The Relevance of the 'h' and 'g' Index to Economics in the Context of a Nation-wide Research Evaluation Scheme: The New Zealand Case

David L. Anderson and John Tressler

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

David L. Anderson
School of Business
Queen's University
Kingston, Ontario K7L 3N6
CANADA
and
Waikato Management School
University of Waikato
Hamilton
NEW ZEALAND

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

Abstract

The purpose of this paper is to explore the relevance of the citation-based 'h' and 'g' indexes as a means for measuring research output in economics. This study is unique in that it is the first to utilize the 'h' and 'g' indexes in the context of a time limited evaluation period and to provide comprehensive coverage of all academic economists in all university-based economics departments within a nation state. For illustration purposes we have selected New Zealand's Performance Based Research Fund (PBRF) as our evaluation scheme. In order to provide a frame of reference for 'h' and 'g' index output measures, we have also estimated research output using a number of journal-based weighting schemes. In general, our findings suggest that 'h' and 'g' index scores are strongly associated with low-powered journal ranking schemes and weakly associated with high powered journal weighting schemes. More specifically, we found the 'h' and 'g' indexes to suffer from a lack of differentiation: for example, 52 percent of all participants received a score of zero under both measures, and 92 and 89 percent received scores of two or less under 'h' and 'g', respectively. Overall, our findings suggest that 'h' and 'g' indexes should not be incorporated into a PBRF-like framework.

1. Introduction

Research output measurement in economics has historically relied on journal weighting schemes, based either on reputational surveys or citation counts (adjusted in various ways). The basic idea behind this approach is that the value of a paper is best captured by the relative importance of the publishing journal. For journal weighting schemes based on citation counts, the premise is that the expected citation flow from a given article is the mean number of citations generated by papers previously published in the same journal. However, recent work (Oswald, 2007; Wall, 2009; and Chang, McAleer and Oxley, 2011) has shown that the distribution of citations is highly skewed, and that many papers in so-called leading journals fail to generate a single cite over periods as long as 25 years. This 'problem' has lead economists to shift their attention to direct citation measures that have been widely used in the sciences. The basic argument in favour of using citation counts (raw or adjusted) is that they purportedly reflect the quality or impact of a paper. That is, more heavily cited papers are said to have a greater impact on the distribution of knowledge than less cited papers. The increasing reliance on citations to reflect impact or quality has been abetted by recent improvements in the range of journal coverage and ease of data manipulation.

In this paper we explore the use of various direct citation schemes to measure research in economics departments across all universities in a nation-state, within the time frame imposed by most research assessment regimes. More specifically, we assess the merits of using direct citation measures in quantifying research in economics departments in New Zealand within the six year time span employed by the nation's research assessment scheme-the Performance Based Research Fund (PBRF). It must be stressed that, at this time, the PBRF scheme is a peer-based exercise that grades every academic at all tertiary education providers in the country. The first version of the scheme was carried out in 2003 (covering research conducted over the period 2000-2002); the second exercise was held in 2006 and assessed research produced over the period 2000-2005; and the next evaluation will be held in 2012 covering research published over the period 2006-2011. It is our view that before another review is launched, there will be serious consideration given to moving, to some degree, toward the use of bibliographic-based research assessment measures (RAMs). This position is based on three factors: the availability of better quality and quantity of citations data, especially with respect to the social sciences; potential cost savings; and the possible loss of confidence in the 'black-box' that lies at the heart of any peer-review system.

2. Limitations of Journal-Based Weighting Schemes

The use of journal weighting schemes has come under increasing criticism. For example, consider the scheme developed by Kalaitzidakis, Mamuneas and Stengos (2003) (this journal weighting scheme will henceforth be denoted as **KMS**) – undoubtedly the most respected of the journal weighting schemes based on the impact factor methodology of Garfield (1972)[1] and the recursive adjustment model

---

[1] The impact factor can be defined as: the total number of citations from journals in year t to articles published in year's t-1 to t-x, divided by the number of articles published by the receiving journal in year's t-1 to t-x, where x can range from 2 to the number of years the journal has been in publication.

developed by Liebowitz and Palmer (1984)[2].   The 2003 version of the **KMS** scheme ranked only ISI[3]-listed journals (143 out of over 1500 currently listed in Econlit as economics or economics related journals). While this may be an appropriate number of journals for ranking upper echelon schools and researchers, it fails to reflect the fact that many economists publish in non-listed ISI journals. Furthermore, the nature of the weights assigned to top and lesser journals is such that they are unlikely to reflect real-life decision-making in most tenure, promotions and merit increase committees. For example, under **KMS**, the 30th, 50th and 90th ranked journals exhibit weights that are only 7.8, 3.9 and 0.8 percent as large as the weight assigned to the top ranked journal. In the New Zealand setting, this results in the top 30 **KMS**-ranked journals utilized by New Zealand researchers generating 85 percent of output (from only 12 percent of refereed papers). The **KMS** scheme has also been criticized for restricting countable citations to those generated by other journals on the ISI-economics list (Kodrzycki and Yu, 2006) and for failing to account for differing citation practices across sub-disciplines within economics (Palacious-Huerta & Volij, 2004).

More generally, all journal-weighting schemes based on the unadjusted or adjusted 'Impact Factor' concept exhibit a severe limitation: 'good' papers can be found in lowly ranked journals and 'bad' papers can be found in highly ranked journals. Recent work by Starbuck (2005), Oswald (2007), Wall (2009), and Chang, McAleer and Oxley (2011) have clearly outlined the limitations of relying on journal- weights to reflect the expected impact of a given paper. For example, Chang, McAleer and Oxley (2011) found that over a 25 year span, more than 26 percent of articles in the top 40 journals, per the Journal Citation Reports (JCR) 2Yr Impact Factor list, fail to receive a single cite, even of the self-generated variety. Furthermore, the non-citation rate exceeds 50 percent for six of the journals on this elite list. Evidence of this sort suggests that reliance on a journal's overall performance may not be a good proxy of the expected life-time impact of an individual article.

3. Direct Citation Schemes

The primary alternative to using journal weighting schemes to measure an article's impact is to take account of the number of cites made directly to the paper over a given period of time. As noted above, the basic argument in favour of using citation counts (raw or adjusted) is that they purportedly reflect the quality or impact of a paper. That is, more heavily cited papers are said to have a greater impact on the distribution of knowledge than less cited papers. When incorporated into a departmental study, the results can be presented either as the total number of cites or, more commonly, as the number of share-adjusted cites per capita. However, such an approach has limitations in ranking academic units and individuals. First, and perhaps the most fundamental problem is the basic premise that underlies

---

2   The basic Liebowitz and Palmer (1984) approach can be characterized as follows: the starting point is a set of impact factors. In the second stage, all cites are weighted by the relevant impact factor of the journal generating a citation. The adjusted impact factors generated by the second iteration are then substituted for the initial impact factors, and the process repeated until a stable solution is found. The final weights are known as adjusted impact factors.

3   Officially known as the Thomson Reuters Institute for Scientific Information Web of Science database, but generally referred to as the ISI database.

citation counting exercises: a 'cite is a cite'. That is, equal weight is given to all cites regardless of the citing source. Second, the distribution of citations per paper is known to be highly skewed; hence it is possible that departments with a few stars will perform very well even if most of the members of the department generate very few or zero cites. This is a special problem in evaluating small departments. Third, the ranking of individuals on the basis of cites per capita is also affected by the 'big hit' issue: for example, should individual A with 200 cites to one paper be ranked equally with individual B with ten papers each of which has been cited twenty times? This is just another example of the well known quantity versus quality problem that researchers face when attempting to measure non-market output.

Arguably the best-known direct citations-based scheme designed to address the quantity/quality problem is the **h**-index developed by Hirsch (2005). The **h**-index is defined as follows: 'A scientist has index **h** if **h** of his or her Np papers have at least **h** citations and the other (Np-**h**) papers have fewer than <= **h** citations each' (Hirsch, 2005, p. 16569).   A simple example will illustrate the concept. From Table1 note that the largest number of papers satisfying the definition is '4'. That is, the fourth ranked paper (where papers are ranked from highest to the lowest number of cites) has five citations, but the fifth ranked paper has only three citations and hence fails to satisfy the definition. The intent of Hirsch's proposal is clear: to simultaneously give credit to researchers who are adept at capturing citations and who produce a large number of refereed papers. Hirsch developed the concept to measure individual performance, but the concept has been extended to measure departmental or group activity.   It should be noted that Hirsch accepted the premise that all cites are of equal value.

Table 1. *An Example of the h and g Index*

Cites per Paper	Paper Rank	Total Cites	Paper Rank Squared
14	1	14	1
9	2	23	4
6	3	29	9
5	**4**	34	16
3	5	37	25
2	**6**	39	36
1	7	40	49

Although the **h**-index makes an attempt to reflect success in both publishing and cite capturing, it has detractors. For example, Ravillion and Wagstaff (2011) provide a theoretical foundation for the consideration of the scholarly influence of citations using a stochastic dominance approach. They describe the influence of a paper as a function of the citations it receives and suggest that for a particular paper this function has two natural properties. First, that influence should be strictly increasing in the number of citations a paper receives, i.e. more citations always implies more influence. Their second, and more contentious assumption, is that the marginal influence of citations declines with increasing citations. They note that the **h**-index is inconsistent with the first of these assumptions since

increasing citations may not increase the **h**-index, while the **g**-index (to be discussed below) violates the second assumption since it implies increasing marginal influence. On the pragmatic front, and related to the above theoretical discussion, is the well known criticism that the **h**-index and its derivatives 'leave cites on the table'. For example, two individuals can have the same **h** score, but the total number of cites to the papers in the **h**-core (the **h** papers with **h** or more cites) may be much higher for one individual than the other. A similar argument applies with respect to cites to papers that lie outside the **h**-core.

To address the 'cites on the table' issue, Egghe (2006) created the **g**-index, defined as: 'A set of papers has a **g**-index **g** if **g** is the highest rank such that the top **g** papers have, together, at least **g** squared citations.' (Egghe, 2006, p.132) We shall use the data in Table1 to illustrate the concept. In this case, refer to the last two columns: total citations (TC) and paper rank (PR) squared. We are now looking for the lowest PR for which the PR squared exceeds the TC. This is seen to be paper '6', as PR squared (36) is less than TC (39), but to move any lower in the ranking (paper7) would result in the PR squared (49) being less than TC (40). It is important to note that **g** exceeds **h** (6 versus 4) as more cites are taken into account (39 versus 34). Although this discussion is based on a simple example, conceptually, **g** can never be less than **h**. New citation-related indexes are constantly being developed and promulgated in an attempt to better capture the quantity and quality or impact dimensions of research output. [4] However, for purposes of this study we shall restrict ourselves to estimating individual and departmental output using three measures based on the ISI database: citation counts, **h**-index, and **g**-index, with the exact specifications to be discussed in the data section of the paper.

Until relatively recently the economics research measurement literature was dominated by the above referenced journal weighting approach (see Macri and Sinha (2006) for an extensive survey of this literature). Indeed, we have only located five studies based on raw citation counts that are of direct relevance to this study; that is, studies based on raw citation counts that attempt to measure research output of all academic economists on a nation-wide basis: Harris (1990), Pomfret and Wang (2003), Anderson and Tressler (2012), Ruane and Tol (2008), and Tol (2008) for Australia, Australia, New Zealand, and the latter two for Ireland,[5] respectively. However, only the first three studies restricted citation collection to a specific period of time, rather than relying on life-time cites. It should also be noted that Hendrekson & Waldenstrom (2011) provide nation-wide estimates for Sweden, but restrict their study to the life-time output of Full Professors.

Recently economists have started to published work on the **h**-index (see Ruane and Tol, 2008; Tol, 2008 and 2009; Courtault et al, 2010; Bodman, 2010; Ellison, 2010; and Hendrekson and Waldenstrom, 2011). A general discussion of the **h**-index, and other citation-based research assessment measures, in the context of economics, can be found in the work of Chang, McAleer and Oxley (2011), and Kim, Min and Zimmermann (2011), and a strong case for utilizing these approaches, especially the **h**-index, can be

---

[4]   For example, see Schreiber (2010) for a study based on 20 variations of the h-index. For a review of the broad range of bibliometric-based RAMs, see Chang, McAleer and Oxley (2011).

[5]   Ruane and Tol (2008) and Tol (2008) provide a partial analysis only. They only consider those with active research records (those for whom they could find ISI or Scopus citations). This led them to exclude 90 academics from a pool of approximately 225.

found in Ellison (2010). Alternatively, and as previously noted, Ravallion and Wagstaff (2011) demonstrate that the **h**-index (and similar measures) are ad-hoc constructs. Studies of economic research using the **g**-index are quite limited. For purposes of this paper, we must stress that save for Ruane and Tol (2008) and Tol (2008), we did not find any nation-wide studies of individual economists based on either the **h** or **g**-index; and for nation-wide departmental ranking, we failed to find a single paper utilizing these techniques. However, we should note that Courtault et al (2010) provide **g** and **h**-index estimates for a select group of universities in France at both the individual and departmental level (based, as are the papers by Ruane and Tol (2008) and Tol (2008), on life-time cites to all researchers in the dataset).

In this study we calculate **h**-index and **g**-indexes in the context of a nation-wide research assessment exercise. For this purpose we have adopted the framework of the New Zealand Performance-Based Research Fund (PBRF). As noted earlier, the PBRF is an assessment exercise that measures the research performance of all university based academics[6] (that is, its basic building block is the individual researcher, not the academic unit as is the case for the UK and Australian research assessment exercises) over a six year period. Economics is one of 42 disciplines covered by the exercise, and the resulting aggregation of individual scores yields a numerical estimate of performance for each academic unit and ultimately each institution (eight universities and a number of polytechnics – we restrict ourselves to the university sector). In 2010, PBRF scores were utilized to distribute approximately 18 percent of the government's total funding to universities (approximately $268 million NZ), while the resulting ordinal rankings are widely used to advertise and promote 'winners' at the academic unit and institutional level. In other words, both cardinal and ordinal measures of performance are important to departments and universities in New Zealand.

4. Data Issues

In order to simulate PBRF-like conditions, we restrict our study to a six year period (2003-2008). Our data-base incorporates all refereed publications in journals recognized by Econlit (as at 15 April 2009). In this study, citations to papers are excluded if they are self-cites, and the remaining cites are share adjusted utilizing the widely employed 1/n rule (where n is the number of authors on the paper). Rephrased, all citation counts in this study are of the non-self, share-adjusted variety. Given the competitive nature of the PBRF, indeed for any nation-wide research assessment exercise with major financial payouts, adjustments for self-citations and multiple authored papers are necessary to prevent game playing tactics that will ultimately destroy the integrity of the scheme. From this database, we construct our three ISI citation-based RAMs: citation counts (**ISIC**), g-index (**ISIg**), and h-index (**ISIh**).The transition from individual cites to departmental output is based on the Stock allocation method. That is, for all academic staff employed on the census date (15 April 2009), all output produced over the period 2003-2008 is attributed to the staff member's employer as at the same date.

---

[6] Strictly speaking, researchers are not required to submit a research portfolio for evaluation, but those failing to do so are, in effect, given a score of zero.

We have made one major deviation from the basic PBRF framework. Under PBRF, the cut-off date for assessment information is 31 December of the last year of review – in our example, this would be 31 December 2008. However, it is widely known that citations lag publication by a considerable period of time, especially in the social sciences (Centre for Science and Technology Studies, 2007). Indeed, in the New Zealand economics context, Anderson & Tressler (2012) found that for publications in 2000 and 2001, only 16 percent of papers had received ISI cites by the third year after release (rising to 32 and 40 percent by the end of year's 5 and 10, respectively). The corresponding figure for ISI cites to ISI listed journals was higher, but still only 29 percent. This suggests that a citation-based scheme, in order to have credibility within the economics community, must include a lag structure. However, this has to be constrained for pragmatic reasons, since governments are attempting to reward relatively current performance and institutions can be expected to wish to minimize the time period between the end of the evaluation period and the announcement of results and the flow of expected funding. In order to balance these competing themes, we have opted for a two year lag structure: that is, citations for all publications produced over the 2003 to 2008 period have been collected up to 31 December 2010.

We have utilized the ISI citation data-base to generate estimates of cites per paper. This data-base has been utilized by virtually all researchers in the bibliographic field; it is generally viewed as the most reliable source of citation data. However, it suffers from a major limitation in the context of a nation-wide review: as at 1 March 2011, only 17 percent of Econlit listed journals are included in the ISI economics group (247 out of approximately 1500). We should also note that ISI assigns equal weight to all cites regardless of origin, although ISI cites can be considered to have passed a crude test in that they must originate from ISI recognized journals across all disciplines.

In order to place our results for direct-citation RAMs (citation counts, **g**-index and **h**-index) in context, we also report departmental and individual results for a select number of traditional RAMs based on journal weighting schemes. For these schemes we adopt the prevailing approach (see Macri and Sinha, 2006): output is measured by share adjusted pages produced by a given researcher, with shares being determined by the 1/n rule (where n is the number of authors on the paper), and word count per page standardized to equal that of an average *AER* page.

Our most basic weighting scheme is based on the premise that all refereed papers are of equal value (if they are listed in Econlit) – we label this scheme as **EQUAL**. In practice this is really a measure of quantity rather than quality; however it does serve as a useful reference point for demonstrating the impact of quality adjustment processes. In order to compare our citation-based regimes to those derived from 'impact-factor' analysis, we have chosen two schemes: the 2011 updated version of the KMS (2003) scheme previously discussed in this paper: we label this set of weights as **KMS;** and the 2008 version of the well known two year journal impact factors for ISI listed journals, denoted herein as **JCR**.[7] Our fourth and final alternative weighting scheme is of the reputational variety. For illustrative purposes, we have selected the regime developed by the Australian Research Council and used in Australia's nation-wide research assessment exercise conducted in 2010 – the Excellence in Research for

---

[7] The ISI document listing the organizations various RAMs, including the 2year impact factor utilized in this study is known as the Journal Citation Report (JCR).

Australia (ERA). Although these journal rankings proved to be controversial, and will not be directly employed in the next round of the ERA, this scheme was never-the-less based on input from the Economic Society of Australia and other academic groups, and provides us with an alternative RAM that should reflect some of the values held by New Zealand researchers and evaluators given the close social, economic and cultural ties between these nations. We label this set of journal weights as **ERA.**

5. Results

Before presenting our results, recall the basic framework we have adopted: the relevant citing period is 1 January 2003 to 31 December 2010 for all refereed papers published in Econlit-listed journals over the period 1 January 2003 to 31 December 2008. Therefore, papers published in 2003 have an eight year window for capturing cites whereas those published early in 2008 have, at most, a three year window. Given that the size of each economics department in our sample has been relatively constant over this time period, the departmental results should not be inherently biased by our citation collection approach. Obviously, the situation at the individual researcher level is different, especially for those beginning their research career post-2002. To address this situation, we have presented results for individual researchers who have been active throughout our assessment period; in other words, those with at least six years of professional experience as at 31 December 2008. Let us now turn to our findings.[8]

In Table 2 we show **ISIg** and **ISIh** results for the 110 researchers in our restricted sample. Save for one star performer, the results are rather compressed: 0-6 and 0-4 for **ISIg** and **ISIh**, respectively. The compression issue can also be observed by noting the percentage of researchers obtaining an ISI score of two or less: 89 percent for **g** and 94 percent for **h**. In fact, 52 percent of researchers exhibit a score of zero for both **ISIg** and **ISIh**. Also in Table1 we provide information on the number of papers published by individual researchers for any given score. For example, researchers obtaining an **ISIh** score of '0', published between zero and 17 papers. Equally interesting is the fact that one economist receiving an **ISIh** score of '1' published 35 Econlit recognized papers over the period 2003-2008. All of this is admittedly descriptive, but nevertheless suggests that the '**h**' and '**g**' indexes are relatively insensitive to varying levels of output as traditionally defined - at least with respect to the ISI database.

In Table 3 we explore the degree of association between our various direct citation and journal-based measures. All combinations of direct citation measures are highly correlated- ranging from 0.86 to 0.97. However, the estimates for pairs of direct and journal-based measures yield variable results. We see that **ISIg** and **ISIh** are moderately correlated with **EQUAL** (0.61/ 0.62), **ERAB** (0.78/ 0.77), and **JCRIF** (0.69/ 0.67), and more weakly correlated with **KMS** (0.48/ 0.46).[9]

---

[8] This restriction reduces the sample size from 135 to 110.

[9] The Spearman Rank Correlation Coefficients have much the same look as do the estimates in the prior table.

Table2. *Share-Adjusted g & h Index Results, ISI Citations*

Score	ISIg		ISIh	
10	0	(0)	0	(0)
9	1	(21)	0	(0)
8	0	(0)	0	(0)
7	0	(0)	0	(0)
6	2	(17-25)	1	(21)
5	3	(6-13)	0	(0)
4	2	(5-24)	2	(17-25)
3	4	(4-7)	4	(6-24)
2	15	(2-13)	12	(4-13)
1	26	(1-35)	34	(1-35)
0	57	(0-17)	57	(0-17)
# of Staff	110		110	

Note1: All Staff with 6 years experience as at 31 December 2008
Note2: In brackets, number of papers (range) by researchers

Table 3. *Pearson Correlation Coefficients; Various Output Measures over the Period 2003-2008*

	ISIC	ISIg	ISIh	EQUAL	ERA	JCR	KMS
ISIC	1.00	0.90	0.86	0.58	0.72	0.59	0.40
ISIg		1.00	0.97	0.61	0.78	0.69	0.48
ISIh			1.00	0.62	0.77	0.67	0.46
EQUAL				1.00	0.91	0.57	0.31
ERAB					1.00	0.79	0.51
JCR						1.00	0.82
KMS							1.00

Note1: All Academic Staff as of 15 April 2009 with at least six years experience (N=110)

One notable feature of our database is the large number of zero output estimates. Not only does this apply to **ISIg** and **ISIh**, but also to **JCR08IF** and **KMS** output measures. For example, 52 percent of researchers possess **ISIh** and **ISIg** scores of 'zero'; for **JCR** and **KMS** the corresponding percentage is 33. To address this issue, we created a new dataset restricted to the top thirty producers of share-adjusted ISI cites (**ISIC**). The results of the exercise are shown in TABLE 4.   It is apparent that all combinations of direct citation measures (**ISIC**, **ISIg** and **ISIh**) are highly correlated (0.92 to 0.94); however, the relationship between direct citation and journal-based output measures is much weaker and more variable than under the original dataset. In particular, note that **ISIg** and **ISIh** are weakly correlated with **EQUAL** (0.36/ 0.40) and even more so with **KMS** (0.29/ 0.26).[10]

---

[10]   The corresponding rank order correlation coefficients exhibit the same relationship patterns as discussed above, save for somewhat larger coefficients from our three ISI RAMs and KMS (0.45, 0.49 and 0.58, respectively).

Table 4. *Pearson Correlation Coefficients for TOP 30 ISI Share Adjusted Cite Performers,*
Various Output Measures over the Period 2003-2008

	ISIC	ISIg	ISIh	EQUAL	ERA	JCR	KMS
ISIC	1.00	0.94	0.92	0.44	0.65	0.45	0.17
ISIg		1.00	0.92	0.36	0.65	0.53	0.29
ISIh			1.00	0.40	0.67	0.50	0.26
EQUAL				1.00	0.86	0.38	0.09
ERAB					1.00	0.73	0.38
JCR						1.00	0.78
KMS							1.00

Note1: Academic Staff as at 15 April 2009, min. six years experience.

The impact of alternative output measures on individual researchers is clearly shown in TABLE 5. Here we list the top 30 performers as ranked by the number of ISI share-adjusted cites (**ISIC**). As expected, the **ISIC** leaders perform well under **ISIg** and **ISIh;** in general, they also perform well under our alternative schemes, with the notable exception of **KMS.** Nevertheless, there is enough variation throughout to suggest that individual researchers have much to lose or gain from the RAM selection process. For example, see TABLE 6. Here we provide a few arbitrarily selected examples of the situation faced by some of our researchers assuming **ISIh** to be the official RAM. In the upper portion of the table we show the research output of six researchers who have obtained an **ISIh** score of '1'. We suggest that most tenure, appointment and promotion committees would not be indifferent between a candidate with a research portfolio consisting of the *Journal of Economic Theory*, *Review of Economics and Dynamics* and the *Journal of Monetary Economics* and a candidate with a single publication from one of the Group B listed journals. Note that the economist with the above portfolio was ranked 7th out of 135 according to the KMS scheme but received the same ISI h-index score as Researchers 4 to 6 who received the lowest possible KMS score (135/135). A similar situation exists with respect to some researchers with an **ISIh** score of 'zero'. For example, it is difficult for us to imagine that a publication in the *Journal of Economic Literature* has zero value, or that it could be judged to be of equal value to a 'null set' (see h=0; Researcher5). Once again, for comparison purposes note that the individual with a portfolio consisting solely of the *Journal of Economic Literature* is ranked 18th by KMS, and receives the same **ISIh** score as 18 academics who failed to produce a single paper in an Econlit listed journal (KMS rank: 135/135).

Let us now turn our attention to the impact of **g** and **h**-index measures on departmental performance.[11] At this point we must remind the reader that under the PBRF scheme, both absolute and relative performance is important to institutions. For example, a university's absolute score is the key variable in determining its share of the research revenue pie, whereas its ranking is the critical factor in the public relations game. That is, if institution A receives a score of 10, and Institution B receives a score of, say, 5, institution A will claim to be better than B and will receive twice as much funding (everything else being

---

[11] Recall that this analysis is based on our full dataset (N=135); that is, we include all academic staff employed by each economics department as at April 2009 without regard to years of professional experience.

equal). If however, Institution B were to improve its score to 9 in the next evaluation round, it would still be without bragging rights, but it would now receive 90% as much funding as Institution A (once again, everything else being equal). Therefore, it is important to look at the impact of **ISIg** and **ISIh** on both cardinal and ordinal performance.

Table 5. *Rank by Various Output Measures of Top30 ISI Share Adjusted Cites Performers 2003-2008*

Researcher ID	ISI Shadj Cites	ISI Shadj g	ISI shadj h	EQUAL	ERA	JCR	KMS
R1	1	1	1	3	2	6	23
R2	2	2	2	6	7	8	20
R3	3	2	2	2	3	7	9
R4	4	4	4	9	6	2	1
R5	4	4	4	4	1	1	2
R6	6	4	4	42	28	28	6
R7	7	7	4	5	5	3	13
R8	8	7	8	40	43	39	25
R9	9	13	8	13	12	12	18
R10	10	9	8	12	9	15	7
R11	11	9	20	29	31	20	27
R12	12	9	8	30	16	5	3
R13	12	9	8	19	21	16	22
R14	14	13	20	39	26	25	41
R15	15	13	8	25	15	4	5
R16	16	28	20	1	4	32	28
R17	17	13	20	35	49	35	34
R18	18	13	8	50	46	23	11
R19	19	13	8	17	18	30	38
R20	20	13	20	18	22	19	53
R21	21	28	20	61	67	63	49
R22	22	13	20	8	11	27	47
R23	23	13	8	63	60	40	60
R24	23	13	8	7	10	29	29
R25	25	13	8	46	35	24	17
R26	25	13	20	37	27	36	44
R27	27	13	20	74	74	50	59
R28	28	13	20	52	30	17	10
R29	28	28	20	28	23	21	14
R30	30	13	8	26	19	11	19

Note1: Citations collected over the period 1 January 2003 to 31 December 2010
Note2: Academic Staff with 6 or more experience as at 31 December 2008

Table 6. *Examples of Outcomes for Individual Economists: ISI-based h-index*
Research Portfolios with an h-Index of 1

Res.	KMS RANK	Publications	Res.	KMS RANK	Publications
1	5/135	Quarterly Journal of Economics Journal of International Econ.	4	135/135	Aust. J of Agr & Res Ec
2	7/135	Journal of Economic Theory Review of Economics and Dynamics Journal of Monetary Economics	5	135/135	Telecommunications Policy
3	31/135	Journal of Applied Economics Economics Letters Manchester School Open Economies Review Journal of Development Studies Plus: 30 additional refereed papers	6	135/135	Economics Letters

Research Portfolios with an h-Index of 0

Res	KMS RANK	Publication	Res.	KMS RANK	Publication
1	18/135	Journal of Economic Literature	5	135/135	18 Researchers with Zero Output
2	55/135	Ec of Education Review Australian J of Labour Ec Australian Ec Papers International Rev of Applied Ec New Zealand Ec Papers International J of Manpower			
3	63/135	International J of Industrial Org			
4	135/135	J of Markets & Morality History of Ec Review (2) European J of Hist. of Ec Thou Journal of Interdisciplinary Ec International J of Social Ec			

Note: 36 economists received a KMS score of zero, and for ranking purposes are denoted as achieving a rank of 135/135.

In Table 7 we present the Pearson Correlation Coefficients associated with all combinations of our seven RAMs. Note that the coefficients for all combinations of our three ISI-based measures range from 0.88 to 0.98. **ISIg** and **ISIh** are strongly correlated with **ERAB** and **JCR**, less so with **EQUAL** and only modestly so with **KMS** (0.46/ 0.38). However, our raw citation measure (**ISIC**) exhibits somewhat different behaviour: strongly correlated with **EQUAL** (0.94) but essentially uncorrelated with **KMS** (0.09). In general, the departmental correlation coefficients tend to be less stable than those associated with our individual results, probably due to the rather small size in the former case (eight observations only). The associated Spearman Rank correlation coefficients are displayed in TABLE 8. We now find that the relationship between the three ISI RAMs and **KMS** is quite similar to that displayed by the former with **EQUAL**, **JCR** and **ERAB**. With one exception, the relevant correlation coefficients range from 0.53 to 0.76.

Table 7. *Pearson Correlation Coefficients for Departmental Output Measures*
2003-2008

	ISIC	ISIg	ISIh	EQUAL	ERA	JCR	KMS
ISIC	1.00	0.89	0.88	0.94	0.88	0.67	0.09
ISIg		1.00	0.98	0.77	0.84	0.91	0.46
ISIh			1.00	0.77	0.82	0.85	0.38
EQUAL				1.00	0.93	0.56	-0.13
ERAB					1.00	0.76	0.25
JCR						1.00	0.74
KMS							1.00

Note: For all academic staff as of 15 April 2009

Table 8. *Spearman Rank Correlation Coefficients for Departmental Output Measures*
2003-2008

	ISIC	ISIg	ISIh	EQUAL	ERA	JCR	KMS
ISIC	1.00	0.85	0.81	0.67	0.71	0.64	0.60
ISIg		1.00	0.85	0.53	0.65	0.87	0.76
ISIh			1.00	0.67	0.76	0.69	0.55
EQUAL				1.00	0.95	0.52	0.48
ERAB					1.00	0.67	0.62
JCR						1.00	0.91
KMS							1.00

Note: For all academic staff as of 15 April 2009

Although somewhat less sophisticated, the importance of the RAM selection process in determining institutional winners and losers can best be seen in Table 9 and Table 10. In the former we present standardized results for each output measure; that is, listed scores represent the percentage attainment of a given department relative to that attained by the leading department. For example, under **ISIC**,

Auckland's per capita output is 24 percent of that produced by Waikato's researchers. The importance of the RAM selection process can be seen by continuing to compare the above noted departments under alternative RAMs. For example, if we substitute **ISIg** and **ISIh** for **ISIC**, Auckland's performance relative to Waikato's improves dramatically in absolute terms: from 23.9 to 66.2 and to 65.0 percent. Auckland also improves its relative ranking (see Table9): from 5th under **ISIC** to 3rd and 2nd for **ISIg** and **ISIh**, respectively. In this case Auckland's financial payout increases substantially but its bragging rights improve modestly (it is unusual for institutions to refer to their ranking unless they are '#1'). At this point we should stress that if either of our impact factor-based RAMs (**JCR** or **KMS)** were to be the official weighting scheme, Auckland's economics department would assume first place in the rankings.

Table 9. *Normalized Departmental Scores: Weighted Pages per Capita Various Output Measures 2003-2008*

	ISI Shadj Cites	ISI Shadj h	ISI Shadj g	EQUAL	ERA	JCR	KMS
Auckland	23.9	66.2	65.0	39.8	55.9	100.0	100.0
AUT	4.4	20.0	14.3	45.8	51.8	25.5	5.0
Canterbury	27.2	60.0	60.7	49.3	59.5	82.5	54.7
Lincoln	7.3	36.4	26.0	27.9	17.9	14.6	1.8
Massey	5.2	25.6	20.6	25.4	25.1	17.9	2.7
Otago	25.8	70.6	57.1	46.2	57.8	64.5	17.8
Victoria	26.3	56.5	60.5	36.2	50.1	88.5	44.7
Waikato	100.0	100.0	100.0	100.0	100.0	99.6	18.0

Table 10. *Departmental Rankings: Weighted Pages Per Capita, Various Citation Measures 2003-2008*

	ISI Shadj Cites	ISI Shadj h	ISI Shadj g	EQUAL	ERA	JCR	KMS
Auckland	5	3	2	5	4	1	1
AUT	8	8	8	4	5	6	6
Canterbury	2	4	3	2	2	4	2
Lincoln	6	6	6	7	8	8	8
Massey	7	7	7	8	7	7	7
Otago	4	2	5	3	3	5	5
Victoria	3	5	3	6	6	3	3
Waikato	1	1	1	1	1	2	4

The argument that the RAM selection process matters can be clearly seen by referring to the situation facing AUT. Under **ISIC**, its per capita performance is only 4.4% of Waikato's; however, under **ISIg** and **ISIh** it improves three and four fold respectively (to 14.3 and 20.0 percent respectively). Although AUT's financial position improves greatly (everything else being equal) if either **ISIg** or **ISIh** is substituted for **ISIC**, they experience no change in relative ranking- 8th in each and every case. At this point all we can say is that **ISIg** and **ISIh** yield somewhat similar results with respect to absolute performance and relative ranking, but in many cases rather dissimilar results from those generated by alternatives, especially **KMS** and, surprisingly, with respect to absolute performance, **ISIC**.

6. Policy Implications and Conclusion

Our findings are such that it is difficult to argue that the **g** and **h**-index should be formally incorporated into a New Zealand-like research assessment scheme (PBRF). Technically, the **g** and **h** RAMs yield individual and departmental rankings and scores that are somewhat consistent with those generated by our low powered journal weighting schemes (**EQUAL** and **ERA**), less so with the scheme based on the basic impact factor methodology (**JCR**) and far less so with the weighting scheme based on adjusted impact factors (**KMS**). Hence, one is ultimately faced with the same issues encountered in the journal weighting scheme selection process: which journals should be included in the citation exporting and importing lists; and is it appropriate to assume that all cites to and from selected journals are of equal value.

Our primary reason for recommending that **h** and **g**-indexes not be utilized in the PBRF relates to pragmatic matters: the results by-and-large fail the test of reasonableness. For example, it is difficult to imagine that most academic economists could have faith in a scheme that places 52 percent of members in the 'no measureable output' category, and 94 and 89 percent of participants in the three lowest categories for **ISIh** and **ISIg**, respectively. Or consider the case previously outlined in which an individual placed 7th out of 135 under the KMS ranking scheme (arguably the most respected of all journal based ranking schemes in economics) and yet received an **ISIh** score of '1' along with 33 other researchers. All of these examples relate to the lack of differentiation in scores and rankings associated with **h** and **g** indexes in the context of a time limited, all inclusive (with respect to both individuals and institutions) research assessment scheme. However, we must acknowledge that this outcome may be New Zealand-specific or attributable to data collection and timing issues. Let us now turn our attention to these matters.

Recall that our objective was to assess the relevance of the **h** and **g**-indexes to economics in the context of a nation-wide research evaluation scheme. For this purpose we chose New Zealand's PBRF. This meant that we only counted citations to work generated over a six year period, although the citation collection period was extended by two years to yield a maximum citation generation period of eight years. Alternatively, virtually all prior work, as noted earlier in the paper, has been based on life time citations to the life time output of selected researchers. Obviously, this approach generates many more citations per researcher, everything else being equal, and undoubtedly results in a much greater spread of **g** and **h** scores, and far fewer zero scores that characterize our results. However, the life time citation counting approach is not consistent with research assessment exercises that attempt to reward current (or recent) output.

The PBRF is all inclusive; that is, all academics are part of the assessment group. Hence, the study includes many academics that produce very few, if any, refereed papers over the applicable time period. On the other hand, most existing economics papers on **g** and/or **h** are based on a selective sample of researchers. For example, some studies include only Full Professors (Hendrekson and Waldenstrom, 2011), some attempt to identify elite producers (Bodman, 2010); others restrict the sample to academics in a select group of economics departments (Ellison, 2010; and Courtault et al, 2010); and some that attempt to provide nation-wide coverage utilize life-time cites to generate **g** and **h** scores

(Ruane and Tol, 2006; Tol, 2006). Without belabouring the point, all previous empirical studies of the **g** and **h**-index, in the context of the economics profession, have been based on restrictive criteria. They have attempted to identify elite researchers and/or elite departments. This is not to detract from their findings, but merely to suggest that the **g** and **h**-index may be very good at identifying 'stars', but may not be suited to a world populated by some very good researchers, many average performers, and some non-performers, especially if only a portion of their life time work is included in the study.

Before concluding, we must provide a number of qualifications to our work. First, we have utilized the ISI database throughout the study. It is possible that the use of a competing scheme such as Google Scholar (GS) or Scopus would lead to different results. For example, GS has the advantage of normally yielding a much larger number of citations per paper, and hence it should lead to a broader range of scores and far fewer zero scores that plague our study. On the other hand, GS is far less discriminating with respect to citing sources, as it covers working papers, conference papers, books and reports. We leave it to others to argue the relative merits of GS, ISI and Scopus. Second, we have followed convention and have restricted countable research to refereed journal articles only.   We acknowledge that other forms of scholarship ideally should be included such as citations to books, conference papers and working papers; however, they have been ignored for pragmatic reasons and in this regard we have once again followed convention. Third, our findings are based on the performance of New Zealand's academic economists over the period 2003 to 2008. Hence, our results may be time specific and are definitely discipline specific. Fourth, New Zealand is a very small country with only eight university-based economics departments; therefore the transferability of our findings to larger domains is open to question and further study.

References

Anderson, D.L. and Tressler, J. (2012), 'Using Citations to Measure Research Output in New Zealand Economics Departments: The Problem of 'Long and Variable' Lags', *Agenda,* **19**(1), forthcoming.

Bodman, A.R. (2010), 'Measuring the Influentialness of economic geographers during the 'great half century': an approach using the h index', *Journal of Economic Geography,* **10**(1), 141-156.

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

Chang, C.L., McAleer, M., and Oxley, L. (2011), 'What Makes a Great Journal Great in economics? The Singer not the Song', *Journal of Economic Surveys,* **25**(2), 326-361.

Courtault, J.M., Hayek, N., Rimbaux, E., and T. Zhu (2010), 'Research in economics and management in France: A bibliometric study using the h-index', *Journal of Socio-Economics*, **39**(2), 329-337.

Egghe, L. (2006), 'Theory and practice of the g-index', *Scientometrics,* **69**(1), 131-152.

Ellison, G. (2010), 'How does the Market Use Citation Data?, The Hirsch Index in Economics', NBER, Working Paper 16419, http://www.nber.org/papers/w16419.

Garfield, E. (1972), 'Citation Analysis as a Tool in Journal Evaluation', *Science,* **178**(4060), 471-479.

Harris, G.T. (1990), 'Research Output in Australian University Economics Departments: An Update for 1984-1988', *Australian Economic Papers,* **29**(3), 249-259.

Henrekson, M. and Waldenstrom, D. (2009), 'How Should Research Performance be Measured? A Study of Swedish Economists', *Manchester School,* **79**(6), 1139-1156.

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', C*anadian 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.

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.

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

Macri, J. and Sinha, D. (2006), 'Rankings Methodology for International Comparisons of Institutions and Individuals: An Application to Economics in Australia and New Zealand', *Journal of Economic Surveys,* **20**(1), 111-156.

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

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

Pomfret, R. and Wang, L.C. (2003), 'Evaluating the Research Output of Australian Universities' Economics Departments', *Australian Economic Papers,* **42**(4), 418-441.

Ravallion, M. and Wagstaff, A. (2011), 'On Measuring Scholarly Influence by Citations', *Scientometrics,* **88**(1), 321-337.

Ruane, F. and Tol, R.S.J. (2008), 'Rational (successive) h-indices: An application to economics in the Republic of Ireland', *Scientometrics,* **75**(2), 395-405.

Schreiber, M. (2010), 'Twenty Hirsch index variants and other indicators giving more or less preference to highly cited papers', *Annals of Physics (Berlin),* **522**(8), 536-554.

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

Tol, R.S.J. (2008), 'A rational, successive g-index applied to economics departments in Ireland', *Journal of Informetrics,* **2**(2), 149-155.

Tol, R.S.J. (2009), 'The h-index and its alternatives: An application to the 100 most prolific economists', *Scientometrics,* **80**(2), 317-324.

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