

On Measuring Scientific Influence

Martin Ravallion

Adam Wagstaff

The World Bank
Development Research Group
Director's office
July 2010



Abstract

Bibliometric measures based on citations are widely used in assessing the scientific publication records of authors, institutions and journals. Yet currently favored measures lack a clear conceptual foundation and are known to have counter-intuitive properties. The authors propose a new approach that is grounded on a theoretical “influence

function,” representing explicit prior beliefs about how citations reflect influence. They provide conditions for robust qualitative comparisons of influence—conditions that can be implemented using readily-available data. An example is provided using the economics publication records of selected universities and the World Bank.

This paper—a product of the Director’s office, Development Research Group—is part of a larger effort in the department to assess the impact of World Bank research. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at mravallion@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

On Measuring Scientific Influence

Martin Ravallion and Adam Wagstaff¹

*Development Research Group, World Bank
1818 H Street NW, Washington DC, 20433, USA*

¹ Our thanks to Qinghua Zhao for his programming work to help retrieve Google Scholar citation data, and to Imran Hafiz for help retrieving the bibliographic metadata from SCOPUS. The views expressed here are those of the authors and need not reflect those of the World Bank, its Executive Directors, or the countries they represent. Correspondence: mravallion@worldbank.org and awagstaff@worldbank.org.

With the huge expansion in bibliographic data and easier (electronic) access to those data, most researchers have become very aware of how much others have cited their work. Bibliometric measures, such as Hirsch's (2005) famous *h*-index, are now routinely calculated from these data and used to assess the performance of individual researchers, universities and journals. Recruiting, promotion, tenure, employer choice and funding decisions are increasingly influenced by these measures.

As economists coming fresh into bibliometrics, we were surprised how little attention has been given to the logically-prior conceptual issue of what the theoretically ideal measure would look like. Citations are clearly only of interest as an observable indicator for a latent, but more important, concept of "scholarly influence" (more often called "scientific impact" in the literature, but we prefer our more modest term). Yet it is rare to find any explicit discussion in this literature of how one expects citations to reflect scholarly influence. And without being explicit about that relationship it is hard to understand what exactly the *h*-index (or the variations on it since Hirsch's paper) is measuring.

This paper proposes a new class of measures of the impacts of a publication record. Our point of departure from past work is that we build the bibliometric measure on a theoretical "influence function" that embodies our priors about how citations reflect influence. The assumed properties of this function then determine the bibliometric measure, thus making the measure's conceptual foundation transparent.²

While the properties we postulate for the influence function appear to be quite natural, we find that widely-used bibliometric measures have one or more properties that are inconsistent with those assumptions. We argue that these are counter-intuitive properties for a measure of the overall impact of a publication record.

We also recognize explicitly that reasonable people need not agree on the properties of the influence function. Then the question arises as to whether one can derive robust orderings of two or more publication records without knowing the precise way in which citations reflect

² The idea that measures of performance should be built on a precise formulation of the objective function is not, of course, new. The most prominent example we know of has been in a strand of the economics literature on the measurement of income inequality, in which the measure is defined as the loss of aggregate social welfare due to inequality, where social welfare is the sum of individual "utilities," each of which is a stable function of own income; the utility function is taken to be unobserved, and so a matter for prior judgment. The properties of the inequality measure are thus derived from the prior (ethical) assumptions made about social welfare. Dalton (1920) first suggested this approach. An influential formalization and development was provided by Atkinson (1970).

influence. Adapting ideas from the economic theory of stochastic dominance,³ we provide criteria for robust qualitative comparisons of publication records, based on our assumptions about the theoretical properties of the influence function.

The arguments for and against the h -index, as reviewed in the next section, provide a motivation for our approach. We then present our approach in an intuitive (largely non-mathematical) way, before providing a more formal exposition. This is followed by examples.

For and against the h -index

We define a publication record as a list of the publications by a given author (or other entity being compared). Each element of this record has its own citation count. The list of the citations received ranked in descending order of citations can be called the citation profile. These are the data provided by standard sources such as SCOPUS, the *Social Science Citation Index* (SSCI), and Harzing's *Publish or Perish* (POP), which uses Google Scholar.

There are a number of ways one can use these data to assess the scientific impact of a publication record. One way is to calculate the average number of citations per publication. This has the advantage that it measures quality independently of scale (not penalizing small university departments for example). But this reflects a disadvantage too: a researcher or institution that published just one well-cited paper could hardly be considered very productive.

A consensus appears to have emerged in the biological, physical and social sciences that the Hirsch (2005) index (the h -index hereafter) is a useful comprehensive measure of the scientific impact of a publication record, based on its citation profile. An h -index of x means that x is the highest rank in the citation profile such that the first x items received at least x citations. Hirsch argues that this index is a robust and relevant measure of “..the importance, significance and broad impact of a scientist’s cumulative research contributions.” Hirsch’s paper has been widely cited (1200 times in Google Scholar), and the h -index has its own (substantial) [Wikipedia](#) entry. The index has also been popularized by citation software such as POP, Scopus and Web of

³ The theory of stochastic dominance has been mainly used for comparing risky portfolios and in comparing income distributions in terms of social welfare or poverty. An important early contribution in the context of portfolio choice was Hadar and Russell (1971). Applications to social welfare and the measurement of poverty and inequality include Atkinson (1970, 1987), Dasgupta et al. (1973) and Shorrocks (1983). A difference to past applications of dominance theory (that we know of) is that in the present case one cares about the number of objects being compared (the number of publications) as well as their “results” (citations in our case, returns to an investment or incomes in other applications).

Science (which implemented the *h*-index within two years of the publication of Hirsch's paper). Alonso et al. (2009) provide a useful review of the literature on the *h*-index—a review that covers some 90 papers—in just four years since Hirsch (2005).

While the *h*-index has clearly been a major contribution, and appears to have gained wide acceptance, the literature has pointed to some concerns. It is known that the *h*-index can be deceptive in comparing different scientific disciplines, with different referencing cultures. (Hirsch warned against using his index for such comparisons.) One needs to normalize for these differences, and there have been some proposals, as reviewed in Alonso et al. (2009).

The (often-heard) claims about the “robustness” of the *h*-index appear to rest on its insensitivity to lowly-cited papers (see, for example, Vanclay, 2007). However, robustness to other differences in citation profiles is more problematic. Woeginger (2008a) shows that adding a publication with above average cites need not increase the *h*-index for that record. He gives the example of two records: person A has six papers with citations 1,2,3,4,5,6 and B has five papers, with 1,2,3,4,5. Nobody could doubt that A has the better record, yet the *h*-index is 3 for both.

Though we have not seen it done in practice, this type of example can be easily avoided by re-defining the index in a continuous (rather than integer) form. For this purpose, we can define a citation curve as the continuous representation of the citation profile; the citation curve can be obtained by interpolating between the discrete points on the citation profile. On a graph, one simply ranks publications in descending order of citations, and plots citations on the vertical axis and the (cumulative) number of publications on the horizontal axis, and then joins the successive points with (say) straight lines. The continuous version of the *h*-index is then found at the point where the citation curve cuts a 45-degree line from the origin, as illustrated in Figure 1. In the case of Woeginger's example, the continuous *h*-index is 3.5 for A and 3 for B.

However, even the continuous *h*-index can violate the seemingly plausible requirement that extra citations for a given set of papers should yield a higher measure of success for a publication record. Consider the shift in the citation curve indicated by the dashed line in Figure 1; despite the higher citations, the *h*-index is unchanged.

Such observations prompted Egghe (2006) to propose an alternative measure, the *g*-index, which gives higher weight to more highly cited papers. When a publication record has a *g*-index of x it means that x is the highest rank such that the top x papers have at least x^2 citations.

Egghe argues that his measure better reflects the “visibility” of scientists. (We will return to the g -index.)

This is an instance of a broader set of concerns about how the h -index does, or does not, reflect differences in citation profiles. Two identical profiles will naturally have the same h -index. And if citation curves never touch each other then the higher curve always has a higher continuous h -index. However, intersecting citation curves may well be common, given that the way citation counts vary across papers can differ greatly between authors. In particular, the density of citations at any point (including at the value taken by the h -index), and hence the slope of the citation curve at that point, will vary. One author might have a very steep curve, with a high concentration of citations amongst a small number of publications, while it is relatively flat for another author, with a more even profile of the citations received. There are infinitely many citation curves consistent with a given h -index.

An alternative approach

We postulate the existence of an influence function, which gives the degree of scholarly influence implied by any given level of citations. So the influence function can be thought of as a valuation function for citations. Each publication has its own influence, as reflected in its own citations. The properties of the influence function determine the bibliometric measure.

We assume that the aggregate influence of a publication record is the sum of the influences of its constituent publications, each of which depends solely on its own citations. This additivity property is not beyond question. For example, it might be conjectured that the marginal influence of extra citations for a given paper may be lower for an author with many other well cited papers than for an author for very few other cited papers. There is scope for relaxing additivity in the following analysis, to allow for a more general measure of aggregate influence.⁴ However, this complicates the analysis considerably, without much obvious gain. We will stick to the simple additive idea of aggregate influence.

While the influence function is a theoretical concept (in that it cannot be directly observed) we can postulate certain properties that appear plausible on *a priori* grounds. We shall

⁴ The analysis for a non-additive but increasing and quasi-concave (or, more generally, Schur-concave) aggregate influence function would have a number of formal similarities with the analysis in Dasgupta et al. (1973), in the context of measuring income inequality.

assume that the influence function is stable across the units being compared. In other words, the same level of citations implies the same level of influence in each of the publication records. This assumption is implicit in past citation comparisons, such as those using the *h*-index. Here we will only consider comparisons within a given discipline. However, stability can in principle be relaxed by introducing an explicit scaling factor for inter-disciplinary differences in the influence function.

A second assumption about the influence function is monotonicity, meaning that the function is strictly increasing, in that a higher citation count for a publication implies greater influence. This is a natural assumption (though, as we have noted, the *h*-index need not order records consistently with that assumption).

We will say that there is citation-curve dominance when two citation curves do not overlap; one curve is somewhere above the other and nowhere below. When combined with stability, monotonicity implies that a robust comparison of aggregate influence can be deduced from citation-curve dominance. The publication record with the higher citation curve will have had greater aggregate influence, as assessed by any stable increasing influence function.

However, this does not get us far given that citation curves will often intersect. We can still draw conclusions about the aggregate influence of the publication record, but only by making stronger assumptions about the influence function. A powerful, but potentially contentious, additional assumption is diminishing marginal influence, whereby the first citation to a given publication has the highest impact, followed by the second, and so on. This is equivalent to assuming that the influence function is concave.

A simple example of a measure of the overall influence of a publication record satisfying these three assumptions—stability, monotonicity, and concavity—is the quadratic-influence index (*qi*-index), which we define more precisely in the next section. This is the special case of the general class of indices satisfying our assumptions; the key characteristic of this special case is that the marginal influence of extra citations is linear in citations.

We show in the next section that a concave influence function (along with other technical assumptions) implies robust orderings in terms of aggregate influence if the area under the inverse citation curve of one record is everywhere greater than that of the other.⁵ Our empirical

⁵ The inverse citation curve is obtained by flipping the ordinary citation curve—swapping the axes.

example later will illustrate how powerful this assumption can be in resolving otherwise ambiguous rankings.

The assumption of diminishing marginal influence does not hold for the aforementioned g -index, as proposed by Egghe (2006) as an alternative to the h -index. Woeginger (2008b) provides an axiomatic derivation of the g -index, in which one of the axioms is essentially the opposite assumption, namely rising marginal influence. As Woeginger notes, the g -index implies a preference for inequality, in that it rates higher a publication record with a more unequal citation profile (in the sense that citations are transferred from publications with low-cites to those with higher ones at the same mean).

However, we do not think this is a plausible property. The gain in influence indicated by the first citation received by a publication that was previously ignored must surely be larger than the gain in influence that is implied by an extra citation received by the most cited, and hence most influential, paper in a discipline. To put the point another way, imagine two publication records, each containing two publications. In A's record, both papers received 50 citations, while in B's, one paper received 100 citations and the other paper received none. Only one of B's papers is known to have had any influence, while both of A's have demonstrated influence. The concave influence function we postulate here implies that A's record has been more influential, while the convex influence function implies that B's record is the more influential.

Egghe's motivation for the g -index is to reward highly-cited papers, which may carry little or no weight in the h -index, as in Figure 1. However, the approach we take here allows us to address Egghe's concern without building in the (seemingly counter-intuitive) preference for inequality of citations. We can allow that higher citations always imply higher influence (arguably the main feature Egghe wants) but that they do so at a declining rate. In short, we interpret the core motivation for Egghe as avoiding the potential violations of monotonicity that can arise using the h -index, not a desire for a convex influence function *per se*.

So far we have focused on the aggregate influence of a publication record. One might well argue that average influence is also relevant to assessing a record—as an indicator of the average quality of the papers. In comparisons of journal quality, for example, it is not obvious that one solely wants a measure of aggregate influence, given the often large variation between journals in the number of articles they publish per year and in the number of years they have been publishing. We can assess average quality by normalizing the citation curve by the total

number of publications (so that the horizontal axis gives percentiles of total publications). We can define a normalized h -index, whereby the highest $x\%$ of publication in terms of citations received at least x citations.⁶ We can also show that dominance in the normalized curve implies that the higher curve has higher influence per publication, for any influence function. Similarly, if the normalized curves intersect, one may still be able to make a robust comparison for the subset of influence functions exhibiting diminishing marginal influence.

This overview has asserted a number of claims that require a more precise mathematical demonstration. Next we outline our assumptions and define the qi -index, after which we provide the dominance criteria for robust orderings. We then turn to the empirical examples.

A more formal exposition

Let the influence of c citations for a given publication be $I(c)$. (For analytic convenience we treat citations, c , as a continuous variable.) We make the following assumptions throughout:

Core assumptions: The influence function, $I(c)$ for c in $[0, c^{\max}]$, is: (i) stable across the units being compared;⁷ (ii) continuous and differentiable in $[0, c^{\max}]$; (iii) monotonically increasing, i.e., $I'(c) > 0$ for all c in $[0, c^{\max}]$;⁸ and (iv) normalized such that $I(0)=0$ and $I(c^{\max}) = 1$.

The core assumptions describe a general class of influence functions. When citations curves intersect, we are also interested in seeing whether robust comparisons are possible for a subset of influence functions satisfying the following additional assumptions:

Extra assumptions: The influence function is also: (i) strictly concave, i.e., $I''(c) < 0$ for all c in $[0, c^{\max}]$; and (ii) with marginal influence going to zero in the limit as c^* is reached, i.e., $I'(c^{\max}) = 0$.⁹

⁶ This should not be confused with the h -index per publication, which is also called the “normalized h -index” in some of the literature (Alonso et al., 2009).

⁷ The following results can be modified to allow for a multiplicative scaling factor, γ_i , so that the influence function for records of type i is $\gamma_i I(c)$.

⁸ This can be weakened to allow $I'(c) = 0$ for some c .

⁹ Condition (ii) is not essential for our main results, though it simplifies things, and by allowing a sufficiently large c^{\max} it does not seem unduly restrictive to set $I'(c^{\max}) = 0$.

The overall influence of a record is taken to be the sum of the influences across all publications in that record. The aggregate influence of the i 'th record is then given by:

$$I_i \equiv N_i \int_0^{c^{\max}} I(c) f_i(c) dc \quad (1)$$

Here we let N_i denote the number of publications in i 's record (treated as continuous) while $f_i(c)$ denotes the (continuous) density function. Note that the marginal influence function, $I'(c)$, can be interpreted as the weight attached to c citations. Setting $I(0)=0$ and $I(c^{\max})=1$ (condition (iv) in the core assumptions) is equivalent to normalizing the weights to sum to unity:

$$\int_0^{c^{\max}} I'(c) dc = 1 \quad (2)$$

To help interpret (1) it is useful to re-write it in the following form:

$$I_i = (1 - D_i) N_i \quad (3)$$

Here D_i can be thought of as a discount factor, and the aggregate influence is essentially a discounted measure of the number of publications.¹⁰ After integrating by parts and re-arranging, we can write the discount factor as:

$$D_i \equiv \int_0^{c^{\max}} I'(c) F_i(c) dc \quad (4)$$

in which $F_i(c)$ is the cumulative distribution function for citations (the proportion of publications receiving no more than c citations). Thus the discount factor is a weighted mean of the points on the cumulative distribution function of citations. A weaker publication record, in the sense of having a larger proportion of lowly cited papers, will be discounted more heavily. At one extreme, none of the publications get any citations, and so we have $F_i(c) = 1$ for all c in $[0, c^{\max}]$. Then $D_i = 1$ and aggregate impact is zero. At the other extreme, the aggregate influence of a record in which every paper attained c^{\max} citations would simply be the number of publications, which is the maximum value of I across all possible records.

¹⁰ D has some similarity to a measure of inequality, but note that (unlike an inequality measure) D does not go to zero when all publications have the same citations unless that is at the level c^{\max} .

The *qi*-index: An example of an influence function satisfying both our core and extra assumptions is the quadratic influence function: $I(c) = (2 - c/c^{\max})c/c^{\max}$. (The marginal influence function is: $I'(c) = (1 - c/c^{\max})2/c^{\max}$.) Inserting this into (1) we obtain our *qi*-index, for which the discount factor takes the form:

$$D_i = \frac{2}{c^*} \int_0^{c^{\max}} \left(1 - \frac{c}{c^*}\right) F_i(c) dc \quad (5)$$

In other words, the weights are the proportionate deficits relative to maximum citations. On integrating by parts a second time, we obtain a very simply formula for the *qi*-index:

$$I_i = \frac{2N_i}{(c^{\max})^2} \int_0^{c^{\max}} G_i(c) dc \quad (6)$$

where:

$$G_i(c) \equiv \int_0^c [1 - F_i(x)] dx \quad (7)$$

To further interpret (6), notice that $N_i [1 - F_i(c)]$ is the number of papers that received more than c citations, and that this is the inverse of the citation curve, $c_i = c_i(x)$. (The continuous *h*-index is the solution to $h_i = c_i(h_i)$.) $N_i G_i(c)$ is the cumulative of the inverse citation curve. Thus the *qi*-index is twice the total area under this curve, normalized by $(c^{\max})^2$.

Similarly, if we normalize the citation curve by the total number of publications (so that the horizontal axis gives the percentile of publications rather than number), we can readily calculate the cumulative values of the inverse of this normalized citation curve. For the quadratic influence function, average influence is directly proportional to the total area under this curve.

The *qi*-index can be encompassed within a larger class of parametric influence functions:

$$I(c) = (2 - \alpha c/c^{\max})c/c^{\max} \quad (0 \leq \alpha \leq 1) \quad (8)$$

This relaxes one of our assumptions, by allowing the possibility that $I'_\alpha(c^{\max}) = 2(1 - \alpha)/c^{\max} > 0$ (for $\alpha < 1$). Also, by setting $\alpha < 1$ one can adjust the curvature of the influence function (making the function “more linear” as α falls). Thus α parameterizes how rapidly marginal influence declines as the level of citations rises. At the other extreme to the *qi*-index ($\alpha = 1$), the linear

version ($\alpha = 0$) is equivalent to measuring aggregate influence by simply counting total citations. Average influence is then measured by average citations per article.

Dominance tests: Given that any precise functional form for the influence function involves an arbitrary choice, it is of interest to ask how far we can get in ranking publication records using only our theoretical assumptions. Invoking the core assumptions, we have:

First-order citation dominance: *Comparing the records of A and B, if*

$N_A[1 - F_A(c)] \geq N_B[1 - F_B(c)]$ for all c in $[0, c^{\max}]$, with strict inequality for some c , then A has higher aggregate influence than B for any influence function satisfying our core assumptions.

This follows from the fact that (on integrating (1) by parts):

$$I_i = N_i \int_0^{c^{\max}} I'(c)[1 - F_i(c)]dc \quad (9)$$

Note that the normalized citation curve can be used to rank records in terms of average influence per publication for any influence function satisfying our assumptions. (This follows from (9).) On also noting that the normalized citation curve— c plotted against $1 - F_i(c)$ —is the mirror image of the inverse of the distribution function, $F_i(c)$, it is plain that if the normalized citation curve for publication record A is everywhere above that for B then influence per publication is higher for A for any influence function satisfying our core assumptions.

If the citation curves intersect then a robust comparison in terms of aggregate influence might still be possible by invoking our extra assumptions, using the following result:

Second-order dominance: *Under the core and extra assumptions, if $N_A G_A(c) \geq N_B G_B(c)$ for all c in $[0, c^{\max}]$ with a strict inequality somewhere, then we can still conclude that record A has the higher aggregate influence than B.*

To verify this claim, we simply integrate (9) by parts, giving:¹¹

$$I_i = -N_i \int_0^{c^{\max}} I''(c)G_i(c)dc \quad (10)$$

¹¹ Note that $G_i(0) = 0$ and that $I'(c^*) = 0$ (under the extra assumptions). However, our second-order dominance condition also holds for $I'(c^*) > 0$.

Note that, if the normalized citation curves intersect we may still be able to establish a robust ordering in terms of average influence by imposing declining marginal influence. By a similar argument, the corresponding second-order dominance condition for this case is that $G_A(c) \geq G_B(c)$ for all c , with strict inequality for some c . If this holds then one can conclude that the publication record of A has greater average influence per publication for any influence function satisfying both our core and extra assumptions.

Empirical examples

We illustrate these ideas using comparisons of the economics publications records of selected universities and the World Bank.

We rely on [Google Scholar](#) for citation counts. Compared to other bibliographic databases, Google Scholar casts a broad net, including citations by books, working papers, conference proceedings, open-access journals, new and less well-established journals. Google Scholar is also more “global” in its reach, as it includes research outputs from everywhere in the world and all languages. Google Scholar is also timelier than the bibliographic databases.

No time limit was set for the data on citations, since all are old institutions; in practice the records we use date back to 1964, but only five percent of them are before 1980. We searched in SCOPUS for articles where any author was affiliated with one of the institutions we consider; multiple records are therefore possible in the dataset where collaboration has taken place across two or more of the institutions. Only articles in the discipline “Economics, Econometrics and Finance” were included in our analysis. Google Scholar citation data were then obtained for each article in our database.

Table 1 gives the summary statistics on citations for Berkeley, Harvard, London School of Economics (LSE), Oxford, Princeton and Yale, plus the World Bank. The h -index ranking is Harvard (first), Berkeley, World Bank, Princeton, Yale, LSE, and Oxford. This is exactly the same ranking as our qi -index and the two series have a correlation coefficient of 0.98. The ranking in terms of mean citations per article is Harvard, Berkeley, Princeton, World Bank, Yale, LSE, Oxford, which is similar to the ranking in terms of our normalized h -index.

However, not all these rankings are robust to allowing all possible influence functions satisfying our core assumptions. Figure 2 shows the citation curve; panel (b) enlarges the area around the origin to show more clearly the relative positions of the curves. There are a number of

intersections, although some robust comparisons are still possible. For any increasing influence function satisfying our core assumptions we can robustly claim that both Berkeley and Harvard have had more influence in the field of economics than Princeton, Oxford or LSE. The rankings of Harvard, Berkeley, Yale and the World Bank are not so robust, as they depend on the precise form of the influence function. A function that gives a sufficiently high weight to high citations would favor Berkeley, Yale and the World Bank over Harvard, while this reverses for measures that put lower weight on high citations.

We can go a long way toward resolving the rankings if we confine attention to influence functions that exhibit declining marginal influence. Figure 3 gives the cumulative areas under the (inverse) citation curves, to test for second-order dominance. The curves are indistinguishable at low levels but fan out markedly above that. The overall ranking that emerges is exactly the same as that implied by the h -index and the qi -index (Table 1).

When we calculated the normalized citation curves we also found that robust comparisons of average influence per publication are elusive over all possible influence functions satisfying our core assumption. However, the ranking is clearer when one tests for second-order dominance, which is given in Figure 4. The ranking is Harvard, Princeton, Berkeley, World Bank, Yale, LSE, Oxford. This is very close to the ranking in terms of mean citations in Table 1 (noting that Berkeley is very close to Princeton in this dimension).

Conclusions

We have proposed an approach to bibliometric measurement that has a clearer conceptual foundation than past, essentially ad hoc, approaches. The core of the idea is an “influence function,” representing how citations reflect scholarly influence. Only by building bibliometric measurement on an explicit influence function can we be sure that the measure is consistent with our priors about how citations reflect influence. However, the function is taken to be a theoretical concept, which (unlike citations) cannot be observed.

We have outlined what we believe to be plausible assumptions about the influence function. Rankings of publication records using standard bibliometric measures, such as average citations, the h -index and the g -index, need not conform to the rankings implied by our theoretically ideal measure. We have proposed a simple bibliometric measure satisfying all our assumptions, namely the qi -index. We have also shown that, depending on the assumptions made

about the influence function, robust qualitative comparisons can be derived by exploiting ideas from the theory of stochastic dominance. All the proposed tests use readily available data on the citation profile of publication records.

We have illustrated these ideas using data on the citation records in economics of six prominent universities and the World Bank. We find that the h -index rankings of aggregate influence are not robust to allowing any stable influence function that rewards higher citations. However, more robust rankings emerge if one also assumes that the marginal influence of extra citations declines with the level of citations. This turns out to be a powerful assumption in assessing publication records. Rankings in terms of average influence per publication based on mean citations are similarly sensitive to the underlying weights, though robust rankings emerge if one imposes declining marginal influence.

While the h -index has questionable theoretical properties (as the literature has noted) it is striking that it ranks institutions in our empirical example identically to our theoretically ideal measure. And, across institutions, the cardinal values of the h -index turn out to be very highly correlated with our qi -index. For this empirical application at least, the rankings based on the h -index are reliable. Our theoretically ideal rankings of average influence also turn out to be very close to those based on average mean citations per paper. It remains to be seen whether these two standard bibliometric measures perform as well in other applications.

Table 1: Summary data on citations to economics papers for selected universities and World Bank

	N	Median citations	Mean Citations	h -index	Normalized h -index	qi -index
Harvard University	2191	27	96.16	240	41.0	66.6
London School of Economics	1634	13	43.59	124	30.2	22.8
Oxford University	1117	11	33.45	94	28.2	12.1
Princeton University	1299	26	78.09	165	40.5	32.6
Univ. California, Berkeley	2612	22	78.13	206	37.0	62.5
World Bank	2448	21	67.50	189	37.7	52.2
Yale University	1470	17	65.33	140	34.4	29.8

Notes: Only articles indexed in SCOPUS and in ‘Economics, Econometrics and Finance’ were included. Multi-disciplinary articles were excluded. Bibliographic metadata were collected from SCOPUS in May-June 2010, and Google Scholar citation data were collected in June 2010. The qi -index uses $c^*=6,063$, the highest citation count for any paper.

Figure 1: Citation curves and the continuous h -index

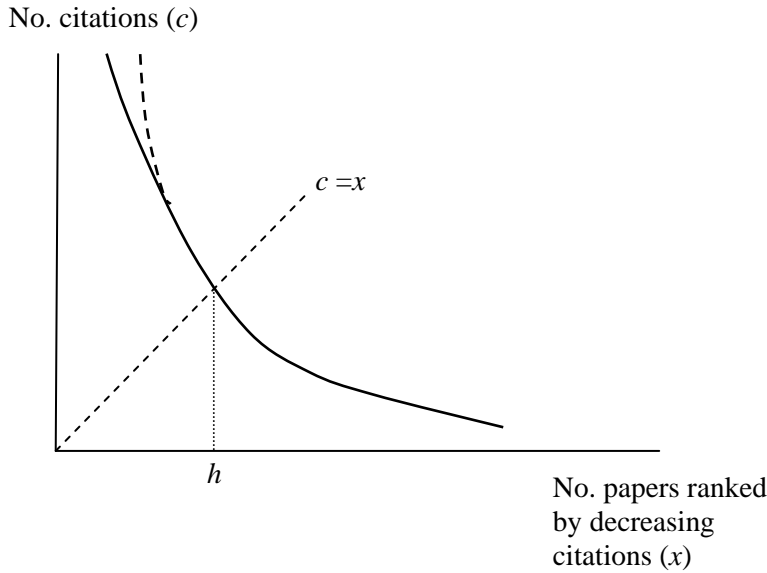
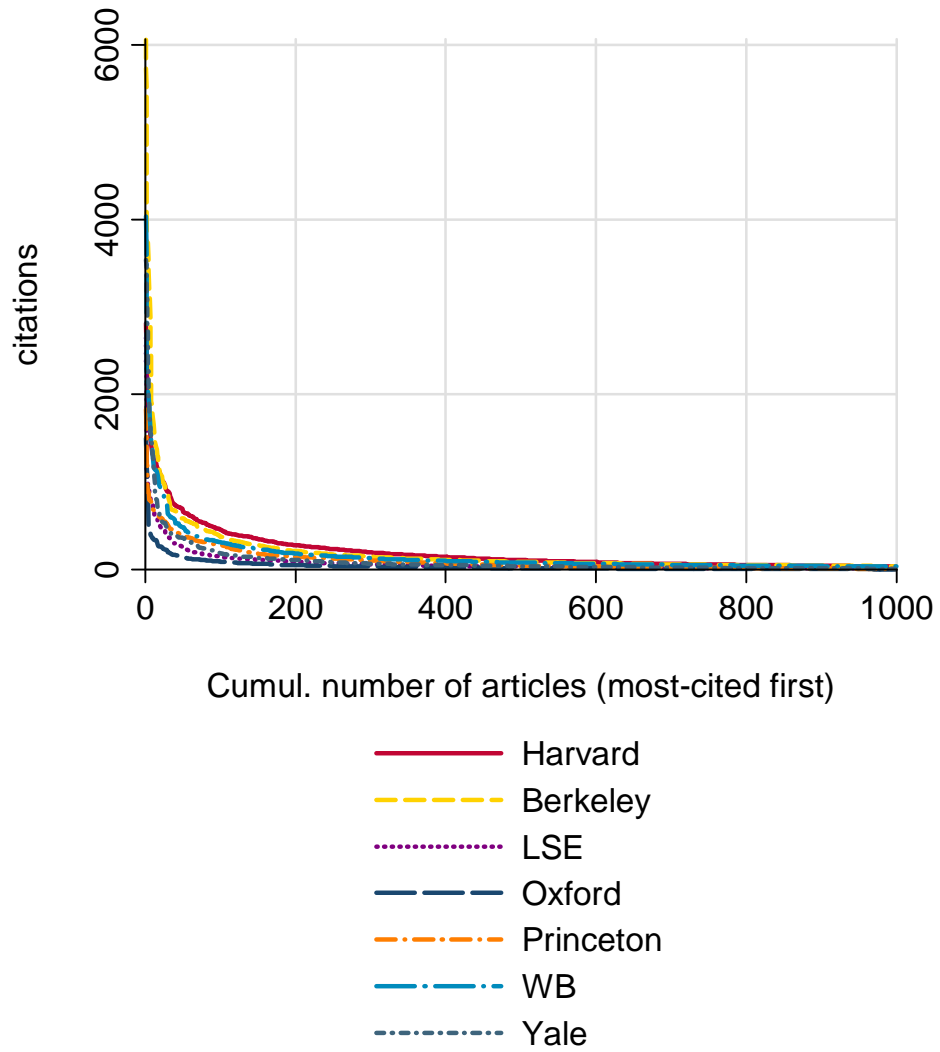


Figure 2: Citation curves for various universities and the World Bank (WB)

(a) Full curves



(b) Blow up of lower segment

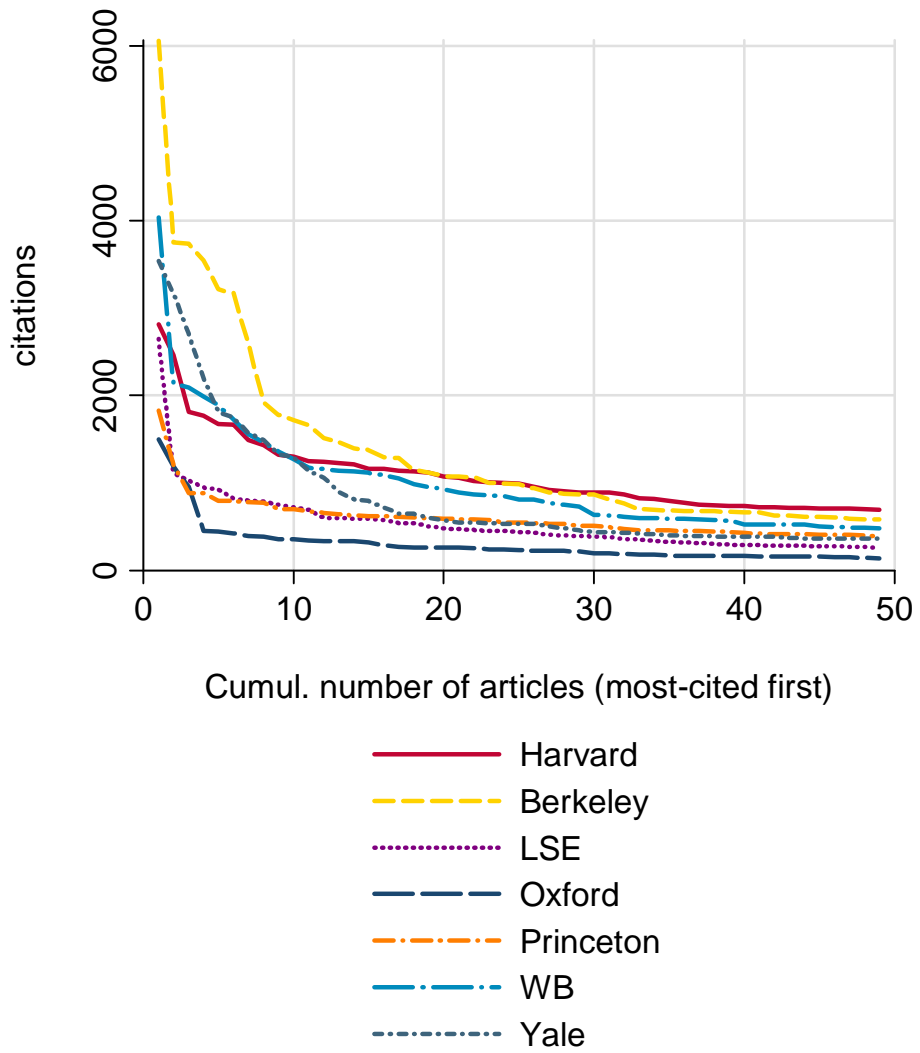


Figure 3: Second-order dominance test for aggregate influence

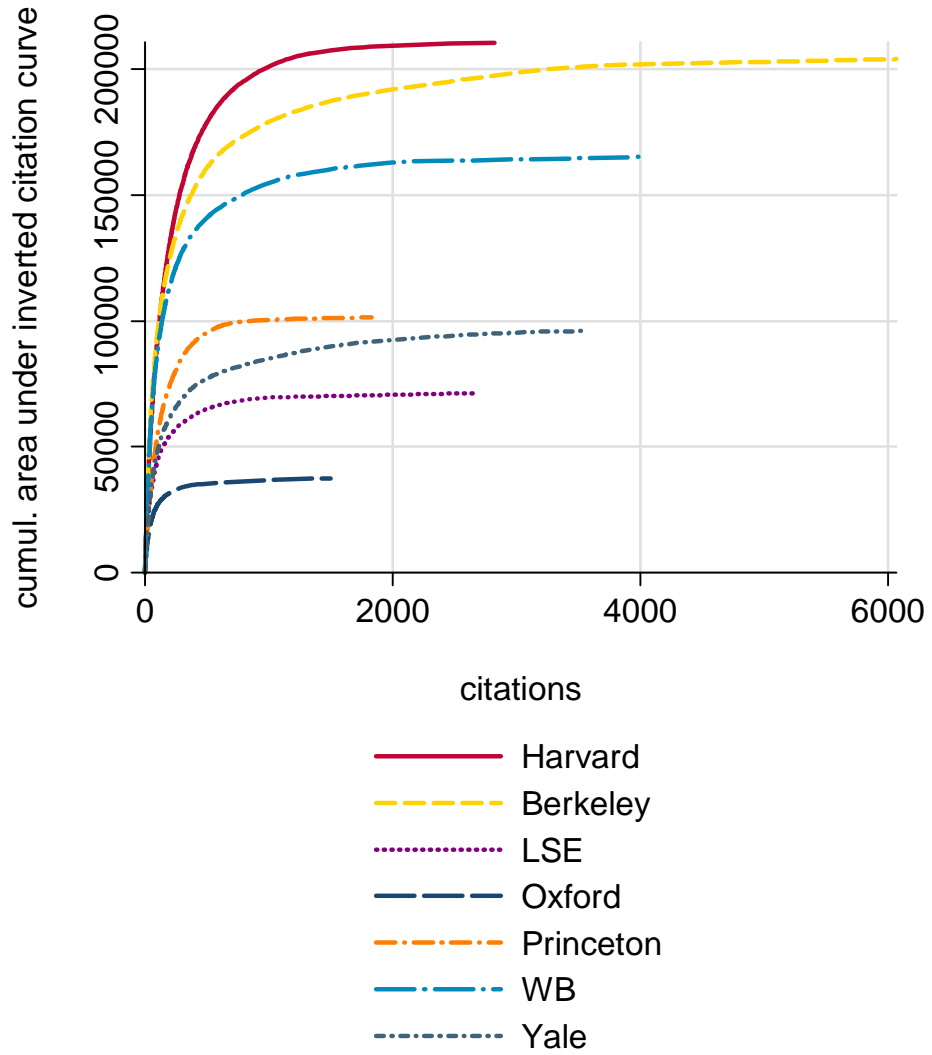
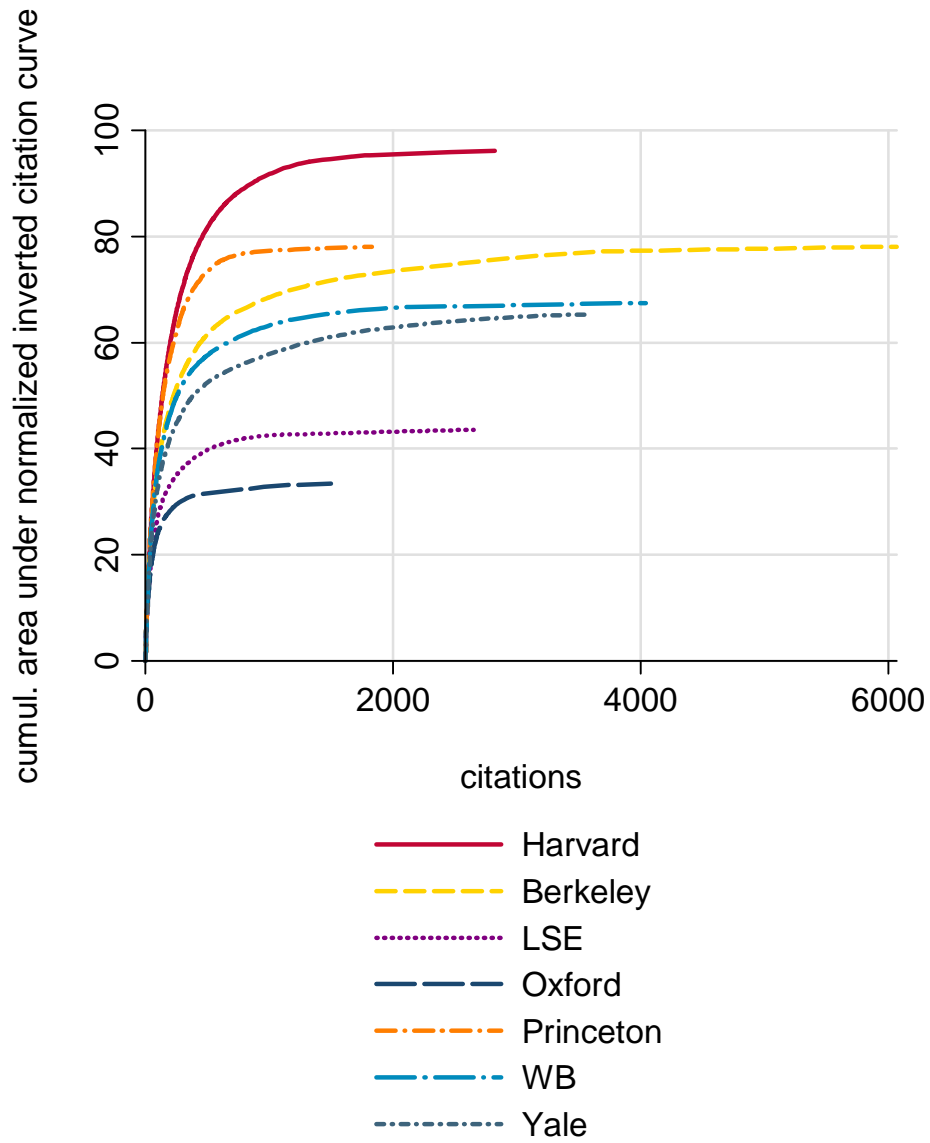


Figure 4: Second-order dominance test for average influence per publication



References

- Alonso, S., F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera. 2009. "h-Index: A Review Focused in its Variants, Computation and Standardization for Different Scientific Fields," *Journal of Informetrics* 3: 273-289.
- Atkinson, Anthony B. 1970. "On the Measurement of Inequality," *Journal of Economic Theory* 2: 244-263.
- _____. 1987. "On the Measurement of Poverty," *Econometrica* 55: 749-64.
- Dalton, Hugh. 1920. "The Measurement of the Inequality of Incomes." *Economic Journal* 30: 348-361.
- Dasgupta, Partha, Amartya Sen and David Starrett. 1973. "Notes on the Measurement of Inequality." *Journal of Economic Theory* 6: 180-187.
- Egghe, Leo. 2006. "Theory and Practice of the g-index," *Scientometrics* 69(1): 131-152.
- Hadar, Josef and William R. Russell. 1971. "Stochastic Dominance and Diversification," *Journal of Economic Theory* 3: 288-305.
- Hirsch, Jorge E. 2005. "An Index to Quantify an Individual's Scientific Research Output." *Proceedings of the National Academy of Sciences* 102(46): 16569-16572.
- Shorrocks, Anthony F. 1983. "Ranking Income Distributions," *Economica* 50: 3-17.
- Vanclay, Jerome. 2007. "On the Robustness of the h-index," *Journal of the American Society for Information Science and Technology* 58(10): 1547-1550.
- Woeginger, Gerhard J. 2008a. "An Axiomatic Characterization of the Hirsch-Index," *Mathematical Social Sciences* 56: 224-232.
- _____. 2008b. "An Axiomatic Analysis of Egghe's g-index," *Journal of Informetrics* 2: 364-368.