA</i> | <i>Gibson</i> |
| Base Case¹ | 0.509 | 0.435 | 0.458 | 0.542 | 0.571 | 0.451 |
| NZEP=Economic Record | 0.507 | 0.427 | 0.431 | 0.540 | 0.569 | 0.440 |
| Weighted Papers | 0.511 | 0.462 | 0.469 | 0.527 | 0.559 | 0.437 |
| Top 4 – Weighted Pages | 0.562 | 0.482 | 0.527 | 0.564 | 0.582 | 0.500 |
| Top 4 – Weighted Papers | 0.572 | 0.524 | 0.534 | 0.554 | 0.559 | 0.527 |
| Co-Authors 1/nth of 150% ² | 0.540 | 0.458 | 0.490 | 0.561 | 0.603 | 0.486 |
| All Co-Authors 100% | 0.589 | 0.497 | 0.535 | 0.609 | 0.571 | 0.532 |

¹ Base case: values of a from Table 3, NZEP=actual, weighted pages, all journal publications and co-author allocation 1/n.

² A solo author receives 100% of the weight of the paper in this scenario not 150%.

Conclusions

This paper considers the relationship between the evaluations of research by peer/expert panels with bibliographic assessments based on journal based weighting schemes using the research of New Zealand academic economists between 2000 and 2005 and corresponding PBRF Research Output (RO) scores. It also provides an indirect indication of the way in which the PBRF panel in economics has evaluated journal based research. However, with a relatively small sample, and noting the self-selection issues that arise from use of data from the partial 2006 assessments, only tentative conclusions are possible.

¹⁸ ESA, with weighted papers, is the exception here.

It seems clear that bibliographic research assessments based on journal weighting schemes do not produce results that can be regarded as close to PBRF RO evaluations. This is not surprising given the much broader scope of research considered by PBRF processes and the additional quality related information available to panels. Citation analysis also suggests that there may be a substantial difference in the quality of articles published in the same journals which might be picked up by the additional information available to PBRF panels.¹⁹ This of course does not imply that the more systematic use of bibliographic methods and the standardisation of bibliographic information provided by researchers could not add to the PBRF evaluation process.

The analysis does suggest that the most aggressive of the international journal based weighting schemes do not reflect PBRF practice in economics. PBRF practice in assessing research in economics is not consistent with the quality standards implied by these schemes. After the weighting schemes considered were adjusted, the resulting implied weights suggest a moderate degree of aggressiveness in assessing quality. It is not the case that only publications in the profession's leading journals matter.

The results also indicate that none of our selected journal weighting schemes generates results that parallel those of the PBRF evaluation. However, there is some indication that schemes with a broader range of coverage are more closely related, but the differences are small. Overall, the ERA and ESA journal evaluation schemes best mimic the PBRF RO scores. This is especially the case if the numerical weights applied to journals in the different schemes are moderately aggressive in their recognition of the five quality classifications.

Many of the aspects of research considered in this paper are relevant to the on-going decisions that academics must make about their research programmes. How much emphasis should be placed on publications in the top journals? Are there costs associated with undertaking research relevant to policy issues in New Zealand? How important is research quantity relative to quality? How is team-based research viewed? More information that can be used in considering these questions could increase the efficiency of the resulting decisions. The use of known bibliographic methods as part of research evaluation exercises such as PBRF is one way of adding transparency and providing information relevant to these decisions.

References

- Abelson P. (2009). The Ranking of Economics Journals by the Economic Society of Australia. *Economic Papers*. 28(2), 176-180.
- Adams J. (2008). Strategic Review of the Performance-Based Research Fund: The Assessment Process. Report Prepared for the TEC by Evidence Ltd.
- Anderson, D. L and Tressler, J. (2011). Ranking Economics Departments in Terms of Residual Productivity: New Zealand Economics Departments, 200-2006. *Australian Economic Papers*. 50(4), 157-168.
- Anderson, D.L. and Tressler, J. (2010). The Merits of Using Citation-Based Journal Weighting Schemes to Measure Research Performance in Economics: The Case of New Zealand. Working Papers in Economics, 10/03, University of Waikato, Department of Economics, www.repec.org.

¹⁹ See for example Chang, McAleer and Oxley (2011).

- 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), 7-25.
- Anderson, D.L. and Tressler, J. (2008). Research Output in New Zealand Economics Departments 2000-2006. *New Zealand Economic Papers*, 42(2), 155-189
- 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)
- 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.
- Franceschet, M. and Costantino, A. (2011). The First Italian Research Assessment Exercise: A Bibliometric Perspective. *Journal of Informetrics*. 5(2), 275-291.
- Geary J., Marriott, L. and Rowlinson, M. (2004). Journal Rankings in Business and Management and the 2001 Research Assessment Exercise in the UK. *British Journal of Management*. 15(2), 95-141.
- Gibson, J. (2000). Research Productivity in New Zealand University Economics Departments: Comments and Update. *New Zealand Economic Papers*, 34(1), 73-87.
- 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. (2010). An Updated Ranking of Academic Journals in Economics. Working Paper 9/2010, Economics Department, University of Guelph, Guelph, Canada.
- Kodrzycki, Y.K. and Yu, P. (2006). New Approaches to Ranking economics Journals. *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy* 5(1), Article 24.
- Laband, D. and Piette, M. (1994). The Relative Impact of Economics Journals. *Journal of Economic Literature*, 32(2), 640-666.
- 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 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.
- Mingers, J., Watson, K. and Scaparra, P. (2009). Estimating Business and Management Journal Quality from the 2008 Research Assessment Exercise in the UK. Working Paper No. 205, Kent Business School, United Kingdom.
- Norris, M. and Oppenheim, C. (2003). Citation Counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE. *Journal of Documentation*, 59(6), 709-730
- OECD (2010). Performance-based Funding for Public Research in Tertiary Education Institutions. Workshop Proceedings, OECD Publishing, <http://dx.doi.org/10.1787/9789264094611-en>
- Sinha, D. and Macri, J. (2002). Rankings of Australian Economics Departments, 1988-2000, *Economic Record*, 78(241), 136-146.
- Smart, W. (2007). Quality vs impact: A comparison of Performance-Based Research Fund quality scores with citations, Ministry of Education, Wellington.