

# How Does the Market Use Citation Data? The Hirsch Index in Economics<sup>1</sup>

Glenn Ellison  
MIT and NBER

August 2010

<sup>1</sup>I thank Sarah Venables, Cathy Yao, and especially Hongkai Zhang for outstanding research assistance. Financial support was provided by the NSF (SES-05508897), the Toulouse Network for Information Technology, and Microsoft Research.

## **Abstract**

A large literature following Hirsch (2005) has proposed citation-based indexes that could be used to rank academics. This paper examines how well several such indexes match labor market outcomes using data on the citation records of young tenured economists at 25 U.S. departments. Variants of Hirsch's index that emphasize smaller numbers of highly-cited papers perform better than Hirsch's original index and have substantial power to explain which economists are tenured at which departments. Adjustment factors for differences across fields and years of experience are presented.

Keywords: Hirsch index, citation data, economics profession

# 1 Introduction

In academic fields there is a strong desire for “objective” numerical indexes that can be used to quantify research output. A recent paper by Hirsch (2005) proposing that scientists could be ranked using a simple index  $h$  computed from citation data has received a remarkable amount of attention. This includes scores of papers proposing alternate indexes that their authors argue are even better (mostly on intuitive grounds). In this paper I propose that a useful criterion for assessing competing indexes (and for assessing whether the indexes have any value at all) is to examine whether they are consistent with labor market outcomes. I carry out such an exercise for the economics profession using data on the citation records of young tenured economists at 25 U.S. departments. Among the findings are that Hirsch-like indexes are strongly correlated with labor market outcomes in economics (when subfield and age controls are used) and that variants that place more emphasis on a smaller number of highly cited papers are preferable to Hirsch’s original index. A comparison with the long-standing practice of citation counting indicates that Hirsch-like indexes are an advance over previous practice, but not a revolutionary one.

Hirsch’s (2005) paper has a remarkable citation record – it has already accumulated over 1000 citations according to Google scholar – in part because it came at a time when bibliometric data are gaining prominence. For example, the U.K. announced its intention to transform its Research Assessment Exercise (RAE) from a process relying on peer review to a process relying on bibliometric data. Many scholars and professional societies have decried such moves and it is clear that careful expert evaluation can provide much that citation data will never capture.<sup>1</sup> But it is also clear that there is great demand for statistical evaluation tools: screening devices are needed to select candidates for closer examination among the multitude who could potentially be considered; groups providing peer review must themselves be monitored by less expert administrators; and the quality of real-world peer review can be far less than the ideal. Within the bibliometric space, citation data seem to be gaining prominence relative to publication count data. A number of factors may be contributing to this: technological advances including the development of Google Scholar have made citation data more accessible; the proliferation of journals and muddling of journal hierarchies make the signal that a paper was published by a particular journal less meaningful; and the reduced role of journals in information dissemination creates a greater need to assess papers not published in traditional peer-reviewed journals (Ellison, 2010).

Hirsch’s  $h$  index is a simple clever construction: the index is defined to be the largest

---

<sup>1</sup>See, for example, Adler et al. (2008).

number  $h$  such that the researcher has at least  $h$  papers with  $h$  or more citations. The index has some attractive features. One is that it deemphasizes the number of citations to a researcher's most-cited paper. Citations have a very fat-tailed distribution so a single paper can contribute a large share of a researchers' total citations. This is especially problematic when there is uncertainty about how much credit each author should receive for a paper, e.g. when the most-cited paper is a paper coauthored as a student with an advisor. Another is that the focus of the index shifts in a natural way when comparing researchers at different levels. When comparing young researchers it emphasizes whether they have written a few papers that have had some impact, and when comparing distinguished senior researchers it ignores minor papers and considers only papers that have a substantial number of citations.

Despite these attractive features, the  $h$  index is unappealing on its face when applied to economists. Economists write fewer papers than do physicists and individual papers get many citations. As a result, the  $h$  index starts to look uncomfortably like a publication count. An illustrative example is that Nobel laureate Roger Myerson's  $h$  index is an unexceptional 32 – twelve of the 166 much younger economists in my sample already have  $h$  indexes at least as large.<sup>2</sup> The  $h$  index is poorly aligned with the profession's view of Myerson because the profession cares much about the tremendously important papers Myerson has written than about the impact of his 33rd most cited paper.

Hirsch's original paper argued that an attractive feature of the  $h$ -index is that it is not "arbitrary" as would be, for example, counting the number of papers with at least 100 citations. I would make a doubly opposite argument: Hirsch's index is highly arbitrary, but this arbitrariness should be seen as an attractive feature. Hirsch's index is arbitrary because he could have chosen many variants. For example, we can define the  $(a, b)$  *generalized  $h$ -index*  $h_{(a,b)}$  to be the largest number  $h$  such that a researcher has at least  $h$  papers with  $ah^b$  citations each. From this perspective, Hirsch's index arbitrarily sets multiple parameters equal to one. I see this arbitrariness as attractive because variants of the  $h$  index may better capture how a record of contributions would be evaluated in a given field. For example, if we count papers on a coauthorship-adjusted basis and use a  $h_{(120,2)}$  index to rank economists, then Myerson would rank above everyone in my sample: he would get a three because he has two sole-authored papers and two co-authored papers with at least  $120 \times 3^2 = 1080$  citations. No one in my sample matches this.

In this paper I explore which generalized Hirsch indexes seem most appropriate for

---

<sup>2</sup>The 12 economists are Joshua Angrist, Daron Acemoglu, Christopher Barrett, Menzie Chinn, David Cutler, Edward Glaeser, Jonathan Gruber, Gordon Hanson, Guido Imbens, Michael Kremer, Steve Levitt, and John List.

assessing economists and offer guidance on how indexes might be adjusted for differences in citation patterns across fields and other factors by examining how the various indexes correlate with labor market outcomes. For this purpose, I gathered citation records from Google Scholar for all tenured economists at 25 U.S. economics departments who received their Ph.D. in the last 20 years.<sup>3</sup> The motivation is that the identity of the department in which an economist is tenured can serve as a proxy for the labor market’s assessment of his or her quality. I then assess various Hirsch-like indexes in terms of how well they can be used to predict which economists work at which departments.

The assessment of which Hirsch-like indexes perform best is done by comparing the goodness-of-fit obtained when using various  $h_{(a,b)}$  indexes to predict which economists work in which departments. I present two versions of this analysis. First, I order economics departments according to the 1995 NRC rankings and examine how the  $h_{(a,b)}$  index predicts the NRC rank of each economist’s department in an ordered probit model. Second, I avoid imposing any *a priori* ranking of the departments by estimating a logit model in which each department’s “quality” is an estimated parameter. The latter estimation can be thought of asking whether the Hirsch-like index is a variable on which economists in each department appear to be clustered. (It also yields a ranking of economics departments.) In each case I report the goodness of fit obtained for a wide variety of  $(a, b)$  combinations rather than treating  $a$  and  $b$  as parameters to be estimated. My primary observation is that the best fits are obtained with indexes that depart from Hirsch’s in the intuitively appealing direction of increasing  $a$  and/or  $b$ . A rough characterization is that the best fits come from models which in practice assess economists who are 15 years post-Ph.D. on their 5 to 10 most influential papers. I offer the suggestion that the  $h_{(5,2)}$  index might be adopted as a standard reference tool.

A second observation is that a variety of choices of  $a$  and  $b$  give models with similar predictive power. The primary reason for this is that in practice many indexes are highly correlated. A comparison with old-fashioned citation counts shows that these also perform worse than the  $h_{(5,2)}$  index but not that much worse. Accordingly, Hirsch indexes can be thought of as an improvement over citation counting, but not as a revolutionary advance.

To provide practical guidance for those who wish to use generalized Hirsch indexes as an evaluation tool and to provide a better sense of the goodness-of-fit obtained with these

---

<sup>3</sup>The faculty rosters, which were obtained from Christopher Snyder, are from the 2006-2007 academic year and cover the top 25 departments in Coupe’s (2003) ranking. Only faculty who were tenured at that time and had received their Ph.D. in 1989 or later are included. Economists in the sample who moved to another of the 25 departments prior to 2009 are coded as being tenured at their new institution.

indexes I present also estimates from nonlinear least squares models with each author's  $h_{(5,2)}$  index as the dependent variables and department dummies, field dummies, and the researcher's experience as explanatory variables. The inclusion of field dummies improves the fit and would be recommended if one wanted to use the indexes as an assessment tool. The estimated school dummies could be interpreted as an index of department quality, but also suggest that there may be some systematic differences between what some parts of the profession value and what is highly cited. Plots of the field-adjusted indexes vs. years post-Ph.D. provide a sense for how indexes tend to grow over time and also illustrate that even within the elite set of departments considered in this paper one can readily see differences in the citation records of economists at the very best and just very good departments.

A final section of results discusses how the market treats coauthored papers. In most of the paper I impose the restriction that authors get  $1/n$  credit for an  $n$ -authored paper. Models in which authors get full credit for coauthored papers usually fit the data slightly better, but differences are not large and estimates of a coauthorship-credit parameter do not yield statistically significant results.

## 1.1 Related literature

This paper contributes to a literature that is growing at a remarkable pace. The frequency with which papers are appearing is such that I imagine that the class of generalized  $h$ -indexes studied in this paper have been proposed somewhere even though I have not found such a paper.<sup>4</sup> The most closely related paper in terms of relating the citation indexes to labor market outcomes is Jensen et al. (2008) which examines whether the  $h$ -index and other measures predict which CNRS researchers are promoted in a sample containing 586 candidates from many disciplines. They report that the  $h$  index outperforms the other measures considered, but that the predictive power of the regressions is low. Other papers have attempted to validate the  $h$ -index by showing that it is correlated with peer assessments and decisions of grantmaking bodies.<sup>5</sup> I have not seen other papers calculating adjustment factors that could be applied to improve comparisons between subfields of any other field,

---

<sup>4</sup>Kosmulski (2006) appears to have been the first to propose a member of the class of what I refer to as  $(a, b)$  generalized Hirsch indexes in proposing the  $(1, 2)$  version. He saw the primary advantage as being that it was easier to compute because it involved substantially fewer papers per author. Wu (2009) proposes the  $(10, 1)$  version. Lehmann et al. (2006) is mostly critical as opposed to advocating variants on Hirsch's index, but does mention that Hirsch's index is arbitrary in that  $(a, b)$  indexes are also possible. Egghe (2008) and Schreiber (2008) appear to have independently proposed the fractional counting of coauthored papers I maintain through most of the paper.

<sup>5</sup>van Raan (2006) examined the correlation between the  $h$  index and peer ratings (on a three point scale) of 147 chemistry research groups in the Netherlands. Bornmann et al. (2008) examine grants to biomedical researchers.

but Igelesias and Pecharrroman (2007) does propose that one could compare  $h$ -indexes across fields, e.g. between chemistry and mathematics, by multiplying the raw  $h$ -index by a field specific correction factor. They discuss what these factors should be if citations in different fields follow different power laws and present correction factors for 21 fields derived from an analysis of citation distributions.<sup>6</sup> My observation that many citation measures are highly correlated mirrors findings reported in various other fields.<sup>7</sup>

Other noteworthy papers in the literature include Hirsch (2007), which compares the  $h$ -index and other measures (including variants that place more weight on highly cited papers) in terms of their ability to predict future success (measured with the  $h$ -index and other ways) and Lehmann et al. (2006) which argues that one can derive an upper bound on the ability to distinguish between scientists using any index by examining how much the index varies when one resamples from the distribution of a scientists' citations and that the  $h$  index fares worse in this test than do some other measures.

At least two papers have previously computed  $h$ -indices for economists. Ruane and Tol (2008) compute individual and successive  $h$  indexes to rank economics departments in Ireland.<sup>8</sup> Tol (2009) computes  $h$  indices for the 100 economists with the largest number of papers listed in IDEAS/REPEC and finds that the  $h$ -index is highly correlated with several other citation and publication indices.

The most closely related paper is one that does not compute Hirsch indexes. Using a dataset containing information on tenured economists at 88 economics departments, Hamermesh and Pfann (2009) examine how the rank of the department in which an economist works is related to the economist's total citations, the number of citations to his or her most cited paper, and to the number of papers he/she has published. They find that total citations are a significant determinant of job outcomes, but that the other two variables are not. They also examine other dependent variables including salaries and whether economists have won honors such as being elected as a fellow of the Econometric Society. The regressions examining honors similarly find that total citations matter and indicate that total publications may even have a negative effect. The salary regressions, in contrast, find that salaries are positively related to the total number of publications.

---

<sup>6</sup>The list includes multiplying economics  $h$  indexes by 1.32 to make them comparable to physics.

<sup>7</sup>See, for example, Bornmann et al. (2008) and Kosmulski (2006).

<sup>8</sup>The successive  $h$  index for a department is the largest  $h$  such that a department has at least  $h$  economists with a Hirsch index of  $h$  or higher.

## 2 The Hirsch Index

Hirsch’s  $h$  index is a simple construction: the index is defined to be the largest number  $h$  such that the researcher has at least  $h$  papers with  $h$  or more citations. Informally, it can be thought of as a count of the number of “good” papers that a researcher has written, with the clever addition that the index raises the bar for what “good” means when comparing more accomplished researchers. The Hirsch index can also be motivated as a measure that is akin to a citation count, but which deemphasizes the number of citations to a researcher’s most-cited papers.<sup>9</sup> Deemphasizing extremely highly-cited papers is seen attractive for a couple reasons. First, it is not clear how much credit each author of a coauthored paper should receive. Second, the distribution of citations per paper has been reported to have a power law distribution with an infinite variance, which would make a researcher’s total citations a noisy estimate of an underlying parameter.<sup>10</sup> Despite these attractive features, the  $h$  index is unappealing on its face when applied to economists. Economists write fewer papers than do physicists and individual papers get many citations. As a result, the  $h$  index unappealingly similar to a publication count.

Through most of this paper I will consider indexes that depart from Hirsch’s in two ways. First, I will usually consider coauthorship-adjusted versions of the index in which an author only gets  $\frac{1}{n}$  credit for a paper with  $n$  authors. Second, I will examine indexes in which a researcher’s index is  $h$  if the last paper counted has at most  $ah^b$  citations.<sup>11</sup> With larger values of  $a$  and  $b$ , the  $h_{(a,b)}$  index focuses on a smaller set of papers. For example, with  $(a, b) = (10, 1)$  an economist achieves an index of 3 when he or she has 3 papers with at least 30 citations, rather than 3 papers with 3 citations each. Indexes with larger values of  $b$  remain focused on a small number of papers even when comparing highly accomplished researchers. For example, the  $(6, 1)$  and  $(2, 2)$  indexes both count a researcher’s third most-cited paper if it has at least 18 citations, but the former is influenced by a researcher’s tenth most cited paper if it has at least 60 citations, whereas the latter ignores the tenth most-cited paper unless it has 200 citations.

---

<sup>9</sup>Hirsch (2005) reports that  $h$  is approximately proportional to the square root of the total number of citations.

<sup>10</sup>See Price (1965) for an early study.

<sup>11</sup>For example, if a researcher has two-authored papers with 85, 60, 35, 25, 15, 10, and 8 citations, then the researcher’s  $(10, 1)$  index would be 2 because he has 2 coauthorship-adjusted papers (the first four) with at least 20 citations each. The researcher’s  $(1, 2)$  index would be 3 because he has 3 coauthorship-adjusted papers with at least 9 citations.

### 3 Data

This paper examines the citation records of the 166 economists who held tenured positions at one of 25 highly-regarded U.S. economics departments in the 2006-2007 academic year and who received their Ph.D.'s in 1989 or later.<sup>12</sup> The motivation for restricting the sample to relatively young tenured economists is that labor market outcomes are intended to serve as a proxy for the market's assessment of an economists' contributions, and this seems most defensible for young tenured faculty.

Citation data for each economist were collected from Google Scholar. The primary variables collected were the number of citations and the number of coauthors for up to 50 papers by each researcher.<sup>13</sup> The other variables used in the analysis are the affiliation of each economist and dummies for their primary field of interest. The affiliation is set to be where the economist had tenure in 2006-2007 unless the economist subsequently moved to another of the departments in our sample.<sup>14</sup> Economists were classified into fields using their descriptions of their interests on their webpages and by looking at the titles of their most-cited papers.<sup>15</sup> I chose to divide two traditional fields (International and Econometrics) into subfields because my impression from previous analyses of citations I have carried out at the paper-level suggest that there might be substantial differences.<sup>16</sup>

Table 1 presents summary statistics on the author-level database (which contains 166 observations). The first row indicates that the average economist in the sample is approximately 14.2 years post-Ph.D. in 2008. The next two rows give the number of papers for which data is available and the average number of authors per paper. The number of papers

---

<sup>12</sup>The departments the top 25 U.S. schools in Coupe's (2003) ranking. The decision to use this ranking over many others was entirely for convenience. Chris Snyder graciously provided me with faculty rosters he had collected for use in Snyder and Zidar (2009).

<sup>13</sup>More precisely, a page containing up to 50 references was downloaded by typing "au:"Firstname Last-name"" into the search box with advanced search parameters restricting the search to the "Business, Administration, Finance, and Economics" and "Social Sciences, Arts, and Humanities" subject areas. When the resulting page appeared to contain a substantial number of papers not by the author in question searches were refined to try to eliminate these incorrect references, e.g. by including negative keywords omitting articles by authors with different middle initials. The resulting lists were then screened by hand to eliminate additional papers that were not by the author in question. The papers Google Scholar returns are roughly the 50 most cited papers, although Google sometimes has papers slightly out of order. Most pages were downloaded in December of 2008. Some pages for which the initial search was deemed to have not been properly refined were replaced with pages downloaded in June of 2009.

<sup>14</sup>Economists who subsequently moved to schools outside our sample, e.g. Michael Keane, are entered as affiliated with their former institution.

<sup>15</sup>Most economists were classified as being in a single field. For thirty five economists with substantial research interests in two fields I set the field dummy for what appeared to be the primary field to 0.7 and the field dummy for the secondary field to 0.3. In eight cases where research topics were even more diverse other classifications were made.

<sup>16</sup>Some results from these analyses appear in Ellison (2002b) and Ellison (2010).

can be less than 50 for two reasons: some economists have written fewer than 50 papers; and some papers were removed from Google’s 50 paper lists because they did not belong to the economist under consideration.<sup>17</sup> An attractive feature of the Hirsch framework is that it is usually completely unaffected by such data issues – most indices we consider will count many fewer than 50 papers.

The next section of Table 1 contains a number of citation measures. They are computed on a coauthorship-adjusted basis: the total citation counts give each author  $1/n$  credit for each citation to an  $n$ -authored paper; and a coauthored paper with 100 citations will add 0.5 to the count of papers with at least 100 citations. The average number of citations per economist is 1253.<sup>18</sup> The standard deviation is large and the distribution is quite skewed, but the economists in the sample are highly accomplished and the median number of citations is still 801.5.<sup>19</sup> The count data indicate that the average author has 3.2 papers with at least 100 citations and 13.2 papers with at least 10 citations.<sup>20</sup> The gap between the average economist’s having 1253 and only having 3.2 papers with at least 100 citations gives some sense of importance of the few most cited papers to a typical economists’ total citations.

The table also includes various generalized Hirsch indexes,  $h_{(a,b)}$ . (Again, we focus mainly on indexes with fractional counts of co-authored papers.) The mean value of the  $h_{(1,1)}$  index is 12.3. Note that this implies that the index ends up counting a large number of papers per author including some that don’t have many citations: recall that authors are getting fractional credit for coauthored paper so an index of 12 can be thought of as a publication count indicating that the researcher has authored or coauthored 20+ papers with 12 or more citations. Comparing the  $h_{(a,1)}$  indexes with the  $h_{(1,1)}$  index one notes the obvious relationship that the index ends up counting fewer papers when  $a$  is larger. For the range of values given, the indexes still end up counting a substantial number of papers for the economists in our sample. For example, the  $h_{(10,1)}$  index has a mean of 5.2. This corresponds to about 9 papers before adjusting for coauthorship. I would guess is still larger

---

<sup>17</sup>Google usually returns 50 items even for economists with fewer than 50 papers listed on their CV because Google is unable to completely aggregate citations to different versions of each paper. It is, however, much better at this than are other sources for citation data.

<sup>18</sup>These totals, of course, miss any citations to papers not in a researcher’s top 50.

<sup>19</sup>Those who are used to seeing data on ISI citations should note that Google Scholar returns many more citations per paper than does ISI for recent papers. Two primary reasons for this are that Google Scholar includes citations that occur in unpublished papers found on the web not just in published papers and that Google aggregates together citations to all versions of a paper, whereas ISI only aggregates well citations made to the published version of paper.

<sup>20</sup>The latter number is downward biased because I only collected data on each author’s fifty most-cited papers, but the bias should be fairly small: only 14 of the 166 authors had at least 10 citations to all included papers.

Variable	Mean	St. Dev	Min	Max
YearsPostPhD	14.2	3.0	8	19
NumPapers	46.8	6.4	16	50
AvgAuthors	2.0	0.3	1.3	2.8
Coauthorship-adjusted Counts and Citation Measures				
TotalCites	1253	1303	139	13737
AdjNumPapers	27.4	5.5	12.3	43.3
Num > 200	1.3	2.2	0	18.1
Num > 100	3.2	3.6	0	24.4
Num > 50	5.8	5.1	0	30.9
Num > 10	13.2	7.8	2.3	34.6
Coauthorship-adjusted Hirsch indexes				
$h_{(1,1)}$	12.3	5.6	3.8	31.1
$h_{(2,1)}$	9.9	4.6	3	29.6
$h_{(5,1)}$	7.0	3.4	1.8	23.4
$h_{(10,1)}$	5.2	2.6	1.2	19.3
$h_{(20,1)}$	3.7	1.9	0.8	12.8
$h_{(1,2)}$	6.2	2.1	2.5	14.8
$h_{(2,2)}$	5.0	1.7	2.1	11.4
$h_{(5,2)}$	3.7	1.2	1.5	7.8
$h_{(10,2)}$	2.9	0.9	1.1	6.2
$h_{(1,3)}$	4.0	0.9	2.1	6.9
$h_{(2,3)}$	3.3	0.8	1.8	5.8
$h_{(5,3)}$	2.6	0.6	1.3	4.4
Unadjusted Counts and Citation Measures				
TotalCites	2183	2212	139	13737
Num > 200	2.3	3.8	0	28
Num > 100	5.4	6.0	1	39
Num > 50	10.0	8.7	0	50
Num > 10	22.8	12.8	4	50
$h_{(1,1)}$	18.0	8.4	5	50
$h_{(10,1)}$	7.0	3.4	1.7	23
$h_{(1,2)}$	7.7	2.4	3.3	17.2
$h_{(5,2)}$	4.5	1.4	1.8	9
$h_{(1,3)}$	4.5	1.0	2.3	7.5

Table 1: Summary Statistics on Author-level Database

than the number of papers that are considered important in a typical hiring or tenure case. The indices with  $b > 1$  have smaller means. It is also noteworthy that there is a reduction in the standard deviation relative to the mean as  $b$  increases. For example, the  $h_{(20,1)}$  index has a mean of 3.7 and a standard deviation of 1.9, the  $h_{(5,2)}$  index has a mean of 3.7 with a standard deviation of 1.2 and the  $h_{(1,3)}$  index has a mean of 4.0 and a standard deviation of 0.9. One can think of the higher  $b$  indexes having the property that the number of papers considered changes less as one moves from evaluating young/less distinguished economists to older/more distinguished economists. Roughly, the  $h_{(5,2)}$  index can be thought of as focusing on about five coauthored papers (or equivalent) for less accomplished economists and focusing on about ten papers for the more accomplished economists in our sample.

The bottom part of the table reports citations counts and generalized Hirsch indexes computed without coauthorship adjustments. Hirsch's (2005) paper computed indexes in this manner, so the  $h_{(1,1)}$  index in this table corresponds directly with his proposal. This variable has a mean of 18 in our sample. A cutoff of 18 citations is larger than that which one would effectively be using in computing a coauthorship-adjusted  $h_{(1,1)}$  index, but still seems low relative to the threshold I would have chosen if I had been asked to construct an ad hoc measure that counted the number of important papers an economist had written.

## 4 Which Hirsch-like Indices Predict Employment Patterns?

In this section I present estimates from two models examining which Hirsch-like models are most predictive of labor market outcomes. The first model uses a traditional ranking of an economist's department as the dependent variable. The second does not impose any a priori ranking: each department's quality is estimated as part of the model. I find that variants of the Hirsch index that count smaller numbers of papers tend to do better than the  $h_{(1,1)}$  index and have substantial correlation with labor market outcomes.

### 4.1 Estimates of an ordered probit model using NRC rankings

My first analysis uses the 1995 NRC rankings of the department at which each economist works as a proxy for the market assessment of an economist's quality. Note that the NRC rankings were conducted before just about anyone in my sample was tenured and before many were at their current positions. Hence, the motivation for the use of the NRC rankings is that they may reflect persistent differences in departmental prestige/resources which may have allowed the higher-ranked departments to attract or retain the most sought after young senior faculty members.

To explore which Hirsch indexes appear to most closely reflect labor market outcomes I estimate ordered probit models using various Hirsch indexes: economist  $i$  is assumed to have an unobserved quality

$$\log(y_i^*) = \log(h_{(a,b),i}) + \alpha YearsPostPhD_i + \sum_f \gamma_f Field_{if} + \epsilon_i,$$

and to work in the  $j^{th}$  ranked department if  $y_i^* \in [\mu_j, \mu_{j-1})$ . Note that the model includes a linear adjustment for years post Ph.D. and field adjustments are made by including the variables  $Field_{if}$  classifying economists into fields.

To give a sense for how well a variety of Hirsch-like indexes align with employment outcomes, Table 2 lists the maximized per observation log-likelihood for various values of  $a$  and  $b$ . The primary observation I'd make from this table is that employment outcomes are best explained by Hirsch-like indexes that count a smaller number of papers than does Hirsch's original index. For example, the first column indicates that the Hirsch-like indices counting the  $h^{th}$  paper if it has  $ah$  citations work better if we choose values of  $a$  that are substantially larger than one. The  $h_{(10,1)}$  index provides the best fit among the  $h_{(a,1)}$  indexes considered. A second observation is that appears not to matter much whether one constructs an index that focuses on fewer papers by increasing the  $a$  or  $b$  parameters or by some combination of the two: the models using the  $h_{(10,1)}$ ,  $h_{(5,2)}$ ,  $h_{(2,2)}$ , and  $h_{(0.5,3)}$  indexes all fit about as well. A third observation is that the fit is less sensitive to the choice of the  $a$  parameter when the  $b$  parameter is larger.

	Maximized log likelihood using $h_{(a,b)}$		
	Value of $b$		
Value of $a$	1	2	3
0.5	-2.94	-2.89	-2.86
1	-2.93	-2.87	-2.87
2	-2.91	-2.86	-2.87
5	-2.88	-2.86	-2.88
10	-2.86	-2.87	-2.89
20	-2.88	-2.88	-2.88

Table 2: Goodness of fit for various  $h_{(a,b)}$  in ordered probit model using 1995 NRC rankings

An obvious limitation of this analysis is that it relies on the 1995 NRC rankings of departments as a proxy for the market's assessment of researchers' contributions. These ratings are quite old by now, and even in 1995 they were not designed solely to reflect

research contributions.

## 4.2 Estimates with estimated department quality

My second analysis does not employ any a priori information about rankings of different economics departments. Rather than examining whether Hirsch indexes are correlated with any a priori quality measure, the idea is to examine whether a given Hirsch-like index brings order to the data in terms of making it appear that departments are clusters of economists of similar quality.

Specifically, I estimate a logit-like model in which the probability that economist  $i$  works in department  $j$  is a decreasing function of the mismatch  $d_{ij}$  between the economist and the department,

$$\text{Prob}\{y_{ij} = 1|X_i, Z_j\} = \frac{e^{\beta_{0j} - \beta_1 d_{ij}}}{\sum_{j'} e^{\beta_{0j'} - \beta_1 d_{ij'}}$$

where the mismatch function  $d_{ij}$  is the squared difference between economist  $i$ 's Hirsch-like index (adjusted for age and field effects) and the "quality"  $\delta_j$  of department  $j$ :

$$d_{ij} = (\delta_j - h_{(a,b)}i(1 + \sum_f \gamma_f \text{Field}_{if})(1 + \alpha(\text{YearsPostPhD}_i - 19)))^2.$$

The estimated parameters are the  $\beta$ 's, quality dummies  $\delta_j$  for the 25 departments, multiplicative field dummies  $\gamma_f$  for 15 fields, and the factor  $\alpha$  to control for differences in time-since-Ph.D.<sup>21</sup>

Table 3 gives information on goodness-of-fit of this model for various specifications of the Hirsch-like index. Again, the rows correspond to different choices of  $a$ , the columns to different choices of  $b$ , and each entry gives the per observation log likelihood of the model when  $h_{(a,b)}$  is used as the citation index. The results are qualitatively similar to those of the ordered probit model using NRC rankings: the best fit is obtained from indexes that count fewer papers than does the  $h_{(1,1)}$  index; and it appears to not matter much whether one does this by increasing  $a$  or  $b$ . The  $h_{(5,2)}$  index, which counts the  $h^{\text{th}}$  paper if it gets at least  $5h^2$  citations, has the best fit of any of the indexes considered.<sup>22</sup> Recall that the  $h_{(5,2)}$  index has a mean of 3.7 and a standard deviations of 1.2. Taking coauthorship into account, this means that it can be thought of as usually counting an economist's top 5 to 10 papers.

<sup>21</sup>To limit the number of parameters to be estimated I assume that the  $\beta_{0j} = \beta_0 N_j$  where  $N_j$  is the number of economists in my sample in department  $j$ .

<sup>22</sup>The likelihoods are -2.696 for the  $h_{(5,2)}$  model and -2.703 for the  $h_{(1,3)}$  model.

	Maximized log likelihood using $h_{(a,b)}$		
	Value of $b$		
Value of $a$	1	2	3
0.5	-2.82	-2.76	-2.72
1	-2.82	-2.73	-2.70
2	-2.78	-2.73	-2.71
5	-2.75	-2.70	-2.72
10	-2.71	-2.72	-2.73
20	-2.74	-2.73	-2.73

Table 3: Goodness of fit for various  $h_{(a,b)}$  in model with estimated department quality

The fact that the relative rankings of the  $h_{(a,b)}$  are similar across the two models is comforting. The estimated department qualities  $\hat{\delta}_j$  end up being highly aligned with NRC rankings. For example, in the estimates using the  $h_{(5,2)}$  index, seven of the eight schools with the largest  $\hat{\delta}_j$  are also in the top eight in the NRC rankings and seven of the eight schools with the lowest  $\hat{\delta}_j$  are among the eight with the lowest NRC rankings.<sup>23</sup>

One way to get a feel for the goodness of fit is to ask what fraction of economists are working at the  $k$  departments that the model considers most likely. For the model based on the  $h_{(5,2)}$  index the answer is that 13.3% of economists work in the single most likely department and 55.4% work in one of the five departments that are predicted to be most likely for them.

### 4.3 A proposal: adopt the coauthorship-adjusted $h_{(5,2)}$ index

It is useful for a profession to standardize on a citation metric: this makes it easier to make comparisons with results reported elsewhere; and it reduces the scope for manipulating analyses to support a favored candidate or hypothesis. I would propose that economists capture these benefits by adopting the coauthorship-adjusted  $h_{(5,2)}$  index as a standard metric.

The original Hirsch index would have some practical advantages that derive from its growing popularity in other fields, but I would not advocate that economists use it: it does not line up with my informal sense for how the profession values contributions; and the above analyses indicate that it is not as correlated with labor market outcomes as are other indexes.

---

<sup>23</sup>The eight schools with the highest estimated quality are MIT, Harvard, Stanford, Chicago, Pennsylvania, NYU, Princeton, and UC-Berkeley.

At the other extreme, I also would not advocate estimating the  $a$  and  $b$  parameters to several decimal places using one of the models above and standardizing on a “best-fit” metric. Three reasons for this are that the best fit values of  $a$  and  $b$  will vary from dataset to dataset; that models involving whole numbers are easier to work with when computing indexes manually, and that a range of choices for  $a$  and  $b$  seem to produce models that fit about as well.<sup>24</sup>

The  $h_{(5,2)}$  index seems to offer a good combination of fitting the data well and being easy to compute. Acting on this suggestion, I will focus on this index in subsequent analyses.

#### 4.4 Is the Hirsch index a revolutionary advance in scientometrics?

The Hirsch index has attracted an extraordinary amount of attention in the past five years. As one recent contributor (Prathap 2009) puts it,

The  $h$  index (Hirsch 2005) has rapidly captured the imagination of scientometricians and bibliometricians to such an extent that one can now divide the history of the subject virtually into a pre-Hirsch and a post-Hirsch period.

One source of interest is hope that the Hirsch index may prove to be a revolutionary advance in measuring research output that enables a host of new analyses. From this perspective it is instructive to ask whether the Hirsch index is more powerful than previous tools for the purpose of understanding the economics labor market.

In the pre-Hirsch period one could have explored whether labor market outcomes were correlated with citation records. One natural method would have been to simply compute total citation counts for each researcher. Another would have been to pick some arbitrary cutoff level like 100 citations and count the number of papers by each economist with citations exceeding this threshold. To assess whether the Hirsch index is an advance over such methods I reestimated the model of section 4.2 using old-fashioned metrics in place of the  $h_{(a,b)}$  index: one estimation used  $\log(\text{TotalCites})$ ;<sup>25</sup> the other include the number of papers by each author that have at least 100 citations.<sup>26</sup> The maximized log likelihoods were -2.71 from the model using  $\log(\text{TotalCites})$  and -2.77 for the model using the counts of papers with 100 citations. The former is only a slightly worse fit than the  $h_{(5,2)}$  index and is better than the  $h_{(1,1)}$  index. The latter model is still better than the  $h_{(1,1)}$  index.

---

<sup>24</sup>Finding best-fit parameters for the model is difficult because the likelihood is not smooth in  $a$  and  $b$ . The best parameters I have found (2.70, 2.48) give a log likelihood of -2.691 vs. -2.696 for the (5, 2) combination.

<sup>25</sup>Recall that this citation measure uses data only on each authors’ 50 most cited papers and gives partial credit for citations to coauthored papers.

<sup>26</sup>Technically, this latter method is a generalized Hirsch index: the  $h_{(100,0)}$  index.

Accordingly, I would say that the Hirsch index is clearly not a revolutionary tool from the perspective of allowing us to better account for labor market outcomes in economics.

To give a better sense of why the fits are similar for a number of different indexes, Table 4 lists correlation coefficients for several generalized Hirsch indexes and for the old-fashioned measures. The  $h_{(5,2)}$  index has a correlation of about 0.9 with both old-fashioned measures. Intuitively, the correlations are high because  $h_{(5,2)}$  index is similar to counting total citations (but with a limitation on how much credit is given for the most cited papers) and similar to counting papers with at least 100 citations (but varying the procedure to count more/fewer papers for more/less distinguished researchers). Even the measures that are not as similar by construction can end up being fairly highly correlated in practice because most real-world researchers who a number of highly cited papers have also written many not as highly cited papers. For example, the correlation between the  $h_{(1,1)}$  index (in practice roughly a count of all papers with at least 10-20 citations) and the  $h_{(100,0)}$  index (exactly a count of all papers with at least 100 citations) is 0.82. I note also that the generalized Hirsch indexes that were found to have similar fits,  $h_{(10,1)}$ ,  $h_{(5,2)}$ , and  $h_{(1,3)}$ , are extremely highly correlated: pairwise correlations range from 0.953 to 0.990. The fact that such correlations are so high is another reason why I feel that adopting some easy-to-compute index as the norm is most appropriate.

Index	Correlation Coefficients					
	$h_{(1,1)}$	$h_{(10,1)}$	$h_{(5,2)}$	$h_{(1,3)}$	$\log(Cites)$	$h_{(100,0)}$
$h_{(1,1)}$	1.000					
$h_{(10,1)}$	0.889	1.000				
$h_{(5,2)}$	0.862	0.959	1.000			
$h_{(1,3)}$	0.858	0.953	0.990	1.000		
$\log(Cites)$	0.799	0.860	0.899	0.900	1.000	
$h_{(100,0)}$	0.823	0.944	0.915	0.887	0.793	1.000

Table 4: Correlations between Various Citation Indexes

Although the indexes are highly correlated, the modified Hirsch indexes do fit the data better than the old-fashioned techniques and hence their adoption seems warranted. It should also be kept in mind that the fact that the modified Hirsch indexes are similar to old-fashioned citation counting techniques does not mean that they are not a powerful tool – it may mean in part that old-fashioned citation counting was underappreciated.

## 5 Using Hirsch-like indexes: Patterns Across Schools, Fields, and Time

There is a tremendous inherent demand for tools that can be used to evaluate researchers. Nonexperts like deans and granting organizations need some source of information even if it is just as a check what experts are telling them. And experts also value statistical measures for various reasons: some method is needed to select an initial list of candidates to evaluate more carefully; faculty members with less expertise in a subfield need help in assessing what those in the subfield are telling them; and reading papers is a costly activity that economists would say should only be undertaken if the benefits justify the cost. I present here some results that offer guidance on how Hirsch-like indexes should be used if they are to be used. The results will also provide more of a sense of the degree to which labor market outcomes are aligned with citation records.

People using citation indexes as an evaluation tool typically want to do one of two things: compare the accomplishments of two researchers; or assess whether a given researcher seems appropriate for a particular position. For the former one needs to be able to correct for differences across fields and years of experience. For the latter one also needs some idea of what would be typical for the department in question. The first goal of this section is to present estimates that inform such comparisons. I will focus on the  $h_{(5,2)}$  index and estimate a simple model of how the index varies across school, fields, and age cohorts by nonlinear least squares:

$$h_{(5,2),i} = \sum_s \delta_s School_{is} e^{\alpha(YearsPostPhD_i - 14) + \sum_j \gamma_j Field_{ij}} + \epsilon_i$$

Alternatively, I also estimate the parameters via the related OLS regression:

$$\log(h_{(5,2),i}) = \sum_s \log(\delta_s) School_{is} + \alpha(YearsPostPhD_i - 14) + \sum_j \gamma_j Field_{ij} + \epsilon_i$$

### 5.1 Typical values of $h_{(5,2)}$ by school

The first item of interest in the regression are the school dummies. They can be interpreted as giving an average value of  $h_{(5,2)}$  for a macroeconomist (the omitted field) who is fourteen years post-Ph.D. and has tenure in the department in question. Estimates obtained by estimating this relationship via nonlinear least squares are given the first column of Table 5. The highest estimated school dummy is 5.47 or MIT. This would correspond to having 5 papers (or 10 co-authored papers) with more than  $5 \times 5.47^2 \approx 150$  citations, and an additional paper with exactly 150 citations. The next two schools, Harvard and Stanford, have

estimated dummies of 4.87 and 4.75. The next few schools have dummies close to 4, which corresponds with having 4 papers (or eight coauthored papers) with at least 80 citations. The twentieth ranked school has an estimated dummy of around 3, which corresponds to having 3 papers (or six coauthored papers) with at least 45 citations. The last five schools have values between 2 and 3.

School	NLLS Model		1995 NRC
	Coef. Est.	St. Err.	Score (rank)
MIT	5.47	0.43	4.93 (3)
Harvard	4.87	0.37	4.95 (1)
Stanford	4.75	0.61	4.92 (4)
Chicago	4.06	0.33	4.95 (2)
Princeton	3.91	0.24	4.84 (5)
NYU	3.87	0.32	3.62 (17)
UC-Berkeley	3.79	0.32	4.55 (7)
Penn	3.69	0.44	4.43 (8)
UCLA	3.64	0.36	4.12 (11)
Columbia	3.62	0.28	4.07 (12)
Wisconsin	3.51	0.53	3.93 (15)
Northwestern	3.46	0.37	4.39 (9)
Duke	3.41	0.19	3.36 (22)
Yale	3.29	0.26	4.70 (6)
Rochester	3.29	0.24	4.01 (14)
UC-Davis	3.27	0.31	2.75 (38)
Minnesota	3.25	0.47	4.22 (10)
UC-San Diego	3.20	0.44	3.80 (16)
Michigan	3.14	0.24	4.03 (13)
Maryland	3.03	0.29	3.46 (20)
Ohio State	2.77	0.41	2.83 (35)
Cornell	2.69	0.32	3.56 (18)
Texas	2.41	0.34	2.91 (31)
Southern Calif	2.24	0.17	2.66 (40)
Illinois	2.00	0.08	3.07 (28)

Table 5: Estimated school dummies from NLLS model of  $h_{(5,2)}$

The estimated school dummies are strikingly similar the 1995 NRC rankings. In part this is due to a coincidence – typical  $h_{(5,2)}$  values turn out to be similar in magnitude of the numerical scores given by the NRC, which ranged from 4.95 (Harvard) to 2.66 (Southern California) for the schools in our sample. But the coincidence is just making apparent that the correlation of 0.79 is quite high.

Among the schools that do well on the citation metric relative to their 1995 NRC scores

are NYU, Duke, and UC-Davis. Presumably this reflects that they have hired good young senior faculty since that time. Yale and Northwestern, on the other hand, have estimated school dummies that are lower than their NRC scores. These schools are also known for having a strong presence in theory, econometrics, and industrial organization, and particularly in the more technical sides of these fields. One limitation of the analysis that might be responsible for this coincidence is that the regressions have estimated field dummies, but do not have variables to look for within-field variation (e.g. it may be that papers in the more technical parts of these fields are less cited). Another potentially relevant limitation is multicollinearity: the number of observations is not very large relative to the number of schools and fields, and this makes it hard to say whether researchers in theory, econometrics, and industrial organization are generally less cited, or whether Yale and Northwestern have not been as successful in the recent past at the young senior level.

## 5.2 Variation in $h_{(5,2)}$ across fields

There is substantial variation in citations across fields of economics. Views differ with regard to whether economists working in areas that receive many citations should be lauded for this or whether we should “correct” for citation differences when comparing them to researchers working in less-cited fields. But there is broad agreement that it is desirable to be cognizant of what is typical in a given field when assessing a candidate for promotion/appointment in that field.

To provide some guidance of this sort, the first column of Table 6 reports the estimated field effects from our base NLLS regression of  $h_{(5,2)}$  on school dummies, field dummies, and years post-Ph.D. The omitted field is macroeconomics, so the coefficient estimates can be thought of as measuring the extent to which economists in the field in question tend to have a higher or lower  $h_{(5,2)}$  indexes than macroeconomists in the same department. Three coefficients are significantly positive: those working in international finance, behavioral/experimental, and finance tend to be more widely cited than their macro colleagues. Only the micro theory coefficient is both negative and significant.

The fact that macroeconomists have relatively low citation indexes is perhaps surprising. Macroeconomics is thought of as a field in which papers are widely cited and traditional lists of highly cited economists are very macro heavy. For example, REPEC’s list of most-cited economists has six macroeconomists in its top ten.<sup>27</sup> Some factors that may be involved are field life-cycle effects (early rational expectations papers garnered many citations) and

---

<sup>27</sup>The six are Robert Barro, Robert Lucas, Mark Gertler, Olivier Blanchard, Edward Prescott, and Ross Levine. The second ten has three macroeconomists: Greg Mankiw, Jordi Gali, and Alberto Alesina.

a shift in the topics that young macroeconomists have been writing about that has taken them away from topics like growth and monetary policy in which many highly-cited papers are found.

The estimated field effects are of sufficient magnitude to make them practically important. For example, an economist at the 20th ranked school working in international finance would be expected to have a higher  $h_{(5,2)}$  index than a macroeconomist at Princeton. The fact that so few of the estimated field effects are significant reflects the limited sample size – there are only 166 economists across the 15 fields. The coefficients are jointly significant at the 0.1% level.

Field	Estimated Field Effects in Models of $h_{(5,2)}$				Field Effects in a	
	Base NLLS Model		No School Dummies		Citation Regression	
	Coef. Est.	Std. Err.	Coef. Est.	Std. Err.	Coef. Est.	Std. Err.
Behavioral/Exper.	0.24	0.09	0.23	0.07	-0.21	0.11
Development	0.12	0.08	0.12	0.08	0.14	0.19
Finance	0.17	0.08	0.13	0.08	-0.13	0.09
History	-0.13	0.15	-0.10	0.17	-0.33	0.22
Industrial Org.	0.01	0.08	-0.07	0.08	-0.25	0.10
Int'l Finance	0.25	0.09	0.27	0.10	0.32	0.18
Int'l Trade	0.18	0.10	0.17	0.11	-0.10	0.12
Labor	0.11	0.07	0.13	0.07	-0.05	0.09
Metrics - Cross Sec.	-0.11	0.13	-0.04	0.10	-0.28	0.15
Metrics - Time Ser.	0.11	0.12	0.07	0.14	0.12	0.15
Micro Theory	-0.15	0.07	-0.13	0.09	-0.28	0.07
Public Finance	0.04	0.06	0.10	0.08	-0.19	0.10
Political Econ.	0.05	0.10	0.10	0.12	-0.35	0.14
Other	0.18	0.19	0.24	0.21	-0.20	0.11

Table 6: Estimated field dummies from NLLS model of  $h_{(5,2)}$  and from a paper-level citation regression

As noted earlier, one other aspect of the dataset that makes estimation difficult is that there is some collinearity between the school and field dummies. The second part of Table 5 reports estimated field dummies from a model that includes a function of the NRC index rather than school dummies. Most of the coefficients are similar. The switch to using NRC rankings does change the estimated industrial organization effect so that researchers in that field are predicted to have lower citation indexes. The estimates on the cross-section econometrics, public finance, and political economy variables become larger.

The final two columns of the table present quantitatively noncomparable estimates of field effects from a different regression/data source. The dataset is one from Ellison (2002b).

It contains information on 1393 papers published in the top five general interest journals in 1990-1998. The regression has the log of the total citations to each paper as the dependent variable and includes a number of controls in addition to field fixed effects as explanatory variables.<sup>28</sup> There are a number of similarities across columns. The three fields in which papers are more cited than macro papers in this regression are international finance, development, and time series econometrics. These are all fields in which researchers were estimated to have relatively large Hirsch indexes. The fields in which papers get the fewest citations according to this regression are political economy, history, micro theory, cross section econometrics and industrial organization. These are mostly fields where we estimate that researchers have relatively small Hirsch indexes. I view the similarity of the two sets of numbers as giving some confidence that the differences in citation indexes across fields are a real phenomenon which would show up more strongly if I had a larger sample of economists.

The most striking dissimilarity between the field effects from the two sources is that for macroeconomics: most field dummy coefficients in the  $h_{(5,2)}$  regressions are positive indicating that macroeconomists have relatively low citation indexes, whereas most coefficients in the paper-level regression are negative indicating that macro papers in top general interest journals received a relatively large numbers of citations.

It is clear that at least part of what makes citations differ across fields are factors one would want to control for, such as the fact that papers in some fields (and some journals) tend to have longer reference lists.<sup>29</sup> Nonetheless, it is rare in practice to field-adjust citations. There are probably several reasons: some firmly believe that work that is getting cited more should be encouraged; some might like to adjust citations but are too lazy; and others would definitely want to adjust their calculations but are deterred by the fact that it is hard to find good advice on how to do field adjustments. I suspect that the latter reason is the most important and hope that Table 6 leads to a change in practice.

To get some sense of the importance of adjusting for differences across fields I would note that the log-likelihood of a  $h_{(5,2)}$ -based model similar to those of section 4.2 but omitting the field fixed effects is -2.81. Hence, the improvement in fit from using field adjustments is approximately the same as the improvement in fit one gets from using the  $h_{(5,2)}$  index instead of the  $h_{(1,1)}$  index.

---

<sup>28</sup>Citations were collected from ISI in February of 2000. Other regressors include journal dummies, journal-specific time trends, several author characteristics, the length of the paper, and the order in which the paper appeared in its journal issue.

<sup>29</sup>See Palacios-Huerta and Volij (2004) for a discussion of the impact of such differences on journal rankings.

### 5.3 Variation in $h_{(5,2)}$ over time

An economist’s Hirsch index will increase over time for two reasons: he or she will write more papers; and existing papers will accumulate more citations. Hirsch’s (1995) original paper included a model of these two processes under which the Hirsch index would be directly proportional to the number of years for which a researcher has been active.

My base NLLS model included a multiplicative age effect:  $e^{\alpha(YearsPostPhD-14)}$ . In this model the estimate for  $\hat{\alpha}$  is 0.043 (s.e. 0.006) indicating that the  $h_{(5,2)}$  index increases by a little more than 4% per year. I investigated a variety of alternative specifications for experience effects both in the NLLS estimations and in linear regressions of  $\log(h_{(5,2)})$  on school and field dummies and found little evidence that more complicated age corrections are needed. A quadratic term  $(YearsPostPhD-14)^2$  is insignificant when added to the model. A full set of dummies for years post-Ph.D. is not jointly significant. It should be kept in mind, however, that all economists in the dataset were between 8 and 19 years post-Ph.D. in 2008. Nonlinearities may be important outside this range.

### 5.4 A graphical look at school and time effects and goodness-of-fit

Figure 1 presents a graph designed to give a sense of the variation in the  $h_{(5,2)}$  index across schools and over time as well as the goodness-of-fit of the model. Each point on the graph corresponds with an economist in our sample.  $YearsPostPhD$  is on the  $x$  axis. And the  $y$ -coordinate is the economist’s “field-adjusted” (5,2)-Hirsch index:  $h_{(5,2),i} e^{-\sum_f \hat{\gamma}_f Field_{if}}$ . The colors of the datapoints provide information on where each economist has tenure: economists from the departments with the four highest estimated school dummies are colored red; economists from the next four departments are colored orange; and so on following the order of the rainbow. The six lines in the graphs are predicted values for economists from each category of schools.

The differences across schools are clearly visible in the figure. The top four departments stand out for having most of the economists with extremely high field-adjusted citation indexes. Many of the economists who are just below the top people are in the 5th through 8th departments. The next three groups of departments (9th-12th, 13th-16th, and 17th-20th) are not as cleanly separated. At the lower extreme, most of the datapoints along the bottom edge are the violet squares used to mark economists from the 21st through 25th schools. Quantitatively, the root mean-square error of the additive error in the NLLS model is 0.81. The root mean-square error of the log-linear model is 0.22.

The time dimension of the graph gives a sense of why the log-linear model for experience

Adjusted H(5,2)-index by School and Career Stage

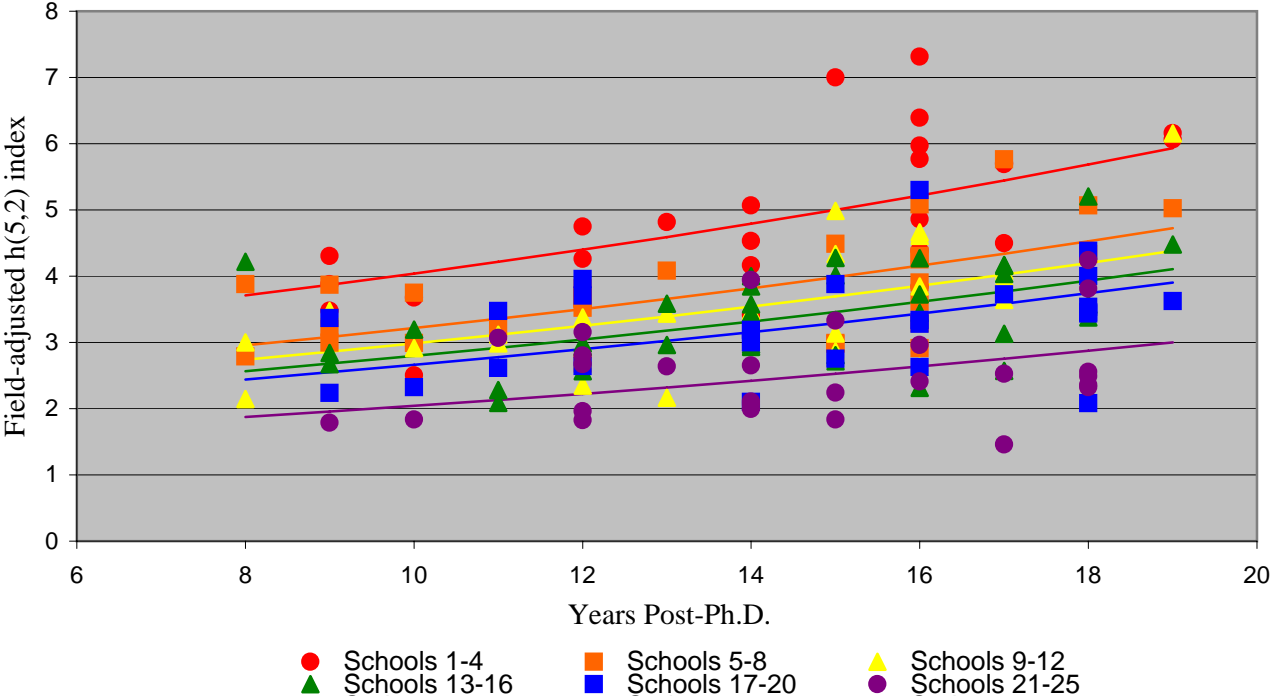


Figure 1: Field-adjusted  $h_{(5,2)}$  indexes for economists at various schools

is not rejected in favor of more complicated models. Visually it seems to fit pretty well over most of the range of the data.

## 5.5 Some simple comparison lists

In this section I present a few “top ten” lists to illustrate the potential utility of citation indexes and to illustrate the effects of age and field adjustments.

First, the top left panel of Table 7 presents a simple highly-cited economist list consisting of the ten economists in the sample with the highest  $h_{(5,2)}$  indexes. Unsurprisingly, most are from older cohorts, and many work in fields where citation indexes tend to be high.

If one wanted to quickly generate a list of people who could be considered for a faculty slot or an award and wanted to give equal consideration to economists regardless of age and field, then one could instead construct a list of people who are highly cited relative to their age and field by using the estimates of the previous sections to rank economists on the basis of  $h_{(5,2)}e^{-\hat{\alpha}(YearsPostPhD_i-14)-\sum_j \hat{\gamma}_j Field_{ij}}$ . The upper right panel of Table 7 lists the top ten economists in our sample in this metric. Among those who move up into the top ten are two Clark medal winners, Esther Duflo and Steven Levitt, who have not had as much time to amass citations as have the economists in the left panel.<sup>30</sup>

Estimated school effects could be useful both to schools wanting to get a quick impression of how a particular economist’s citation record compared with what was typical for their faculty or if one wanted to generate a list of “underplaced” economists who might be targets to try to hire away from their current institution. The bottom panel of Table 7 contains an illustrative list of the twenty economists in our sample who rank most highly if one divides each economist’s field- and experience-adjusted citation index by the estimated school dummy for his or her school. This doubly adjusted index can be interpreted as the ratio of the economists’ index to what would be expected given his or her experience, field, and place of employment.

## 6 How Does the Market Treat Jointly Authored Work?

In all of the models estimated so far I have maintained the assumption that authors get fractional credit for  $n$ -authored papers. I do so largely because of a belief that this is the correct thing to do: the total output of the profession is not larger when all papers have

---

<sup>30</sup>The other two younger Clark medal winners also score highly in this metric: Susan Athey and Emmanuel Saez would be ranked 11th and 20th, respectively.

Most Cited Economists			Most Cited Relative to Field/Age			
Name	Ph.D.		Name	Ph.D.		Fld & age
	Year	$h_{(5,2)}$		Year	$h_{(5,2)}$	adj. $h_{(5,2)}$
Daron Acemoglu	1992	7.84	Daron Acemoglu	1992	7.84	6.72
Edward Glaeser	1992	7.71	Chad Jones	1993	7.00	6.71
Chad Jones	1993	7.00	Edward Glaeser	1992	7.71	5.87
Bruce Hansen	1989	6.87	Glenn Ellison	1992	5.83	5.48
Joshua Angrist	1989	6.43	Edward Vytlacil	2000	3.77	5.45
Matthew Rabin	1989	6.37	Esther Duflo	1999	4.84	5.33
Gordon Hanson	1992	6.32	Jonathan Gruber	1992	6.00	5.30
Jonathan Gruber	1992	6.00	Ilya Segal	1996	4.07	5.17
David Cutler	1991	5.92	Stephen Morris	1991	5.18	5.07
Glenn Ellison	1992	5.83	Steven Levitt	1994	5.62	5.07

Most "Underplaced" Economists				
Name	School	Ph.D.		Fld, age, & sch.
		Year	$h_{(5,2)}$	adj. $h_{(5,2)}$
Edward Vytlacil	Yale	2000	3.77	1.65
Gordon Hanson	UC-San Diego	1992	6.32	1.52
Christopher Barrett	Cornell	1994	4.43	1.47
Bruce Hansen	Wisconsin	1989	6.87	1.44
Chad Jones	Stanford	1993	7.00	1.41
Mark Duggan	Maryland	1999	3.50	1.37
Jeffrey Smith	Michigan	1996	4.42	1.37
Stephen Donald	Texas	1990	3.41	1.33
Michael Keane	Yale	1990	5.50	1.33
Roni Michaely	Cornell	1990	5.04	1.33
Enrico Moretti	UC-Berkeley	2000	4.33	1.32
Dora Costa	UCLA	1993	4.38	1.31
Stephen Morris	Princeton	1991	5.17	1.30
Lutz Kilian	Michigan	1996	4.00	1.29
Ted O'Donoghue	Cornell	1996	4.00	1.28
Emmanuel Saez	UC-Berkeley	1999	4.02	1.26
William Dupor	Ohio State	1997	3.06	1.25
Steven Levitt	Chicago	1994	5.62	1.25
Judith Chevalier	Yale	1993	4.83	1.24
Aviv Nevo	Northwestern	1999	3.49	1.24

Table 7: Lists of highly cited economists with and without controls

multiple authors and not giving fractional credit creates incentives to artificially bolster rankings by adding coauthors.

I cannot deny, however, that this argument is somewhat at odds with the overall motivation for this paper. Indeed, my informal impression is that the “market” does not fully discount joint work. For example, I think a candidate coming up for tenure with just four sole-authored publications would be seen as having a very thin record, whereas a candidate with six co-authored papers and three tri-authored papers would not be seen this way. Accordingly, while I advocate using indexes based on fractional credit, I think it is interesting to investigate how the market credits joint work.

The top panel of Table 8 reports maximized log-likelihoods from models which are identical to those of Table 3, but which use generalized Hirsch indexes which give full credit for each  $n$ -authored paper. In general these models seem to fit slightly better than the fractional credit models of Table 3. The improvement is fairly small for  $h_{(5,2)}$  model. The improvement is larger for several other models, including the  $h_{(1,1)}$  model which now corresponds exactly with Hirsch’s (2005) definition. The best-fit model is that using the  $h_{(0.5,3)}$  index with the  $h_{(20,1)}$ -based model close behind.

	Maximized log-likelihood using $h_{(a,b)}$ with full credit for joint work		
	Value of $b$		
Value of $a$	1	2	3
0.5	-2.80	-2.71	-2.67
1	-2.77	-2.69	-2.69
2	-2.73	-2.68	-2.69
5	-2.71	-2.68	-2.72
10	-2.70	-2.70	-2.73
20	-2.67	-2.73	-2.71

	Log-likelihoods using $\log(Cites)$ for various $c$						
$c$	1	0.8	0.6	0.4	best fit (0.36)	0.2	0
log-likelihood	-2.713	-2.709	-2.707	-2.707	-2.707	-2.707	-2.708

Table 8: Goodness of fit for various models giving full or intermediate credit for joint work

The fact that the fits are better in the top panel of Table 8 than they were in Table 3 can be interpreted as saying roughly that the “market” looks more like it gives full credit than  $1/n$  credit for  $n$ -authored papers. Giving full credit and  $1/n$  credit for  $n$ -authored papers are obviously not the only two possibilities. For example, we could construct indexes

under the assumption that the market gives  $1/n^c$  credit for  $n$ -authored papers for various  $c$ . Here,  $c = 1$  would correspond to the fractional credit I have assumed throughout the paper,  $c = 0$  gives authors full credit for joint work, and intermediate values are intermediate. The bottom panel of Table 8 explores whether we can say much about what value of  $c$  the market uses given our data. It presents the maximized log-likelihoods from models which use  $\log(\text{Cites})$  as the primary explanatory variable, but which construct the *Cites* variable using fractional counting schemes with different  $c$ 's.<sup>31</sup> The primary conclusion I take away from this table is that the data do not allow us to say much about how the market credits economists for joint work. The point estimate is that the market gives more than  $1/n$  credit for  $n$  authored papers, e.g.  $0.5^{0.36} = 0.78$  credit for a co-authored paper. But the log-likelihoods are similar over the full range of values of  $c$  and the improvement with the best-fit parameter ( $c = 0.36$ ) is sufficiently small so that we would not reject the  $1/n$  credit model ( $c = 1$ ) in a likelihood-ratio test.

## 7 Conclusion

In this paper I have examined the degree to which Hirsch index citation metrics are aligned with market outcomes for economists. The original Hirsch index is poorly suited to the economics profession. But variants of the Hirsch index that focus on more highly cited papers do appear to be an advance on traditional citation counting techniques.

Models that adjust the Hirsch index in this direction and correct for differences across fields can do a fairly good job of accounting for labor market outcomes. Measures of productivity are central to many questions one would like to ask in studying academic or scientific research. For example, studies of the extent to which scientific progress is predictable, studies of the effects of grants or other resources on subsequent research success, studies of the effects of incentive schemes, and studies of discrimination. I hope that future work on the economics profession will be improved by researchers' adopting the generalized Hirsch indexes recommended here. Such studies may also provide positive spillovers to the literature on citation indexes. For example, it would be useful to know how well different citation indexes work in accounting for initial tenure decisions. I hope also that studies similar to this one will explore how success should be measured in other disciplines if they are to reflect market judgments.

Because I know that many economists would like to apply citation metrics in connection

---

<sup>31</sup>I focus  $\log(\text{Cites})$  based indexes rather than a generalized Hirsch index because the best-fit  $c$  will differ for different values of  $a$  and  $b$  and maximizing over  $a$  and  $b$  is difficult numerically in the Hirsch framework.

with hiring and promotion decisions and not just in academic studies of the economics profession, I have presented both field-adjustment factors and typical values of the co-authorship adjusted  $h_{(5,2)}$  index for economists in each department in my dataset. I hope that this makes it easier for people who want to use citation indexes to do it fairly well.<sup>32</sup> I would, however, like to caution potential users against viewing citations as an “objective” method for comparing economists. Generalized Hirsch indexes may be a good predictor of employment outcomes in part because employment outcomes have a causal effect on citations (Ellison 2010).

While I am concerned that people may rely too much on the easy-to-compute citation measures discussed in this paper and too little on careful assessments of work, I see this concern mostly as continuing the existing concern that people may rely too much on counting publications instead of reading them. Indeed, with this comparison in mind I find it comforting that citation analyses seem to work fairly well. In Ellison (2002a, b) I noted that the economics publication process has slowed dramatically over the last few decades. I am unaware of any research showing that journal-directed revisions improve papers in any way (let alone to a degree commensurate with the resources expended). While one hopes that the peer-review system can improve, there must be a chance that we will instead see it collapse (Ellison, 2010). This is worrying because it would destroy the quality signals that publication has provided. In this eventuality, it would be comforting if citation metrics could provide a viable alternative. Whether citations would continue to provide useful information in such a world, of course, is a further question – the citation metrics studied here are unavoidably based on citations received in a world where peer-reviewed journals exist.

Given the inherent desire for impartial measures to help resolve disagreements between researchers in different fields, I would like to emphasize that the estimates presented here also cannot be regarded as indicating how researchers in different fields *should* be treated. The field-adjustments in my model will reflect both cross-field differences in citations holding “quality” fixed and cross-field differences in the “quality” thresholds that schools apply when hiring and promoting researchers. I have not attempted to separate these two effects. The field adjustments presented here just tell us what the economics labor market is doing, not what one should do if one wanted to maximize some objective or treat researchers equitably.

While I am recommending that the  $h_{(5,2)}$  index be adopted by the economics profession

---

<sup>32</sup>One word of caution is that the figures in this paper are derived from Google Scholar citation counts from December of 2008. Subsequent expansions in Google’s database have increased the citation counts that it reports.

as a standard tool for quantifying research output, I do not wish to suggest that the search for better metrics should cease. The indexes used here are convenient in that they only use data that can be collected from Google Scholar in a matter of seconds, but by doing so they are ignoring a great deal of potentially valuable information. Just as Google's search engine revolutionized web search by paying attention to which websites were citing which other websites, I imagine that citation indexes that pay attention to where papers are being cited could be much more powerful.<sup>33</sup> There are many potentially informative characteristics of each citation: the citing journal, the citing author, citations to the citing paper, and perhaps some day text-derived estimates of how central the cited paper is to the citing paper. The potential seems vast and this could be an exciting area for many years to come.

---

<sup>33</sup>See Liebowitz and Palmer (1984), Palacios-Huerta and Volij (2004), and West et al. (2010) for discussions of ranking journals by such approaches.

## References

- Adler, Robert, John Ewing, and Peter Taylor (2008): "Citation Statistics: A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)," mimeo.
- Althouse, Benjamin M., Jevin D. West, Ted C. Bergstrom, and Carl T. Bergstrom (2008): "Differences in Impact Factor Across Fields and Over Time," University of California, Santa Barbara Department of Economics Working Paper 2008-4-23.
- Bornmann, Lutz, Gerlind Wallon and Anna Ledin (2008): "Is the  $h$  index Related to (Standard) Bibliometric Measures and to the Assessments by Peers? An Investigation of the  $h$  Index by Using Molecular Life Sciences Data," *Research Evaluation* 17, 149-156.
- Coupe, Tom (2003): "Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990-2000," *Journal of the European Economic Association* 1(6), 1309-1345.
- Egghe, L. (2008): "Mathematical Theory of the  $h$ - and  $g$ -index in Case of Fractional Counting of Authorship," *Journal of the American Society of Information Science and Technology* 59 (10), 1608-1616.
- Ellison, Glenn (2002a): "Evolving Standards for Academic Publishing: A  $q$ - $r$  Theory." *Journal of Political Economy* 110 (5), 994-1034.
- Ellison, Glenn. (2002b): "The Slowdown of the Economics Publishing Process." *Journal of Political Economy* 110 (5), 947-993.
- Ellison, Glenn, (2010): "Is Peer Review in Decline," *Economic Inquiry*, forthcoming.
- Gans, Joshua (2000): *Publishing Economics: An Analysis of the Academic Journal Market in Economics*. London: Edward Elgar.
- Hamermesh, Daniel S. and Gerard A. Pfann (2009): "Markets for Reputation: Evidence on Quality and Quantity in Academe," NBER Working Paper 15527.
- Hirsch, J. E. (2005): "An Index to Quantify an Individual's Scientific Research Output," *Proceedings of the National Academy of Sciences* 102, 16569-16572.
- Hirsch, J. E. (2007): "Does the  $h$  Index Have Predictive Power," *Proceedings of the National Academy of Sciences* 104, 19193-19198.
- Iglesias, Juan E. and Carlos Pecharroman (2007): "Scaling the  $h$ -index for different scientific ISI fields," *Scientometrics* 73 (3), 303-320.
- Jensen, Pablo, Jean-Baptiste Rouquier, and Yves Croissant (2009): "Testing Bibliometric Indicators by Their Prediction of Scientists Promotions," *Scientometrics* 78 (3), 467-479.
- Johnson, Dan. "Getting Noticed in Economics: The Determinants of Academic Citations,"

*American Economist*, 41 (1), 43–50, 1997.

Kosmulski, Marek (2006): “A New Hirsch-type Index Saves Time and Works Equally Well as the Original h-Index,” *ISSI Newsletter* 2, 4-6.

Laband, David N. and Robert D. Tollison. “Intellectual Collaboration,” *Journal of Political Economy*, 108 (3), 632–662, 2000.

Lehmann, Sune, Andrew D. Jackson, and Benny E. Lautrup (2006): “Measures for Measures,” *Nature* 444, 1003-1004.

Liebowitz, Stanley J. and John P. Palmer (1984): “Assessing the Relative Impacts of Economics Journals,” *Journal of Economic Literature* 22, 77-88.

Medoff, Marshall H. “Collaboration and the Quality of Economics Research,” *Labour Economics*, 10, 597–608, 2003.

Möbius, Markus and Tanya Rosenblat, “Getting Closer or Drifting Apart,” *Quarterly Journal of Economics* 119 (3), 971–1009, 2004.

Palacios-Huerta, Iganacio and Oscar Volij (2004): “The Measurement of Intellectual Influence,” *Econometrica* 72 (3), 963-977.

Prathap, Gangan (2009): “Is There a Place for a Mock  $h$ -index?,” *Scientometrics*, forthcoming

Price, Derek J. de Solla (1965). “Networks of Scientific Papers.” *Science* 149 (July 30 1965), 510-515.

Ruane, Frances, and Richard S. J. Tol (2008): “Rational (Successive)  $h$ -indices: An Application to Economics in the Republic of Ireland,” *Scientometrics* 75, 395-405.

Schreiber, Michael (2008): “A Modification of the  $h$ -index: the  $h_m$ -index Accounts for Multi-authored Manuscripts,” *Journal of Informetrics* 2 (3), 211-216.

Snyder, Christopher M. and Owen Zidar (2009): “Resume Padding Among Economists,” mimeo, Dartmouth College.

Tol, Richard S. (2009): “The  $h$ -index and its Alternatives: An Application to the 100 Most Prolific Economists,” *Scientometrics* 80, 317-24.

van Raan, Anthony F. J. (2006): “Comparison of the Hirsch-index with Standard Bibliometric Indicators and with Peer Judgment for 147 Chemistry Research Groups,” *Scientometrics* 67 (3), 491-502

West, Jevin D, Theodore C. Bergstrom, and Carl T. Bergstrom (2010): “The Eigenfactor Metrics: A Network Approach to Assessing Scholarly Journals,” *College & Research Libraries* 71 (23), 236-244.

Wu, Qiang (2009): “The  $w$ -index: A Significant Improvement on the  $h$ -index,” *Journal of the American Society for Information Science and Technology* 61 (3), 609-614.