UNIVERSITY
OF SUSSEX

ECONOMICS

Working Paper Series

No. 01-2018

How to Count Citations If You Must: Comment

David I. Stern[1] and Richard S.J. Tol[2,3,4,5,6]

[1] Crawford School of Public Policy, The Australian National University
[2] Department of Economics, University of Sussex, UK
[3] Institute for Environmental Studies, Vrije Universiteit, Amsterdam
[4] Department of Spatial Economics
[5] Tinbergen Institute, Amsterdam
[6] CESifo, Munich

david.stern@anu.edu.au; r.tol@sussex.ac.uk

Abstract: Perry and Reny (2016) show that the Euclidean length of a citation list satisfies five axioms including "depth relevance". We explore "breadth relevance", which favors consistent achievers over one-hit wonders. A convex combination of depth and breadth relevant citation metrics does not satisfy the independence axiom, but violations are rare. We estimate the parameters of this metric using two datasets and three rankings, controlling for cohort effects. We find that simply counting citations--neither breadth nor depth--maximizes the correlation between citation index and department rank. However, depth may explain the allocation of researchers across lower ranked departments.

JEL classification: A14; C43

Key words: research assessment; citations

How to Count Citations If You Must: Comment

David I. Stern[a] and Richard S.J. Tol[b c d e f]

[a] Crawford School of Public Policy, The Australian National University, 132 Lennox Crossing, Acton, ACT 2601, Australia, david.stern@anu.edu.au
[b]Department of Economics, University of Sussex, BN1 9SL Falmer, United Kingdom, r.tol@sussex.ac.uk
[c]Institute for Environmental Studies, Vrije Universiteit Amsterdam, The Netherlands
[d]Department of Spatial Economics, Vrije Universiteit Amsterdam, The Netherlands
[e]Tinbergen Institute, Amsterdam, The Netherlands
[f]CESifo, Munich, Germany

December 21, 2017

Abstract

Perry and Reny (2016) show that the Euclidean length of a citation list satisfies five axioms including "depth relevance". We explore "breadth relevance", which favors consistent achievers over one-hit wonders. A convex combination of depth and breadth relevant citation metrics does not satisfy the independence axiom, but violations are rare. We estimate the parameters of this metric using two datasets and three rankings, controlling for cohort effects. We find that simply counting citations—neither breadth nor depth—maximizes the correlation between citation index and department rank. However, depth may explain the allocation of researchers across lower ranked departments.

JEL Classification A14 C43
Keywords Research assessment, citations

1 Introduction

Perry and Reny (2016) propose a new way to count citations: The Euclidean length of a citation list. The innovation in their approach is not so much the new index—new citation indices are proposed regularly—but rather their axiomatic approach.[1] One of their axioms—***Depth Relevance***—implies that citations should be concentrated on a few papers. We show that their empirical support for this axiom weakens when we control for cohort effects. 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 exclusively neither on *breadth* nor *depth*—best maximizes the correlation between researchers' citation indices and department ranks. However, *depth* may explain the allocation of researchers across lower ranked departments.

Perry and Reny (2016) show that the Euclidean index is the only one that satisfies five desirable properties that they argue a citation index should have. One of these five axioms is *depth* relevance, where splitting the citations for a single paper into two separate papers cannot increase the citation index. 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. Its 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 metric does not formally satisfy Perry and Reny (2016)'s *Independence* axiom, but changes in rank due to adding a paper with equal citations to all authors are rare.

Following Perry and Reny (2016), we estimate the parameters of our index using the data compiled by Ellison (2013). Ellison (2013) finds that Hirsch-like indices 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

[1] Note that there are post-hoc axiomatisations of existing citation indices (Woeginger, 2008; Deineko and Woeginger, 2009; Quesada, 2011).

[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. We find that a simple count of citations best

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 and study heterogeneity.

We find that maximum likelihood estimates converge to a function that weights only *breadth* or *depth* relevance. Non-parametric analysis shows no or little 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 Index. There is support for the Euclidean Index in the distribution of researchers across lower-ranked international universities.

In the remainder of the paper, we first present the theory of *breadth* relevant citation indices and indices that are both *breadth* and *depth* relevant. We then turn to the empirical application.

2 Citation Indices

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

Monotonicity The value of the index 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 equally often.

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) show that the above three axioms together imply

explains Ellison's data and so it is not surprising that a tighter focused Hirsch-like index 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., 2017), a violation of monotonicity.

that the citation index C of researcher r with citation numbers $c_{r,i}$ is

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

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

Perry and Reny (2016) add a fourth axiom:

Depth Relevance The value of the index 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 1,000 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 research 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 index 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 $\sigma < 1$. It immediately follows that:

Theorem A citation index that satisfies *Monotonicity*, *Independence*, and *scale Invariance*, cannot simultaneously satisfy *Depth Relevance* and *Breadth Relevance*.

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

No citation index can be both *depth* and *breadth* relevant. As argued in Footnote 2, the Hirsch index is *depth* relevant in some cases and *breadth* relevant in others. Another compromise index 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 subindices differently. Perhaps for top research universities, $\theta \approx 1$, while for teaching-oriented universities and colleges that simply want some evidence of scholarly activity, $\theta \approx 0$. We test and reject the hypothesis for top research universities below.[6]

Proposition L A *linear* combination of citation indices that each satisfy *Monotonicity*, *Independence*, and *Scale Invariance*, satisfies *Monotonicity* and *Scale Invariance* but not *Independence*.

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

[6] We also have data on lower ranking research universities but did not collect data on teaching-oriented universities and colleges.

Proof See Appendix.

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 index prefers researcher A over B, and the *breadth* relevant index 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 simple form that finds a compromise between *depth* and *breadth* relevance involving no additional parameters is the Cobb-Douglas function.

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

This form assumes that those institutions that place some weight on both indices ($0 < \theta < 1$) treat both of them as essential. Taking logarithms, (3) becomes a monotone transformation of a linear combination of monotone transformations of the original metrics.

Remark A logarithmic transformation of the original index satisfies the same axioms as the original index does (Perry and Reny, 2016).

This *remark* and *Proposition L* together imply that index C'' satisfies both *Monotonicity* and *Scale Invariance*.

More generally, we can consider a CES aggregate[7] of the two metrics that allows institutions to consider the two to be substitutable of varying degrees:

$$C'''(c_r) = \left(\theta \left(\sum_i c_{r,i}^{\sigma}\right)^{\frac{\phi}{\sigma}} + (1-\theta) \left(\sum_i c_{r,i}^{\tau}\right)^{\frac{\phi}{\tau}}\right)^{\frac{1}{\phi}} \tag{4}$$

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'''.

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

Proof See Appendix.

A counterexample shows that Equation (4) 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.

[7]Equation (2) is a CES function with an infinite elasticity of substitution at the top level. Equation (3) is a CES function with an elasticity of substitution of unity.

The counterexamples illustrate two key features of the violation of *Independence*: First, the two indices must disagree on the ranking. Second, the additional paper must have a small effect on the dominant index and a large effect on the dominated index. 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 index. 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, we cannot yet judge whether junior researchers will 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.

3 Application

3.1 Data

We use two data sets:[8]

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. 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. Many 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 number of authors in Ellison's file on 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 set these citation numbers to zero. We then removed the records of the six researchers with no citations. After cleaning the data, we have 1,523 authors and 106,016 papers.

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

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. 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 etc. who may not have the same potential level of impact as regular faculty in their departments.

QS ranks universities 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 five 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 index used in the ranking.

We extracted the *CitEc* data between 23 and 25 October 2017 when there were 50,624 researchers with publications registered in *RePEc*. We only include papers with non-zero citations in our analysis as a paper with zero citations has no effect on the citation indices. We, therefore, also removed researchers with no citations. We also removed researchers whose first paper was published before 1956,[9] as there are no researchers in the 1955 and 1954 cohorts and data beyond that point are sparse and particularly noisy (Ellison, 2013) and entirely reflect historical rather than current realities. The number of researchers who we could assign to the 400 universities ranked by QS is 16,420 with 284,886 cited publications.

3.2 Methods

As a first step, we use OLS to regress department ranks on a *depth* relevant citation index ($\sigma = 2$) and *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 4 using parametric maximum likelihood regression and non-parametric 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 regression by their field weights. We divide the citations for each of the authors publications by their field deflator. The citation indices are then computed using these deflated data.

2. Experience adjustment: We deflate the relevant citation index by the mean value of the index 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 received by a cohort 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 researchers 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. As described above, we removed researchers whose first publication was prior to 1956.

[9] This includes Robert Solow whose first publication was in 1953.

We estimate the following regression models:

$$R_r = \alpha + \beta \ln(C_r/C_j) + e_r \tag{5}$$

$$R_r = \alpha + \beta \frac{(C_r/C_j)^\lambda - 1}{\lambda} + e_r \tag{6}$$

$$\frac{R_r^\theta - 1}{\theta} = \alpha + \beta \frac{(C_r/C_j)^\lambda - 1}{\lambda} + e_r \tag{7}$$

where R_r is the rank of the department to which researcher r is affiliated, C_r is their citation index, α is the intercept, β is a regression parameter to be estimated reflecting the elasticity of rank with respect to the citation index, and e_r is a random error term. C_j is the mean citation index for cohort j.

We found that the residuals of the semi-log specification had good properties for the Ellison dataset, while a log-log specification was highly heteroskedastic. Equation (5) is a special case of (6), 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 indices implied by (5). On the other hand, β has a simpler interpretation in (5). Finally, the most general model that we estimate is (7), where the dependent variable is also Box-Cox transformed. Equation (7) has the best residual properties for the *CitEc* dataset. [10]

We estimate the model by maximum likelihood (ML) using the log likelihood function concentrated with respect to the standard deviation. We concentrated out α_j by demeaning R_r in (5) and R_r and the transformed citation index in (6) and (7).

We choose an initial parameter vector and then compute the index C_r for all researchers 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 index and for the CES aggregate (4) and Cobb-Douglas function (3). We estimate the model for 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.

[10] Ellison (2013) estimates an ordered probit model that relaxes the assumption that there is a simple cardinal functional relation between rank and the citation index. This requires estimating department specific parameters. Ellison does not attempt to estimate the parameters of the citation index 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.

For the *CitEc* data we divide the sample into the top 50 universities, which have individual ranks, and the bottom 350 universities, where only bands of ranks are reported. For the Ellison data, we also carried out all analyses on the subsample of 836 researchers who have 100 or fewer publications and, therefore, have untruncated publication lists.

As the regression analysis makes an assumption of cardinality of the NRC or QS ranks, which may not be appropriate, as well as further assumptions underlying maximum likelihood estimation, we also carry out a non-parametric analysis using rank correlation coefficients. The correlation analysis computes the Spearman rank correlation coefficient for the cohort demeaned log index and 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 market outcomes.

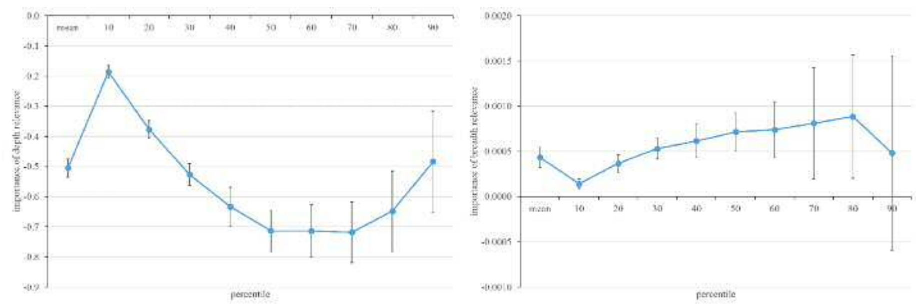
For the single index analysis we compute all the correlations for the cohort-demeaned log citation index, $\ln(C_r/C_j)$, with NRC or QS rank for values of τ or σ from 0.01 to 4. We also do a grid search for the cohort-demeaned log CES function, $\ln(C_r''''/C_j'''')$, computing rank correlations with NRC or QS rank for all permutations of σ, τ, θ, and ϵ over the following ranges 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 (3) 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.

3.3 Results

3.3.1 Exploratory Analysis

Figure 1 shows the impact of a *depth relevant* citation index ($\sigma = 2$) and a *breadth relevant* one ($\tau = 0.5$) on department rank as found by OLS. Departments here include all departments at CitEc with 10 or more registered staff,[11] including teaching-oriented departments. Departments are ranked on the average number of citations per publication. This rank-correlates best with the NRC ($\rho = 0.70$) and QS ($\rho = 0.58$) rankings (among the possible rankings based on a single index contained in the CitEc data), and punishes those who increase quantity without quality. That said, the estimated relationship is part tautological as the data used for the right-hand side is used for the left-hand side too. The result in Figure 1 is an association, rather than a causal relationship. The

[11] Note that "department" is here used to describe departments, schools, colleges, institutes, centres and whatever name people use to denote their main affiliation.



(a) Depth relevance (b) Breadth relevance

Figure 1: Estimated coefficients of the influence of author citation indices (demeaned by cohort) on department rank (CitEc, institutions with 10 or more registered economists) for mean and quantile regression. Panel (a) shows the impact of a citation index that is *depth* relevant ($\sigma = 2$) and Panel (b) the impact of one that is *breadth* relevant ($\sigma = 0.5$).

results are more easily interpreted as the chance that an economist with particular characteristics is hired by a department of a certain rank, but of course the characteristics of individual economists affect the standing of their department. We see that *depth relevance* ($p < 0.0005$) and *breadth relevance* ($p < 0.0005$) 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 departments.

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

Figures 6 and 7 repeat the analysis without demeaning by cohort and for all departments at CitEC—not just those with more than 10 registered researchers. Figures 8 and 9 show the results for the NRC and QS rankings, which are much smaller sets of departments. The findings are qualitatively similar: Depth is appreciated but *breadth* is punished (or insignificant in case of the NRC ranking); higher-ranked departments evaluate citations differently than lower-ranked departments.

3.3.2 Field Effects for Ellison Data

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.

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)

Heteroskedasticity robust standard errors in parentheses

3.3.3 Regression Analysis: Ellison Data

Estimates of all the CES models converge onto an estimate with a single citation index i.e. θ=1 or 0. We present the estimates for the optimal citation index in Tables 2 and 3:

For the full sample, the best fit for σ ranges from 1.32 for the semilog model to 1.10, using the Double Box-Cox model. When we restrict the sample to only those authors with untruncated citation lists we get a range of 1.26 to 0.95. This is very close to simply counting citations, i.e. $\sigma = 1$. All these estimates are statistically insignificantly different from unity but significantly less than 2, which is the power for the Euclidean Citation Index. These values are much lower than Perry and Reny (2016) found because we take cohort effects into account. Their approach treats 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 market outcomes.

We estimate that β is negative and statistically significant in both the full sample and for the top 25 universities. This makes sense, as the lower a department is ranked, the higher the value of its rank. Researchers with higher citation indices are likely to be located at higher ranked departments. However, the measured effect is quite small: A researcher with one standard deviation more than the mean value of the relative citation index for their cohort—which is a little more than doubling the index—is likely to be placed 4 to 5 ranks better than the mean researcher.

Estimates of σ for the top 25 universities are between 1.21 and 1.01. Results for the next 25 universities are not as satisfactory, as neither the power of the citation index nor the slope parameter can be estimated precisely. For the sample of all authors, there is also a local maximum of the likelihood function for $\sigma = 4.43$. For the sample with untruncated citation lists the estimate converges to $\tau = 0$. There is little correlation between rank of department and number

Table 2: Maximum Likelihood Estimates: Ellison Data, Semi-log Model

	All Universities		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 Index 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	

Standard errors for regression coefficients and p-values for Breusch-Pagan test in parentheses.
Significant at the 1% level.

Table 3: Maximum Likelihood Estimates: Ellison Data, Box-Cox Model

	All Universities		Top 25	
	All Authors	$\leq$ 100 Papers	All Authors	$\leq$ 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 Index 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

Standard errors for regression coefficients and p-values for test statistics in parentheses. Significance levels: 5%, 1%.

of citations for this group. 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.

Estimates for all universities using the sample with untruncated citation lists have a slightly lower estimate of σ/τ. This confirms our intuition that ignoring the bottom part of researchers' citation lists would bias the results in favor of depth.

Table 4: Maximum Likelihood Estimates: Ellison Data, Double Box-Cox Model

	All Universities		Top 25	
	All Authors	$\leq$ 100 Papers	All Authors	$\leq$ 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)
θ	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 Index 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

Standard errors for regression coefficients and p-values for test statistics in parentheses. Significance levels: 5%, 1%.

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.[12] Interestingly, the transformation for

[12]Due to excess kurtosis we always reject the normality assumption using the Jarque-Bera test.

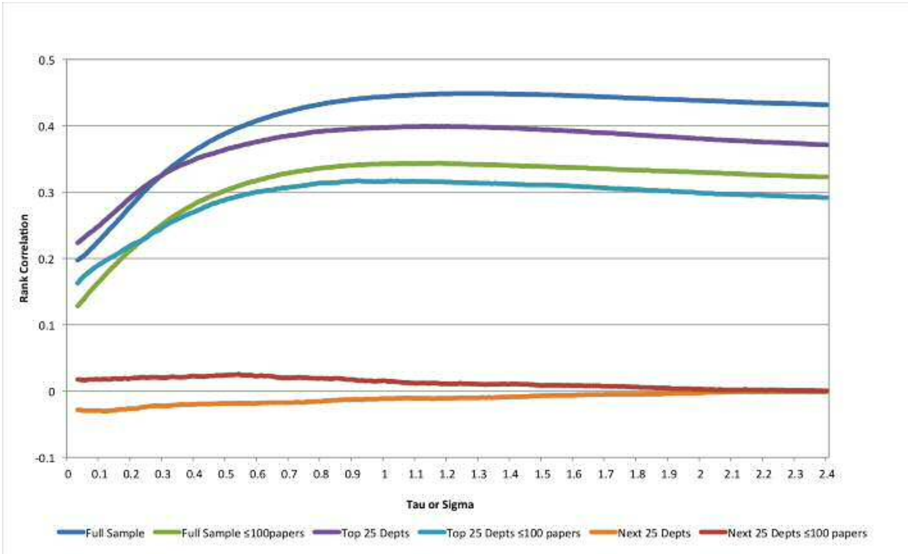


Figure 2: Rank Correlations for a Single Index: Ellison Data

dependent and explanatory variables is the same and closer to the square root function than to logarithms.

3.3.4 Correlation Analysis: Ellison Data

Figure 2 presents the correlations for the full sample of 1523 researchers and subsamples of researchers at the top 25 departments and the next 25 departments. For each of these we computed the correlations for all researchers and for researchers 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. Restricting this sample to authors with $\leq$100 papers reduces the correlations and the maximum now occurs for $\sigma = 1.17$. Eliminating the 25 lower ranked departments also reduces the correlations. The maximum for the full sample is $\sigma = 1.18$ and for the untruncated sub-sample 0.91. These maxima are again clearly quite far from the value of 1.85 obtained by Perry and Reny (2016).

For the next 25 universities the results here also show a lack of correlation between citations and market outcomes. 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 researchers 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* index is almost equal to the number of publications. However, the correlation is only very slightly

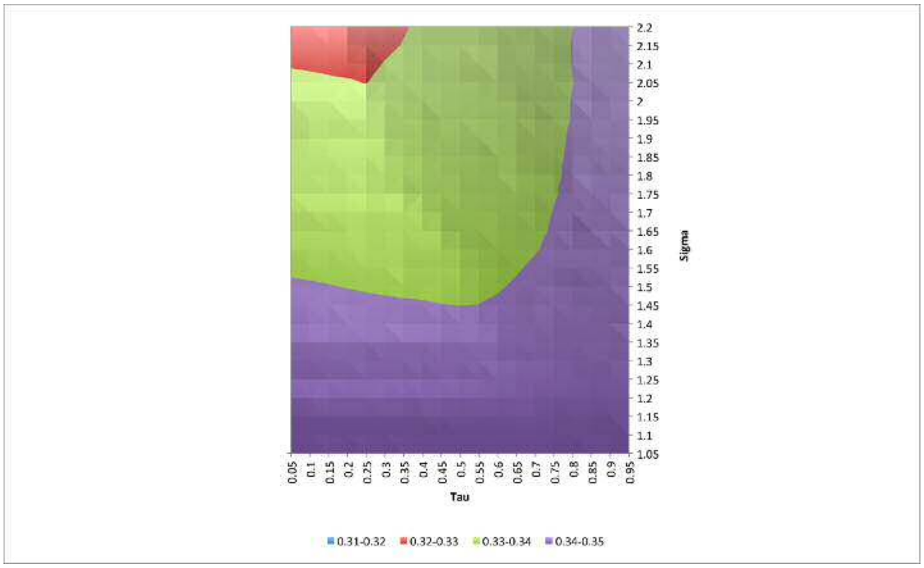


Figure 3: Correlations for the CES Function: Ellison Data, All Universities, ≤ 100 Papers

higher (0.001) than the maximum for the single index 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.

3.3.5 CitEc Data: Regression Analysis

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

For the full sample, the best fit is $\sigma = 2.09$ using the semilog model, $\sigma = 1.57$, using the 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 Index—though the first is not. On the other hand, all three estimates are significantly greater than unity. When we restrict the sample to the top 50 universities we obtain estimates of 1.68 using the semi-log model, 1.24 using the 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 bottom 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 semilog model to 1.08 to 3.05 for the 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.

Table 5: Maximum Likelihood Estimates: CitEc Data

	All Universities			Top 50			Bottom 350		
	Semi-log	Box-Cox	Double Box-Cox	Semi-log	Box-Cox	Double Box-Cox	Semi-log	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)
θ			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 Index 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

Standard errors for regression coefficients and p-values for test statistics in parentheses. Significant at the 1% level.

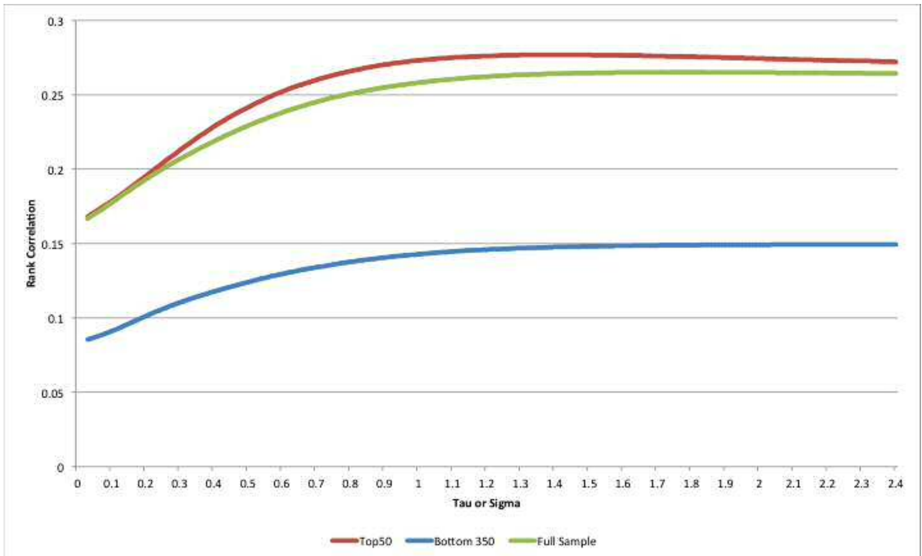


Figure 4: Rank Correlations for a Single Index: CitEc Data

Here, based on the full sample and double Box-Cox model, a researcher with twice the mean value of the citation index for their cohort would be placed 32 ranks higher than the mean researcher. However, this level of relative outperformance 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 bottom 350 universities.

3.3.6 Correlation Analysis: CitEc Data

Figure 4 presents the correlations for the full sample of 16,240 researchers at the top 400 universities, and the subsamples of researchers at the top 50 and bottom 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 the bottom 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 10 shows correlations between the CES function and QS for the full sample. Higher values of σ are associated with higher correlations. Here, the maximum correlation is the same as in the single index 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.

4 Discussion and Conclusions

Ellison (2013) suggests that it is better to have a few highly-cited papers than several well-cited ones. Perry and Reny (2016) axiomatize this. We introduce *breadth relevance* to complement their concept of *depth relevance*. A *breadth relevant* citation index favors consistent achievers over one-hit wonders. The *breadth relevant* citation index is a CES aggregate with an elasticity of substitution τ less than unity, whereas *depth relevant* citation indices have elasticities of substitution (σ) greater than unity. It immediately follows that there can be no citation index that is both *breadth* and *depth* relevant if that citation index is also *monotone*, *independent*, and *scale invariant*. A citation index can be both *breadth* and *depth* relevant at the expense of *independence*.

Using the same dataset as Perry and Reny (2016), we find little empirical support for *depth relevance*. This result is radically different from that of Perry and Reny (2016) because we control for cohort effects in our analysis. Rather, the distribution of researchers across 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.

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 index. 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* index 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 index moves towards *breadth relevance*. Here, adding lower-impact publications to a publication list that contains high-impact ones is seen as a negative (Powdthavee et al., 2017).

The axiomatic approach to citation indexing by Perry and Reny (2016) substantially advanced the field of scientometrics. While the theorist works out the mapping from axioms to performance indices, 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 other axioms. 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.

References

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

Vladimir G. Deineko and Gerhard J. Woeginger. A new family of scientific impact measures: The generalized kosmulski-indices. *Scientometrics*, 80(3): 819–826, Sep 2009. ISSN 1588-2861. doi: 10.1007/s11192-009-2130-0. URL https://doi.org/10.1007/s11192-009-2130-0.

Glenn Ellison. How does the market use citation data? the Hirsch index in economics. *American Economic Journal: Applied Economics*, 5(3):63–90, 2013. doi: 10.1257/app.5.3.63. URL http://www.aeaweb.org/articles.php?doi=10.1257/app.5.3.63.

J. E. Hirsch. 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, 2005. doi: 10.1073/pnas.0507655102. URL http://www.pnas.org/content/102/46/16569.abstract.

J. P. Ostriker, C.V. Kuh, and J.A. Voytuk. *A Data-based Assessment of Research-Doctorate Programs in the United States*. Committee to Assess Research-Doctorate Programs, National Research Council, 2011.

Motty Perry and Philip J. Reny. How to count citations if you must. *American Economic Review*, 106(9):2722–41, September 2016. doi: 10.1257/aer.20140850. URL http://www.aeaweb.org/articles?id=10.1257/aer.20140850.

N. Powdthavee, Y.E. Riyanto, and J.L. Knetsch. *Impact of Lower Rated Journals on Economists Judgements of Publication Lists: Evidence from a Survey Experiment*, volume 10752 of *IZA Discussion Paper*. Institute of Labour Economics, 2017.

Antonio Quesada. Axiomatics for the Hirsch index and the Egghe index. *Journal of Informetrics*, 5(3):476 – 480, 2011. ISSN 1751-1577. doi: https://doi.org/10.1016/j.joi.2011.01.009. URL http://www.sciencedirect.com/science/article/pii/S1751157711000101.

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

Richard S.J. Tol. The matthew effect for cohorts of economists. *Journal of Informetrics*, 7(2):522 – 527, 2013. ISSN 1751-1577. doi: https://doi.org/10.1016/j.joi.2013.02.001. URL http://www.sciencedirect.com/science/article/pii/S1751157713000163.

Gerhard J. Woeginger. An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*, 56(2):224 – 232, 2008. ISSN 0165-4896. doi: https://doi.org/10.1016/j.mathsocsci.2008.03.001. URL http://www.sciencedirect.com/science/article/pii/S0165489608000322.

A Appendix A: Proof of Propositions

Proposition L A linear combination of citation indices that each satisfy *Monotonicity*, *Independence* and *Scale Invariance*, satisfies *Monotonicity* and *Scale Invariance* but not *Independence*.

Monotonicity
If $C(c_r;\sigma_j)$ is monotone, then $\sum_j \theta_j C(c_r;\sigma_j)$ is monotone $\forall \theta_j > 0$.

Scale Invariance
If $C(\lambda c_r) = \lambda C(c_r)$, then $\sum_j \theta_j C(\lambda c_r;\sigma_j) = \lambda \sum_j \theta_j C(c_r;\sigma_j)$ as $\sum_j \theta_j \lambda = \lambda$ for $\sum_j \theta_j = 1$.

Independence
Distinguish two cases. Suppose that
$C(c_A;\sigma) > C(c_B;\sigma)\ \forall \sigma > 1$
and
$C(c_A;\tau) > C(c_B;\tau)\ \forall 0 < \tau < 1$.
Then
$\theta C(c_A;\sigma) + (1-\theta)C(c_A;\tau) > \theta C(c_B;\sigma) + (1-\theta)C(c_B;\tau)\ \forall 0 < \theta < 1$.
That is, both citation indices rank A above B.
Independence implies that
$C(c_A,\delta;\sigma) > C(c_B,\delta;\sigma)$
and
$C(c_A,\delta;\tau) > C(c_B,\delta;\tau)$.
Therefore
$\theta C(c_A,\delta;\sigma) + (1-\theta)C(c_A,\delta;\tau) > \theta C(c_B,\delta;\sigma) + (1-\theta)C(c_B,\delta;\tau)\ \forall 0 < \theta < 1$
In words, if researcher A outranks researcher B on two independent citation indices, then a linear combination of these citation indices is independent.

Now suppose that

$C(c_A;\sigma) > C(c_B;\sigma)$
and
$C(c_A;\tau) < C(c_B;\tau)$

That is, one citation index ranks A over B and the other B over A. A linear combination may rank A over B or B over A. Independence implies that one citation index continues to rank A over B and the other B over A. The ranking according to a linear combination continues to be ambiguous. Adding a paper with δ citations may increase the gap between the two researchers according to one index but close the gap according to the other. Positions may reverse for a linear combination. See the example in the main text. □

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

Monotonicity

If $C(c_r;\sigma_j)$ is monotone then $\left(\sum_j \theta_j C(c_r;\sigma_j)^\phi\right)^{\frac{1}{\phi}}$ is monotone $\forall\theta_j > 0$.

Scale Invariance

If $C(\lambda c_r) = \lambda C(c_r)$, then $\left(\sum_j \theta_j C(\lambda c_r;\sigma_j)^\phi\right)^{\frac{1}{\phi}} = \left(\sum_j \theta_j \lambda^\phi C(c_r;\sigma_j)^\phi\right)^{\frac{1}{\phi}} = \lambda\left(\sum_j \theta_j C(c_r;\sigma_j)^\phi\right)^{\frac{1}{\phi}}$.

Independence

Distinguish two cases. As for *Proposition L*, if researcher A outranks researcher B on two independent citation indices, then a CES combination of these citation indices is independent.

However, if one citation index ranks A over B and the other B over A, a CES combination may rank A over B or B over A. Independence implies that one citation index continues to rank A over B and the other B over A. The ranking according to a CES combination continues to be ambiguous. Adding a paper with δ citations may increase the gap between the two researchers according to one index but close the gap according to the other. Positions may reverse. See the example in the main text. □

B Appendix B: Violations of Independence

As shown above, the proposed citation index violates the axiom of independence. The proposed index is the weighted sum of two indices that satisfy independence, so independence is satisfied if the two indices agree on the ranking. In our sample of 1554 economists, for $\sigma = 2$ and $\tau = 0.5$, the *depth* relevant ranking disagrees with the *breadth* relevant ranking in only 5.4% of cases. This number steadily rises for larger σ and smaller τ, reaching 8.9% for $\sigma = 10$ and $\tau = 0.1$.

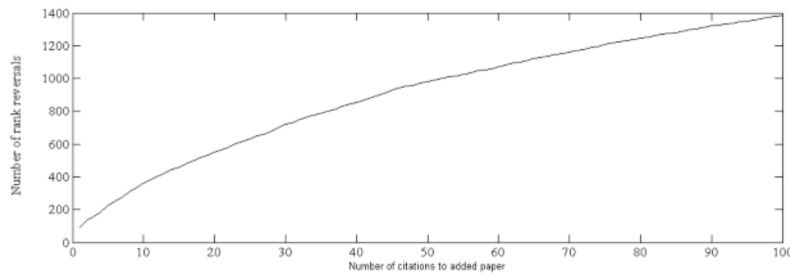


Figure 5: Number of rank reversals for a sample of 1554 economists, for $\sigma = 2$, $\tau = 0.5$, and $\theta = 0.8$, if a paper is added with the number of citations displayed on the horizontal axis.

The citation index C of researcher r with citation numbers $c_{r,i}$ is

$$C(c_r) = \left(\sum_i c_{r,i}^\sigma \right)^{\frac{1}{\sigma}} \Leftrightarrow C(c_r)^\sigma = \sum_i c_{r,i}^\sigma \tag{8}$$

Adding a new paper with δ citations

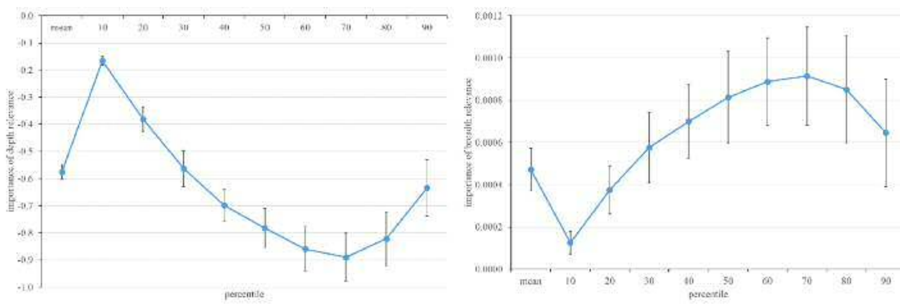
$$C(c_r, \delta) = \left(\sum_i c_{r,i}^\sigma + \delta^\sigma \right)^{\frac{1}{\sigma}} = \left(C(c,r)^\sigma + \delta^\sigma \right)^{\frac{1}{\sigma}} \tag{9}$$

This is a recursive expression, and the recursion carries over to the composite citation index.

For $\sigma = 2$, $\tau = 0.5$, and $\theta = 0.8$, for our sample of 1554 economists, adding a paper that is cited once leads to a rank reversal in 93 cases. There are 2.4 million bilateral rank comparisons, and 130,708 bilaterals that disagree on *depth* and *breadth* ranking. In other words, 93 is a small number. Figure 5 shows that number of rank reversals if the additional paper is cited up to 100 times. That number increases to 1,385, slightly more than 1% of possible rank reversals and a miniscule fraction of all rank comparisons.

The pattern does not change if we instead use $\sigma = 1.15$, $\tau = 0.95$, and $\theta = 0.95$: The number of rank reversals is 1 for one additional citation, increasing to 37 for 100 extra citations. For the optimized Cobb-Douglas function, we find 5 rank reversals for adding a paper that is cited once, rising to 68 for an extra paper cited 100 times. For the optimized CES function, we find no rank reversals.

In sum, while violations of the independence axiom can happen in theory, this is rare in practice.

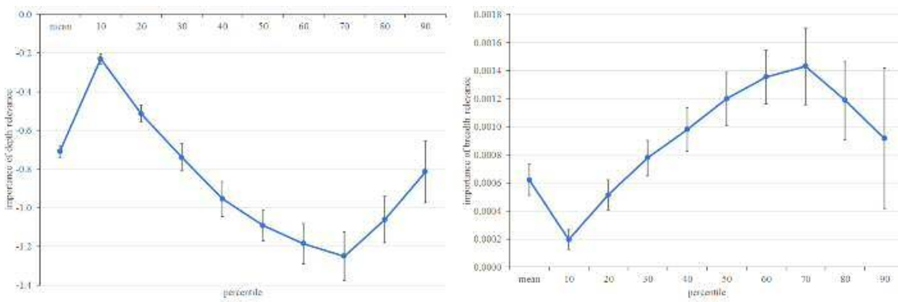


(a) Depth relevance

(b) Breadth relevance

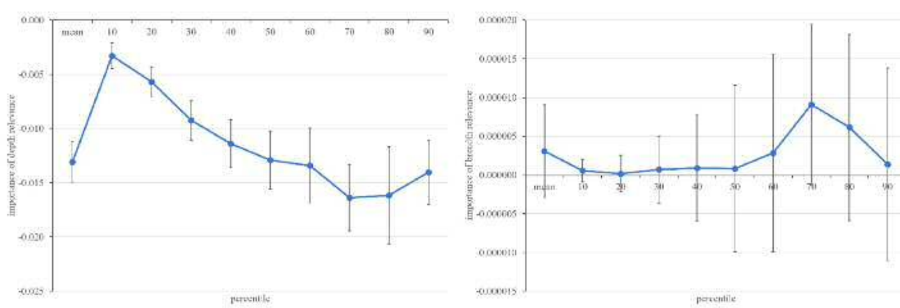
Figure 6: Estimated coefficients of the influence of author citation indices on department rank (CitEc, institutions with 10 or more registered economists) for mean and quantile regression. Panel (a) shows the impact of a citation index that is *depth* relevant ($\sigma = 2$) and Panel (b) the impact of one that is *breadth* relevant ($\sigma = 0.5$).

C Appendix C: Additional results



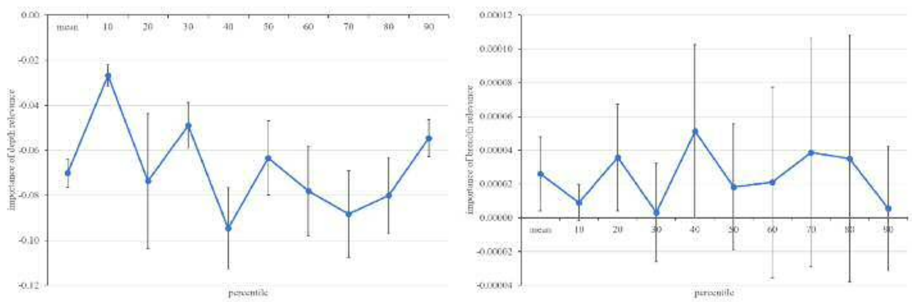
(a) Depth relevance (b) Breadth relevance

Figure 7: Estimated coefficients of the influence of author citation indices on department rank (CitEc, all institutions) for mean and quantile regression. Panel (a) shows the impact of a citation index that is *depth* relevant ($\sigma = 2$) and Panel (b) the impact of one that is *breadth* relevant ($\sigma = 0.5$).



(a) Depth relevance (b) Breadth relevance

Figure 8: Estimated coefficients of the influence of author citation indices on department rank (NRC) for mean and quantile regression. Panel (a) shows the impact of a citation index that is *depth* relevant ($\sigma = 2$) and Panel (b) the impact of one that is *breadth* relevant ($\sigma = 0.5$).



(a) Depth relevance (b) Breadth relevance

Figure 9: Estimated coefficients of the influence of author citation indices on department rank (QS) for mean and quantile regression. Panel (a) shows the impact of a citation index that is *depth* relevant ($\sigma = 2$) and Panel (b) the impact of one that is *breadth* relevant ($\sigma = 0.5$).

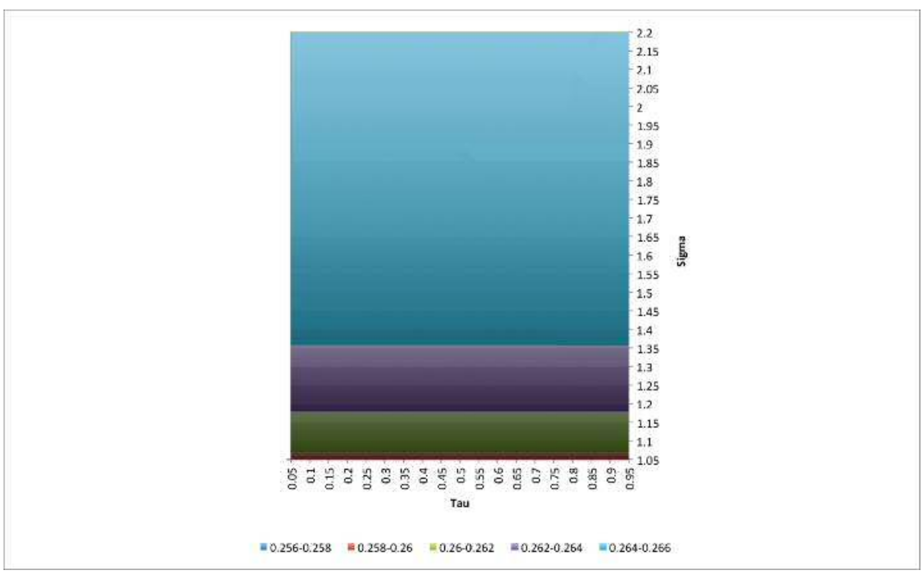


Figure 10: Rank Correlations for CES Function: CitEc Data