Received: 27 June 2019 | Accepted: 14 February 2021

DOI: 10.1111/ecin.12994

ORIGINAL ARTICLE

Economic Inquiry

Depth and breadth relevance in citation metrics

David I. Stern, Professor[1] | **Richard S. J. Tol, Professor**[2,3,4,5,6,7]

[1]Crawford School of Public Policy, The Australian National University, Acton, ACT, Australia

[2]Department of Economics, University of Sussex, Falmer, UK

[3]Institute for Environmental Studies, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[4]Department of Spatial Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[5]Tinbergen Institute, Amsterdam, The Netherlands

[6]CESifo, Munich, Germany

[7]Payne Institute for World Resources, Colorado School of Mines, Golden, Colorado, USA

Correspondence
Richard S. J. Tol, Department of Economics, University of Sussex, BN1 9SL Falmer, UK.
Email: r.tol@sussex.ac.uk

Abstract

The Euclidean length of a citation list is "depth relevant": the metric increases when citations are transferred from less to more cited papers. We introduce "breadth relevance," which favors consistent achievers over one-hit wonders. The exponent of the CES aggregator then is less than unity rather than greater than unity, as for depth relevance. Using two datasets on citations of economists for the top 50 US and global universities, simply counting citations maximizes the correlation between the citation metrics of researchers and the peer-reviewed rank of their department. However, citation depth may explain the allocation of researchers across lower-ranked departments.

KEYWORDS

citations, rankings, research assessment

JEL CLASSIFICATION

A14; C43; J24

1 | INTRODUCTION

The academic pecking order is partly set by performance rankings. Rankings are torn between what matters and what can be measured, but most would agree that citations are a reasonable proxy for peer appreciation of research. Perry and Reny (2016) propose a new way to count citations: The Euclidean length of a citation list. They claim that this explains the distribution of researchers across the top 50 US economics departments. We show that their empirical support for this axiom weakens when we control for cohort—the year of PhD or first publication—effects, as one should (Hamermesh, 2018). Perry and Reny (2016) argue that citation metrics should satisfy *Depth Relevance*, which implies that citations should be concentrated on a few papers. We offer an alternative axiom, which we dub *Breadth Relevance*, which rewards researchers who have consistent citation performance across their publications. Our empirical analysis shows that simply counting citations—focusing neither on *breadth* nor *depth*—maximizes the correlation between researchers' citation metrics and department ranks. However, *depth* relevance may explain the allocation of researchers across lower-ranked departments.

Abbreviation: ML, maximum likelihood.

Versions of this article were presented at seminars at the University of Göttingen and the University of Sussex.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Economic Inquiry* published by Wiley Periodicals LLC on behalf of Western Economic Association International.

Perry and Reny (2016) show that Euclidean length is the only metric that satisfies five desirable properties that they argue a citation metric should have. The innovation in their approach is not so much the new metric—new citation metrics are proposed regularly—but rather their axiomatic approach.[1] Together, their axioms imply that the ideal citation metric is a CES aggregate with exponent two. One of these five axioms is *depth* relevance, where splitting the citations for a single paper into two separate papers cannot increase the citation metric. This implies that shifting a citation from one of a researcher's less cited papers to one of their more cited papers always increases the metric. In the extreme, a "one hit wonder," with one highly cited paper and no citations to their other papers, would be valued above a researcher with the same number of citations but a more even citation record. Our alternative axiom, *breadth* relevance, favors consistent achievers over one-hit wonders. We show that a metric cannot have both *depth* and *breadth* relevance (if other axioms are satisfied at the same time). We also show that the *breadth* relevant metric has the form of a CES aggregator, just like the *depth* relevant metric. Our *breadth* relevant metric also satisfies Perry and Reny (2016)'s *Monotonicity*, *Independence*, and *Scale Invariance* axioms.

Breadth relevance's weakness is that it always favors splitting papers or shifting citations from higher-ranked papers to lower-ranked papers. Therefore, we also propose a metric that is a convex combination of the *depth* and *breadth* relevant citation metrics. Different institutions may put different weights on the two metrics in hiring and promotion, and these weights may vary systematically with the characteristics of these institutions. Our new combined metric does not formally satisfy Perry and Reny (2016)'s *Independence* axiom, but changes in rank are rare when adding an equally-cited paper to all authors.

We follow Perry and Reny (2016), who follow Ellison (2013), estimating the parameters of our metric by finding the best match between individuals and departments. Gibson et al. (2017) instead match individuals' ranks and salaries, which can only be done for some institutions. Ellison (2013) finds that Hirsch-like metrics that focus on a smaller number of more highly cited papers perform better, suggesting that *breadth* relevance is not much valued in this sample of data.[2] We also use a much larger dataset that we scraped from *CitEc* for economists at the 400 universities ranked by *QS*. The latter sample allows us to assess a wider variety and larger number of institutions in order to obtain more precise estimates. As Gibson et al. (2017) found differences between higher- and lower-ranked departments, we focus on that heterogeneity.

We find that maximum likelihood (ML) estimates converge to a function that weights only *breadth* or *depth* relevance. Nonparametric analysis also shows little or no advantage to considering both *breadth* and *depth* together. However, we find that, among top universities, assignment of researchers to departments is more closely related to total citations rather than the Euclidean Metric. There is support for the Euclidean Metric in the distribution of researchers across lower-ranked international universities.

In the remainder of the article, we first present the theory of *breadth* relevant citation metrics and metrics that place weight on both *breadth* and *depth* relevance. We then turn to the empirical application.

1.1 | Citation metrics

Perry and Reny (2016) propose five axioms.[3] We reorder them. The first two axioms are:

Monotonicity: The value of the metric does not fall if a new paper with sufficiently many citations is added.

Independence: The ranking of two authors does not change if both publish a new paper that is cited the same number of times.

The appeal of these two axioms is intuitive.[4] Note that the H-index violates the *Independence* axiom. The third axiom is:

Scale Invariance: The ranking of two authors does not change if all citation numbers of both are multiplied by the same, positive number.

This axiom is intuitive too. If all citation numbers are doubled, the ranking should not change. This axiom has a practical implication as well. If the aim is to rank researchers from different disciplines, then citation numbers should be corrected for differences in citation habits between disciplines. *Scale Invariance* allows for multiplicative corrections to rank *between* disciplines without changing the ranking *within* disciplines. The same holds for fields within disciplines; and for cohorts.

Perry and Reny (2016), echoing Burk (1936), show that these three axioms together imply that the citation metric C of researcher r with citation numbers $c_{r,i}$ to papers i is

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$C(c_r) = \left(\sum_i c_{r,i}^{\sigma} \right)^{\frac{1}{\sigma}} \tag{1}$$

for any $\sigma > 0$. This is the well-known CES function (Arrow et al., 1961; Solow, 1956).

Perry and Reny (2016) add a fourth axiom:

Depth Relevance: The value of the metric weakly increases if the citations of two papers are all attributed to either.

Perry and Reny (2016) show that *Depth Relevance* implies that $\sigma > 1$.[5] Is *depth relevance* a good axiom? It emphasizes quality over quantity, a sentiment widely shared among economists. Essentially, it says that researcher *A* with one paper that is cited 1000 times is more valuable than researcher *B* with 10 papers that are each cited 100 times. However, *depth relevance* is a global property. Researcher *C* with one paper that is cited 100 times is ranked above researcher *D* with 10 papers that are each cited 10 times. One may argue that researcher *C* is a one-hit wonder, and prefer the broader experience of researcher *D*. The corresponding axiom is:

Breadth Relevance: The value of the metric weakly increases if the citations to one paper are divided over two papers.

As a corollary to Theorem 1 of Perry and Reny (2016), *Breadth Relevance* implies $\sigma < 1$.[6] It immediately follows that:

Theorem 1. *A citation metric that satisfies* Monotonicity, Independence, *and* Scale Invariance, *cannot simultaneously satisfy* Depth Relevance *and* Breadth Relevance.

Proof. $\nexists \sigma$ such that $\sigma < 1 \wedge \sigma > 1$.

As argued in Footnote 2, the Hirsch index is *depth* relevant in some cases and *breadth* relevant in others. Following the example of Hirsch (2007)[7] and combining two metrics, another compromise metric is

$$C'(c_r) = \theta \left(\sum_i c_{r,i}^{\sigma} \right)^{\frac{1}{\sigma}} + (1-\theta) \left(\sum_i c_{r,i}^{\tau} \right)^{\frac{1}{\tau}} \tag{2}$$

for any $\sigma > 1$, $0 < \tau < 1$, and $0 < \theta < 1$.

Different institutions may weight the two sub-metrics differently. Perhaps for research universities, $\theta \approx 1$, while for teaching-oriented universities and colleges that simply want some evidence of scholarly activity, $\theta \approx 0$.[8] A combination of low σ (close to one) and high θ means that departments are close to simply preferring more citations to less without caring much about breadth or depth. This is approximately what we found for top research universities. On the other hand, if θ were low and σ high, even though institutions treat breadth and depth as infinitely substitutable, given (2), they wish faculty to distinguish themselves strongly in one or the other direction—either by receiving very high citations for a small number of articles, or showing consistent performance across the corpus of their work.

The linear function (2) assumes that *depth* and *breadth* are infinitely substitutable. If, instead, institutions that place some weight on both metrics treat both of them as essential, we would have a Cobb–Douglas aggregate. More generally, we can consider a CES aggregate[9] of the two metrics that allows institutions to consider the two to be substitutable of varying degrees:

$$C'''(c_r) = \left(\theta \left(\sum_i c_{r,i}^{\sigma} \right)^{\frac{\phi}{\sigma}} + (1-\theta) \left(\sum_i c_{r,i}^{\tau} \right)^{\frac{\phi}{\tau}} \right)^{\frac{1}{\phi}} \tag{3}$$

where $\phi = (\epsilon - 1)/\epsilon$ and ϵ is the elasticity of substitution. Importantly, if $\epsilon < 1$, then there are minimum necessary levels of both *depth* and *breadth* required to achieve a given level of C'''. This nested CES function maintains all of the desirable properties of the CES function *within* and *between* the composite indices (Keller, 1976; Sato, 1967).

Only a CES function satisfies the first three axioms of Perry and Reny (2016). A nested CES function therefore does not. However,

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proposition 1. *A CES combination of citation metrics that each satisfy* Monotonicity, Independence, *and* Scale Invariance, *satisfies* Monotonicity *and* Scale Invariance *but not* Independence.

Proof. See Appendix.

That is, it is the axiom of *Independence* that implies, and is implied by, a CES aggregation. Note that Equation (2) is a special case of Equation (3). The Proposition 1, therefore, applies to a linear combination as well.

A counterexample suffices to show that Equation (2) violates *Independence*. Set $\sigma = 2$, $\tau = 0.5$. Researcher A has two papers, one cited twice, one cited not at all. Researcher B has two papers, each cited once. The *depth* relevant metric prefers researcher A over B, and the *breadth* relevant metric has B over A. For $\theta = 0.8$, *depth* relevance dominates *breadth* relevance. Now both researchers publish an additional paper, cited once. *Depth* relevance continues to prefer A, and *breadth* relevance still prefers B. The additional paper makes both researchers broader and deeper, but the increase in *breadth* is larger than the increase in *depth*. *Breadth* relevance now dominates *depth* relevance, and researcher B outranks researcher A.

Another counterexample shows that Equation (3) violates *Independence*. Set parameters and citation records as in the above counterexample. For $\theta = 0.9$ and $\phi = 2$, *depth* relevance dominates *breadth* relevance. If both publish one more paper, cited once, *breadth* relevance dominates *depth* relevance.

These counterexamples illustrate two key features of the violation of *Independence*: First, the two sub-metrics must disagree on the ranking. Second, the additional paper must have a small effect on the dominant metric and a large effect on the dominated metric. Interpreting *Independence* as a criterion to judge the evolution of researchers over time, for an established researcher, with a large portfolio of papers, an additional paper will not have an out-sized effect on any citation metric. Interpreting *Independence* as a criterion to compare researchers, it is rare to encounter two large portfolios of papers which differ in one citation number only. We, therefore, argue that violations of *Independence* are likely to be rare for established researchers.

At the same time, it is uncertain whether an individual junior researcher will turn out to be *deep* or *broad*. New papers are, therefore, particularly informative and rank reversals reveal the extra insight gained into their relative strengths.

Appendix B shows that, for reasonable parameter choices and the data discussed below, violations of *independence* are indeed rare if not very rare. The proposed metric satisfies *independence* for all practical purposes. For almost all researchers in our sample, our proposed metric does not change the ranking of two authors if both publish a new paper that is cited the same number of times.

2 | APPLICATION

2.1 | Data

We use two data sets[10]:

1. Ellison (2013)'s dataset on economists at the top 50 US economics departments.
2. Data that we scraped from CitEc for all economists registered with *RePEc* based at the 400 universities ranked by the QS World University Ranking for 2017.

Ellison (2013)'s data consists of lists of papers and their citations downloaded from *Google Scholar* and author data collected separately for the top 50 US economics departments. A shortcoming of Ellison (2013)'s dataset is that authors' citation lists are truncated to a maximum of 100 publications. This is not a problem for Ellison (2013)'s purpose of computing the Hirsch index and its variations, but does distort computation of our metrics and will conceivably result in an underestimate of the weight placed on *breadth* relevance for more senior authors. Also, as the data were retrieved from *Google Scholar* there is quite a bit of noise. As is usually the case with *Google Scholar* data (Halevi et al., 2017), some publications have erroneous or missing publication dates or are clearly assigned to the wrong author. We do not use the publication dates in our analysis and so did not attempt to fix any of these problems. A small number of authors in Ellison's file of individual paper data did not have corresponding author data and vice versa even though they were not on his provided list of dropped authors. We deleted these authors. A few papers had −1 citations in the database. We

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

set these citation numbers to zero. We then removed the records of the six authors with no citations. After cleaning the data, we have 1,523 authors and 106,016 papers.

The 50 top departments are defined by their rank in the 1995 report of the National Research Council on research-doctorate programs in the United States—the ranks are provided in Ellison's dataset. This report was based on a survey of faculty members who were asked to assess the scholarly quality and effectiveness in education of individual doctoral programs in their own fields (Ostriker et al., 2011). It was, therefore, based purely on peer review or reputation. Our analysis, therefore, attempts to determine the function of citations that best matches this peer assessment.

We extracted data from *CitEc* using a modified version of the webscraper of Tol (2013). The Matlab code is included in the online material. The data have a similar structure to the Ellison data set. The main difference is that *CitEc* does not include data on fields of specialization and we did not attempt to retrieve that data from other *RePEc* services. Therefore, we are unable to adjust raw citations for field of specialization. Variation in citation counts across fields is much smaller than that across cohorts, so it is more important to normalize for cohort. Also, while the Ellison data assigns researchers to cohorts based on the number of years since they received their PhD, we used the year of first publication in the *RePEc* database to assign researchers to cohorts.

As in the case of *Google Scholar*, *CitEc* has limitations as an accurate source of citation data. First, *RePEc* relies on individuals registering for the service and on volunteers and publishers uploading data on publications. Publications are only assigned to authors if either authors claim them or archive managers include each author's *RePEc* handle in their metadata. It is plausible that less engaged or underperforming researchers are less likely to register. Similarly, younger researchers who may be happier to interact with online services might be more likely to register. Second, *CitEc* has so far only successfully extracted citations from 74% of documents in *RePEc* because of technical issues in accessing or parsing the documents. Extraction of citations is most efficient for working paper series such as *MPRA* and *NBER*. Third, the data include researchers who are retired or even dead and PhD students, post-docs, and so forth. who may not have the same potential level of impact as regular faculty in their departments. However, Hamermesh (2018) shows that there are strong rank correlations between *CitEc* citation counts and *Google Scholar* and *Web of Science* citation counts.

QS ranks universities rather than departments using six criteria:

1. Academic reputation, 40% weight. Based on responses from "70,000 individuals in the higher education space regarding teaching and research quality."
2. Employer reputation, 10% weight. Based on 30,000 responses to the *QS Employer Survey*.
3. Faculty/student ratio, 20% weight.
4. Citations per faculty, 20% weight. Field normalized citations for the previous 5 years provided by *Scopus*.
5. International faculty ratio, 5% weight.
6. International student ratio, 5% weight.

Thus, while citations are included in the measure, they are only assigned a 20% weight. Assessment by peers and employers (of students) accounts for 50% of the metric used in the ranking.

2.2 | Methods

As a first step, in order to test whether *depth* and *breadth* relevance are indeed relevant, we use OLS to regress the institution ranks of authors on a *depth* relevant citation metric ($\sigma = 2$) and on a *breadth* relevant one ($\sigma = 0.5$). We further control for age and tenure. As importance of *depth* and *breadth* relevance may vary with school rank, we repeat this using quantile regression.

Next, we find optimal values for σ, τ, θ, and ϕ in Equation (3) using parametric maximum likelihood regression and nonparametric rank correlation approaches.

We adjust for field and experience as follows:

1. Field adjustment: Following Perry and Reny (2016), we deflate citations in the Ellison dataset by average citations in that field (as defined by Ellison) relative to the mean citations in economic history. First, we compute average citations per publication for each author and regress this variable on the field weights for each author (which sum to one for each author). We compute a field deflator for each author by multiplying the regression coefficients from this

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

regression—the "field effects"—by that author's field weights. We divide the citations for each of the author's publications by their field deflator. The citation metrics are then computed using these deflated data.

2. Experience adjustment: We deflate the relevant citation metric by the mean value of the metric in the cohort. The advantage of this approach, over using an adjustment similar to the one that we use for fields, is that it takes into account both the varying average citations per paper across cohorts but also the varying length of publication lists across cohorts. Perry and Reny (2016) do not adjust for experience. There are 54 cohorts in the Ellison data, based on years since receiving the PhD. The seven authors with more than 54 years of post-PhD experience form a single cohort. For the *CitEc* data, we have 62 cohorts based on year of first publication. We removed authors whose first publication was prior to 1956.

Table 1 presents estimates of the field effects, again using OLS. Economic history has the lowest average number of citations per publication at 26 and finance the highest at 88. These results differ slightly from those of Perry and Reny (2016), because we use a regression method that assigns fractional weights in fields to individuals. All the estimates are highly statistically significant. The effect for OTHER is least precisely estimated; this includes a variety of fields including agricultural economics and law and economics.

We estimate the following regression models:

$$R_r = \alpha_1 + \beta_1 \ln\left(C_r/C_j\right) + \epsilon_{1,r} \tag{4}$$

$$R_r = \alpha_2 + \beta_2 \frac{\left(C_r/C_j\right)^{\lambda} - 1}{\lambda} + \epsilon_{2,r} \tag{5}$$

$$\frac{R_r^{\mu} - 1}{\mu} = \alpha_3 + \beta_3 \frac{\left(C_r/C_j\right)^{\lambda} - 1}{\lambda} + \epsilon_{3,r} \tag{6}$$

where R_r is the rank of the department (Ellison data) or university (CiTEc data) to which author r is affiliated, C_r is their citation metric, the α_k are intercepts, the β_k are regression parameters to be estimated reflecting the elasticity of rank with respect to the citation metric, and the ϵ_k, r are random error terms. C_j is the mean citation metric for cohort j. $|\lambda$ and μ are Box-Cox parameters, also to be estimated. Obviously, we could use cohort fixed effects in (4) instead of deflation of the citation metric by the cohort mean, but as we can see the cohort adjustment is nonlinear in (5) and (6).

As shown by the skewness test and the Breusch–Pagan test for heteroskedasticity, we found that the residuals of the semi-log specification had good properties for the Ellison dataset, while those of a log–log specification were highly heteroskedastic. Equation (4) is a special case of (5), which uses a Box-Cox transform instead of natural logarithms. This model allows us to somewhat relax the strong parametric restrictions on the relationship between ordinal rank and cardinal citation metrics implied by (4). On the other hand, β has a simpler interpretation in (4). Finally, the most general model that we estimate is (6), where the dependent variable is also Box-Cox transformed. We refer to this as the "double

TABLE 1 Estimates of field effects for the full sample

BEHAV_EXP	43.29	ECONHIST	26.09	PUBLIC	45.95
	(11.53)		(6.66)		(5.61)
DEVELOPMENT	44.68	IO	35.13	POLECON	44.55
	(9.34)		(5.25)		(12.30)
METRICSCS	43.35	INTFIN	41.72	MICROTHEORY	42.07
	(7.71)		(16.06)		(6.31)
METRICSTS	76.09	INTTRADE	79.27	OTHER	59.68
	(16.54)		(20.67)		(24.19)
FINANCE	88.20	LABOR	56.69	MACRO	71.35
	(19.26)		(8.95)		(8.86)

Notes: Heteroskedasticity robust standard errors in parentheses.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Box-Cox model" and (5) as the "single Box-Cox model." Equation (6) has the best residual properties for the *CitEc* dataset.[11]

We estimate (4) to (6) by maximum likelihood using the log likelihood function concentrated with respect to the standard deviation. We concentrated out the α_k by demeaning R_r in (4) and R_r and the transformed citation metric in (5) and (6).

We choose an initial parameter vector and then compute the metric C_r for all authors using the database of papers. We then compute C_j for each cohort. Next the log likelihood function is evaluated over the observations of authors. These three steps are then repeated for other parameter values using the BFGS algorithm to minimize the negative of the concentrated log likelihood. We compute standard errors using the BHHH estimate of the variance–covariance matrix of the parameters evaluating the necessary derivatives numerically. We use the unconcentrated likelihood for this, substituting in the ML estimate of the standard deviation.

We estimate the model for both the individual citation metric and for the CES aggregate (3) We estimate the model for authors from the full sample of departments and for two subsamples for each dataset. For the Ellison data we divide the sample into the top 25 and next 25 departments. For the *CitEc* data we divide the sample into the top 50 universities, which each have their own rank, and the next 350 universities, where only bands of ranks are reported. For the Ellison data, we also carried out all analyses on the subsample of 836 authors who have 100 or fewer publications and, therefore, have untruncated publication lists.

As the regression analysis treats the NRC or QS ranks as cardinal rather than ordinal and makes further assumptions underlying maximum likelihood estimation, we also carry out a nonparametric analysis using rank correlation coefficients. The correlation analysis computes the Spearman rank correlation coefficient for the cohort demeaned log metric of each author and their rank. The correlation coefficient is not a smooth function of the parameters as ranks change discretely as the parameters change. It also does not appear to be a perfectly convex function of the parameters. We address this issue by computing the correlations for many different vectors of parameters and then visualizing the results to understand which range of parameters best corresponds to the observed allocation of economists to departments.

For the single metric analysis, we compute all the correlations for the cohort-demeaned log citation metric, $\ln(C_r/C_j)$, with NRC or QS rank for values of τ or σ from 0.01 to 4. For the cohort-demeaned log CES function, $\ln\left(C_r'''/C_j'''\right)$ we compute all rank correlations with NRC or QS rank for all permutations of σ, τ, θ, and ϵ over the following ranges of the parameters with increments of 0.05: $1.05 \geq \sigma \leq 2.2$, $0.05 \geq \tau \leq 0.95$, $0.05 \geq \theta \leq 0.95$, and $0.05 \geq \epsilon \leq 2$, using a Cobb–Douglas function for $\epsilon = 1$. This results in 346,560 permutations. We found the maximum rank correlation coefficient for each σ, τ pair by sorting over the values of θ and ϵ. Sometimes multiple θ, ϵ pairs are associated with the same correlation coefficient. In these cases, we first sort by ϵ selecting the minimum ϵ and then by θ selecting the minimum θ. We also evaluated the linear function (2) over the same ranges of σ, τ, θ, but it was always inferior to a CES function with a lower elasticity of substitution.

3 | RESULTS

3.1 | Exploratory analysis

Figure 1 shows the impact of a *depth relevant* citation metric ($\sigma = 2$) and a *breadth relevant* one ($\tau = 0.5$) on the institution ranks of authors. The mean estimate is the OLS estimate, while the percentile estimates are quantile regression estimates. The analysis includes authors from all institutions in CitEc with 10 or more registered staff, including teaching-oriented universities. Institutions are ranked on the average number of citations per publication. This rank-correlates well with the NRC ($\rho = 0.85$) and QS ($\rho = 0.65$) rankings[12], and punishes those who increase quantity without quality. That said, the estimated relationship is part tautological as both the right-hand side and the left-hand side are based on the same citation data. The result in Figure 1 is an association, rather than a causal relationship. The results are more easily interpreted as the chance that an economist with particular characteristics is hired by an institution of a certain rank, but of course the characteristics of individual economists affect the standing of their institution. We see that *depth relevance* ($p < .0005$) and *breadth relevance* ($p < .0005$) are highly significant, in contrast to Perry and Reny's assumption that only *depth relevance* matters. However, the estimated coefficient is positive, so that is, broader researchers are more likely to be found at lower-ranked institutions.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

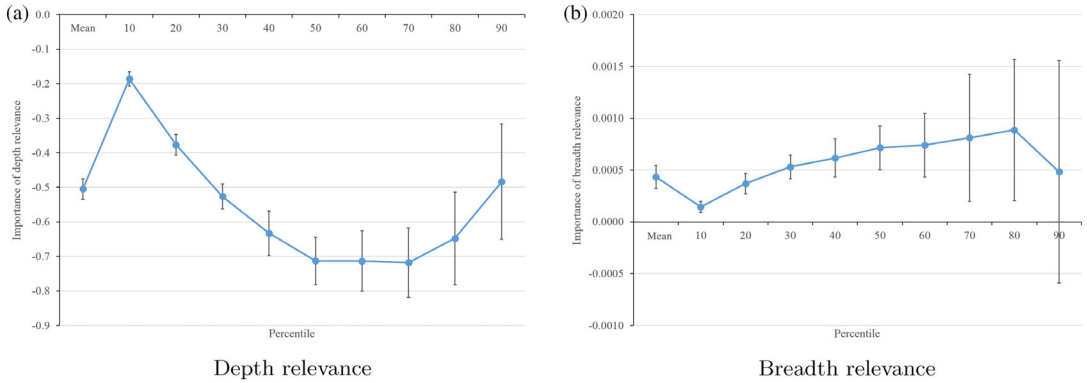


FIGURE 1 Estimated coefficients (vertical axis) of the influence of author citation metrics (demeaned by cohort) on their institution rank (CitEc, institutions with 10 or more registered economists) for authors from all institutions and for authors from each institution rank decile (horizontal axis). Panel (a) shows the impact of a citation metric that is *depth* relevant ($\sigma = 2$) and panel (b) the impact of one that is *breadth* relevant ($\sigma = 0.5$)

Using quantile regression, we find that the impact of both *depth* and *breadth* is less important at higher-ranked institutions. On the other hand, both *depth* and *breadth* explain the distribution at middle and lower-ranked institutions with deeper researchers having higher ranks and broader researchers lower ranks.

Figures C1 and C2 in Appendix C repeat the analysis without demeaning by cohort and including authors from all institutions at CitEC—not just those with more than 10 registered authors. Figures C3 and C4 (AppendixC) show the results for the NRC and QS rankings, including only authors from universities appearing in these rankings. The findings are qualitatively similar: Depth is appreciated but *breadth* is punished (or insignificant in case of the NRC ranking); higher-ranked universities evaluate citations differently than lower-ranked universities.

3.2 | Ellison data: Regression analysis

Estimates of all the CES models (3) converge onto an estimate with a single citation metric (1) i.e. $\theta = 1$ or 0.[13]We present the estimates for Equations (4), (5), and (6) for single citation metrics in Tables 2, 3, and 4, respectively:

For the full sample, the best fit for σ ranges from 1.32 for the semi-log model (Table 2) to 1.10, using the Double Box-Cox model (Table 4). When we restrict the sample to only those authors with untruncated citation lists we get a range of 1.26 to 0.95 across the different regression models. This is very close to simply counting citations, that is, $\sigma = 1$. All these estimates are statistically insignificantly different from unity but significantly less than 2, which is the exponent for the Euclidean Length Citation Metric. These values are much lower than Perry and Reny (2016) found, because we take cohort effects into account. Perry and Reny (2016) treat an assistant professor with a short publication list as a low quality researcher and senior academics with long publication lists and large numbers of citations as inherently better. We do not. When we correct for this, we find that the restriction implied by the *Directional Consistency* axiom of Perry and Reny (2016) is not particularly supported by the observed allocation of economists to departments.

We estimate that β is negative and statistically significant in both the full sample and for authors from the top 25 departments for all models. This makes sense, as the higher an author's department is ranked, the smaller R is. Authors with higher citation metrics are likely to be located at higher-ranked departments. However, the measured effect is quite small: An author with one standard deviation more than the mean value of the relative citation metric for their cohort—which is a little more than doubling the metric—is likely to be placed 4–5 ranks better than the mean author.

Estimates of σ for authors from the top 25 departments range from 1.01 to 1.21 depending on specification. Results for the next 25 departments are not as satisfactory, as neither the power of the citation metric nor the slope parameter can be estimated precisely (Table 2). For the sample of all authors, there is also a local maximum of the likelihood function for $\sigma = 4.43$. For the sample of authors with untruncated citation lists, the estimate converges to $\tau = 0$. There is

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Maximum likelihood estimates: Ellison data, semi-log model

	All departments		Top 25		Next 25	
	All authors	≤100 papers	All authors	≤100 papers	All authors	≤100 papers
β	−5.4108	−4.4498	−2.7382	−2.1416	0.0726	
	(0.3206)	(0.4693)	(0.2333)	(0.3516)	(0.3569)	
σ or τ	1.3211	1.2632	1.1195	1.2076	0.4598	0
	(0.1492)	(0.2723)	(0.1697)	(0.3890)	(2.2654)	
B-P test	0.0602	0.1892	4.6129	2.8476	1.5123	
	(0.8061)	(0.6636)	(0.0317)	(0.0915)	(0.2188)	
Skewness	0.5104	0.3439	0.1835	0.0454	0.0063	
	(0.0000)	(0.0000)	(0.0226)	(0.6846)	(0.9499)	
N authors	1,523	836	929	483	594	353
N papers	106,016	37,316	66,306	21,706	39,710	15,610
Relative citation metric statistics and effect size of 1 standard deviation increase at mean citations						
Min	0.002	0.010	0.010	0.014	0.000	
Max	9.179	7.529	7.138	6.845	11.937	
stdev	1.116	0.996	0.964	0.914	1.428	
Effect size	4.1	3.1	1.8	1.4	−0.1	

Notes: Standard errors for regression coefficients and *p*-values for Breusch–Pagan test in parentheses.

TABLE 3 Maximum likelihood estimates: Ellison data, single box-cox model

	All departments		Top 25	
	All authors	≤100 papers	All authors	≤100 papers
β	−6.4354	−5.3776	−3.1180	−2.5113
	(0.4054)	(0.5540)	(0.2805)	(0.4012)
σ or τ	1.1735	0.9828	1.0416	1.0333
	(0.1072)	(0.1683)	(0.1419)	(0.2725)
λ	0.2482	0.4348	0.2235	0.3225
	(0.0773)	(0.1601)	(0.1097)	(0.2092)
B-P test	0.0052	0.0123	4.4214	2.2455
	(0.9424)	(0.9118)	(0.0355)	(0.1340)
Skewness	0.4831	0.3248	0.1780	0.0359
	(0.0000)	(0.0000)	(0.0270)	(0.7479)
N authors	1,523	836	929	483
N papers	106,016	37,316	66,306	21,706
Relative citation metric statistics and effect size of 1 standard deviation increase at mean citations				
Min	0.002	0.005	0.008	0.011
Max	9.512	7.813	7.317	6.888
stdev	1.127	1.028	0.980	0.940
Effect size	5.3	4.5	2.3	2.0

Notes: Standard errors for regression coefficients and *p*-values for test statistics in parentheses.

little correlation between rank of department and number of citations for this group. Single Box-Cox estimates converged to $\tau = 0$ with large standard errors and so we do not report them in Table 3. We did not try to estimate the double Box-Cox models for this data given these results.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Maximum likelihood estimates: Ellison data, double box-cox model

	All departments		Top 25	
	All authors	**≤ 100 papers**	**All authors**	**≤ 100 papers**
β	−1.5698	−1.5261	−1.4447	−1.4139
	(0.0760)	(0.1377)	(0.1144)	(0.2149)
σ or τ	1.1047	0.9515	1.0086	1.0108
	(0.0825)	(0.1410)	(0.1253)	(0.2539)
λ	0.3363	0.4817	0.2940	0.3390
	(0.0712)	(0.1440)	(0.1125)	(0.2075)
μ	0.4835	0.5583	0.6473	0.7541
	(0.0062)	(0.0083)	(0.0104)	(0.0140)
B-P test	35.4994	8.4822	26.0552	24.9799
	(0.0000)	(0.0036)	(0.0000)	(0.0000)
Skewness	0.0166	−0.0850	−0.0794	−0.1561
	(0.7916)	(0.3163)	(0.3240)	(0.1627)
N authors	1,523	836	929	483
N papers	106,016	37,316	66,306	21,706
Relative citation metric statistics and effect size of 1 standard deviation increase at mean citations				
Min	0.001	0.005	0.008	0.010
Max	9.882	8.125	7.609	6.891
stdev	1.138	1.036	0.989	0.946
Effect size	5.4	4.6	2.4	3.3

Notes: Standard errors for regression coefficients and *p*-values for test statistics.

Estimates for authors from all departments using the sample with untruncated citation lists have a slightly lower estimates of σ than for the full sample, across all specifications. This confirms our intuition that ignoring the bottom part of authors' citation lists would bias the results in favor of depth.

The double Box-Cox results in Table 4 show more consistency across samples than the results in Tables 2 and 3. This suggests that by also transforming the dependent variable we have achieved a more linear specification. On the other hand, the heteroskedasticity properties are not better, though there is no significant skewness for the Double Box-Cox residuals in either the full sample or subsample, while there is for the Single Box-Cox model in the full sample just as there is for the semi-log model.[14] Interestingly, the transformation for dependent and explanatory variables is the same and closer to the square root function —$\lambda = \mu = 0.5$ —than to logarithms.

3.3 | Ellison data: Correlation analysis

Figure 2 presents the correlations between the citation metric (1) adjusted for experience (C_r/C_j) and department rank for the full sample of 1,523 authors and subsamples of authors at the top 25 departments and the next 25 departments. For each of these we computed the correlations for all authors and for authors with 100 or fewer papers. The curves are smoother the larger the sample. For the full sample, the maximum correlation is achieved for $\sigma = 1.31$, though there is a broad range of parameter values with similar correlations. Note that the correlation between rank and the cohort demeaned citation metric here is 0.449, while the rank correlation using Perry and Reny (2016)'s approach is only 0.262. Controlling for cohort increases the goodness of fit. Restricting this sample to authors with ≤100 papers reduces the correlations and the maximum now occurs for $\sigma = 1.17$. Eliminating authors from the 25 lower-ranked departments also reduces the correlations. The maximum for the full sample is $\sigma = 1.18$ and for the untruncated subsample 0.91. These maxima are again clearly quite far from the value of 1.85 obtained by Perry and Reny (2016).

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

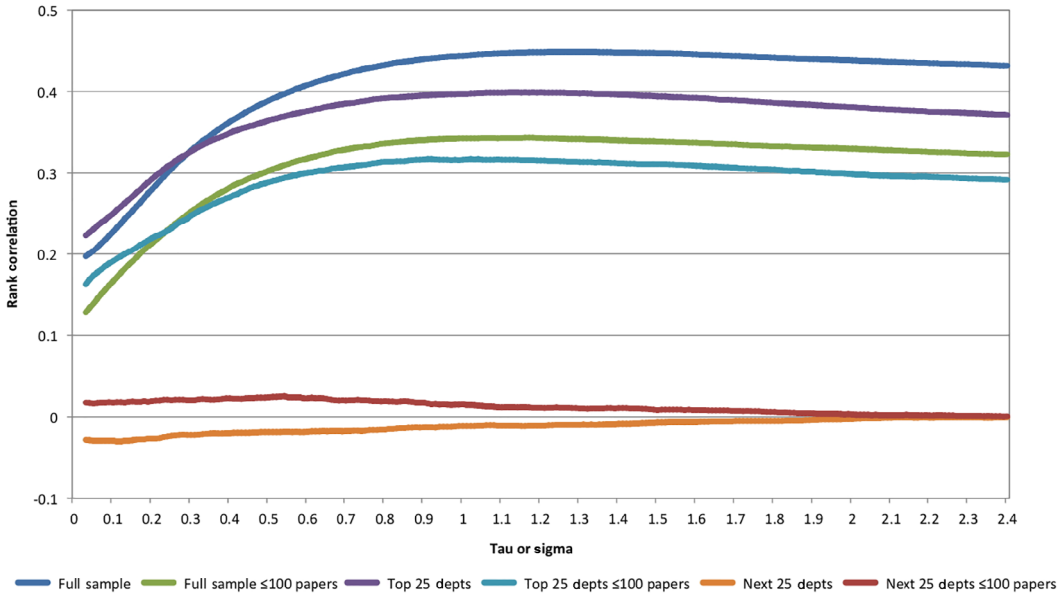


FIGURE 2 Rank correlations for a single metric: Ellison data

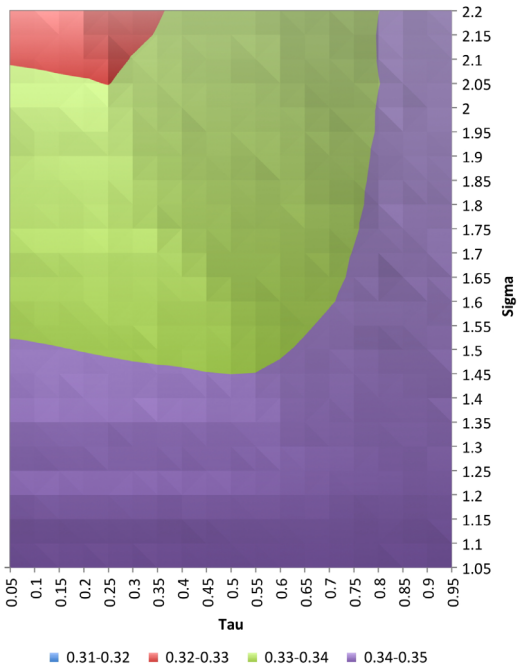


FIGURE 3 Correlations for the CES function: Ellison data, all departments, ≤100 papers

For the next 25 departments, the results here also show a lack of correlation between citations and the observed allocation of economists to departments. For the full sample the correlations are all negative with a minimum at $\tau = 0.12$. The correlation for the sample of untruncated publication lists is at a maximum for $\tau = 0.53$.

Figure 3 shows correlations between the CES function and NRC rank for the subsample of 836 authors who have 100 or fewer publications. High values of σ combined with low values of τ clearly have lower correlations. The maximum correlation is for $\sigma = 1.15$ and $\tau = 0.05$—the *breadth* metric is almost equal to the number of publications. However, the correlation is only very slightly higher (0.001) than the maximum for the single metric analysis. The value of ϵ associated with this point is 0.55 and the value of θ is 0.10. The low values of τ and θ may seem surprising, but the results are very insensitive to the choice of τ or θ for this low value of σ.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.4 | CitEc data: Regression analysis

Estimates of all the CES models again converged onto an estimate with a single citation metric. We present the estimates for the optimal single citation metric in Table 5.[15]

For the full sample, the best fit is $\sigma = 2.09$ using the semi-log model, $\sigma = 1.57$, using the single Box-Cox model, and $\sigma = 1.43$ using the double Box Cox model (Table 5). The latter two estimates are significantly lower than 2—the power for the Euclidean Citation Metric—though the first is not. On the other hand, all three estimates are significantly greater than unity. When we restrict the sample to authors from the top 50 universities, we obtain estimates of 1.68 using the semi-log model, 1.24 using the single Box-Cox model, and 1.13 using the double Box-Cox model. The latter two estimates are not significantly different from unity. On the other hand, when we restrict the sample to the next 350 universities, we obtain 3.25, 2.06, and 2.15. None of the estimates are significantly different from 2, indeed have wide 95% confidence intervals ranging from 0.15 to 6.35 for the semi-log model to 1.08 to 3.05 for the single Box-Cox model. Again, the correlation between rank and citations is lower in the lower part of the distribution.

We again estimate that β is negative and highly statistically significant. Here, based on the full sample and double Box-Cox model, an author with twice the mean value of the citation metric for their cohort would be placed 32 ranks higher than the mean author. However, this level of relative out-performance moves the needle much less within the top 50 universities with an improvement of 4 ranks. β is more consistent across samples for the double Box-Cox model but the variation is greater than for the Ellison dataset showing that it is harder to fit a single, simple model across this broader spectrum of universities. On the other hand, for the top 50 universities the double Box-Cox model fits well based on the heteroskedasticity test and skewness is much reduced for all samples, though the latter is always statistically significant given the large samples. The double Box-Cox model also appears to fit better than the other options for the next 350 universities.

TABLE 5 Maximum likelihood estimates: CitEc data

	All universities			Top 50			Next 350		
	Semi-log	Single Box-Cox	Double Box-Cox	Semi-log	Single Box-Cox	Double Box-Cox	Semi-log	Single Box-Cox	Double Box-Cox
β	−21.3834	−27.1755	−2.1162	−3.0272	−4.0625	−0.8059	−11.0956	−13.3456	−0.1826
	(0.6525)	(0.8198)	(0.0487)	(0.1975)	(0.2676)	(0.0469)	(0.6744)	(0.8669)	(0.0113)
σ or τ	2.0929	1.5659	1.4371	1.6796	1.2384	1.1301	3.2523	2.0627	2.1506
	(0.2842)	(0.1313)	(0.0826)	(0.3569)	(0.1715)	(0.1306)	(1.5837)	(0.5031)	(0.5465)
λ		0.3339	0.4265		0.4195	0.4581		0.2580	0.2934
		(0.0303)	(0.0240)		(0.0771)	(0.0721)		(0.0660)	(0.0661)
μ			0.4389			0.4408			0.1761
			(0.0012)			(0.0043)			(0.0013)
B-P test	82.1579	95.2773	103.5564	54.6125	69.5996	0.5976	65.0617	65.3043	19.165
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.4395)	(0.0000)	(0.0000)	(0.0000)
Skewness	0.5281	0.5240	−0.1394	0.3428	0.3248	−0.1796	0.4284	0.4274	−0.0525
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0164)
N authors	16,420	16,420	16,420	3,872	3,872	3,872	12,548	12,548	12,548
N papers	284,886	284,886	284,886	97,050	97,050	97,050	187,836	187,836	187,836
Relative citation metric statistics and effect size of 1 standard deviation increase at mean citations									
Min	0.002	0.001	0.00	0.002	0.001	0.001	0.005	0.004	0.004
Max	41.523	36.748	35.354	17.761	16.188	16.166	60.134	53.482	54.241
stdev	1.620	1.607	1.612	1.319	1.340	1.361	1.621	1.563	1.569
Effect size	32.9	30.7	32.0	2.5	4.1	3.9	10.7	14.2	12.9

Notes: Standard errors for regression coefficients and *p*-values for test statistics in parentheses.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

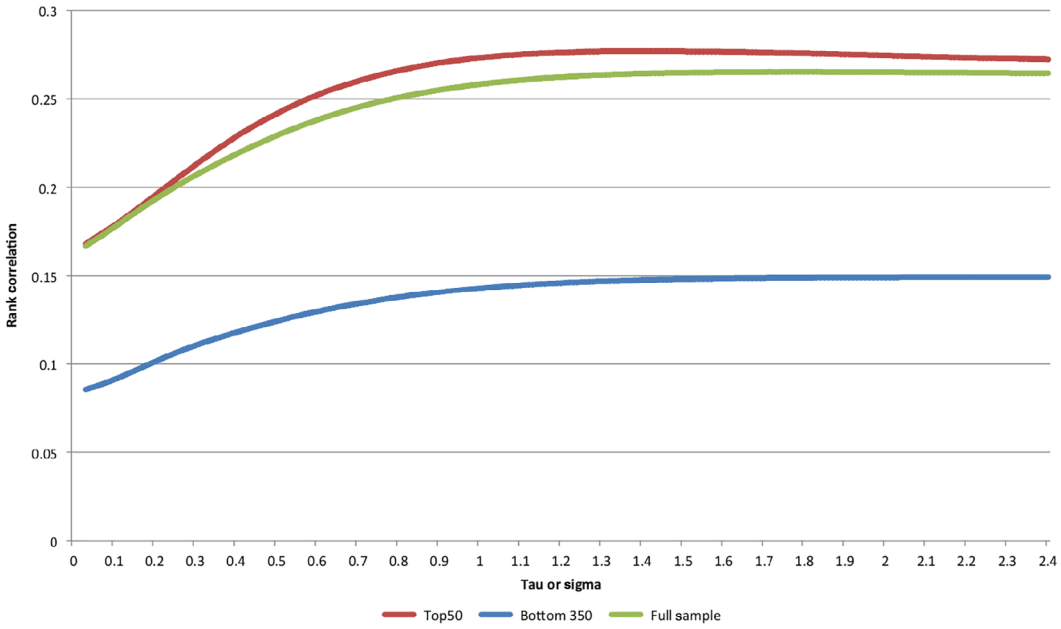


FIGURE 4 Rank correlations for a single metric: CitEc data

3.5 | CitEc data: Correlation analysis

Figure 4 presents the correlations for the full sample of 16,240 authors at the top 400 universities, and the subsamples of authors at the top 50 and next 350 universities. For the full sample, the maximum correlation is achieved for $\sigma = 1.79$, though there is a broad range of parameter values with similar correlations. The correlations are actually higher for the sample of the top 50 universities. Here the maximum correlation is for $\sigma = 1.34$. The correlations are much lower for authors from the next 350 universities and the maximum occurs for $\sigma = 2.22$. But there is a very wide range of values of σ with almost identical correlations.

Figure C.5 shows correlations between the CES function and QS rank for the full sample. Higher values of σ are associated with higher correlations. Here, the maximum correlation is the same as in the single metric analysis and τ has almost no effect on the results. The optimal value of θ ranges from 0.85 to 0.95 and ϵ is uniformly 0.05. The parametric and nonparametric results together suggest that depth becomes increasingly relevant and breadth less relevant when we included authors from lower-ranked universities in the sample.

4 | DISCUSSION AND CONCLUSIONS

Ellison (2013) suggests that it is better to have a few highly-cited papers than a larger number of well-cited ones. Perry and Reny (2016) axiomatize this. We introduce *breadth relevance* to complement their concept of *depth relevance*. A *breadth relevant* citation metric favors consistent achievers over one-hit wonders. The *breadth relevant* citation metric is a CES aggregate with an elasticity of substitution τ less than unity, whereas *depth relevant* citation metrics have elasticities of substitution (σ) greater than unity. A citation metric can place weight on both *breadth* and *depth* relevance at the expense of *independence*; however, we found that such indicators were statistically over-parameterized.

Using the same dataset as Perry and Reny (2016), we find weak empirical support for *depth relevance* among the top 50 US departments. Instead, the best fit is for $\sigma = 1$, which is equivalent to counting total citations and is neither breadth nor depth relevant. This result is different from that of Perry and Reny (2016) because we control for cohort effects in our analysis. Controlling for cohort significantly increases the correlation between department rank and authors' citation metrics. The analysis of Perry and Reny (2016) suffers from omitted variable bias, with cohort the omitted variable. Mechanically, older authors have had more time to accumulate citations (Hamermesh, 2018) and typically have a larger number of papers. It is relatively more likely though that more junior authors have at least one highly cited paper rather than having a large number of papers. If we do not control for cohort effects in our analysis, then

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

focusing on a small number of highly cited papers levels the playing field between more junior and more senior authors a little, providing a better fit to the distribution of authors across departments. Support for this idea comes from regressing the log of the raw citation metric on the number of years post-PhD. For $\sigma = 1$ the elasticity is 0.091 while for $\sigma = 2$ it is 0.076. So, increasing σ does reduce the difference in citation metrics between junior and senior authors.

When we do control for cohort, the distribution of researchers across the top 50 US departments is best explained by total citations, which emphasizes neither *depth* nor *breadth*. Using a much larger and more varied sample of universities based on *CitEc* data and QS rankings, we find qualified support for *depth relevance*. But assignment of researchers to top universities appears to be more closely related to the simple sum of total citations, while assignment to second-tier universities gives more weight to high-impact papers. Similarly, Hamermesh and Pfann (2012) find that in terms of impact on salaries, "a citation is a citation ... having citations concentrated on a few 'home runs' [does not] have an extra impact on salary" (Hamermesh, 2018, p. 144). In fact, controlling for total citations, the number of citations to the most cited paper has a negative effect on the salaries of economists at 43 US public universities. Importantly, Hamermesh and Pfann (2012) do control for cohort effects.

A possible speculative explanation of behavior across the spectrum of universities could be as follows. Lowest-ranked universities, outside of the 400 universities ranked by QS, might simply care about publication without worrying about impact. Having more publications would be better than having fewer at these institutions, suggesting a *breadth relevant* citation metric. Our exploratory analysis that includes universities outside of those ranked by QS supports this. We found that breadth was inversely correlated with average citations in the lower percentiles.

Middle-ranked universities, such as those ranked between 400 and 50 in the QS ranking, care about impact; having some high-impact publications is better than having none and a *depth-relevant* metric describes behavior in this interval.

Finally, among the top-ranked universities such as the QS top 50 or NRC top 25, hiring and tenure committees wish to see high-impact research across all of a researcher's publications and the best-fit metric moves toward *breadth relevance*. Alternatively, researchers at top universities are so good that they produce a large quantity of high-quality work while the journals in which they publish are widely read and thus cited. It may also be that adding lower-impact publications to a publication list that contains high-impact ones is seen as a negative (Powdthavee et al., 2018).

The axiomatic approach to citation metrics by Perry and Reny (2016) substantially advanced the field of scientometrics. While the theorist works out the mapping from axioms to performance metrics, the labor market decides which axioms are important, reflecting the preferences of academics, funders, and students. We here tested three axioms—*depth relevance* vs *breadth relevance* with *independence* as an externality—against each other. It should be clear that the net should be cast wider to include alternative axioms to rank individuals' impact, that the axiomatic approach should be extended to productivity, and to departments and journals—all so that rankings can be improved (Broecke, 2015; Combes & Linnemer, 2003; Coupé, 2003; DuBois & Reeb, 2000; Dusansky & Vernon, 1998; Gibbons et al., 2015; Goldstein & Spiegelhalter, 1996; Kalaitzidakis et al., 1999; Kalaitzidakis et al., 2003; Lubrano et al., 2003; Malsch & Tessier, 2015; Rowlinson et al., 2015; Scott & Mitias, 1996; Stern, 2013; Van Fleet et al., 2000). One particular drawback of the metric used here (and many other metrics) is that all citations are treated equally. That is, we assume that a paper in the *American Economic Review* that is cited 50 times is as important to a career as a paper in *Manchester School* that has 50 citations—even though the latter journal has fewer readers. Similarly, we assume that a citation in a top journal is as important as a citation in a lower-tier journal. Our framework should also be tested with alternative performance measures, such as salaries (Gibson et al., 2017), the placement of young stars (Bryan, 2019), and technology transfer (Showalter & Jensen, 2019). We show that different segments of the market—cohort, field, and rank—view different aspects of academic performance differently, but did not test for other characteristics, such as gender and nationality. All this is deferred to future research.

The results presented here are useful to academics, as it allows them to match their (current or aspirational) citation profile against the profile sought after by economics departments of different ranks. It helps department heads understand what they need to do to move up the ranks, or to prevent a downward slide. Perhaps most importantly, we show considerable heterogeneity between departments of different ranks. Research evaluations that use the same criteria across the rank order may therefore well lead to a mismatch of national and local expectations and a suboptimal allocation of resources. Rank-dependent evaluation criteria or, perhaps more practically, different criteria for different leagues would alleviate this.

ACKNOWLEDGMENTS

Seminar participants had many welcome suggestions for improvement. Ted Bergstrom and two anonymous referees made excellent comments on earlier versions.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ORCID

David I. Stern https://orcid.org/0000-0001-6595-4268
Richard S. J. Tol https://orcid.org/0000-0002-8012-3988

ENDNOTES

[1] Note that there are post hoc axiomatizations of existing citation metrics (Deineko & Woeginger, 2009; Quesada, 2011; Woeginger, 2008).

[2] A researcher has a Hirsch index of h if she has published h papers that are cited h times or more (Hirsch, 2005). A researcher has a $h_{(a,b)}$ index of h if she has published h papers that are cited ah^b times or more (Ellison, 2013). Ellison assumes $b \geq 1$, which favors *depth relevance*. The original Hirsch index, $a = b = 1$, and Ellison's generalization can be *breadth* relevant—moving excess citations $> ah^b$ from publications in the H-core to those just outside it can increase the index. It can also be *depth* relevant—moving citations from the least cited papers to those near the H-core can also increase the index. For greater b, it is more attractive to sacrifice lesser-cited papers for better-cited ones. However, the generalized Hirsch index is never globally *depth* or *breadth* relevant. The new C5 metric proposed by Hamermesh (2018), which sums the citations of the five most-cited papers, is also depth relevant in spirit, although not strictly so. We find that a simple count of citations best explains Ellison's data and so it is not surprising that a tighter focused Hirsch-like metric performs better than the original one.

[3] We ignore one of their axioms, *Directional Consistency*, which states that if two researchers are equally ranked and are still equally ranked after adding a common vector of additional citations to their citation vectors—the additional vector adds the most citations to their most cited paper and then progressively fewer citations to their less and less cited papers—then they will still be equally ranked if a scalar multiple of that vector is added instead. Perry and Reny do not provide any intuitive appeal for this axiom, and we could not detect any ourselves.

[4] Some people judge their colleagues on their worst work (Powdthavee et al., 2018), a violation of monotonicity.

[5] The axiom of *Directional Consistency* additionally implies that $\sigma = 2$.

[6] Perry and Reny (2016) show that monotonicity, independence, and scale invariance together imply CES. They show that, for a CES function, depth relevance implies $\sigma > 1$. It follows that for breadth relevance $\sigma < 1$.

[7] Hirsch (2007) tests whether a combination of the Hirsch (2005) index and the total number of citations $h_\alpha = \sqrt{h^2 + \alpha N_C}$ o a researcher's future papers better than either indicator alone. Using a correlation graph similar to Figures 2 and 4 in this paper he finds the optimal value of α is −0.1. But based on the information in the paper this does not seem to be statistically different from zero, which is similar to what we find below for combinations of metrics. Note that the Hirsch index may be either breadth or depth relevant. The total number of citations is neither breadth nor depth relevant. A combination of these two indicators is not informative about breadth or depth.

[8] Our data includes lower-ranking research universities but we did not collect data on teaching-oriented universities and colleges.

[9] Equation (2) is a CES function with an infinite elasticity of substitution at the top level, while the Cobb–Douglas function is a CES function with an elasticity of substitution of unity.

[10] All data and code can be found here. The file with the complete citation record of all economists on CitEc is large.

[11] Ellison (2013) estimates an ordered probit model that relaxes the assumption that there is a simple cardinal functional relation between rank and the citation metric. This requires estimating department specific parameters. Ellison does not attempt to estimate the parameters of the citation metric jointly with these departmental parameters. Joint estimation would greatly complicate the optimization and likely reduce the precision of the estimated parameters and so we have not pursued this direction here. Instead, we complement our regression type estimates with a rank correlation analysis.

[12] Spearman rank correlations for a university ranking based on total citations in CitEc are 0.86 for NRC and 0.64 for QS. Rank correlations for total number of publications are 0.39 for NRC and 0.43 for QS. The rank correlation between the NRC and QS rankings is 0.93.

[13] Following the suggestion of one of the editors, we also estimated the linear function (2) with fixed values of $\tau = 1$ and $\sigma = 2$ using the semi-log regression specification (4). This represents a linear combination of total citations and Euclidean length. Comparing these estimates to an unrestricted estimate of (3), we cannot reject the restriction using a likelihood ratio test and θ is insignificantly different to zero, reflecting that the optimal value of σ is closer to one than two in Table 2.

[14] Due to excess kurtosis, we always reject the normality assumption using the Jarque–Bera test.

[15] Again, we also estimated the linear function (2) with fixed values of $\tau = 1$ and $\sigma = 2$ using the semi-log regression specification (4). Comparing these estimates to an unrestricted estimate of (3) we cannot reject the restriction using a likelihood ratio test and θ is insignificantly different to one, reflecting that the optimal value of σ is closer to two than one for the semi-log estimates in Table 5.

REFERENCES

Arrow, K.J., Chenery, H.B., Minhas, B.S. & Solow, R.M. (1961) Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics*, 43(3), 225–250.

Broecke, S. (2015) University rankings: do they matter in the UK? *Education Economics*, 23(2), 137–161. https://doi.org/10.1080/09645292.2012.729328.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bryan, K.A. (2019) Young "stars" in economics: what they do and where they go. *Economic Inquiry*, 57(3), 1392–1407. https://doi.org/10.1111/ecin.12762.

Burk, A. (1936) Real income, expenditure proportionality, and Frisch's "new methods of measuring marginal utility". *The Review of Economic Studies*, 4(1), 33–52. https://doi.org/10.2307/2967658.

Combes, P.-P. & Linnemer, L. (2003) Where are the economists who publish? Publication concentration and rankings in Europe based on cumulative publications. *Journal of the European Economic Association*, 1(6), 1250–1308. https://doi.org/10.1162/154247603322752548.

Coupé, T. (2003) Revealed performances: worldwide rankings of economists and economics departments, 1990–2000. *Journal of the European Economic Association*, 1(6), 1309–1345. https://doi.org/10.1162/154247603322752557.

Deineko, V.G. & Woeginger, G.J. (2009) A new family of scientific impact measures: the generalized Kosmulski-indices. *Scientometrics*, 80 (3), 819–826. https://doi.org/10.1007/s11192-009-2130-0.

DuBois, F.L. & Reeb, D. (2000) Ranking the international business journals. *Journal of International Business Studies*, 31(4), 689–704. https://doi.org/10.1057/palgrave.jibs.8490929.

Dusansky, R. & Vernon, C.J. (1998) Rankings of U.S. economics departments. *Journal of Economic Perspectives*, 12(1), 157–170. https://doi.org/10.1257/jep.12.1.157.

Ellison, G. (2013) How does the market use citation data? The Hirsch index in economics. *American Economic Journal: Applied Economics*, 5 (3), 63–90. https://doi.org/10.1257/app.5.3.63.

Gibbons, S., Neumayer, E. & Perkins, R. (2015) Student satisfaction, league tables and university applications: evidence from Britain. *Economics of Education Review*, 48, 148–164. https://doi.org/10.1016/j.econedurev.2015.07.002.

Gibson, J., Anderson, D.L. & Tressler, J. (2017) Citations or journal quality: which is rewarded more in the academic labor market? *Economic Inquiry*, 55(4), 1945–1965. https://doi.org/10.1111/ecin.12455.

Goldstein, H. & Spiegelhalter, D.J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 159(3), 385–443. https://doi.org/10.2307/2983325.

Halevi, G., Moed, H. & Bar-Ilan, J. (2017) Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—review of the literature. *Journal of Informetrics*, 11(3), 823–834. https://doi.org/10.1016/j.joi.2017.06.005.

Hamermesh, D.S. (2018) Citations in economics: measurement, uses, and impacts. *Journal of Economic Literature*, 56(1), 115–156. https://doi.org/10.1257/jel.20161326.

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

Hirsch, J.E. (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102.

Hirsch, J.E. (2007) Does the h index have predictive power? *Proceedings of the National Academy of Sciences*, 104(49), 19193–19198. https://doi.org/10.1073/pnas.0707962104.

Kalaitzidakis, P., Mamuneas, T.P. & Stengos, T. (1999) European economics: an analysis based on publications in the core journals. *European Economic Review*, 43(4–6), 1150–1168. https://doi.org/10.1016/S0014-2921(99)00019-7.

Kalaitzidakis, P., Stengos, T. & Mamuneas, T.P. (2003) Rankings of academic journals and institutions in economics. *Journal of the European Economic Association*, 1(6), 1346–1366. https://doi.org/10.1162/154247603322752566.

Keller, W.J. (1976) A nested CES-type utility function and its demand and price-index functions. *European Economic Review*, 7(2), 175–186. https://doi.org/10.1016/0014-2921(76)90057-X.

Lubrano, M., Bauwens, L., Kirman, A. & Protopopescu, C. (2003) Ranking economics departments in Europe: a statistical approach. *Journal of the European Economic Association*, 1(6), 1367–1401. https://doi.org/10.1162/154247603322752575.

Malsch, B. & Tessier, S. (2015) Journal ranking effects on junior academics: identity fragmentation and politicization. *Critical Perspectives on Accounting*, 26, 84–98. https://doi.org/10.1016/j.cpa.2014.02.006.

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

Perry, M. & Reny, P.J. (2016) How to count citations if you must. *American Economic Review*, 106(9), 2722–2741. https://doi.org/10.1257/aer.20140850.

Powdthavee, N., Riyanto, Y.E. & Knetsch, J.L. (2018) Lower rated publications do lower academics' judgements of publication lists: evidence from a survey experiment of economists. *Journal of Economic Psychology*, 66, 33–44. https://doi.org/10.1016/j.joep.2018.04.003.

Quesada, A. (2011) Axiomatics for the Hirsch index and the Egghe index. *Journal of Informetrics*, 5(3), 476–480. https://doi.org/10.1016/j.joi.2011.01.009.

Rowlinson, M., Harvey, C., Kelly, A., Morris, H. & Todeva, E. (2015) Accounting for research quality: research audits and the journal rankings debate. *Critical Perspectives on Accounting*, 26, 2–22. https://doi.org/10.1016/j.cpa.2013.05.012.

Sato, K. (1967) A two-level constant-elasticity-of-substitution production function. *The Review of Economic Studies*, 34(2), 201–218. https://doi.org/10.2307/2296809.

Scott, L.C. & Mitias, P.M. (1996) Trends in rankings of economics departments in the U.S.: an update. *Economic Inquiry*, 34(2), 378–400. https://doi.org/10.1111/j.1465-7295.1996.tb01383.x.

Showalter, D. & Jensen, R. (2019) University startup intensity and faculty quality. *Economic Inquiry*, 57(2), 855–875. https://doi.org/10.1111/ecin.12752.

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Solow, R.M. (1956) A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1), 65–94. https://doi.org/10.2307/1884513.

Stern, D.I. (2013) Uncertainty measures for economics journal impact factors. *Journal of Economic Literature*, 51(1), 173–189. https://doi.org/10.1257/jel.51.1.173.

Tol, R.S.J. (2013) The Matthew effect for cohorts of economists. *Journal of Informetrics*, 7(2), 522–527. https://doi.org/10.1016/j.joi.2013.02.001.

Van Fleet, D.D., McWilliams, A. & Siegel, D.S. (2000) A theoretical and empirical analysis of journal rankings: the case of formal lists. *Journal of Management*, 26(5), 839–861. https://doi.org/10.1016/S0149-2063(00)00060-X.

Woeginger, G.J. (2008) An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*, 56(2), 224–232. https://doi.org/10.1016/j.mathsocsci.2008.03.001.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Stern, D.I. & Tol, R.S.J. (2021) Depth and breadth relevance in citation metrics. *Economic Inquiry*, 59:961–977. https://doi.org/10.1111/ecin.12994

14657295, 2021, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecin.12994 by Australian National University, Wiley Online Library on [26/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License