The h-index and its alternatives:
An application to the 100 most prolific economists

RICHARD S. J. TOL[a,b,c,d]

[a] *Economic and Social Research Institute, Dublin, Ireland*
[b] *Institute for Environmental Studies, Vrije Universiteit, Amsterdam, The Netherlands*
[c] *Department of Spatial Economics, Vrije Universiteit, Amsterdam, The Netherlands*
[d] *Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA*

The h-index is a recent but already quite popular way of measuring research quality and quantity. However, it discounts highly-cited papers. The g-index corrects for this, but it is sensitivity to the number of never-cited papers. Besides, h- or g-index-based rankings have a large number of ties. Therefore, this paper introduces two new indices, and tests their performance for the 100 most prolific economists. A researcher has a t-number (f-number) of t (f) if t (f) is the largest number for which it holds that she has t (f) publications for which the geometric (harmonic) average number of citations is at least t (f). The new indices overcome the shortcomings of the old indices.

Introduction

A researcher has a *h*-index of *h* if *h* is the highest number for which holds that she has *h* publications that are cited at least *h* times [HIRSCH, 2005]. Research quantity and quality are measured with a single, simple index, that has taken the world of research assessment by storm. The appeal of the *h*-index is intuitive. To increase your *h*-number, you need more and better papers.

There is one shortcoming of the *h*-index, and one drawback. The shortcoming is that it is a measure of lifetime achievement. Therefore, normalised values have been proposed that adjust for the stage of career [EGGHE, 2007; SIDIROPOULOS & AL., 2007]. The drawback is that the *h*-index is determined by the number of citations of a single paper. Joe Stiglitz has the highest *h*-number in the dataset used in this paper. It is 53, as he has 53 papers that are cited at least 53 times. However, 3 of his papers are cited more than 1,000 times. The *h*-index ignores this, even though prizes are won with a few exceptional papers.

Therefore, EGGHE [2006] independently introduced the *g*-index, while JIN [2006] introduced the numerically equivalent *A*-index. A researcher has a *g*-index of *g* if *g* is the highest number for which holds that she has *g* publications that are cited at least *g* times on average.

The problem with the g-index is that the same as with all arithmetic averages: It ignores the distribution of citations. A researcher with one highly-cited paper, perhaps written with a co-author, and a large number of unremarkable publications would still have a high g-number. JIN & AL. [2007] therefore introduce the r-index, which is the product of the square roots of g and h. Here, I propose a different solution, namely to use the harmonic and the geometric averages rather than the arithmetic one.

Data and methods

The 100 most prolific economists were identified from *IDEAS/REPEC*, using the "number of distinct works" in May 2007.[1] As *IDEAS/REPEC* still has only a poor coverage of citations, publication and citation data for these 100 economists were downloaded from the *ISI Web of Science*. See in Appendix, Table A1.

The h-index solves

$$\max_h c_h \geq h \qquad (1)$$

where c_i is a series of publications, denoted by their number of citations, in declining order. Both c_i and h are natural numbers; see Ruane and Tol (forthcoming) for a rational h-index. The h-index is not larger than the number of cited publication. Papers that are cited at least h times, are said to be in the h-core [LIANG, 2006].

The g-index solves

$$\max_g \sum_{i=1}^{g} c_i \geq g^2 \Leftrightarrow \max_g \frac{1}{g} \sum_{i=1}^{g} c_i \geq g \qquad (2)$$

The first formulation is due to EGGHE [2006], the second is inspired by JIN [2006].[2] The average number of citations is a real number, but the number of papers is a natural number and therefore the g-number is natural too. The g-index is not larger than the number of publications, but may be larger than the number of cited publication.

The f-index, here introduced, solves

$$\max_f \frac{1}{\frac{1}{f} \sum_{i=1}^{f} \frac{1}{c_i}} \geq f \qquad (3)$$

The f-index is not larger than the number of cited publications, as the harmonic means goes to zero at the point where the number of citations goes to zero. As with the

---

[1] See http://ideas.repec.org/top/top.person.dnbworks.html for the latest ranking.
[2] Note that JIN'S [2006] *A*-index is the average number of citation in the h-core, while EGGHE'S [2006] g-index is the average number of citations in the g-core.

h-index, the least-cited paper is the most important one, but in contrast with the h-index, any citation to any paper in f-core increases the f-index.

The t-index, here introduced, solves

$$\max_t \exp\left[\frac{1}{t}\sum_{i=1}^t \ln(c_i)\right] \geq t \Leftrightarrow \max_t \prod_{i=1}^t c_i^{1/t} \geq t \tag{4}$$

The t-index shares many of its properties with the f-index. However, the first partial derivative of the harmonic mean is proportional to c^{-2}, while the first partial derivative is the geometric mean is proportional to $c^{-1/t}$. In both cases, an additional citation to a not-so-often cited paper counts more than an additional citation to an often-cited paper. However, this effect is stronger in case of the harmonic mean. The f-index is more egalitarian than the t-index. Furthermore, the t-index is more egalitarian than the g-index, and f-index is less egalitarian than the h-index. The following relationship holds

$$h \leq f \leq t \leq g \tag{5}$$

One can also graphically solve the above indices. Rank the publications in the number of citations. Plot the number of citations, or the respective averages for the first n papers. Plot the rank. The index is where the lines cross, rounded down. Figure 1 illustrates this for the publications of Joe Stiglitz.
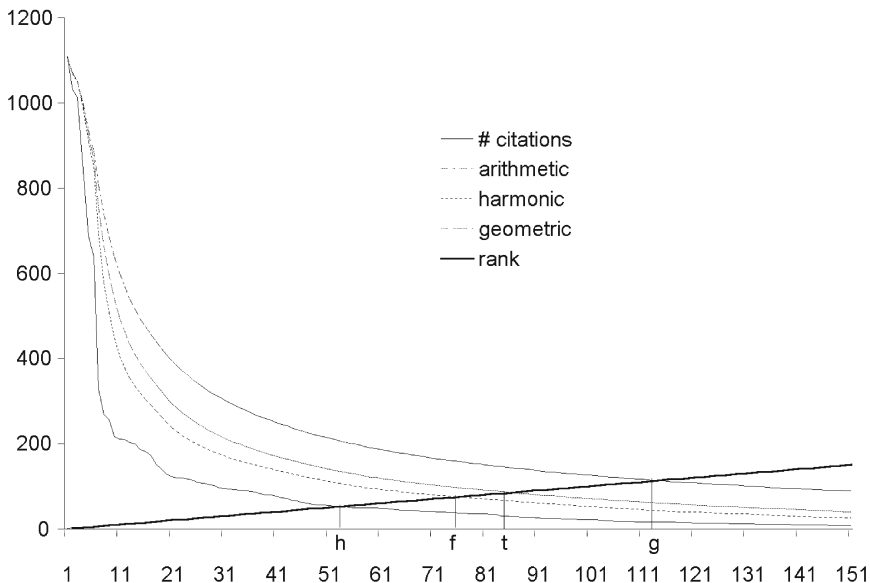

Figure 1. The number of citations per paper by Joe Stiglitz, the arithmetic, geometric, and harmonic averages of the n most cited papers, and the citation rank

Results

Figure 2 plots the *g*-, *f*- and *t*-indices against the *h*-index. Table A1 has the numbers. Clearly, there is a strong correlation between the indices. Table 1 confirms this.

Table 1. Rank correlations

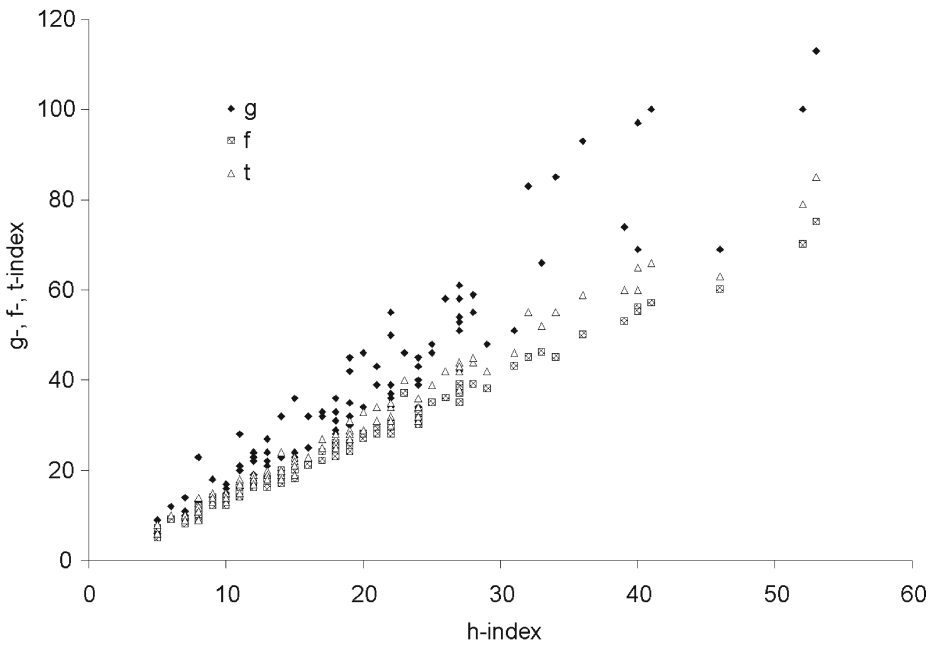|	c	c/p	h	g	f	t	r
p	0.662	0.294	0.650	0.592	0.645	0.633	0.624
c		0.893	0.970	0.990	0.977	0.985	0.991
c/p			0.867	0.925	0.880	0.898	0.907
h				0.960	0.994	0.989	0.985
g					0.971	0.983	0.990
f						0.997	0.990
t							0.995

Figure 2. The *g*-, *f*-, and *t*-index as a function of the *h*-index

It shows the rank correlation. For completeness, the ranks based on the number of publications, the number of citations, and the average number of citations per paper is also added, as is the *r*-index for completeness. The rank correlation between the indices is 96% or higher. With citations or citations per paper, the rank correlation is 87% or higher. Only the ranking based on the number of papers stands out: The rank correlation may be as low as 59%. Therefore, the various indices do add information about research quality relative to publication and citation numbers, but they add little information relative to one another.

The added value of the *g*-, *f*- and *t*-indices therefore does not lie in a different overall ranking. The details of the ranking are different, however. The *h*-index produces a large number of ties: 86 out of 100 economists are tied with at least one other person. The *g*-index improves on this, but only slightly: there are 77 ties. The *f*-index does much better with 46 ties, and the *t*-index is best with 40 ties. Because of the non-linearity of the harmonic and geometric means, the *f*- and *t*-indices are more sensitive to small differences between researchers.

Above, I noted that the *g*-index is not constrained in principle by the number of cited papers. In practice, this holds for 3 of the 100 economists: Robert Barro, Robert Engle, and James Poterba. For each of these, it is true that if they publish one more paper, their *g*-index automatically increases by 1, regardless of whether that paper is ever cited; indeed their *g*-number is constrained by the number of publication. In fact, Engle's *g*-number is 10 greater than his number of cited papers, and for Barro the gap is 14. Poterba has no papers that are uncited. The *h*-, *f*- and *t*-indices do not have this problem, by construction.

Conclusion

I propose two alternatives to the *h*-index. Like the *h*-index, both alternatives have the advantage that research quantity and quality are combined in a single number. Unlike the *g*-index proposed earlier, the indices proposed here are constrained by the number of cited papers. Furthermore, the new indices have more discriminatory power than the *h*- and *g*-indices. However, the broad ranking of researchers is by and large the same for the four alternative indices. Computing all four indices adds work but no insight.

That leaves the choice between the *f*- and the *t*-index. The *t*-index has more discriminatory power, but that may be true only for the particular set of researchers assessed here. Both indices are maximum if every paper is cited the same number of times, but the *f*-index deviates much faster from this maximum than does the *t*-index. As a few papers get cited much more than others, it seems reasonable not to place too much weight on the distribution of citations. This argument favours the *t*-index, but admittedly this is a matter of taste.

*

I had useful discussions on this subject with Frances Ruane.

References

EGGHE, L. (2007), Dynamic h-index: The Hirsch index in function of time, *Journal of the American Society for Information Science and Technology*, 58 (3) : 452-454.

EGGHE, L. (2006), Theory and practise of the g-index, *Scientometrics*, 69 (1) : 131–152.

HIRSCH, J. E. (2005), An index to quantify an individual's scientific research output, *Proceedings of the National Academy of Science*, 102 : 16569–16572.

JIN, B. (2006), H-index: An evaluation indicator proposed by scientist, *Science Focus* (in Chinese*)*, 1 (1) : 8–9.

JIN, B., L. LIANG, R. ROUSSEAU, L. EGGHE (2007), The R- and AR-indices: Complementing the h-index, *Chinese Science Bulletin*, 52 (6) : 855–863.

LIANG, L. (2006), h-index sequence and h-index matrix: Constructions and applications, *Scientometrics*, 69 (1) : 153–159.

RUANE, F. P., R. S. J. TOL, Rational (successive) h-indices: An application to economics in the Republic of Ireland, *Scientometrics* (forthcoming).

SIDIROPOULOS, A., D. KATSAROS, Y. MANOLOPOULOS (2007), Generalized Hirsch h-index for disclosing latent facts in citation networks, *Scientometrics*, 72 (2) : 253–280.

Appendix

Table A1. The number of the papers, cited papers and citations; the average number of citations per paper; and the *h*-, *g*-, *f*-, *t*- and *r*-numbers of the 100 most prolific economists according to IDEAS/REPEC in May 2007. The economists are ordered according to the number of publications counted in IDEAS/REPEC.

Name	papers	cited	citations	cit/par	h	g	f	t	r
Joseph E. Stiglitz	190	183	13565	71.4	53	113	75	85	77
Peter C. B. Phillips	160	128	7405	46.3	34	85	45	55	54
Stephen J Turnovsky	169	150	1972	11.7	24	33	30	31	28
Martin Shubik	149	103	1205	8.1	17	32	24	27	23
Barry Julian Eichengreen	101	78	800	7.9	15	24	20	21	19
Martin S. Feldstein	186	164	5473	29.4	40	69	56	60	53
M Hashem Pesaran	92	85	2361	25.7	25	46	35	39	34
Sebastian Edwards	78	63	1202	15.4	18	33	24	27	24
John Whalley	130	92	1335	10.3	17	33	22	25	24
Jean-Jacques Laffont	160	135	2838	17.7	29	48	38	42	37
Bruno S. Frey	140	123	2216	15.8	27	42	35	38	34
Richard B. Freeman	96	80	2757	28.7	27	51	39	43	37
Dermot James Hayes	62	39	635	10.2	13	24	18	20	18
Werner Gueth	79	48	853	10.8	11	28	15	18	18
Jason Shogren	154	119	1360	8.8	19	30	25	27	24
James J. Heckman	138	116	10081	73.1	41	100	57	66	64
David A. Peel	189	121	769	4.1	11	21	16	18	15
Christopher F Baum	28	15	104	3.7	5	9	7	8	7
Jeffrey Alexander Frankel	69	62	2125	30.8	24	45	32	36	33
Bruce Alan Babcock	40	34	449	11.2	14	19	18	18	16
Jean Tirole	120	112	5068	42.2	46	69	60	63	56
William Poole	44	27	551	12.5	8	23	12	14	14
Andrew Hughes Hallett	27	17	60	2.2	5	6	5	6	5
Laurence J. Kotlikoff	58	49	1370	23.6	18	36	25	28	25
James Poterba	12	12	250	20.8	8	12	11	12	10
J. Scott Armstrong	74	65	2146	29.0	19	45	25	29	29
Carl Chiarella	55	29	162	2.9	7	11	9	10	9
John C. Quiggin	119	94	1217	10.2	16	32	21	23	23
Helen H. Jensen	26	16	99	3.8	7	9	8	9	8
Mark P. Taylor	99	91	1815	18.3	24	40	31	34	31
Alan B. Krueger	67	62	3024	45.1	27	54	38	42	38
Lawrence H Summers	110	96	5644	51.3	39	74	53	60	54
Wolfgang Karl Haerdle	86	74	2057	23.9	24	43	33	36	32
Clive W. J. Granger	136	104	8792	64.6	36	93	50	59	58
Christian S. Gourieroux	56	45	1075	19.2	14	32	20	24	21
John Christopher Beghin	39	26	153	3.9	8	10	9	9	9
Marcel Boyer	39	31	243	6.2	8	13	11	12	10
Robin W. Boadway	68	61	778	11.4	15	24	20	22	19
Jere Richard Behrman	113	99	1617	14.3	24	34	30	32	29
Ronald MacDonald	92	69	696	7.6	14	23	19	20	18
Michael McAleer	123	87	708	5.8	15	23	18	19	19
Joshua Aizenman	64	50	364	5.7	10	16	13	14	13
Bennett McCallum	83	69	1706	20.6	24	39	31	34	31
Paul A. Samuelson	204	152	4848	23.8	33	66	46	52	47
Lars E. O. Svensson	94	84	2397	25.5	23	46	37	40	33
Richard J. Arnott	73	67	1311	18.0	22	34	28	30	27
Olivia S. Mitchell	35	25	442	12.6	14	20	18	19	17
David F. Hendry	93	82	3534	38.0	26	58	36	42	39
Rudiger Dornbusch	96	71	2552	26.6	22	50	29	34	33
Andrei Shleifer	111	107	10022	90.3	52	100	70	79	72
Stephen P. Jenkins	56	49	628	11.2	14	23	19	20	18
David Neumark	94	78	1444	15.4	20	34	27	29	26
David M Newbery	53	44	642	12.1	12	24	17	19	17
Robert F. Engle	83	73	10373	125.0	32	83	45	55	52

Table A1. (cont.)

Name	papers	cited	citations	cit/par	h	g	f	t	r
Robert J. Barro	97	83	9501	97.9	40	97	55	65	62
Daniel L Thornton	35	26	219	6.3	7	14	8	9	10
Jeffrey Marc Wooldridge	39	30	765	19.6	13	27	18	20	19
Andre de Palma	92	71	1203	13.1	19	32	26	28	25
Ngo Van Long	94	56	319	3.4	9	14	12	13	11
Philippe Michel	86	60	398	4.6	10	17	13	14	13
David B. Audretsch	81	68	1851	22.9	19	42	26	31	28
Richard S.J. Tol	64	48	525	8.2	13	21	18	19	17
Robert D. Tollison	182	137	1412	7.8	18	31	24	26	24
Philip Hans Franses	142	105	695	4.9	14	19	17	18	16
Daron Acemoglu	59	54	1559	26.4	22	39	29	32	29
David Matthew Levinson	41	24	142	3.5	7	10	8	9	8
Bruce D. Smith	114	89	1129	9.9	18	28	23	25	22
W Kip Viscusi	163	146	3394	20.8	31	51	43	46	40
Stephen M. Miller	76	57	586	7.7	12	22	17	19	16
Thomas J. Sargent	84	76	3516	41.9	27	58	38	44	40
Murray C. Kemp	129	98	917	7.1	16	25	21	23	20
Alberto Alesina	72	61	2832	39.3	27	53	37	42	38
Carl Walsh	52	44	605	11.6	12	23	16	18	17
Michael David Bordo	63	45	400	6.3	9	18	14	15	13
James Tobin	103	63	3133	30.4	22	55	30	35	35
David A. Hennessy	61	28	196	3.2	6	12	9	10	8
Myrna Wooders	45	37	458	10.2	11	20	16	17	15
Andrew Rose	64	56	1476	23.1	22	37	28	32	29
William A. Brock	60	53	1904	31.7	21	43	29	34	30
Frederick (Rick) van der Ploeg	80	68	567	7.1	12	19	16	17	15
John M. Hartwick	66	52	1335	20.2	15	36	20	23	23
Stijn Claessens	42	34	568	13.5	13	22	16	18	17
Rik Hafer	40	32	304	7.6	10	15	12	13	12
Daniel Hamermesh	103	80	1219	11.8	18	31	26	28	24
William Arnold Barnett	58	39	920	15.9	18	29	23	25	23
Ray C. Fair	59	54	1586	26.9	21	39	28	31	29
Jaime A.P. de Melo	52	48	398	7.7	10	16	14	15	13
Lester Ingber	52	41	1266	24.3	19	35	24	27	26
Eric S. Maskin	74	70	3169	42.8	28	55	39	44	39
Walter Erwin Diewert	62	57	2195	35.4	20	46	28	33	30
Stephen John Nickell	78	70	2467	31.6	25	48	35	39	35
Timothy J. Besley	62	54	1390	22.4	22	36	28	31	28
Olivier Blanchard	75	64	3566	47.5	28	59	39	45	41
Eric Ghysels	59	46	669	11.3	15	24	20	22	19
Gilles Saint-Paul	34	23	338	9.9	9	18	12	14	13
Avinash Kamalakar Dixit	91	84	3891	42.8	27	61	37	43	41
Walter Bossert	69	56	415	6.0	11	16	14	15	13
Philip Arestis	62	40	237	3.8	8	12	10	11	10
Ping Wang	52	42	371	7.1	10	17	13	14	13
John Roemer	77	60	648	8.4	15	22	20	21	18