UNIVERSITY OF WAIKATO
Hamilton
New Zealand

Citations or Journal Quality:
Which is Rewarded More in the Academic Labor Market?

John Gibson, David L. Anderson and John Tressler

Department of Economics

Working Paper in Economics 13/15

November 2015

Corresponding Author

John Gibson
Waikato Management School
University of Waikato
Private Bag 3105
Hamilton, 3240
NEW ZEALAND

Email: jkgibson@waikato.ac.nz
Phone:64-7-838-4289
Fax: 64-7-838-4331

David L. Anderson
School of Business
Queen's University
Kingston, Ontario K7L 3N6
CANADA

Email: dla@queensu.ca

John Tressler
Waikato Management School
University of Waikato
Private Bag 3105
Hamilton, 3240
NEW ZEALAND

Email: tressler@waikato.ac.nz

1

Abstract

Research quality can be evaluated using citations or from the prestige of the journal that publishes the research. Recent studies advocate for more weight on citations, which measure actual impact, while the journal where an article publishes is merely a predictor of whether it was thought likely to have an impact. Yet there is little comprehensive evidence on the role of citations versus journal quality in evaluating research. In this paper we use data on tenured economists in the University of California system to relate their salary to their lifetime publications of 5500 articles in almost 700 different academic journals and to the 140,000 citations to these articles. The results show little role for citations in affecting faculty salary, with an impact only one-seventh that of a measure of journal publications. The distribution of citations, whether using an h-index or the generalized h-index proposed by Ellison (2013), is also not a significant predictor of salary.

Keywords

academic salary
citations
h-index
journal rankings
research evaluation

JEL Codes

A14, J44

1. Introduction

The evaluation of research quality is important for funding agencies, publishers, and academic departments that make labor market decisions about hiring, promotion and tenure. The rise in performance-based university research funding systems (Hicks 2012) gives one impetus for systematically evaluating research quality. The competition between publishers to attract authors and library subscriptions provides another motive for measuring and publicizing research impact, especially of scholarly journals. For individual academics, the expected or revealed quality of their research has always had a profound impact on careers, as noted by Liebowitz and Palmer (1984, p.77) in one of the seminal studies of economics journal rankings:

> 'Where articles are published can affect one's promotion, tenure, and salary at one's present job; it can also affect one's brand name and the ability to change jobs.'

Notably, it is *where* the article is published that the quotation suggests has an effect. While the improved coverage, timeliness, and ease of use of citations databases is facilitating direct evaluation of individual research outputs, the use of journal rankings to proxy for the value of published research remains the basis for many assessments of the research outputs of economics departments and of individual economists (for example, Coupé 2003, Macri and Sinha 2006).

But there is a growing chorus of normative arguments for relying more on citations. First, the prestige of the journal that publishes an article should become less informative over time, since the editor decided to accept the article based on their prediction of whether it was likely to have an impact, while measures of actual impact, such as citations, become available as time elapses. This argument is formalized by Sgroi and Oswald (2013) who show how the build-up of citations gives new information, in a Bayesian modelling sense, and that as time passes the weight given to this information should rise while the weight given to the quality of the journal should dwindle towards zero. This falling weight on where an article publishes also recognizes that article quality varies within a journal; an important article may be published in a lesser ranked journal while some articles in highly ranked journals are ignored by the literature (Oswald, 2007). A further argument is that citations may be less affected by nepotism and unprofessional behavior; according to Liebowitz (2013, p.1):

> 'Because a journal's decision depends on the opinion of only an editor and a few referees chosen by the editor, there is a great deal of latitude for gratuitous decisions. By way of contrast, the number of citations that a paper receives is determined by the entire academy, likely reducing the influence of gratuitous behavior.'

Furthermore, the argument that citations accumulate too slowly in economics and social sciences to be of practical use for evaluating recent research – compared with in the natural sciences – is

also increasingly questioned since long-run citation counts in economics appear to be fairly predictable from early citations (Stern 2014, Burns and Stern 2015).

Despite normative arguments about citations there is little comprehensive evidence on their actual role compared with the role of journals. That is, the positive question of how much weight is given to citations versus to journal rankings is rarely studied, either in the context of research assessment exercises or in the labor market decisions made by academic departments. Peer evaluations like the Research Excellence Framework (REF) in the United Kingdom may ask panelists to ignore journal rankings and to treat each submission on its merits (where those merits may include citations) but reviewers will be generally aware of rankings and it is possible to *ex post* find the journal rankings that appear most similar to the peer evaluations (Anderson, Smart and Tressler, 2013). Some direct, albeit limited, evidence on citations versus journals is from a survey of the chairs of 47 economics departments (all but one in the United States) who were asked about the factors used to evaluate the articles of Associate Professors who were seeking promotion to full Professor; only nine chairs (19%) gave the citations to an article equal or more weight than they gave to the rank of the publishing journal (Liebowitz 2014).

In this paper we study one type of academic labor market outcome, the salary of tenured professors, to see what the market reveals about the information coming from citations versus the information from the journals where the articles are published. Specifically, we relate the salary of tenured economists in the University of California (UC) system to their lifetime publications of 5500 articles in almost 700 different academic journals, and to the 140,000 citations to these articles. The basis of our approach is that for research intensive universities, salary is likely to be directly related to perceived research quality, and so can help to sift between those explanatory variables that are derived from the journals where articles are published and those that measure the citations to the articles. We focus on the UC system as the largest research-intensive public university system in the United States (so there is good external validity) and because it has high quality salary data from a public disclosure database.

We build on the seminal contribution of Ellison (2013) in studying how the academic labor market uses citation data – while he looks at different citation indices we examine the broader question of citations versus journals. Of course, citations and publications have been used in prior studies of academic salary (which we review in Section 2) but not in the comprehensive manner used here. Specifically, we use nine different journal ranking schemes, allow for the quantity and quality of journal articles, consider citation distributions, compare highly-ranked with less highly-ranked departments, and search for the optimally weighted combination of citations and journals for articles of different vintages.

The results show little role for citations in affecting salary. Averaging over nine journal ranking schemes, a standard deviation increase in quality-adjusted journal output has a seven times larger impact on salary than does a standard deviation increase in citations. Non-nested tests favor models using journal output over those using citations, regardless of whether using total citation counts or indices that consider their distribution, such as the *h*-index (the author has written *h* papers that are each cited at least *h* times). In contrast, the measure of quality-adjusted journal output is always a highly significant predictor of salary for economists in the UC system. Searching for the optimal weights on citations and journals shows that up to 99% of the weight would go on journals and just one percent on citations. Even ignoring newer journal output while letting citations to older articles keep accumulating, as may occur under a very conservative rule of waiting for article-specific quality to be revealed by citations before revising salary, leads to just a slight rise in the revealed weight placed on citations.

These results are contrary to the normative argument for putting more weight on citations when evaluating research. It could just be that self-interested academics have a vested interest in ignoring information provided by citations (Liebowitz 2014). But it also may be that the signal provided by citations is not as strong as suggested by their proponents. Indeed, Sgroi and Oswald (2013) note exactly this; the more that citations are emphasized as a criterion for success the less useful they become because of the incentive to manipulate them. One manipulation is that editors may coerce authors to add citations to their journals; Wilhite and Fong (2012) find 175 coercing journals in their survey of researchers in economics, sociology, psychology and business. While instances of coercive citation may be reported, authors who pre-emptively expand bibliographies in response to an expected desire of editors and potential referees to be cited are harder to detect.[1] Spiegel (2012) shows the 'bibliographical bloat' that results from a focus on citation metrics – over a 30-year period the median number of articles cited by papers in the *Journal of Finance* almost quadrupled, from 13 in 1980 to 44 in 2010, while the length of Introductions where many of these citations are inserted more than quadrupled.

2.   Previous Literature

The relationship between academic salary for economists and various metrics of research output and quality has been studied since the 1970s. Research productivity is an important determinant of academic salary, so these metrics also are in models of a wide range of other phenomena, such

---

[1]   Cross-citation clubs can also be formed by individual authors and by journals. For example, *Technological and Economic Development of Economy* rose to third place in the economics category of the ISI Journal Citation Reports in 2010 (behind *JEL* and *QJE* and ten places ahead of the *AER*) by having 60% of citations coming from five journals from the same university and another 23 percent of citations from journals published by a neighboring university.

as: returns to seniority, co-authorship, returns to quality, journal rankings, and idea splitting.[2] This literature typically uses classifications of journals in broadly defined excellence tiers, and citation-based metrics, to represent research quality. In the same era, a related literature on journal rankings in economics, with corresponding quality weights, developed from the seminal work of Liebowitz and Palmer (1984) based on impact adjusted citation measures.[3] But few academic labor market studies use quality adjusted productivity measures that can be derived from these quality weights. In this section we review existing work on the relationship between research metrics and academic salary, to provide a background to the role of citations versus journal rankings in influencing labor market outcomes.

Hamermesh, Johnson and Weisbrod (1982) were the first to use citations to represent research quality in an academic salary model. They argued that '…one scholar's productivity should be measured by the sum of direct and indirect influences on other producers as well as by direct contributions (publications) (p.473).' Using five years of citation data from the Social Science Citation Index (now the *Web of Science* (WoS)) for 158 full professors of economics from seven large public universities, citations alone were shown to provide a useful measure of research productivity for six of the schools.

Following Ransom's (1993) suggestion that earnings fell with seniority, several studies of this apparent anomaly used a variety of indicators of research productivity, including simple counts of articles and books, journal tiers to represent excellence, and WoS citation counts.[4] In terms of the representation of research productivity, the most comprehensive of these studies was Bratsberg, Ragan and Warren (2010), who considered models where all articles listed in *EconLit* were classified into four quality tiers based on Scott and Mitias (1996) and others, quantity was represented using *AER* equivalent pages deflated by the number of authors, and a quadratic in the number of WoS citations, and book categories, were considered.

In addition to the use of citations, the practice of using journal excellence tiers to denote quality has been a feature of much of this literature.[5] Generally, the classification of journals

---

[2] See, for example, on seniority: Ransom (1993), Moore *et al*. (1998), Bratsberg, Ragan and Warren, (2003, 2010), and Hilmer and Hilmer (2011); co-authorship: Sauer (1988), Moore Newman and Turnbull (2001), Hilmer and Hilmer (2005); returns to quality: Sauer (1988); journal rankings Gibson (2000) and Gibson, Anderson and Tressler (2014); and idea splitting Gibson (2014).

[3] Widely used journal rankings and weights are provided by: Mason, Steagall and Fabritius (1997), Combes and Linnemer (2010), Coupé (2003), Kodrzcki and Yu (2006), Kalaitzidakis, Mamuneas and Stengos (2010) and Laband and Piette (1994).

[4] For example, Moore, Newman and Turnbull (1998), Bratsberg, Ragan and Warren (2003) and Hilmer and Hilmer (2011).

[5] Hilmer and Hilmer (2005), Moore, Newman and Turnbull (2001), Hilmer, Hilmer and Lusk (2012) and Hilmer, Ransom and Hilmer (2015) represent further examples.

into tiers is based on informal attempts to represent views on broad journal ranks at a point in time. Within each of these tiers all journals are treated as being of equal quality. This contrasts with the journal ranking literature, where formal processes are used to rank and weight journals based on citations, impact-adjusted citations, perception surveys, or labor market models. To illustrate implications of quality tiers, consider the top two tiers of journals as classified by Bratsberg, Ragan and Warren (2010), and two widely used journal rankings, the Combes and Linnemer (2010) medium convexity CLm scheme based on citation measures and the Kalaitzidakis, Mamuneas and Stengos (2010) KMS scheme with weights assigned based on impact-adjusted citations.

In the CLm scheme, journals included in the Bratsberg et al (2010) top tier have quality weights that vary between 100 for the *QJE* to 27.2 for *Economica*, or if *Economica* is excluded between 100 and 54.7 for the *International Economic Review*, and in the KMS scheme between 100 for the *AER* and 2.2 for *Economica*, or between 100 and 12.4 for the *International Economic Review* if *Economica* is excluded. For the second tier, the CLm weights range from 80.6 for the *Journal of Financial Economics* to 13.6 for the *Journal of Regional Science*. Under the KMS scheme, several journals in the Bratsberg *et al.* second tier are not ranked, and of the ranked journals, the weights range from 27.8 for the *Journal of Monetary Economics* to 0.4 for the *Journal of Regional Science*. Irrespective of views on differences in journal weighting schemes and the appropriate convexity of the weightings, it is clear that a lot of information is lost through the treatment of all articles in a tier as representing the same research quality level.

Sauer (1988) is a rare example of using a journal weighting scheme to study academic labor markets. He applies Liebowitz and Palmer (1984) journal weights to standardised pages and citation counts, and estimates the convexity of the journal weights that is best supported by the data. Both quality adjusted pages and total citations are shown to be important research productivity indicators. Sauer uses these estimates to determine a quality gradient based on returns to journal outputs and the resulting expected citations. Relatedly, Gibson, Anderson and Tressler (2014) determine which of nine commonly used journal rankings schemes best fit labor market returns for economists in the University of California system. Estimates show that more inclusive schemes, where weights do not decline sharply with the rank of the journal, such as the Combes and Linnemer CLm scheme, have the greatest congruence with academic salaries. A similar modelling framework is used by Gibson (2014) to show that there is a significant positive effect on salary from publishing more articles, conditional on the total number of size-, quality-, and co-author-adjusted pages ever published, i.e. there is an incentive for idea-splitting.

Three recent papers using citation based metrics as measures of research quality are directly related to the current paper. Hamermesh and Pfann (2012) consider how research quantity and quality affect reputation and salary. Indicators of reputation are based on awards

and honours, and rankings of departments. Simple publication counts and WoS citations are used to represent research quantity and quality. The estimates suggest that research quantity has a negative impact on reputation, but a positive impact on salary. To explain this anomaly, Hamermesh and Pfann (2012) report that a survey of departments indicates that all had access to faculty publication records, but less than one-fifth collected citations data, and, moreover, estimates of the determinants of salary are not significantly different for the schools that collect citation data. Thus, publications records may allow departments to use journal prestige as sufficient indicators of research quality.

Ellison (2013) considers variants of the citation based h-index of Hirsch (2005), and introduces a class of generalized h-like measures, $h_{(a,b)}$ defined as the number h such that one has at least h papers with ah^b citations. Using the rank of the departments in which economists are located to indicate labor market outcomes, Ellison finds that, using *Google Scholar*, a wide range of values of the parameters a and b outperform $h=h_{(1,1)}$ and suggests $h_{(15,2)}$ as a reasonable indicator for younger scholars and $h_{(5,2)}$ for a larger sample that includes seasoned academics. Using $h_{(15,2)}$ rather than $h_{(1,1)}$ concentrates attention on a much smaller range of publications for most researchers. A potential caveat to Ellison's results is that *Google Scholar* includes a wider range of research outputs, and far higher citations counts than those from WoS, so it is possible that for *Google Scholar* concentrating counts on a smaller number of outputs is an advantage.

Hilmer, Ransom and Hilmer (2015) use the rank of departments and salary to indicate labor market outcomes, and a variety of research productivity indicators including total WoS citations, the original h-index, generalized $h_{(a,b)}$ measures, and the number of articles in the quality excellence tiers proposed by Scott and Mitias (1996). The article counts are not adjusted for length, and include non-refereed material (e.g. editorials and conference proceedings). A model with only the h-index and its square explains more that 52 percent of the variance in log salaries in their data. In more comprehensive models, the h-index outperforms total citation counts and the $h_{(15,2)}$ index proposed by Ellison. In the three preferred models, only articles in the 'elite' tier of Scott and Mitias are statistically significant in one case, 'elite' and 'excellent' are in a second case, and 'excellent' and 'other' articles are significant in a third model. The use of excellence tiers poses the same problems as noted above for Bratsberg, Ragan and Warren (2010). Thus this research leaves open the question of how the h-index and total citations might perform when more complete representations of journal quality are used.

3. Data Description

The sample is all Professors and Associate Professors in economics departments of University of California (UC) campuses, except those with adjunct, affiliate, or part-time positions and those with primarily teaching or administrative roles. This group of $n=167$ is a subset of the sample

used by Gibson, Anderson and Tressler (2014); in contrast to that study, Assistant Professors are excluded because many have not accumulated WoS citations. The findings are, thus, most germane for comparing the role of citations and of journals on salary of seasoned scholars with tenure. The focus is on the UC system because of external validity – this is the largest research-intensive public university system in the United States – and also due to the high quality salary data in a public disclosure database (http://www.sacbee.com/statepay/). Briefly, the dependent variable is the (log of) base salary for the 2010 academic year, and the regressions include a dummy for individuals not on a standard 9-month academic year and pay scale (a few Berkeley economists are on law school scales). Other control variables include quadratics in seniority and experience, and dummies for gender, whether holding a named chair, whether a Nobel Prize winner, and fixed effects for each UC campus.

The citations are to the lifetime articles published by these 167 tenured professors up to the end of 2010. The articles were found by searching *EconLit*, RePEc, WoS, and curriculum vitae. In total there were about 5,500 articles in almost 700 different academic journals, and 140,000 citations to these articles. The citations are from the WoS, which is the most established citations database and is stricter in coverage than others such as *Google Scholar*, which includes citations to and from a variety of unpublished works. Some journals that the UC economists published in were not covered by WoS at the time (noting that coverage rises over time), and so citations to those articles appear to be zero even if they may have citations in other databases. We include an indicator for the proportion of articles that were published by each economist in journals that were not in WoS, at the time, so as to account for this potential under-coverage.

In order to compare citations with a comprehensive measure of where articles publish we need journal rankings and weights to convert output to a constant quality (set at the level of the highest ranked journal, which is the *QJE* for most ranking schemes). A wide range of journal ranking and weighting schemes have been proposed by economists, with no consensus on which is best. We therefore use nine different schemes to ensure that results do not depend on the particular weights used to calculate quality-adjusted journal output. The full descriptions for each scheme are in Gibson *et al*. (2014), with their brief details as follows:

- Mason, Steagall and Fabritius: [MSF] reputational weights for 142 journals from a survey of economics department chairs. This is the least aggressive in down-weighting lower ranked journals but excludes many economics journals.

- Coomes and Linnemer: [CLm, CLh] is the most comprehensive, covering 1168 journals by using a *Google Scholar h*-index to extrapolate from citations for *EconLit* journals to all journals. They use two different rates of down-weighting lower ranked journals, with their medium variant (CLm) the second least aggressive, and their high variant (CLh) the fifth most aggressive of the nine schemes used here.

- RePEc is an impact factor from unweighted citations, covering 984 journals (when we collected the data in May 2012), and is the fourth least aggressive of the nine schemes.

- Coupé is an average of 2-year impact factors for 1994-2000 from the *Journal Citation Reports* for 273 economics journals; this is the third least aggressive of the nine schemes.

- Kodrzycki and Yu: [K&Y_all, K&Y_econ] is an 'eigenfactor' approach where a journal is deemed influential if cited often by other influential journals. Sub-discipline citing intensity is adjusted for, with cites from all social science journals [K&Y_all] and just from economics [K&Y_econ]. These are the third and fourth most aggressive in down-weighting lower ranked journals.

- Kalaitzidakis, Mamuneas and Stengos: [KMS] is an eigenfactor approach, using the average of citations each year from 2003-2008 to articles published in the previous 10 years. This is the second most aggressive scheme, and ranks 209 economics journals.

- Laband and Piette: [LP] is an eigenfactor approach using citations to economics journals over 1985-89 by articles published in 1990. This is the least permissive, covering just 130 journals, and most aggressively down-weights lower ranked journals.

Using each of these sets of weights (which include zero for unranked journals) the pages for each article ever published by sample members are multiplied by each journal's quality weight, adjusting for the number of authors and standardizing to the size of a typical page in the *AER*:

$$Article\ Pages \times Size\ Correction \times (1/number\ of\ authors) \times Journal\ Assessment\ Weight.$$

The definitions and summary statistics for our various citations metrics, our measures of lifetime (to 2010) quality-adjusted journal output, and the control variables are in Table 1. The lifetime journal output ranges from an average of 44 pages using LP journal weights (which are the most discriminating) to 174 pages using MSF weights (the least discriminating), from an average of 33 articles published in the career to date. The total citations to these articles for each economist average just over 800, with a maximum of over 15,000, while the most highly cited article per economist averages 200 citations. The average *h*-index for these academics is just over 11, while the generalized *h*-index proposed by Ellison (2013) as a better predictor of labor market outcomes in economics averages just 2.5. This $h_{(a,b)}$ index is that one has *h* articles each cited ah^b times, and we use $h_{(5,2)}$ because Ellison estimates *a*=5 and *b*=2 for a sample from top-50 departments who average 22 years post-PhD (about the same as the current sample). Ellison fitted this generalized *h*-index using *Google Scholar*, whose permissiveness in counting citations from working papers (and other documents on the internet) to working papers in addition to articles is well-known. For example, compared to the similarly-aged current sample, his sample has many more items and citations (and a higher *h*-index), averaging 67 papers, 4,500 citations and an *h*-index of 22.[6]

---

[6] Another example of *Google Scholar* permissiveness, that avoids issues from differences in sample composition, is that the lead author of the current paper has an *h*-index of 35 in *Google Scholar* but of

Table 1: Variable Definitions, Means and Standard Deviations

Variable	Mean	Std Dev	Description
Salary	173.35	55.89	Base salary in 2010 ($,000)
Number of articles	32.57	24.80	Career articles (in *EconLit*, *RePEc*, or *Web or Science*)
Citations to those articles	827.47	1464.55	Total WoS citations to those articles to end of 2010
Citations to most-cited	201.08	49.22	Total WoS citations to the most cited article
h-index	11.33	6.81	Economist has *h* articles with *h* or more citations
$h_{(5,2)}$-index	2.53	1.21	Generalized *h*-index; *h* articles with $5h^2$ citations
Not-ISI	0.12	0.12	Share of journals not in WoS (so zero cites recorded)

Quality-, size and co-author-adjusted journal output (lifetime AER-sized pages) using journal weights from:

Variable	Mean	Std Dev	Description
MSF	174.47	136.14	Mason, Steagall and Fabritius reputational ranking
CLm	129.41	98.33	Combes-Linnemer medium convexity weights
CLh	85.35	71.88	Combes-Linnemer high convexity weights
RePEc	79.98	71.05	RePEc Simple Impact Factor
Coupé	55.89	49.10	Average of 2-year impact factors for 1994-2000
K&Y_all	48.73	46.21	Kodrzycki and Yu eigenfactor ranks, cites from all journals
K&Y_econ	45.15	42.78	Kodrzycki and Yu ranks, cites just from econ journals
KMS	47.05	41.35	Kalaitzidakis, Mamuneas and Stengos eigenfactor weights
LP	44.11	42.15	Laband and Piette eigenfactor weights

Control Variables

Variable	Mean	Std Dev	Description
Experience (years)	22.66	10.86	Years since first appointment
Seniority (years)	15.07	9.89	Years of employment at current university
Male	0.86	0.35	Person is male (=1) or female (=0)
Holder of a named chair	0.26	0.44	Has endowed or named position or a distinguished chair
Not standard pay scale	0.04	0.19	Not on a standard, 9-month, academic year pay scale
Nobel prize winner	0.01	0.08	Winner of the Nobel Prize

Note: N=167.

just 16 in *Web of Science*. It is an open question whether Ellison's choice of generalized *h*-index would be as strongly skewed towards highly cited papers if he had used citations and articles in *Web of Science* rather than those in *Google Scholar*.

4. Results

The academic salary equations are reported in Table 2 and these explain about three-quarters of the variation in log salary for these economists. All nine regression equations show a highly significant (at $p<0.01$ level) positive effect on salary of quality-adjusted journal output. In contrast, the total citations to articles published over the career (to the end of 2010) are significant in just three of the nine salary regressions, using the journal weighting schemes from Kodrzycki and Yu, and from KMS. For the other equations there is no significant role for citations in explaining salary, conditional on the journals that published the articles. The lack of independent impact for citations is especially apparent when journals are ranked using the Combes and Linnemer medium convexity journal weights, which provide the highest predictive power for any equation in Table 2.[7]

To enable a direct comparison of the magnitudes of these effects the regressions were re-estimated with standardized variables. The (conditional) effect of a standard deviation rise in the quality-weighted journal output is to increase salary by between 21 and 35 log points; being smallest when using the KMS weights to compare journals and largest when using CLm weights. In contrast, a standard deviation increase in the number of citations is associated with salary that is between zero and 10 log points higher, with the smallest effect found if using CLm weights and the largest effect found if using KMS weights. Taking the mean of the standardized coefficients across all nine equations, and calculating the impact on salary, the magnitude of the citation impact appears to be just one-seventh of the magnitude of the impact of the journals that publish the cited articles.[8]

The other result reported in Table 2 is that if an econometric 'horse race' is run between models that explain salary either by using quality-adjusted journal output or by citation counts to that output (with the control variables common to the two competing models), a Vuong (1989) non-nested test significantly favors the journal articles model in seven of the nine specifications and is just outside conventional significance levels in favoring the journals model in another salary equation. The Vuong test does not presume that either competing model is 'true', and instead determines which has verisimilitude (that is, is closer to the truth). Thus we infer from the results in Table 2 that a model putting weight on journal publications rather than on citations to those publications is closer to the true model of departmental and university decision-making that produces the faculty salaries that we observe. We note that no equation has a Vuong test statistic that suggests that the citations model would be favoured over the journals model (negative values of the test statistic would occur in this case).

---

[7] Model comparison and non-nested tests for a larger sample of UC economists that included Assistant Professors also showed that the CLm journal weights were the ones most congruent with the salary data (Gibson *et al.* 2014).

[8] The comparison of standardized impacts is on salary in dollars rather than in logs, using Duan's (1983) smearing estimator to calculate predicted salary in dollar terms.

Table 2: Salary Regressions for Tenured UC Economists using Lifetime Quality-Adjusted Journal Output and WoS Citations

	Journal Weighting Scheme for Calculating Quality-Adjusted Journal Output Over Lifetime Comes From:								
	MSF	CLm	CLh	RePEc	Coupé	K&Y_all	K&Y_econ	KMS	LP
Quality-adjusted journals	0.008	0.013	0.014	0.014	0.020	0.020	0.021	0.017	0.022
	(5.36)**	(5.71)**	(3.98)**	(4.89)**	(4.79)**	(4.79)**	(4.73)**	(3.32)**	(3.48)**
Number of citations (×1000)	0.008	0.001	0.007	0.013	0.015	0.017	0.017	0.021	0.009
	(0.82)	(0.08)	(0.74)	(1.55)	(1.73)	(2.02)*	(2.03)*	(2.40)*	(0.96)
Seniority (years)	-0.017	-0.018	-0.020	-0.020	-0.018	-0.020	-0.020	-0.021	-0.021
	(2.89)**	(3.10)**	(3.13)**	(3.18)**	(2.90)**	(3.07)**	(3.10)**	(3.14)**	(3.26)**
Seniority squared (/100)	0.032	0.035	0.034	0.036	0.031	0.034	0.034	0.035	0.035
	(1.85)	(2.12)*	(1.95)	(2.00)*	(1.74)	(1.87)	(1.89)	(1.91)	(1.92)
Experience (years)	0.025	0.024	0.029	0.028	0.028	0.030	0.030	0.032	0.033
	(4.15)**	(3.86)**	(4.80)**	(4.72)**	(4.62)**	(5.13)**	(5.09)**	(5.28)**	(5.28)**
Experience sq (/100)	-0.043	-0.040	-0.046	-0.045	-0.047	-0.048	-0.047	-0.049	-0.052
	(3.47)**	(3.26)**	(3.67)**	(3.60)**	(3.59)**	(3.81)**	(3.80)**	(3.93)**	(3.92)**
Male	0.040	0.025	0.036	0.037	0.050	0.041	0.041	0.043	0.046
	(0.86)	(0.55)	(0.75)	(0.79)	(1.03)	(0.86)	(0.86)	(0.85)	(0.95)
Holds a named chair	0.097	0.082	0.084	0.092	0.096	0.090	0.091	0.096	0.082
	(2.88)**	(2.52)*	(2.41)*	(2.74)**	(2.83)**	(2.64)**	(2.68)**	(2.58)*	(2.33)*
Nobel prize winner	0.529	0.543	0.468	0.475	0.495	0.454	0.452	0.414	0.426
	(6.57)**	(7.23)**	(6.43)**	(6.63)**	(6.71)**	(6.54)**	(6.53)**	(5.78)**	(6.28)**
Not standard pay scale	0.176	0.177	0.163	0.175	0.174	0.180	0.179	0.145	0.147
	(3.08)**	(3.17)**	(2.53)*	(2.72)**	(2.73)**	(2.58)*	(2.57)*	(1.96)+	(2.19)*
Constant	11.794	11.791	11.768	11.787	11.762	11.764	11.768	11.777	11.768
	(133.10)**	(131.49)**	(127.39)**	(129.49)**	(127.19)**	(126.35)**	(126.83)**	(126.53)**	(123.29)**
R^2	0.732	0.737	0.717	0.718	0.721	0.715	0.714	0.700	0.707
Vuong non-nested test[a]	2.85**	3.18**	2.29*	2.37*	2.22*	2.01*	2.02*	0.97	1.83

Notes:

Dependent variable is log of base salary for the 2010 academic year, as reported at: http://www.sacbee.com/statepay/ Fixed effects for each UC campus (with UC Berkeley the omitted category) and an indicator for the proportion of published articles not in WoS journals (so citations appear zero) are not reported.

N=167 associate and full professors, robust *t* statistics in parentheses, * significant at 5%; ** at 1%.

In Table 3 we report the results of two sensitivity analyses that are designed to expand upon the comparisons between citation-based measures of research impact and measures based on the journals publishing the research. The first sensitivity analysis adds an extra covariate – the count of articles ever published by each academic. An earlier study using a larger dataset of UC economists shows that citations are higher for academics who publish more articles, conditional upon their total number of co-author-, size-, and quality-adjusted pages (Gibson 2014). In other words, academics who split their published research into more, shorter, articles get cited more than otherwise similar academics who write fewer, longer, articles where both academics end up with the same total output in terms of quality-adjusted pages.[9] Thus in Table 2 the citation counts may act as a proxy for an excluded relevant variable – the total number of articles published. Indeed, if the count of articles is added to the model, the citations variable becomes statistically insignificant in all nine of the salary equations (panel A of Table 3), and the coefficients on citations fall in magnitude by about 60 percent. In contrast, quality-adjusted journal output stays highly significant ($p<0.01$) in all nine equations, with the coefficient on journals falling by less than one-quarter when the count of articles is included as a regressor (while article counts are significant in eight of the nine salary equations).

The second sensitivity analysis exploits the quality variation amongst the nine campuses of the UC system that have economics departments. Liebowitz (2014) finds that in lower ranked departments, decisions about senior promotions rely less on perceived quality of an applicant's articles formed by colleagues reading them, and more on where articles are published and their citations, compared with for the average department in his small ($n=47$) sample.[10] The top four economics departments in the UC system are Berkeley, San Diego, Davis and UCLA and the results of estimating salary equations just for the faculty in these four departments are reported in panel B of Table 3, with the results for the other, lower ranked, economics departments reported in panel C. The tests for differences in the coefficients between the sub-samples of top-ranked and lower ranked departments are in panel D and these suggest that citations have larger effects on salary in the lower ranked departments (albeit with most of the differences imprecisely estimated). In contrast there is no variation across the two sub-samples in the impact of journals on salary. In general, both the top-ranked departments and the other departments pay salaries that seem to vary more with where articles are published and less with how many citations there are to those articles. But if one fixes attention just on the citations, there is weak evidence that these seem to be a little more influential on salary in the lower ranked departments, which is consistent with what Liebowitz argues for.

---

[9] This effect is presumably because there is more chance for one's research to be noticed, and then cited, if it is spread more widely over different journals or over different issues of the same journal.

[10] Liebowitz considers this to be efficient, since faculty in lower ranked departments are presumed less competent to judge article quality than are faculty in higher ranked departments.

Table 3: Sensitivity Analyses for Salary Regressions using Lifetime Quality-Adjusted Journal Output and WoS Citations

	Journal Weighting Scheme for Calculating Quality-Adjusted Journal Output Over Lifetime Comes From:								
	MSF	CLm	CLh	RePEc	Coupé	K&Y_all	K&Y_econ	KMS	LP
	A.	*Including the Count of Articles as a Covariate*							
Quality-adjusted journals	0.006	0.010	0.011	0.009	0.013	0.014	0.015	0.012	0.020
	(3.52)**	(4.13)**	(3.51)**	(3.64)**	(3.03)**	(3.93)**	(3.89)**	(2.70)**	(3.67)**
Number of citations (×1000)	0.007	0.000	-0.000	0.008	0.011	0.007	0.008	0.009	-0.008
	(0.79)	(0.04)	(0.02)	(0.93)	(1.31)	(0.83)	(0.85)	(0.91)	(0.74)
Number of articles (×10)	0.016	0.017	0.030	0.026	0.024	0.030	0.030	0.034	0.037
	(1.92)	(2.20)*	(4.41)**	(3.42)**	(3.01)**	(4.11)**	(4.08)**	(4.47)**	(5.42)**
R^2	0.735	0.741	0.737	0.732	0.730	0.734	0.734	0.728	0.740
	B.	*Using the Sub-sample from the Top Four Departments (Berkeley, UCSD, UCLA, UC Davis)*							
Quality-adjusted journals	0.009	0.013	0.014	0.014	0.022	0.019	0.021	0.014	0.025
	(4.90)**	(4.61)**	(3.35)**	(4.27)**	(4.71)**	(4.36)**	(4.27)**	(2.39)*	(3.27)**
Number of citations (×1000)	0.007	0.002	0.009	0.013	0.015	0.017	0.017	0.024	0.004
	(0.59)	(0.21)	(0.81)	(1.24)	(1.41)	(1.67)	(1.68)	(2.26)*	(0.38)
R^2	0.715	0.714	0.690	0.699	0.707	0.694	0.694	0.668	0.690
	C.	*Using the Sub-sample from Economics Departments at the Other (Lower Ranked) UC Campuses*							
Quality-adjusted journals	0.008	0.014	0.020	0.015	0.028	0.023	0.026	0.026	0.028
	(2.46)*	(2.70)**	(2.04)*	(2.29)*	(2.27)*	(1.83)	(1.97)	(1.80)	(1.56)
Number of citations (×1000)	0.009	-0.001	0.031	0.044	-0.016	0.056	0.054	0.053	0.078
	(0.19)	(0.01)	(0.66)	(1.04)	(0.23)	(1.23)	(1.20)	(1.22)	(2.16)*
R^2	0.713	0.723	0.709	0.694	0.703	0.686	0.687	0.694	0.694
	D.	*Interaction Effects between Top 4 Indicator and Quality-Adjusted Journals and Citations (i.e. test of B-C differences)*							
Quality-adjusted journals	-0.001	-0.002	-0.007	-0.000	-0.010	-0.005	-0.007	-0.008	-0.007
	(0.20)	(0.44)	(0.77)	(0.03)	(0.86)	(0.41)	(0.52)	(0.69)	(0.41)
Number of citations (×1000)	-0.025	-0.025	-0.051	-0.073	-0.003	-0.069	-0.066	-0.060	-0.099
	(0.50)	(0.51)	(1.06)	(1.60)	(0.04)	(1.48)	(1.46)	(1.44)	(2.46)*

Notes:

Each regression also includes all of the other variables from Table 2. *N*=167, robust *t* statistics in parentheses, * significant at 5%; ** at 1%.

The results in Table 2 and 3 do not consider vintage effects, with quality-adjusted journal output aggregated into a single measure, regardless of when articles published. The argument by Sgroi and Oswald (2013) about article vintage when evaluating research is that more weight over time ought to be put on the accrued citations and less weight put on the journal where the article published. Based on this reasoning, one could think of adopting a highly conservative rule for academic labor market decisions, of letting article-specific quality be revealed by citations before committing to costly salary decisions. Indeed, the tenure system where new hires have six years to prove the quality of their research before being offered a permanent appointment might be broadly thought of in these terms. We are not advocating this as a good description of how labor market decisions in economics departments are actually made, since any department that tried such a rule would likely be forced into action by a faculty member who scored a solid hit, such as publishing an article in a top-5 journal, and threatened to leave for another department that makes hiring and promotion decisions based on *expected* research impact without waiting for citations to accrue. Instead, we think of this as a scenario that should give citations their best possible chance to show an influence on academic salary. If little such influence is found even under these highly favorable assumptions, the evidence should count more firmly against the view that citations matter a lot for academic labor market decisions.

In order to explore how vintage effects might alter the role of citations versus journals in informing academic labor market decisions, we recalculated the measure of lifetime production in journals by ignoring articles published in either the prior three or prior six years. However for the articles published prior to those cut-offs (that is, by the end of 2004 or the end of 2007) we count all citations that had accrued by the end of 2010. This calculation attempts to mimic the scenario of a conservative rule where one waits for the quality of an article to be revealed by citations before making labor market decisions. The six year period matches the typical length of time that new hires have on a 'tenure clock' while mid-term reviews at three years are when junior hires who are not making apparent progress may consider moving to a lower ranked department and resetting their tenure clock. The six year time period also corresponds to the typical window used for research assessment exercises like the REF in the United Kingdom. According to Sgroi and Oswald (2013) the earliest articles in that window are ones where the greatest weight should be placed on their accrued citations, and the least weight on where they were published. Such articles would have been published six years prior, corresponding to the cut-off we have set.

The results of re-estimating the salary equations with these new information sets are in panels B and C of Table 4. The baseline results from Table 2 that used all available information on citations and publications to the end of 2010 are repeated in panel A of Table 4 for ease of comparison. Even under unrealistic assumptions about academic labor markets that are highly favorable for maximizing the apparent role of citations, we find that salaries are much more

sensitive to the lifetime record of output in journals (omitting more recent articles) than they are to the accrued citations to those journal articles. Our measure of (truncated) career output in journals remains highly significant in all nine salary equations even when it is calculated by omitting articles published in the previous three or six years. In contrast, the accumulated stock of citations is statistically significant in just four (six) salary equations when omitting journal articles from the prior three (six) years; this is little changed from citations being statistically significant in three salary equations in the baseline results. Moreover, the Vuong non-nested tests still strongly favor the journal articles model over the citations model; there are statistically significant test results for four salary equations and four others are just outside usual significance levels ($p<0.06$) when journal articles from the last three years are ignored.[11] In terms of magnitudes, the standardized coefficients for journals range from 0.23 to 0.40 if the last three years are ignored and from 0.20 to 0.35 if the last six years are ignored. In contrast, the standardized coefficients on citations range from just 0.01 to 0.09 (or from 0.04 to 0.11) if the last three (six) years are ignored. Averaging across the nine salary equations, and using the most favorable assumption for maximizing the role of citations, which is that information on journal articles published in the last six years is ignored but citations to earlier articles that accrue over that period continue to be counted, it is still the case that citations have an impact on salary that is less than one-third of the impact of where an article is published.

We also form weighted averages of lifetime citations and of the lifetime quality-adjusted journal output, in the spirit of Sgroi and Oswald (2013) that research evaluators should use such an approach. Our objective is to find the optimal weights on citations and on journals, in terms of those that maximise the log-likelihood of the salary equations. Since we have nine different journal ranking schemes we just use the two extremes; the CLm weights that best fit the UC salary data and always show the least effect for citations, and the KMS weights that always give the highest effects for citations. Using CLm weights first, and then KMS weights, we re-estimate the salary equations 101 times, incrementally decreasing the weight on citations in a weighted average measure of standardized citations and standardized journal output. The maximized log-likelihoods from this search procedure are illustrated in Figures 1a (CLm weights) and 1b (KMS weights). This iterative search procedure is repeated using the measures of lifetime journal output that omit the articles published in the last three and six years (while citations to earlier articles continue to accumulate), corresponding to the analysis in Table 4.

---

[11] The Vuong tests become less precise if journal articles from the last six years are ignored, with the strongest results for the CLm weights ($p<0.02$), the MSF weights ($p<0.06$) and the CLh weights ($p<0.08$). No Vuong test results favor the citations model over the journals model.

Table 4: Salary Regressions Comparing Impact of Journal Articles and WoS Citations for Articles of Different Vintages

	MSF	CLm	CLh	RePEc	Coupé	K&Y_all	K&Y_econ	KMS	LP

Journal Weighting Scheme for Calculating Quality-Adjusted Journal Output Over Lifetime Comes From:

	MSF	CLm	CLh	RePEc	Coupé	K&Y_all	K&Y_econ	KMS	LP
A. Baseline from Table 2, including Articles Published Until End of 2010									
Quality-adjusted journals	0.008	0.013	0.014	0.014	0.020	0.020	0.021	0.017	0.022
	(5.36)**	(5.71)**	(3.98)**	(4.89)**	(4.79)**	(4.79)**	(4.73)**	(3.32)**	(3.48)**
	[0.35]	[0.39]	[0.32]	[0.30]	[0.30]	[0.28]	[0.28]	[0.21]	[0.28]
Number of citations (×1000)	0.008	0.001	0.007	0.013	0.015	0.017	0.017	0.021	0.009
	(0.82)	(0.08)	(0.74)	(1.55)	(1.73)	(2.02)*	(2.03)*	(2.40)*	(0.96)
	[0.03]	[0.00]	[0.03]	[0.06]	[0.07]	[0.07]	[0.08]	[0.10]	[0.04]
R^2	0.735	0.741	0.737	0.732	0.730	0.734	0.734	0.728	0.740
B. Omitting Articles Published in Prior Three Years (But Citations for Older Articles Counted to End of 2010)									
Quality-adjusted journals	0.008	0.014	0.016	0.013	0.020	0.020	0.021	0.019	0.022
	(5.10)**	(5.67)**	(4.14)**	(4.64)**	(4.41)**	(4.41)**	(4.38)**	(3.58)**	(3.31)**
	[0.34]	[0.40]	[0.34]	[0.28]	[0.28]	[0.27]	[0.26]	[0.23]	[0.27]
Number of citations (×1000)	0.010	0.001	0.006	0.014	0.017	0.017	0.017	0.020	0.010
	(1.16)	(0.13)	(0.60)	(1.85)	(2.05)*	(2.05)*	(2.15)*	(2.28)*	(0.96)
	[0.04]	[0.01]	[0.03]	[0.06]	[0.08]	[0.08]	[0.08]	[0.09]	[0.04]
R^2	0.729	0.737	0.720	0.716	0.717	0.714	0.713	0.706	0.707
C. Omitting Articles Published in Prior Six Years (But Citations for Older Articles Counted to End of 2010)									
Quality-adjusted journals	0.008	0.013	0.014	0.012	0.017	0.016	0.017	0.018	0.019
	(4.38)**	(4.76)**	(3.73)**	(3.90)**	(3.72)**	(3.66)**	(3.67)**	(3.07)**	(2.79)**
	[0.30]	[0.35]	[0.28]	[0.23]	[0.23]	[0.20]	[0.20]	[0.20]	[0.22]
Number of citations (×1000)	0.016	0.010	0.014	0.022	0.023	0.024	0.024	0.024	0.018
	(1.80)	(0.98)	(1.45)	(2.69)**	(2.80)**	(2.95)**	(3.02)**	(2.70)**	(1.98)*
	[0.07]	[0.04]	[0.06]	[0.10]	[0.10]	[0.11]	[0.11]	[0.11]	[0.08]
R^2	0.708	0.713	0.700	0.695	0.696	0.692	0.692	0.690	0.690

Notes:

Each regression also includes all of the other variables from Table 2. N=167, robust t statistics in parentheses, * significant at 5%; ** at 1%. The estimated impact of a standard deviation increase in quality-adjusted journal output (total size-, co-author-, and quality-adjusted article pages) or total citations is in [ ].

With CLm weights the optimal convex combination of citations and quality-adjusted journal output puts one percent weight on citations and 99% weight on journals. This 99% to 1% mix also maximises the log-likelihood if the last three years of journal articles are ignored while citations to older articles still accumulate over that period. Even under the very conservative rule of waiting six years before an article is considered for influencing salary, since that gives time for citations to accumulate, the weight on citations is just 11% while there is a 89% weight on where the articles are published.[12]

The results in Figure 1b that use KMS journal quality weights are a little more favorable for citations; allowing for publications and citations right to the end of 2010 one would form an optimal combination that was 69% journals and 31% citations. Ignoring articles published in the last three or six years changes this mix only slightly, with weights on journals of 72% and 65%. It should also be noted that the results in Figure 1 (and also in Table 4) do not account for the number of articles published, and when this variable was used in Table 3 any statistically significant effect of citations on salary disappeared completely. If the count of articles is included as a covariate, the weight on citations in the optimal combination of journals and citations (when using KMS weights that are most favorable to citations) is 21%, 19%, and 33% for publication-counting windows ending in 2010, 2007 and 2004 (while citations are allowed to accumulate to the end of 2010 in all cases).

The results thus far consider total citations but do not distinguish amongst citations. Yet it may be that it is the most (or a few of the most) heavily cited articles that show research impact that affects salary and other labor market outcomes. Indeed, it is this issue of the distribution of citations that the h-index and its generalized variants are designed to deal with. We therefore carry out a further three analyses that (a) consider total citations and the citations to the most cited article, (b) use the h-index for citations, and (c) use the $h_{(5,2)}$ index that Ellison (2013) finds works best in his full sample of more senior academics (who average 22 years post-PhD, which is almost the same as the current sample). Since the citations variables are being 'unpacked' to consider their distribution, we do a similar unpacking of the journals variable by using both the lifetime total number of articles published and the total quality-, size-, and co-author-adjusted pages in those articles. Thus the baseline model, before the citations variables are unpacked, is the one previously reported in panel A of Table 3.

---

[12] Using this rule also lowers the log-likelihood a lot (Figure 1a), showing the unreality of these assumptions.

**Figure 1: Optimally Weighted Combinations of Citations
and Quality-Adjusted Journal Output**

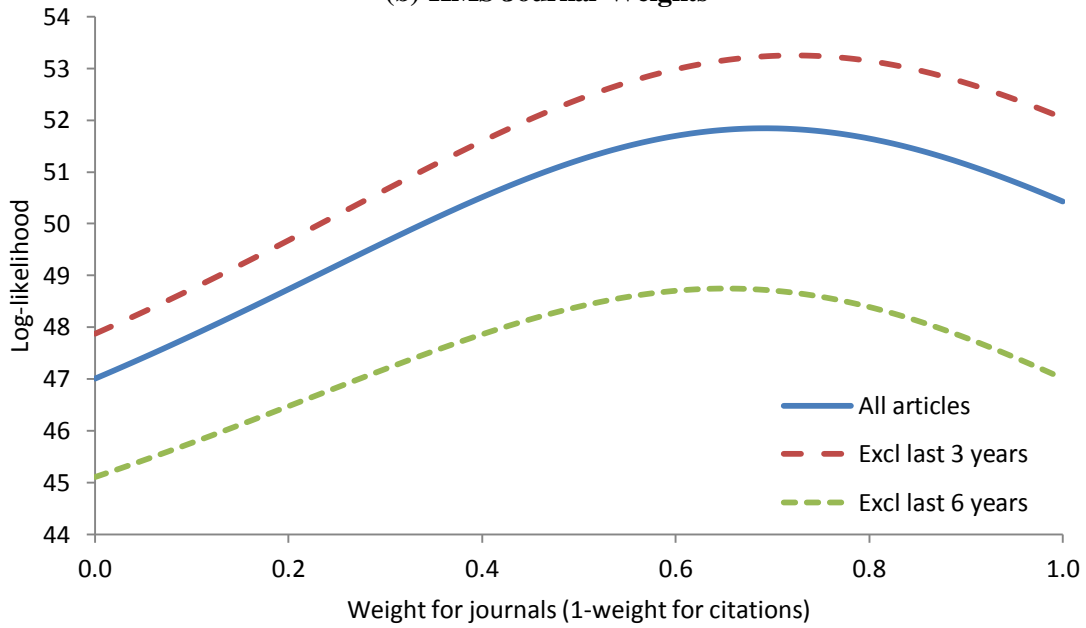**(a) Combes-Linnemer Medium Journal Weights**
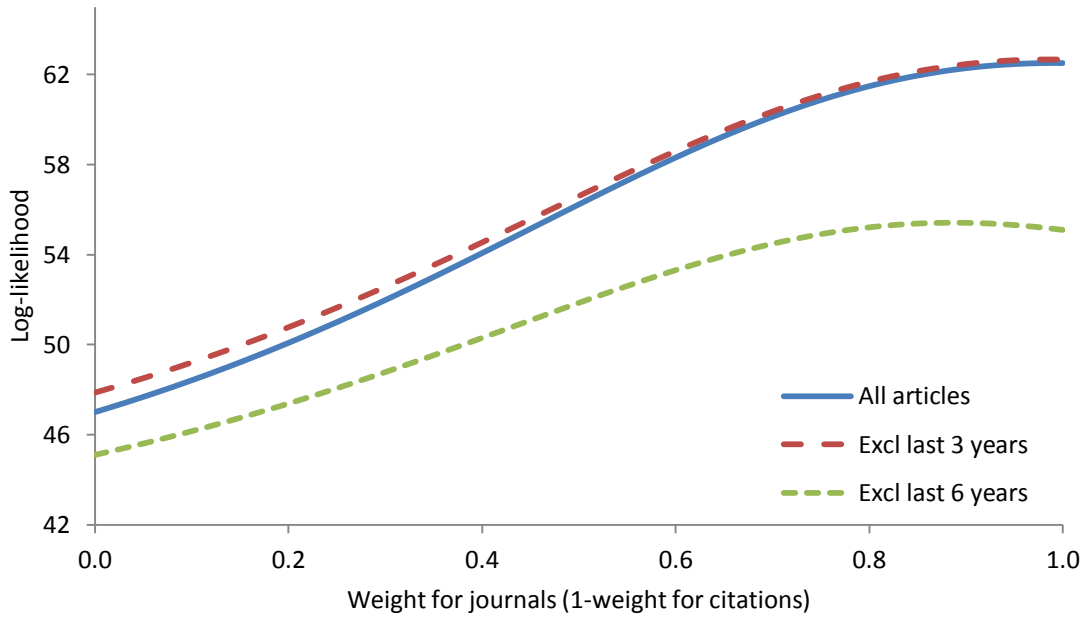
(b) KMS Journal Weights

Table 5: Salary Regression Sensitivity Analyses that Consider the Distribution of Citations

	MSF	CLm	CLh	RePEc	Coupé	K&Y_all	K&Y_econ	KMS	LP
	Journal Weighting Scheme for Calculating Quality-Adjusted Journal Output Over Lifetime Comes From:								

	MSF	CLm	CLh	RePEc	Coupé	K&Y_all	K&Y_econ	KMS	LP
A. Including Total Citations and Citations to the Most-Cited Article									
Quality-adjusted journals	0.007	0.011	0.012	0.010	0.016	0.017	0.018	0.015	0.021
	(3.89)**	(4.52)**	(3.91)**	(4.10)**	(3.69)**	(4.34)**	(4.31)**	(2.95)**	(3.81)**
Number of articles (×10)	0.019	0.021	0.035	0.030	0.027	0.035	0.035	0.039	0.041
	(2.22)*	(2.72)**	(5.10)**	(3.74)**	(3.29)**	(4.68)**	(4.64)**	(4.95)**	(5.82)**
Total citations (×1000)	-0.051	-0.060	-0.060	-0.047	-0.048	-0.056	-0.057	-0.051	-0.050
	(1.94)	(2.30)*	(2.09)*	(1.70)	(1.80)	(2.10)*	(2.10)*	(1.53)	(1.80)
Cites to most-cited (×1000)	0.157	0.164	0.159	0.149	0.162	0.172	0.174	0.160	0.116
	(2.37)*	(2.50)*	(2.29)*	(2.17)*	(2.41)*	(2.56)*	(2.58)*	(1.99)*	(1.72)
R^2	0.739	0.746	0.741	0.735	0.734	0.739	0.739	0.732	0.743
Vuong non-nested test[a]	2.42*	2.95**	2.67**	2.24*	1.97*	2.38*	2.37*	2.03*	2.84**
B. Including the h-index for Citations									
Quality-adjusted journals	0.006	0.009	0.010	0.008	0.011	0.013	0.014	0.010	0.017
	(3.11)**	(3.87)**	(3.23)**	(3.38)**	(2.57)*	(3.64)**	(3.60)**	(2.10)*	(3.70)**
Number of articles (×10)	0.011	0.014	0.025	0.020	0.019	0.024	0.024	0.027	0.030
	(1.33)	(1.69)	(3.15)**	(2.24)*	(2.09)*	(2.86)**	(2.84)**	(2.99)**	(3.58)**
Citations h-index	0.005	0.003	0.004	0.006	0.006	0.005	0.005	0.006	0.004
	(1.15)	(0.72)	(0.81)	(1.42)	(1.36)	(1.19)	(1.20)	(1.30)	(0.81)
R^2	0.737	0.742	0.738	0.734	0.731	0.736	0.736	0.730	0.741
Vuong non-nested test[a]	1.52	2.06*	1.75	1.41	1.20	1.65	1.63	1.01	2.05*
C. Including the Generalized $h_{(5,2)}$ Index for Citations									
Quality-adjusted journals	0.006	0.010	0.011	0.009	0.013	0.015	0.016	0.013	0.018
	(3.67)**	(4.36)**	(3.77)**	(4.04)**	(3.10)**	(4.26)**	(4.23)**	(2.75)**	(4.14)**
Number of articles (×10)	0.016	0.017	0.030	0.027	0.025	0.031	0.031	0.035	0.035
	(1.97)	(2.22)*	(4.50)**	(3.58)**	(3.20)**	(4.29)**	(4.26)**	(4.69)**	(5.46)**
Citations $h_{(5,2)}$-index	0.010	0.000	-0.001	0.008	0.011	0.005	0.005	0.005	0.001
	(0.59)	(0.01)	(0.05)	(0.48)	(0.61)	(0.26)	(0.28)	(0.27)	(0.04)
R^2	0.735	0.741	0.737	0.731	0.729	0.734	0.734	0.728	0.740
Vuong non-nested test[a]	3.22**	3.61**	3.47**	3.34**	3.09**	3.51**	3.53**	3.09**	3.59**

Notes: Each regression also includes all of the other variables from Table 2. *N*=167, robust *t* statistics in parentheses, * significant at 5%; ** at 1%. The Vuong test is of a model using citations variables versus a model using journals variables (both having all control variables), with positive values favoring the journals model.

Allowing for the distribution of citations does not change the finding that salary depends on the quality and quantity of journal articles but is hardly affected by citations. In panel A of Table 5 the results show that salary is higher if there are more citations to the most cited article, conditional on the output in journals, and on total citations (which have a negative effect). But the Vuong tests show that a model using total citations and citations to the most-cited article would be significantly rejected against a model that uses the count of articles and the ranking of the journals that published those articles (with control variables common to both models). In panel B, it is seen that the h-index provides no explanatory power for salary once quality-adjusted journal output and the number of articles is controlled for (and the Vuong tests also reject models that use the h-index for citations against models that use variables measuring output in journals). Similarly, the generalized $h_{(5,2)}$ index has no effect on salary; using this generalized index with more weight on a few highly cited papers gives less explanatory power than using the original h-index.[13] This last result is the opposite to what Ellison (2013) finds, with a different sample and labor market outcome, and using the more permissive *Google Scholar* record of research papers and citations.

5. Discussion and Conclusions

Setting aside the normative question of how much weight *ought* to be put on citations when evaluating research quality, our results answer the positive question of how much weight *is* put on citations. At least for one labor market outcome, for the largest research-intensive public university system in the United States, it seems that hardly any weight is put on citations. If one had to choose between academic salary models with explanatory variables based on where research is published versus models whose variables measure how highly cited that research is, the citations-based models lose out in all of our comparisons. Of course, researchers do not have to choose just one set of variables and can study academic labor markets using variables based on journals along with variables based on citations. Our results suggest that a combination of these types of variables would put most weight on the journals where research is published. This conclusion holds even under conservative rules that follow from Sgroi and Oswald (2013) on optimal information use for evaluating research, which in this context would be to make salary decisions only after waiting some years for the quality of an article to be revealed by the citations it has accrued.

Our results contrast with some recent studies of the academic labor market in economics. For example, Hilmer et al (2015) find various citations variables predict salary for economists in a sample of departments that includes six of the nine studied here (along with 47 others). A likely reason why the results differ is that lifetime journal output is measured less comprehensively by

---

[13] There is a highly significant ($p<0.01$) Vuong test statistic of 2.84 favoring a model using the h-index over a model using the $h_{(5,2)}$ index (with all the control variables common to the two models).

Hilmer et al, who just count the number of articles in 'elite', 'excellent', and 'other' journals. There are three problems with this approach that likely reduce explanatory power of variables measuring output in journals and thereby overstate explanatory power for citations (since the two measures are highly correlated; for example, $r=0.83$ for the h-index and CLm-weighted journal output). First, simple counts of articles explain less of the variation in salary than is explained by the number of size-, co-author-, and quality-adjusted pages (Gibson, 2014). Second, their 'elite' and 'excellent' groups of journals miss many highly ranked journals, perhaps due to being based on a study from 1996; for example, journals in their 'excellent' group can be found at positions between 5th and 155th place in the Combes and Linnemer (2010) journal rankings, and have an average rank position of 45th rather than 21st that would be expected (given the 31 journals in this group and five in the 'elite' group). Third, variation in journal quality within these groups is ignored; for example, the 'excellent' group has journals whose CLm quality weight ranges from just 13.5% of the weight for the *QJE* all the way up to 81% of the *QJE*. If models using our variables to measure the quality and quantity of journal output are compared with models that follow Hilmer *et al.* in measuring journal output, the Vuong tests favor the specification used here, with statistically significant ($p<0.05$) results for the CLm, CLh, and LP weights.[14]

If we move from the broad question of citations versus journals to a narrower question of what sort of citations index to use, our results contrast with Ellison (2013) because we find no evidence in favor of a generalized $h_{(a,b)}$ index over the usual h-index. Since we use a different sample, a different labor market outcome, and a different database for publications and citations, it is unclear which source contributes most to the different findings. It would be a useful study to see what values of a and b for $h_{(a,b)}$ work best when using data from *Google Scholar* compared with using *Web of Science* data to predict a labor market outcome; Ellison's claim that more weight should be put on a few highly cited papers than what the usual h-index puts on them may be due to the permissiveness of *Google Scholar*. That citation counts, the h-index, and publication counts differ so greatly between *Google Scholar* and *Web of Science* suggests that economists should be cautious in applying any rules of thumb about what a 'good' research record looks like since these may be database-specific.

Finally, in terms of the normative issue of whether more weight ought to be put on citations, even though it appears that not much weight actually is put on them, the various manipulations described in Section 1 provide grounds for caution. There are few, if any, consequences for researchers who cite unnecessary papers, compared with the consequences of failing to cite papers that an editor or referee may feel to be relevant. In anticipation of this asymmetry, authors are likely to cite more than is optimal, thereby devaluing the information content of citations. This debasement of citations may become even greater as more emphasis is put on them as supposed measures of the quality of research.

---

[14] This comparison is based on the specifications in Table 6a of Hilmer, Ransom and Hilmer (2015).

References

Anderson, D., Smart, W., and Tressler, J. (2013). Evaluating research–peer review team assessment and journal based bibliographic measures: New Zealand PBRF research output scores in 2006. *New Zealand Economic Papers* 47(2): 140-157.

Bratsberg, B., Ragan, J., and Warren, J. (2003). Negative Returns to Seniority: New Evidence in Academic Markets. *Industrial and Labor Relations Review* 56(2): 306-323.

Bratsberg, B., Ragan, J., and Warren, J. (2010). Does Raiding Explain the Negative Returns to Faculty Seniority? *Economic Inquiry* 48(3): 704-721.

Burns, S., and Stern, D. (2015). Research Assessment Using Early Citation Information. *Research Paper* No 15-01, Crawford School of Public Policy, Australian National University, Canberra.

Combes, P-P and Linnemer L. (2010) *Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of all Journals in Economics*, Groupement de Recherche en Economie Quantitative d'Aix Marseille (GREQAM), Document de Travail, no 2010-28.

Coupé, T. (2003). Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990-2000. *Journal of the European Economic Association* 1(6): 1309-45.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association* 78(383): 605–610.

Ellison, G. (2013). How Does the Market Use Citation Data? The Hirsch Index in Economics. *American Economic Journal: Applied Economics* 5(3): 63-90.

Gibson, J. (2000). Research productivity in New Zealand university economics departments: comment and update. *New Zealand Economic Papers* 34(1): 73-88.

Gibson, J. (2014). Returns to articles versus pages in academic publishing: Do salary-setters show 'article illusion'? *Economics Letters* 125(3): 343-346.

Gibson, J., Anderson, D., and Tressler, J. (2014). Which journal rankings best explain academic salaries? Evidence from the University of California. *Economic Inquiry* 52(4): 1322-1340.

Hamermesh, D., Johnson, G., and Weisbrod, B. (1982). Scholarship, Citations and Salaries: Economic Rewards in Economics. *Southern Economic Journal* 49(3): 472-481.

Hamermesh, D., and Pfann, G. (2012). Reputation and earnings: the roles of quality and quantity in academe. *Economic Inquiry* 50(1): 1-16.

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy* 41(2): 251-261.

Hilmer, C., and Hilmer, M. (2005). How Do Journal Quality, Co-Authorship, and Author Order Affect Agricultural Economists' Salaries? *American Journal of Agricultural Economics* 87(2): 509-523.

Hilmer, C., Hilmer, M., and Lusk, J. (2012). A Comparison of Salary Structures between Economics and Agricultural Economics Departments. *Applied Economic Perspectives and Policy* 34(3): 489-514.

Hilmer, M., and Hilmer, C. (2011). Negative Returns to Seniority and Job Mobility across the Program Quality Distribution: Are Top Public PhD-Granting Programs Different? *American Economic Review: Papers and Proceedings* 101(3): 466-470.

Hilmer, M., Ransom, M., and Hilmer, C. (2015). Fame and the fortune of academic economists: How the market rewards influential research in economics. *Southern Economic Journal* 82(2): 430-452.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science* 101: 16569-165572.

Kalaitzidakis, P., T. Mamuneas, and T. Stengos. 2010. 'An Updated Ranking of Academic Journals in Economics.' Working Paper 9/2010, Economics Department, University of Guelph, Guelph, Canada.

Kodrzycki, Y., and P. Yu. (2006) 'New Approaches to Ranking Economics Journals.' *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy*, 5(1), Article 24.

Laband, D., and M. Piette. (1994) 'The Relative Impact of Economics Journals.' *Journal of Economic Literature*, 32(2), 1994, 640–66.

Liebowitz, S. (2013). Our uneconomic methods of measuring economic research. *Voxeu* 6 Dec http://www.voxeu.org/article/our-uneconomic-methods-measuring-economic-research

Liebowitz, S. (2014). Willful blindness: The inefficient reward structure in academic research. *Economic Inquiry* 52(4): 1267-1283.

Liebowitz, S., and Palmer, J. (1984). Assessing the relative impacts of economics journals. *Journal of Economic Literature* 22(1): 77-88.

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

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

Moore, W., Newman, R., and Turnbull, G. (1998) Do Academic Salaries Decline with Seniority? *Journal of Labor Economics* 16(2): 352-366.

Moore, W., Newman, R., and Turnbull, G. (2001) Reputational Capital and Academic Pay. *Economic Inquiry* 39(4): 663-671.

Oswald, A. (2007). An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. *Economica* 74(293): 21-31.

Ransom, M. (1993). Seniority and Monopsony in the Academic Labor market. *American Economic Review* 83(1): 221-233.

Sauer, R. (1988). Estimates of the Returns to Quality and Co-authorship in Economic Academia. *Journal of Political Economy* 96(4): 855:866.

Scott, L., and Mitias, P. (1996). Trends in rankings of economics departments in the U.S: An update. *Economic Inquiry* 34(3): 378-400.

Sgroi, D., & Oswald, A. (2013). How Should Peer-review Panels Behave? *Economic Journal* 123(570): F255-F278.

Spiegel, M. (2012). Reviewing less – Progressing more. *Review of Financial Studies* 25(5): 1331-1338.

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

Vuong, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses. *Econometrica* 57(2): 307-333.