Scientometrics
DOI 10.1007/s11192-016-1979-1

Research assessment using early citation information

Stephan B. Bruns[1] · **David I. Stern**[2]

Received: 24 February 2016

Abstract Peer-review based research assessment, as implemented in Australia, the United Kingdom, and some other countries, is a very costly exercise. We show that university rankings in economics based on long-run citation counts can be easily predicted using early citations. This would allow a research assessment to predict the relative long-run impact of articles published by a university immediately at the end of the evaluation period. We compare these citation-based university rankings with the rankings of the 2010 Excellence in Research assessment in Australia and the 2008 Research Assessment Exercise in the United Kingdom. Rank correlations are quite strong, but there are some differences between rankings. However, if assessors are willing to consider citation analysis to assess some disciplines, as is the case for the natural sciences and psychology in Australia, it seems reasonable to consider also including economics in that set.

Keywords Citations · Research assessment · Bibliometrics

JEL Classification A14 · H83 · I23

Introduction

In assessments of university research, as carried out in the social sciences in Australia and in all fields in Britain, publications that have already passed through a peer-review process at academic journals and presses are again peer reviewed by the assessment panels. This

✉ Stephan B. Bruns
bruns@uni-kassel.de

David I. Stern
david.stern@anu.edu.au

[1] Meta-Research in Economics Group and INCHER, University of Kassel, Nora-Platiel-Str. 5, 34109 Kassel, Germany

[2] Crawford School of Public Policy, The Australian National University, 132 Lennox Crossing, Acton, ACT 2601, Australia

Published online: 23 May 2016

involves a significant workload for the many academics that are supposed to read these publications in addition to the effort each university must put into selecting the publications that will be reviewed. Farla and Simonds (2015) find that the total economic cost of the 2014 Research Excellence Framework (REF) in the United Kingdom (UK) was £246 million ($384 million). Moreover, as Sayer (2014) argues, this second peer review is inferior to the first. If instead citation-based metrics were used, as is the case for the natural sciences and psychology in Australia, the assessment could be done much faster and cheaper. In this article, we provide evidence on the potential effectiveness of citation analysis as a method of research assessment in economics, a discipline subject to peer review in both the UK and Australia. We hope our results can inform the future development of assessment exercises such as the Excellence in Research in Australia (ERA) and REF. Though the Higher Education Funding Council in England (HEFCE) Metrics Review (Wilsdon et al. 2015) rejected metrics-only assessment, at the time of writing, the British government was again assessing the possibility of reducing the costs of the REF (Department of Business, Innovation & Skills and Johnson 2015).

Stern (2014) shows that long-run citations to articles in economics and political science can be predicted quite well using the first few years of citations to those articles. This predictability in economics and political science is similar to that found in several natural sciences by other researchers (Adams 2005; Waltman et al. 2011; Wang 2013). However, research assessment, as practiced in Australia and the UK, evaluates universities rather than single articles. Here we show that rank correlations are greatly increased when we aggregate over the economics publications of a university and also when we aggregate publications over time. The rank correlation between UK universities ranked by citations received till the end of 2005 for economics articles published in 2003 and 2004 and UK universities ranked by total citations to those articles received through 2014 is 0.97. Rank correlations between university rankings based on the same early citations and university rankings implied by the research assessment exercises in the UK and Australia range from 0.67 to 0.76. Though these rank correlations are quite high, deviations between the peer-review based rankings and our citation-based rankings may be partly explained by the differences between our simulated assessments and how an actual assessment would be conducted. Our results suggest that citation analysis could be useful for research assessment in economics if assessors are willing to use cumulative citations as a measure of research strength, though there do appear to be some systematic differences between peer-review based research assessment and our citation analysis, especially in the UK.

The belief that citations accumulate too slowly in most social sciences such as economics to be useful for short-term research assessment is a reason why citation analysis is less accepted in the social sciences than in the natural sciences (Bornmann and Leydesdorff 2014). Our results show that citations definitely accumulate fast enough in economics at the department or university level in order to be able to predict the longer run citation outcomes of recent publications. However, while the Australian Government's ERA exercise uses citations to recent publications to assess research performance in natural science disciplines and psychology, peer review is used in other social science disciplines. The Australian Research Council (ARC) instructs peer reviewers not to take citation numbers or journal rankings into account when assessing the quality of submissions. The Research Assessment Exercise (RAE) and its successor the REF in the UK uses peer review as the main method of assessment for all disciplines. REF panels can also put some weight on citations data in some disciplines including most natural sciences and economics (Sgroi and Oswald 2013), but only as a positive indicator of academic significance and in

very much a secondary role to peer review (Holgate 2015). This represents a change from the previous RAE, which prohibited the use of citations data by panels.[1]

Previous research (Adams 2005; Levitt and Thelwall 2011; Waltman et al. 2011; Wang et al. 2013; Stern 2014) predicts the future cumulative citations of individual articles based on early cumulative citations and other variables such as journal impact factors. However, in research assessment exercises, such as the REF and ERA, whole universities are assessed in each discipline and a corpus of research outputs covering several years is reviewed.[2] There are a number of reasons why aggregating publications by departments should increase the predictability of citations compared to predicting the citations of an individual article:

1. A portfolio effect—idiosyncratic or random variations in the time paths of citations received by individual articles are averaged away. Aggregation across time will also contribute to the portfolio effect.[3]
2. A size effect—variation in the size of departments introduces an additional correlation between the total numbers of early and cumulative citations.
3. A quality effect—the size of departments is positively correlated with their research quality (Mryglod et al. 2013), which will further increase the correlation between the total numbers of early and cumulative citations across departments.

If we want to allocate research funds according to research *strength*—combining both the quantity and quality of research—then these effects are advantageous.

The next section of the paper reviews recent research on research assessment and citation analysis. Following this, we present our methods, data sources, and results. Finally, we present some discussion and conclusions.

Peer review and citation analysis in research assessment

Existing research finds strong correlations between the rankings produced by UK research assessment exercises and bibliometric analyses for several specific humanities and social science disciplines (e.g. Colman et al. 1995; Oppenheim 1996; Norris and Oppenheim 2003) including economics (Süssmuth et al. 2006; HEFCE 2015). Clerides et al. (2011) compare the 1996 and 2001 RAE ratings of economics departments with independent rankings from the academic literature. They find RAE ratings to be largely in agreement with the profession's view of research quality as documented by independent rankings, although the latter appear to be more focused on research quality at the top end of academic achievement. This is because most rankings of departments in the economics

[1] A variety of research assessment models are in place in different countries (Key Perspectives 2009). Research assessment exercises in other countries use different combinations of peer review and bibliometric analysis. For example, the Italian Evaluation of Research Quality must peer review at least half the submitted research items (Bertocchi et al. 2015). In the U.S., the National Research Council carries out a periodical assessment of doctoral programs. The 2011 assessment (Ostriker et al. 2011) covered 5000 doctoral programs at 212 universities. The most recent assessment aggregated various quantitative metrics, including numbers of citations, using weights derived from a survey of faculty on the importance of the various metrics.

[2] In New Zealand, though, individual researchers are assessed (Anderson and Tressler 2014).

[3] Of course, this effect will also apply to many other ways of aggregating publications including random samples of publications.

literature are based on publications in top journals only, which lower-ranked departments have very few of.

Mryglod et al. (2013) analyze the correlations between the values of the Thomson Reuters Normalized Citation Impact (NCI) indicator and RAE 2008 peer-review scores in several academic disciplines. The NCI computes the normalized (by field) impact factor across a unit of assessment (an academic discipline at a given university) in the RAE based on only the publications actually submitted to the RAE. Mryglod et al. (2013) compute both average and total, or quality and strength (Kenna and Berche 2011), values of the RAE 2008 peer-review scores and NCI for each university. Research strength is the average value, or quality, of each of these two indicators for a unit of assessment multiplied by the number of staff submitted to the RAE. They find very high correlations for the strength indicators for some disciplines and poor correlations for the quality indicators for all disciplines. Moreover, the correlation between peer-evaluated and citation-based scores is weaker for the "soft" sciences. Spearman rank correlation coefficients for their quality indicators range from 0.18 (mechanical engineering) to 0.62 (chemistry). For strength, however, the correlations range from 0.88 (history and sociology) to 0.97 (biology). This is because quality is correlated with size and so the two factors reinforce each other.

Mryglod et al. (2015) attempt to predict the results of the 2008 RAE retrospectively and those of the 2014 REF before they were released. They examined biology, chemistry, physics, and sociology. Of the indicators they trialed, they found that the departmental h-index had the best fit to the 2008 results. Departmental h-index is based on all publications published by a department in the time window assessed by the relevant assessment exercise. The rank correlation ranged from 0.83 in chemistry to 0.58 in sociology. They find that the correlation with the RAE ranking for the immediate h-index is as good as the correlation with the h-index computed in later years for the same set of publications.

However, despite these findings, Bornmann and Leydesdorff (2014) argue that one of the downsides of bibliometrics, as a research assessment instrument, is that citations take time to accumulate while research assessment exercises are designed to assess recent performance:

> This disadvantage of bibliometrics is chiefly a problem with the evaluation of institutions where the research performance of recent years is generally assessed, about which bibliometrics—the measurement of impact based on citations—can say little…. the standard practice of using a citation window of only 3 years nevertheless seems to be too small. (1230)

They argue further that bibliometrics:

> can be applied well in the natural sciences, but its application to TSH (technical and social sciences and humanities) is limited. (1231)

Rather than assuming that peer review is the preferred approach to research assessment and citation analysis should only be used to reduce costs, we can ask whether the review conducted by research assessments such as the REF and ERA meets the normal academic standards for peer review. Research does show that peer review at journals has predictive validity for the citations that will be received by accepted papers compared to those received by rejected papers. However, evidence for the predictive validity of peer review of grant and fellowship applications is more mixed (Bornmann 2011; Gallo et al. 2014), and Sayer (2014) argues that the peer review undertaken in research assessment exercises

does not meet normal standards for peer review. He compares processes used in the REF at the university and national levels to practices of scholarly review found in academic journals, university presses, and North American tenure procedures. He finds that the peer-review process used by the REF falls far short of the level of scrutiny or accuracy of these more familiar peer-review processes. The number of items each reviewer has to assess means that the review cannot be of the same quality as reviews for publication. And reviewers will have to assess much material outside their area of specific expertise. Sayer argues that, though metrics may have problems, a process that gives such extraordinary gatekeeping power to individual panel members is far worse. Moreover, Wooding et al. (2015) find "grade inflation" between RAE 2008 and REF 2014. The number of world-leading (denoted by 4* in RAE) publications increased by 103 % between 2008 and 2014. But the increase in UK research articles among the 10 % most cited globally was only 10–25 %. Therefore, there seems to be a lack of consistency in evaluation over time.

Given the large number of items that panels need to review they are likely to focus on the venue of publication and, at least in business and economics, handy mappings of journals to REF grades exist (Hudson 2013). HEFCE (2015) finds that economics and econometrics is the field with the strongest correlation at the individual publication level between scores in REF 2014 and the impact factors of journals in which the articles were published. This suggests that REF reviewers mostly used journal reputation to assess articles. Regibeau and Rockett (2014) build imaginary economics departments entirely composed of Nobel Prize winners and evaluate them using standard journal rankings geared to the UK RAE. Performing the same evaluation on existing departments, they find that the rating of the Nobel Prize departments does not stand out from other good departments. But if departments composed entirely of Nobel Prize winners perform worse than current departments then it is hard to know what such assessment means. Also, compared to actual recent research evaluations, the Nobel Prize departments' rankings are less stable over time. They argue that this is because the rankings exercise leads departments to hire people with appropriate recent publications to fill any gaps in their performance. They also find that including more than each researcher's best four publications enhances the relative performance of the Nobel Prize departments.[4]

Sgroi and Oswald (2013) examine how research assessment panels could most effectively use citation data to replace peer review. They suggest a Bayesian approach that uses prior information on where an item was published combined with observations on citations to derive a posterior distribution for the quality of a publication. We could then estimate, for example, what is the probability that a publication belongs in the 4* category given where it was published and the early citations it has received. Stern (2014) and Levitt and Thelwall (2011) show that the journal impact factor has strong explanatory power in the year of publication but that this declines very quickly as citations accumulate. So, this approach would be most useful for papers published in the last year or two before the assessment, but for earlier research outputs in the assessment window the added value over simply counting citations would be minimal.

[4] The REF, and previously the RAE, assesses only four publications for each submitted researcher. The submitting university chooses both which researchers and which of their publications to submit.

Data and methods

Data

We used the *Web of Science* to download bibliographic information on all articles that were published by universities in Australia and the UK in 2003 and 2004 in the *Web of Science* Category "Economics", which included 169 journals in 2003 and 172 journals in 2004.[5] The number of journals covered rose steeply after 2006 reaching 333 in 2013. This means that the potential to receive citations that are included in the database has increased substantially over our assessment window. Although this affects cumulative citations, we do not expect that it benefits some universities more than others. We restrict our analysis to the document types "Article", "Article; Proceedings Paper", and "Review". Reviews are articles with more than 100 references. We also separately downloaded all citations to these articles in each year from 2003 to 2014. We downloaded the data in September 2014 and all citations until then are included in our dataset. We combined the separate dataset containing information on the authors and their affiliations with the citations dataset using the name of the article, the name(s) of the author(s), and the beginning and ending pages of each article, which are found in both datasets.

We then selected all articles where at least one author has an affiliation with an address that contains "Australia" or at least one author has an affiliation with an address that contains "England", "Scotland", "Northern Ireland", or "Wales". We also searched for United Kingdom and UK, but no article used these terms in the authors' affiliations. As the names of the universities are not harmonized in the *Web of Science*, many different ways of writing the name of the same university are present in the data. We identified and harmonized the university names by hand to allocate articles to universities and removed articles that did not include a university-based author.

Our sample contains 293 articles in 2003 and 321 articles in 2004 that were published by authors with at least one affiliation to a university in Australia. For the UK, we have 1155 articles in 2003 and 1132 articles in 2004 that were published by authors with at least one affiliation to a university in the UK.

Each article is potentially written by multiple authors, and each author may be potentially affiliated with multiple universities. The *Web of Science* data contains the affiliations listed on the article, which depends on the style used by the journal and, therefore, does not unambiguously identify how many researchers from a given university were authors of a given article. Some journals list each university once irrespective of how many authors of an article are actually affiliated with that university, while others may list the same university multiple times if multiple authors are affiliated with the same university. We count each article only once for each university irrespective of how often the same university may be listed on the article due to these data limitations.

Counting each article once for each university is also reasonable given how actual research assessments are conducted. In the ERA, each publication is counted once for each university. However, a single publication can be submitted in more than one discipline area for a single university in which case it only has a fractional weight in each discipline. The

[5] The vast majority of bibliometric research uses the *Web of Science* as its data source. One reason for this is that it allows researchers to easily download the results of searches as data files. This data includes year-by-year citations to each article. Though *Google Scholar* covers a wider range of citing and cited sources, it is very noisy with many misidentified publications and citations. Constructing a data set for a discipline in a country would be a very labor-intensive process. *Scopus* is also not as user-friendly as the *Web of Science*. For example, one cannot search by discipline in *Scopus*.

RAE usually only allows authors at the same university to submit the same publication if they are submitting under different "units of assessment" or disciplines. Co-authors at different universities can each submit the publication for full credit in the same unit of assessment. Therefore, in most cases our count of articles will match the ERA or RAE.

Methods

We are interested, first of all, in how well a ranking of universities in economics that is based on early citations can predict a university ranking that is based on cumulative citations over a longer period. We also test this predictability for an alternative ranking based on the number of publications of each university among the top 10 % of publications by citations. Second, we are interested in how strongly correlated our rankings are with the outcomes of actual research assessment exercises. We conduct our analysis separately for Australia and the UK.

To compute rank correlations between universities' early and cumulative citations we aggregate each university's citations for articles published in 2003 over the periods 2003, 2003–2004, 2003–2005, and 2003–2014. We then calculate the rank correlation (Spearman) between universities ranked by cumulative citations (2003–2014) and universities ranked by early citations (2003, 2003–2004, or 2003–2005). Spearman's rank correlation coefficient is a special case of the general correlation coefficient. In particular, it uses the average rank if multiple universities have the same number of citations. For example, if the top five universities have 100, 100, 100, 90, and 80 citations, the ranks used in the Spearman rank correlation are 2, 2, 2, 4, and 5. We carry out the same analysis for articles published in 2004, using the early citation windows 2004, 2004–2005, and 2004–2006, and for the set of articles from both 2003 and 2004, using 2003–2004, 2003–2005, and 2003–2006 citation windows.

We also consider the subgroup of universities that actually submitted to the ERA 2010 and RAE 2008. Many UK universities that produced some economics research did not submit the discipline to the RAE (Johnston et al. 2014). Finally, we also compute the rank correlation between individual articles ranked by early citations and ranked by cumulative citations to demonstrate the effect of aggregating articles across a university.

In addition to total citations, we consider university rankings based on each university's number of articles in the top 10 % of articles. The top 10 % of articles are defined by early or cumulative citations for the respective country and publication window. For example, the top 10 % of articles published in 2003 for the citation window 2003–2005 in the UK, are those 10 % of UK articles published in 2003 that have the highest total citations between 2003 and 2005. This indicator is the same as the Leiden size-dependent indicator P(top 10 %) (Waltman et al. 2012). This indicator reduces the effect of outlier articles with very high numbers of citations and should reward consistent quality more.

Finally, we analyze the rank correlations between universities ranked by (early and cumulative) citations and universities ranked by the measures of research quality produced by the ERA 2010 (Australia) and research strength and quality produced by the RAE 2008 (UK). RAE 2008 evaluates universities by determining the fractions of each university's publications that fall into categories 4*, 3*, 2*, and 1*, which reflect research that is "world leading", "internationally excellent", "recognized internationally", and "recognized nationally", respectively. The quality measure of RAE 2008 is calculated as a weighted sum of the shares of research falling into categories 4*, 3*, 2*, and 1* using the weights 4, 3, 2, and 1. This is the method used by UK newspapers to derive a ranking from the RAE and REF results. We compute the strength measure for RAE 2008 by multiplying

the quality measure with the number of full time equivalent employees that were submitted to RAE 2008. We also consider an alternative RAE 2008 ranking based on the funding allocated as a result of the exercise. Here we weight the shares of research falling in the 4*, 3*, 2*, and 1* categories by 9, 3, 1, and 0, respectively, which was the actual weighting used for distributing funding at that time. ERA 2010 provides a rating of universities that directly corresponds to a quality ranking. The information provided in ERA 2010 does not let us compute a strength ranking for Australia.

Results

Table 1 presents the rank correlations between universities ranked by early citations and universities ranked by cumulative citations for Australia and the UK based on articles published in 2003, 2004, and aggregated over time (2003 and 2004). The table also presents the rank correlations between individual articles ranked by early citations and ranked by cumulative citations. Additionally, the table presents rank correlations based on the number of articles each university published in the top 10 % of articles.

With regards to total citations at the article level, the rank correlations are higher for the UK than for Australia but both are lower than those presented by Stern (2014) for the global set of economics articles. Stern (2014) reports a rank correlation of 0.36 between 2006 citations and cumulative 2006–2012 citations for economics articles published in 2006 (0.73 if 2006–2007 is compared with 2006–2012). For articles published in 1999, the rank correlation between 1999 citations and 1999–2012 citations was 0.33 (0.60 if 1999–2000 is compared to 1999–2012). It seems to be easier to predict citations using a global data set than a data set for a single country. It is possible that this is because in a larger sample there are more highly cited articles with individual ranks and gaps in the numbers of citations between them that increase predictability. In Australia, both the mean and standard deviation of citations received by an article is smaller than the world average so that the Australian sample of articles is more alike in the number of citations received than the world sample is. For 2003 articles, the mean and standard deviation of cumulative citations are for Australia 15 and 22, for the UK 23 and 58, and for the world 21 and 50.

More importantly, aggregating citations over universities increases the rank correlations tremendously. The rank correlations are higher for the UK than for Australia, but 3 years of citations (2004–2006) already generates a rank correlation of 0.95 for Australia using 2004 articles alone, so this difference is not so important.

Excluding universities that did not submit to RAE 2008 makes little difference to the results and excluding universities that did not submit to ERA 2010 makes no difference at all to the results. In the first case, those universities that did not submit had few publications, and in the latter case there are very few such universities that published in economics but did not submit.

Aggregating 2003 and 2004 articles together increases the rank correlations relative to using only 2003 citations to 2003 articles but actually slightly reduces the rank correlations compared to using a 2-year window for the 2003 articles alone. In Australia the rank correlation between the university ranking based on citations received in 2003–2004 for 2003 articles and the university ranking based on cumulative citations for 2003 articles is 0.88 but this decreases to 0.82 by adding the 2004 articles, which only have a rank correlation of 0.65 between their 2004 and cumulative citations. This shows that including

Table 1 Rank correlations of universities and articles based on early and cumulative citations

	Australia			United Kingdom		
	2003	2003–2004	2003–2005	2003	2003–2004	2003–2005
2003 Articles						
Total citations						
Article	0.22	0.44	0.65	0.28	0.54	0.70
University	0.66	0.88	0.92	0.82	0.92	0.97
Submitted to RAE 2008/ERA 2010	0.66	0.88	0.92	0.86	0.96	0.98
Number of articles in top 10 %						
University	0.72	0.84	0.84	0.69	0.70	0.89
Submitted to RAE 2008/ERA 2010	0.71	0.83	0.84	0.73	0.77	0.90
	2004	2004–2005	2004–2006	2004	2004–2005	2004–2006
2004 Articles						
Total citations						
Article	0.16	0.42	0.63	0.28	0.58	0.74
University	0.65	0.82	0.95	0.79	0.95	0.98
Submitted to RAE 2008/ERA 2010	0.65	0.82	0.95	0.80	0.94	0.97
Number of articles in top 10 %						
University	0.64	0.54	0.66	0.81	0.76	0.82
Submitted to RAE 2008/ERA 2010	0.64	0.54	0.66	0.85	0.87	0.81
	2003–2004	2003–2005	2003–2006	2003–2004	2003–2005	2003–2006
2003 and 2004 Articles						
Total citations						
University	0.82	0.93	0.97	0.91	0.97	0.98
Submitted to RAE 2008/ERA 2010	0.82	0.93	0.97	0.96	0.98	0.98
Number of articles in top 10 %						
University	0.83	0.91	0.92	0.84	0.88	0.90
Submitted to RAE 2008/ERA 2010	0.85	0.91	0.92	0.90	0.90	0.96

Rank correlations between university and article rankings based on early citations and university and article rankings based on cumulative citations as well as rank correlations between university rankings based on the number of articles in the top 10 % of articles by early citations and by cumulative citations are shown. Cumulative citations are based on citations from 2003 to 2014 ("2003 Articles" and "2003 and 2004 Articles") and on 2004–2014 ("2004 Articles"). The rows labeled "Article" present the rank correlations at the article level. The rows labeled "University" give the rank correlations when articles are aggregated by university. The table also shows the rank correlations excluding universities that did not make submissions in economics to ERA 2010 (Australia) or RAE 2008 (UK). Rank correlations at the article level for the combined articles of 2003 and 2004 are not presented, as this would require a comparable citation window for articles published in 2003 and 2004. However, a citation window of, for example, 2 years uses citations received in 2004–2005 for articles published in 2004 but it only uses citations received in 2003–2004 for articles published in 2003. As such a ranking could be only made in 2005, information on the citations received in 2005 for articles published in 2003 would be ignored

very recent articles, which have had less than a year to accumulate citations, in the assessment can reduce the correlation with long-run citations. In actual research assessment exercises more than 2 years are evaluated and this issue is likely to play a minor role.

Table 2 Rank correlations of universities based on research assessment scores and early and cumulative citations

	2003	2003–2004	2003–2005	2003–2014
2003 Articles				
Total citations				
ERA 2010	0.65	0.68	0.78	0.80
RAE 2008 (strength)	0.65	0.70	0.69	0.63
RAE 2008 (quality)	0.62	0.68	0.67	0.65
RAE 2008 funding (strength)	0.67	0.73	0.72	0.67
RAE 2008 funding (quality)	0.63	0.71	0.70	0.68
Number of articles in top 10 %				
ERA 2010	0.73	0.74	0.80	0.73
RAE 2008 (strength)	0.63	0.63	0.52	0.48
RAE 2008 (quality)	0.61	0.59	0.41	0.41
RAE 2008 funding (strength)	0.66	0.65	0.52	0.48
RAE 2008 funding (quality)	0.62	0.62	0.43	0.44
	2004	2004–2005	2004–2006	2004–2014
2004 Articles				
Total citations				
ERA 2010	0.66	0.63	0.70	0.73
RAE 2008 (strength)	0.57	0.68	0.66	0.64
RAE 2008 (quality)	0.52	0.61	0.58	0.61
RAE 2008 funding (strength)	0.59	0.69	0.67	0.65
RAE 2008 funding (quality)	0.53	0.62	0.60	0.61
Number of articles in top 10 %				
ERA 2010	0.72	0.63	0.64	0.56
RAE 2008 (strength)	0.56	0.56	0.58	0.48
RAE 2008 (quality)	0.51	0.51	0.55	0.42
RAE 2008 funding (strength)	0.59	0.59	0.61	0.49
RAE 2008 funding (quality)	0.54	0.54	0.57	0.44
	2003–2004	2003–2005	2003–2006	2003–2014
2003 and 2004 Articles				
Total citations				
ERA 2010	0.69	0.76	0.77	0.81
RAE 2008 (strength)	0.69	0.69	0.67	0.67
RAE 2008 (quality)	0.68	0.67	0.64	0.69
RAE 2008 funding (strength)	0.73	0.72	0.69	0.70
RAE 2008 funding (quality)	0.71	0.69	0.66	0.71
Number of articles in top 10 %				
ERA 2010	0.76	0.82	0.83	0.78
RAE 2008 (strength)	0.64	0.61	0.56	0.53
RAE 2008 (quality)	0.64	0.58	0.50	0.49
RAE 2008 funding (strength)	0.66	0.64	0.57	0.54

Table 2 continued

	2003–2004	2003–2005	2003–2006	2003–2014
RAE 2008 funding (quality)	0.66	0.61	0.52	0.50

Rank correlations between university rankings based on the score received in ERA 2010 or RAE 2008 and both university rankings based on early and cumulative citations as well as university rankings based on the number of articles in the top 10 % of articles by early and cumulative citations are shown. RAE 2008 (quality) denotes the quality measure of RAE 2008 and RAE 2008 (strength) denotes the strength measure of RAE 2008. RAE 2008 funding denotes the weighting by 9,3,1,0 instead of 4,3,2,1 (please see "Methods" section). For example, 0.65 in the first row is the rank correlation between the university ranking based on ERA 2010 and the university ranking based on citations received in 2003 for articles published in 2003. At the end of this row, 0.80 denotes the rank correlation between the university ranking based on ERA2010 and the university ranking based on cumulative citations (2003–2014) for articles published in 2003

Research assessment exercises might consider discounting the articles published in the last year of the assessment period or using additional information for these articles, e.g. journal impact factors, potentially in a Bayesian framework as suggested by Sgroi and Oswald (2013). In the UK, adding the 2004 articles to the 2003 ones only reduces the rank correlation from 0.92 to 0.91 and the rank correlation for the universities that submitted to RAE 2008 remains constant at 0.96. Therefore, adding this additional information might not be worthwhile.

In most cases, the rank correlations between universities ranked by the number of articles in the top 10 % of articles based on early citations and universities ranked by the number of articles in the top 10 % based on cumulative citations are lower than for total citations, though the gap narrows when we aggregate articles over 2 years. It is particularly noticeable that in some cases widening the citation window increases predictability relatively little for this indicator, though this behavior is quite erratic.

Table 2 presents rank correlations between university rankings based on total citations accumulated over various windows and the rankings implied by ERA 2010 and RAE 2008. These rank correlations are quite low for the UK and are a bit higher for Australia. The ERA process must be more similar to a total citation counting exercise than the UK process despite providing only a quality rather than a strength metric. The correlation between citations and the RAE quality measure is lower than that with the strength measure, as we would expect, but this difference reduces when we aggregate over 2 years of articles.

For the UK, the length of the citation window makes almost no difference to the rank correlation. This is not surprising, as the rank correlation between universities ranked by early citations and universities ranked by cumulative citations is already quite high with just a 1-year early citation window. Hence, the citation-based rankings do not differ substantially with respect to the citations window. For Australia, the rank correlation between the ERA 2010 ranking and the ranking based on citations rises as the citation window is extended. This is consistent with the smaller rank correlations between universities ranked by early citations and universities ranked by cumulative citations for Australia. It seems that the ranking based on cumulative citations better matches the ranking based on ERA 2010.

The Table also presents rank correlations between the rankings implied by the research assessment exercises and the university's number of articles in the top 10 % of articles.

Table 3 Rankings of universities evaluated by the Research Assessment Exercise 2008 (strength) and total citations

University	RAE		2003 and 2004 Articles			2003 Articles 2003–2014	2004 Articles 2003–2014
	Funding	Original	2003–2004	2003–2005	2003–2014		
Univ Oxford	1	1	2	2	2	4	2
LSE	2	4	1	1	1	1	1
Univ Warwick	3	2	6	4	4	7	3
Univ Nottingham	4	3	7	5	7	8	6
UCL	5	7	5	7	5	5	4
Univ Essex	6	6	15	15	16	20	14
Univ Cambridge	7	5	3	3	3	2	5
Univ Manchester	8	9	10	9	10	9	9
Univ York	9	8	8	8	8	10	7
Univ Southampton	10	10	16	15	20	17	23
Queen Mary Univ	11	13	11	18	14	12	25
Univ Leicester	12	11	17	14	19	15	18
Birkbeck Coll	13	14	20	22	23	21	29
Univ Bristol	14	18	19	19	15	19	15
Royal Holloway Univ	15	15	23	24	24	26	21
Univ Glasgow	16	17	21	20	21	18	27
Univ St Andrews	17	12	31	32	31	32	26
Univ Edinburgh	18	20	4	6	6	3	16
Brunel Univ	19	16	25	23	22	23	19
Univ Birmingham	20	19	13	12	13	14	11
Swansea Univ	21	22	31	30	27	29	22
Univ Exeter	22	26	27	29	28	30	24
Univ Aberdeen	23	25	17	17	18	16	17
Univ Sheffield	24	23	9	11	11	13	8
Univ Surrey	25	24	29	30	30	24	32
Univ Loughborough	26	21	23	25	29	28	30
Univ Kent	27	27	28	26	25	31	20
Univ E Anglia	28	29	11	10	9	6	12
City Univ London	29	28	22	21	17	21	13
Univ Dundee	30	30	25	26	32	27	33
Univ Sussex	31	31	14	13	12	11	10
Univ Stirling	32	32	33	26	25	25	28
London Metropolitan Univ	33	33	33	35	34	33	35

Table 3 continued

University	RAE		2003 and 2004 Articles			2003 Articles	2004 Articles
	Funding	Original	2003–2004	2003–2005	2003–2014	2003–2014	2003–2014
Kingston Univ	34	35	30	33	35	35	34
Manchester Metropolitan Univ	35	34	33	33	33	34	31

Funding and Original denote the strength rankings of the Research Assessment Exercise in 2008 based on the weightings 9, 3, 2, 0 and 4, 3, 2, 1, respectively. The remaining rankings are based on cumulative citations for articles published in 2003, 2004, as well as 2003 and 2004. 2003–2004, 2003–2005, and 2003–2014 denote the window over which citations are cumulated

These correlations are lower for the UK and tend to decrease as the citation window is lengthened. This indicates that the ranking of universities based on the number of articles in the top 10 % better matches the RAE 2008 rankings when early rather than cumulative citations are used to calculate the university's number of articles in the top 10 % of articles. This is much less noticeable for Australia. We do not know the reason for this, but can speculate that some articles that eventually accumulate a large number of citations are considered less important by both early citers and research assessment reviewers, who assess articles relatively soon after they are published.

Aggregating over 2 years of publications increases the rank correlations between the rankings of the research assessments and both the university rankings based on citations and the university rankings based on the number of articles in the top 10 %. This seems to be due to year-to-year volatility in the citations rankings of low ranked universities. Table 3 shows the rankings of UK universities based on RAE 2008 and total citations with various citations windows. For example, University of Edinburgh goes from a rank of 3 in 2003 to 16 in 2004 and Queen Mary from 12 to 25. Edinburgh's high ranking for 2003 articles is due to a single econometrics article (Im et al. 2003) that received a very high number of citations.

Table 4 shows the rankings of Australian universities based on ERA 2010 and total citations with various citation windows. Both Table 3 and 4 show that rankings based on early citations are very similar to those based on citations accumulated over a longer period at the university level. The Tables allow us to see which particular universities reduce the rank correlation between the research assessment ranking and our citation-based ranking. We see that the University of Essex was highly ranked in RAE 2008 but low ranked by cumulative citations for articles from 2003 to 2004. Another notable anomaly is St Andrews. The reverse is true of East Anglia, Sussex, Sheffield, and Edinburgh. This may be related to the fields of specialization of these universities. East Anglia specializes in environmental economics, which typically has relatively high citation counts, as well as development economics. Sheffield has a specialization in health economics, which is also likely to get high citation counts due to citing from outside of economics. Essex has strengths in labor economics and micro-theory.

Table 5 provides some information on the research outputs of these universities. The four universities that rank high by citations compared to their RAE ranking all published more articles relative to the number of staff they submitted to the RAE than the two universities that ranked low by citations compared to their RAE ranking. This may be because of different models of publication in different fields—rates of publication per

Table 4 Rankings of universities evaluated by the Excellence in Research for Australia 2010 and total citations

University	ERA 2010	2003 and 2004 Articles			2003 Articles	2004 Articles
		2003–2004	2003–2005	2003–2014	2003–2014	2004–2014
Univ Melbourne	1	2	2	2	2	2
Australian Natl Univ	2	1	1	1	1	1
Monash Univ	2	5	3	4	5	5
Univ New South Wales	2	7	6	5	3	6
Univ Queensland	2	6	7	7	8	3
Univ of Technology Sydney	2	16	11	10	9	12
Univ Western Australia	2	4	5	6	6	7
Bond Univ	8	14	17	11	16	11
Queensland Univ of Technology	8	16	16	16	20	16
Univ Adelaide	8	3	4	8	7	10
Univ Sydney	8	8	8	3	4	4
Curtin University of Technology	12	21	13	15	21	13
Deakin Univ	12	16	24	18	12	26
Edith Cowan Univ	12	11	17	21	22	22
La Trobe Univ	12	10	11	14	11	17
Macquarie Univ	12	11	10	13	14	14
RMIT Univ	12	27	25	23	25	20
Swinburne Univ of Technology	12	27	29	31	NA	28
Univ New England	12	14	9	9	15	8
Univ Tasmania	12	21	23	22	17	27
Univ Cent Queensland	21	16	21	24	18	NA
Charles Sturt Univ	21	21	25	27	31	18
Flinders Univ of South Australia	21	11	17	19	13	25
Griffith Univ	21	27	21	12	23	9
James Cook Univ	21	27	29	29	NA	21
Murdoch Univ	21	27	31	26	29	18
Univ Ballarat	21	27	31	33	29	NA
Univ of Canberra	21	27	31	31	28	30
Univ Newcastle	21	21	17	20	27	15
Univ of South Australia	21	21	14	25	19	29
Univ Western Sydney	21	21	25	30	24	24
Univ Wollongong	21	16	25	28	26	22
Victoria Univ	21	8	14	16	10	NA

ERA 2010 denotes the ranking of the Excellence in Research for Australia in 2010 and the remaining rankings are based on cumulative citations for articles published in 2003, 2004, as well as 2003 and 2004. 2003–2004, 2003–2005, and 2003–2014 denote the window over which citations are cumulated. NA denotes that no economics article was published by this university in the respective year that is listed in the *Web of Science*

Table 5 Characteristics of universities with divergent ranks in the United Kingdom

	RAE 2008 FTE	Number of articles	Ratio	Citations of most cited paper	Field of most cited paper	Most common journal
Univ St Andrews	27.15	27	0.99	52	Finance	*Applied Econ, Econ Letts*
Univ Essex	34.31	51	1.49	57	Labor	*Econ J, J Econ Theory, Manchester School*
Univ Edinburgh	18	31	1.72	1509	Econometrics	*Economy and Society*
Univ Sheffield	15	48	3.20	424	Health	*Health Economics*
Univ E Anglia	13.5	49	3.63	286	Environmental	*World Development*
Univ Sussex	10	46	4.60	158	Health	*Industrial and Corporate Change, World Economy*

The table uses our data on 2003 and 2004 journal articles and the number of full-time equivalents that submitted to the RAE 2008 (RAE 2008 FTE). Ratio denotes the number of articles published per 2008 FTE in 2003 and 2004

academic do seem anecdotally high in environmental and health economics compared to more traditional economics fields. It could also be because these universities only submitted a narrow subset of their staff engaged in economics research to the economics unit of assessment. Edinburgh scored a very high number of citations from the article by Im et al. (2003). The other three universities' most cited papers are in environmental or health economics. Whereas the journals most published in at St Andrews and Essex are traditional economics journals, the journals most published in at the other universities are more specialized or interdisciplinary journals.

Of course, another important reason for divergence between our citation counts and the RAE is likely to be, as we mentioned above, that not all researchers publishing in the economics journals that we capture are affiliated with an economics department and were submitted to the economics unit of assessment in the RAE. This is likely to be especially true for Sussex and East Anglia, two universities where there are large non-economics academic units that produce some publications in economics.

There are fewer obvious outliers in Australia, and due to the longer period between 2003 and 2004 and the research assessment in 2010 there was more opportunity for universities to improve their performance in the interim. This certainly explains the high ranking awarded to UTS and anecdotally could explain some of the other apparent anomalies. For example, UTS recruited Michael Keene in 2006 (left for UNSW in 2011) on an ARC Federation Fellowship and John Geweke—currently the 2nd ranked economist in Australia according to RePEc—in 2009.

Discussion

Our results show that cumulative citations are highly predictable by early citations at the university discipline level in economics. However, the results of our citation-based rankings differ from those based on peer review, though to a lesser degree in Australia than in

the UK. One reason for this, suggested in the previous section, is that we do not normalize by citation potential. Some fields, such as environmental and health economics might have higher citation potential due to greater interest outside of economics. On the other hand, small fields such as economic history have lower citation potential. Source normalized impact factors are now available for journals (Moed 2010) and for universities as a whole (Waltman and van Eck 2013). An actual research assessment exercise could control for citation potential, perhaps by using regression analysis (Bornmann 2015).

Differences between our citation-based rankings and the peer-review rankings can be also explained by differences between the set of publications we analyzed and those used in the ERA 2010 and RAE 2008. First, we only use 2 years of data rather than 5 or 6. Second, universities have considerable leeway regarding the publications they submit in each discipline area. In Australia, the university can, for example, decide whether to submit an article under the economics field of research (FoR 14) or the policy and administration one (FoR 1605) or whether to submit another article under econometrics (FoR 1403) or statistics (FoR 0104).[6] In the UK, a researcher might, for example, be assigned to the economics unit of assessment (UoA 34) or the accounting and finance unit of assessment (UoA 35). We simply decide what is an economics article based on the journal it was published in and some of the journals we included, such as *Economic Geography* and a number of finance journals, would not usually be considered as core economics journals, though they are included in the economics category in the *Web of Science*. Third, we only count publications that are in journals included in the *Web of Science*. In Australia, all research publications published by a university in the field of research are submitted to the ERA. This includes articles in any peer-reviewed journal as well as books, book chapters, and conference papers. Universities have discretion over submitting "grey literature" such as government reports. There is also discretion over submitting the publications of affiliated faculty such as emeritus professors and visiting fellows. In the UK, only the four best publications of each included researcher are submitted and universities can decide not to submit researchers to the RAE. So, compared to the ERA we are probably less comprehensive in coverage but probably more comprehensive in coverage than the RAE. Finally, we assign articles to the universities to which their authors were affiliated at the time of publication. Both the ERA and RAE assign publications to the universities their authors were affiliated with at the time of assessment.[7]

We also computed and evaluated a number of quality indicators including the impact factor of each university's articles and the PP(top 10 %) indicator, which is the proportion (rather than number) of a university's articles published in the top 10 % most cited articles (Waltman et al. 2012). The rankings produced by these indicators were not very intuitive, with some small universities that are not usually regarded as highly ranked appearing ahead of more highly regarded universities such as LSE. We think that there are two reasons for this. First, the number of articles published by some of these universities in *Web of Science* journals is very low and sampling variability means that an occasional

[6] In ERA 2010 and 2012, publications assigned to four-digit fields of research (e.g. economic theory or econometrics) with less than fifty publications in total were not assessed. In ERA 2015, these were assessed as part of the two-digit field of research (e.g. economics) even though the four-digit field was not be assessed. This seems to be a move to reduce gaming of the system by assigning weak publications to four-digit codes that were then not assessed.

[7] RAE 2008 included publications published from 2001 to 2007 inclusively by researchers affiliated with eligible institutions on 31 October 2007 and included by their university in its submission. The 2010 ERA included publications published from 2003 to 2008 inclusively by researchers affiliated with eligible institutions on 31 March 2009.

highly cited article produced by these universities distorts these quality indicators. For example, de Montfort University only has 2 articles in our sample, which received a total of 68 citations through 2014. Second, universities with lower actual research quality publish many articles in journals that are not included in the *Web of Science*. This is especially true in the period we are considering when there were only about 160 economics journals included in the *Web of Science*. The articles included in the *Web of Science* are their better articles. Truncating their output in this way results in an upward-biased estimate of their average quality compared to universities with actually higher average research quality. This is the second main reason why many researchers oppose using citation analysis to assess research quality in the social sciences (Bornmann and Leydesdorff 2014). But as articles in *Web of Science* journals are cited much more on average than articles in other journals, these considerations will have much less effect on strength indicators.

Research assessments may change the structure of the economics discipline. Johnston et al. (2014) show that the total number of economics students increased in UK more rapidly than the total number of all students, but the number of departments offering economics degrees has declined, particularly in post-1992 universities. Also, the number of universities submitting to the REF under economics has declined sharply with only 3 post-1992 universities submitting in the latest round. This suggests that the REF has driven a concentration of economics research in the more elite universities in the UK—the question is whether that means that predicting ranks is easier in the UK than it is in other countries.

There has also been an increased emphasis on research in Australia. Neri and Rodgers (2015) investigate the output of top economics research by Australian academics from 2001 to 2010. They construct a database of 26,219 publications in 45 top journals finding that Australia's output, in absolute and relative terms, and controlling for differences in page size and journal quality, increased and, on a per capita basis, is converging to the levels of the most research-intensive countries. They also find that the historical dominance of the top four universities is diminishing. So the effects of a greater emphasis on research seem to be reversed in Australia. This is probably because, unlike the UK, so far very little spending is determined on the basis of research assessment. Neri and Rodgers also find that the correlation between the number of top 45 journal articles published in 2005–2010 and the ERA 2012 ranking is 0.83 (0.78 for 2003–2008 and ERA 2010).

Conclusions

In this paper, we have shown that when we aggregate individual articles across the economics discipline at a university, the university ranking based on cumulative citations for articles published in 2003 and 2004 can be easily predicted by the university ranking based on citations received until the end of 2004. Enlarging the citation window increases predictability further. Predictability is slightly higher for the UK than for Australia and higher for total citations than for the number of articles a university published among the top 10 % most cited articles. Therefore, if we are interested in research strength, and are prepared to measure that by cumulative citations received, we can certainly use early citations as a reliable predictor of the relative research strength of different universities in economics. The argument that citations accumulate too slowly in the social sciences to be of use is not true at the aggregate level of universities.

However, we found that measures of research quality we constructed using the *Web of Science* data did not produce intuitive results. This is likely because many publications of weaker universities in research terms are not included in the *Web of Science* and so average citation measures for these departments are biased upwards. Today, compared to our study period in 2003–2004, the number of social science journals included in the *Web of Science* has increased and *Scopus* includes even more social science journals. Therefore, a research assessment conducted today may be better able to compute quality measures if they were desired.

We find that our citation-based rankings have moderate correlations with the rankings implied by ERA 2010 and RAE 2008. The correlations are higher for Australia despite only a quality type ranking being available for Australia. The previous section discussed some reasons why our model research assessment will capture a different set of publications than were actually assessed in these exercises. Additionally, we noted that the citation potential of different fields within economics might differ and that this should be taken into account. But, as we noted in the section "Peer review and citation analysis in research assessment", some researchers are very skeptical of the quality of peer review in research assessment exercises. Peer review as conducted in such exercises is not necessarily superior to a well-constructed citation analysis, which our artificial exercise here is certainly not. We believe that the findings in this paper greatly strengthen the case put by Stern (2014) for greater use of citation analysis in research assessment in economics. We predict that similar results would be obtained also for political science given the similarity between the results for that discipline and economics at the article level (Stern 2014).

Acknowledgments We thank Guido Bünstorf and an anonymous referee for valuable comments and Andreas Rehs and Immanuel Bachem for helpful research assistance.

Compliance with ethical standards

Conflict of interest There is no conflict of interest.

References

Adams, J. (2005). Early citation counts correlate with accumulated impact. *Scientometrics, 63*(3), 567–581.

Anderson, D. L., & Tressler, J. (2014). The New Zealand performance based research fund and its impact on publication activity in economics. *Research Evaluation, 23*(1), 1–11.

Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., & Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy, 44*(2), 451–466.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45*, 199–245.

Bornmann, L. (2015). How much does the expected number of citations for a publication change if it contains the address of a specific scientific institute? A new approach for the analysis of citation data on the institutional level based on regression models. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23546.

Bornmann, L., & Leydesdorff, L. (2014). Scientometrics in a changing research landscape. *EMBO Reports, 15*(12), 1228–1232.

Clerides, S., Pashardes, P., & Polycarpou, A. (2011). Peer review vs metric-based assessment: Testing for bias in the RAE ratings of UK economics departments. *Economica, 78*(311), 565–583.

Colman, A. M., Dhillon, D., & Coulthard, B. (1995). A bibliometric evaluation of the research performance of British university politics departments: Publications in leading journals. *Scientometrics, 32*(1), 49–66.

Department for Business, Innovation & Skills and Johnson, J. (2015). *Press release: Government launches review to improve university research funding*. https://www.gov.uk/government/news/government-launches-review-to-improve-university-research-funding.

Farla, K., & Simmonds, P. (2015). *REF accountability review: Costs, benefits and burden*. Report by Technopolis to the four UK higher education funding bodies.

Gallo, S. A., Carpenter, A. S., Irwin, D., McPartland, C. D., Travis, J., Reynders, S., et al. (2014). The validation of peer review through research impact measures and the implications for funding strategies. *PLoS ONE, 9*(9), e106474.

HEFCE. (2015). The Metric Tide: Correlation analysis of REF2014 scores and metrics (Supplementary Report II to the Independent Review of the Role of Metrics in Research Assessment and Management). *Higher Education Funding Council for England*. doi:10.13140/RG.2.1.3362.4162.

Holgate, S. T. (2015). A comment on "Scientometrics in a changing research landscape". *EMBO Reports, 16*(3), 261.

Hudson, J. (2013). Ranking journals. *Economic Journal, 123*, F202–F222.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*, 53–74.

Johnston, J., Reeves, A., & Talbot, S. (2014). Has economics become an elite subject for elite UK universities? *Oxford Review of Education, 40*(5), 590–609.

Kenna, R., & Berche, B. (2011). Critical mass and the dependency of research quality on group size. *Scientometrics, 86*(2), 527–540.

Key Perspectives Ltd. (2009). *A comparative review of research assessment regimes in five countries and the role of libraries in the research assessment process: A pilot study*. Dublin: OCLC Research.

Levitt, J. M., & Thelwall, M. (2011). A combined bibliometric indicator to predict article impact. *Information Processing and Management, 47*, 300–308.

Moed, H. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265–277.

Mryglod, O., Kenna, R., Holovatch, Y., & Berche, B. (2013). Comparison of a citation-based indicator and peer review for absolute and specific measures of research-group excellence. *Scientometrics, 97*, 767–777.

Mryglod, O., Kenna, R., Holovatch, Y., & Berche, B. (2015). Predicting results of the Research Excellence Framework using departmental h-index. *Scientometrics, 102*(3), 2165–2180.

Neri, F., & Rodgers, J. (2015). The contribution of Australian academia to the world's best economics research: 2001 to 2010. *Economic Record, 91*(292), 107–124.

Norris, M., & Oppenheim, C. (2003). Citation counts and the research assessment exercise V: Archaeology and the 2001 RAE. *Journal of Documentation, 59*(6), 709–730.

Oppenheim, C. (1996). 'Do citations count? *Citation indexing and the research assessment exercise', Serials, 9*, 155–161.

Ostriker, J. P., Kuh, C. V., & Voytuk, J. A. (Eds.) (2011) *A data-based assessment of research-doctorate programs in the United States*. Committee to Assess Research-Doctorate Programs, National Research Council.

Regibeau, P., & Rockett, K. E. (2014). *A tale of two metrics: Research assessment vs. recognized excellence.* University of Essex, Department of Economics, Discussion Paper Series 757.

Sayer, D. (2014). *Rank hypocrisies: The insult of the REF*. Thousand Oaks: Sage.

Sgroi, D., & Oswald, A. J. (2013). How should peer-review panels behave? *Economic Journal, 123*, F255–F278.

Stern, D. I. (2014). High-ranked social science journal articles can be identified from early citation information. *PLoS ONE, 9*(11), e112520.

Süssmuth, B., Steininger, M., & Ghio, S. (2006). Towards a European economics of economics: Monitoring a decade of top research and providing some explanation. *Scientometrics, 66*(3), 579–612.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., Van Eck, N. J., et al. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*, 2419–2432.

Waltman, L., & van Eck, N. J. (2013). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics, 96*(3), 699–716.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics, 87*, 467–481.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics, 94*(3), 851–872.

Wang, D., Song, C., & Barabási, A.-L. (2013). Quantifying long-term scientific impact. *Science, 342*, 127–132.

Wilsdon, J., et al. (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management. *Higher Education Funding Council for England*. doi:10.13140/RG.2.1.4929.1363.

Wooding, S., van Leeuwen, T. N., Parks, S., Kapur, S., & Grant, J. (2015). UK doubles its "world-leading" research in life sciences and medicine in six years: Testing the claim? *PLoS ONE, 10*(7), e0132990.