



University of Sussex  
Business, Management & Economics

## Economics Department Working Paper Series

No.55-2013

**The Matthew Effect for Cohorts of Economists**

**Richard S. J. Tol**

**Department of Economics, University of Sussex, Brighton, United Kingdom**

**Institute for Environmental Studies, Vrije Universiteit, Amsterdam, Netherlands**

**Department of Spatial Economics, Vrije Universiteit, Amsterdam, Netherlands**

**Email: [R.Tol@sussex.ac.uk](mailto:R.Tol@sussex.ac.uk)**

**Abstract:** This paper applies the Ijiri-Simon test for systematic deviations from Gibrat's Law to citation numbers of economists. It is found that often-cited researchers attract a new citation numbers that are disproportionate to the quality of their work. It is also found that this Matthew Effect is stronger for economists who started their academic career earlier.

**JEL Classification:** A14

**Key Words:** Matthew Effect

## 1. Introduction

To those that hath shall be given. The Matthew Effect describes self-reinforcing success (Merton 1968). We watch Gangnam Style on YouTube not because we particularly like Psy's music, but because we are curious what everyone is talking about. We sometimes cite an academic paper, not because it is particularly good or relevant, but because everyone else cites it and we want to signal that we are au fait with what is seminal or hot (Hu et al. 2006; Hunt and Blair 1987; Kim et al. 2007; Simkin and Roychowdhury 2005; Strevens 2006). In a previous paper (Tol 2009), I introduce a formal test of the Matthew Effect, and test whether papers that have often been cited attract a disproportionate amount of new citations. That paper used a sample of 100 eminent economists. I here use a more convincing sample of 31,000 economists, and also test whether the Matthew Effect is stronger for older cohorts.

There are a number of papers that offer anecdotic support for the Matthew Effect (Donner-Banzhoff 2011; Keith and Babchuk 1998; Medoff 2006; Merton 1968; Nieri et al. 2007; Serenko et al. 2011; van Dalen and Henkens 2001), but (Tol 2009) was the first to offer a rigorous empirical test – albeit at the expense of replacing Merton's broad concept of "recognition" with the narrow notion of "number of citations". (Bonitz et al. 1997; Bonitz et al. 1999; Bonitz 2005; Brown 2004) use a Matthew indicator, defined as the number of citations in excess of the number of citations than would be expected on the basis of the journal impact factor. This ignores within-journal heterogeneity and therefore cannot be used for individual authors or papers. At an aggregate level (universities, countries), it is more appropriate, but still only defined for the short period after publication considered by the journal impact factor. A number of papers use definitions of the Matthew Effect that deviate from what appears to be Merton's intention. For example, (Schott 1998) shows that there are centers and peripheries in science. He argues that the "centre" benefits from being the centre – a Matthew Effect but cannot exclude, however, that the centre is the centre because it is better. (Lariviere and Gingras 2010) show that papers in higher-ranked journals are better cited; although they control for paper quality, they do not for journal readership. (Asrejadidi et al. 2011; Bazeley 1998; Gonzalez-Brambila and Veloso 2007; Katz 1999; Kenna and Berche 2011; Lange and Frensch 1999; Laudel 2006; Lesueur 2012; van Looy et al. 2004; Weakliem et al. 2012) show that there are (interacting) scale and persistence effects in productivity, research funding and citations, but do not show self-reinforcement, a crucial feature of the Matthew Effect.

The Matthew Effect may be defined as a positive feedback effect (Hu et al. 2007) or as increasing returns to scale (Tol 2009). If a number of firms of different size all grow at the same rate, then the size distribution of those firms will converge to Gibrat's Law (Simon 1955).<sup>1</sup> That is, the log of the firm size is proportional to the log of the rank. Gibrat's Law is also known as Zipf's Law (generalized), and the distribution is the Pareto or Lotka distribution (Egghe 2005).

---

<sup>1</sup> (Bothner et al. 2010; Kenna and Berche 2011; Squazzoni and Gandelli 2012) offer alternative mathematical models of the Matthew Effect.

The paper proceeds as follows. Section 2 describes the method, and Section 3 the data. Section 4 discusses the results. Section 5 concludes.

## 2. Method

Consider a population of objects, each with a different size. If each object grows at the same rate, the distribution of sizes will be a Pareto distribution. The size of each object will be inversely proportional to its rank raised to some power. In natural logarithms

$$(1) \quad \ln S_i = \alpha + \beta \ln R_i$$

where  $S$  is the size of object  $i$  and  $R$  is its size rank. If  $\beta=-1$ , this is Zipf's law. However, Equation (1) will not hold if large objects typically grow faster or slower than small objects. One test (Ijiri and Simon 1974) uses the regression model

$$(2) \quad \ln S_i = \alpha + \beta \ln R_i + \gamma \ln^2 R_i$$

If  $\gamma < 0$  ( $\gamma > 0$ ), large objects grow systematically faster (slower) than small objects. If  $\gamma = 0$  – rather if the estimated  $\gamma$  does not deviate statistically significantly from zero – then the growth rate is independent of the size.

This test has been applied, among other things, to firm size (Ijiri and Simon 1974), to motion pictures (de Vany and Walls 1997; de Vany and Walls 2004), to theatre productions (Maddison 2004), and to citations (Tol 2009). In the last case we recognize the Matthew Effect: often-cited authors receive many new citations – and disproportionately so. If the growth rate is independent of the initial number of citations, then good papers would be cited often and bad papers not so often:  $\gamma = 0$ . If the growth rate is positively correlated with the initial number of citations, then good papers are cited excessively:  $\gamma < 0$ .

I estimate Equation (2) separately for every cohort, where a cohort is defined by the first year of publication. The Matlab script is reproduced in Appendix B. One would expect that the Matthew Effect is weak for young scholars. Famous schools and famous scholars often have the first pick of PhD candidates and presumably select the most promising ones. Some of that fame rubs off on young scholars. Therefore, the Matthew Effect would be initially weak – rather than absent. Over time, however, as scholars build up their own prestige (or not), the Matthew Effect should become stronger.

## 3. Data

IDEAS/RePEc (Krichel and Zimmermann 2005) is an internet service for economists at <http://ideas.repec.org/>. It operates as an archive for working papers (almost 500,000 items from over 3,500 series). Publication lags are substantial in economics. IDEAS/RePEc

serves as the main platform for the early availability of submitted articles. IDEAS/RePEc also operates archives for journal articles (almost 800,000 items from over 1,500 journals), books and book chapters (almost 30,000 items), and software components (almost 3,000 items). Paper series, individuals and departments are ranked on a range of indicators (Seiler and Wohlrabe 2012; Zimmermann 2007). There are nascent activities on curated bibliographies, academic genealogies, and collaboration networks. IDEAS/RePEc has profiles of over 43,000 economists, and a database of over 10 million references, 4 million of which are to items in the publication databases. I here exploit the latter two.

Data were collected on 9 and 10 November 2012. Data were scraped from the simplified citation profiles. The Matlab scripts are reproduced in Appendix A. The citation profiles have information on the number of works (journal articles, working papers, books, chapters), the number of co-authors, total citations, total self-citations (of any co-author), h-index, and years since first publication. There is a separate page for each citation profile. I wrote a Matlab script, reproduced in Appendix A, to visit each page and extract the relevant data. The indicator used here is total citations minus self-citations.

Data can be found at <http://www.sussex.ac.uk/Users/rt220/CohortMatthewPublic.xlsx>

More than 43,000 economists have registered with IDEAS/RePEc. Over 34,000 have listed publications. 24,898 economists have been cited. Figure 1 depicts the numbers per cohort, where cohorts are defined by years since first publication. Table 1 has the actual numbers. “Young” economists are overrepresented, reflecting both a thriving discipline in a growing higher education sector and a greater importance of IDEAS/RePEc.

#### 4. Results

Figure 2 shows the histogram of citations per paper. Most economists are not cited very well: The mode of the distribution lies between 100 and 200 citations, or one citation per paper. The distribution trails off, but the rate of decline is less than exponential, indicating that the distribution is not a Pareto one – a first sign of the Matthew Effect. At the log-scale of Figure 2, there is a slight curve at the bottom ranks. At the top ranks, the curvature is more pronounced: Top ranking economists have *more* citations than suggested by the Pareto distribution, an indication of the Matthew Effect.

Figure 3 shows the natural logarithm of citations as a function of the natural logarithm of its rank. The violation of Gibrat’s law is obvious at the bottom ranks. More subtly, at the top ranks, the curve occasionally ticks up: Extrapolation from the citation record of mid ranking economists to top ranking ones would thus *underestimate* their success, another indication of the Matthew Effect.

Eyeballing data plots is instructive but not conclusive. Regression results are shown in Table 1. Figure 2 depicts the parameter of interest,  $\gamma$ , for each cohort. As expected, the Matthew Effect is weak for young scholars. In fact, for the first cohort,  $\gamma > 0$ . That is, the opposite of the Matthew Effect is observed. This is probably because the best schools in

economics tend to withhold their students from publication, as the first paper is supposed to make a big splash.

For all cohorts bar the first, the Matthew Effect is observed. It is small but significant for young researchers, and increases in strength for older scholars. As the number of observations is much lower for the oldest cohorts (see Figure 1), confidence in the evidence declines – but the Matthew Effect remains significantly different from zero.

## 5. Discussion and conclusion

This paper formally tests the Matthew Effect for citations based on systematic deviations from Gibrat's Law. Using citation data for almost 25,000 economists, it is shown that the citation distribution is more skewed than can be expected from quality differences alone. This is clear from plots of the data, and confirmed by regression analysis. It is also found that the Matthew Effect is stronger for older researchers.

This paper thus provides further evidence for the Matthew Effect, at least for citations to economists. Future research should test the Matthew Effect for other disciplines and for other indicators of success.

There are a number of caveats. Self-registration, as practiced by IDEAS/RePEc, may imply sample selection bias. The quality of citation data on IDEAS/RePEc has never been tested. Citations are limited to citations in economic journals and working papers, which implies that citation numbers are underreported for those researching on the boundaries of economics. Most importantly, the Ijiri-Simon test relies on a mathematical model to draw conclusions about a dynamic process from a static cross-section. Future research will have to test the Matthew Effect using appropriate statistical models of the citation histories of a large and representative sample of researchers.

## References

- Asrejadidi, S.J., M.H.Biglu, and A.Rafi (2011), 'The study of "Matthew Effect" on science productions of Iranian Medical Universities on the basis of "power-law relationship" in a five year period in the Web of Science', *Information Sciences and Technology*, **26**, (4), pp. 67-81.
- Bazeley, P. (1998), 'Peer review and panel decisions in the assessment of Australian Research Council project grant applicants: what counts in a highly competitive context?', *Higher Education*, **35**, 435-452.
- Bonitz, M. (2005), 'Ten Years Matthew Effect for Countries', *Scientometrics*, **64**, (3), 375-379.
- Bonitz, M., E.Bruckner, and A.Scharnhorst (1997), 'Characteristics and Impact of the Matthew Effect for Countries', *Scientometrics*, **40**, (3), 407-422.

- Bonitz, M., E.Bruckner, and A.Scharnhorst (1999), 'The Matthew Index -- Concentration Patterns and Matthew Core Journals', *Scientometrics*, **44**, (3), 361-378.
- Bothner, M.S., R.Haynes, W.Lee, and E.B.Smith (2010), 'When do Matthew Effects occur?', *Journal of Mathematical Sociology*, **34**, (2), pp. 80-114.
- Brown, C. (2004), 'The Matthew Effect of the Annual Reviews Series and the Flow of Scientific Communication through the World Wide Web', *Scientometrics*, **60**, (1), 25-36.
- de Vany, A. and W.D.Walls (1997), 'The Market for Motion Pictures: Rank, Revenue, and Survival', *Economic Inquiry*, **35**, 783-797.
- de Vany, A. and W.D.Walls (2004), 'Motion Picture Profit, the Stable Paretian Hypothesis, and the Curse of the Superstar', *Journal of Economic Dynamics & Control*, **28**, 1035-1057.
- Donner-Banzhoff, N. (2011), 'The Matthew-effect, superstars and the impact factor: An international comparison of family practice research output', *Zeitschrift fur Allgemeinmedizin*, **87**, (9), pp. 366-370.
- Egghe, L. (2005), 'A Characterization of the Law of Lotka in Terms of Sampling', *Scientometrics*, **62**, (3), 321-328.
- Gonzalez-Brambila, C. and F.M.Veloso (2007), 'The determinants of research output and impact: A study of Mexican researchers', *Research Policy*, **36**, 1035-1051.
- Hu, M.-B., R.Jiang, Q.-S.Wu, and Y.-H.Wu (2007), 'Simulating the wealth distribution with a Richest-Following strategy on scale-free network', *Physica*, **A381**, 467-472.
- Hu, M.-B., W.-X.Wang, R.Jiang, Q.-S.Wu, B.-H.Wang, and Y.-H.Wu (2006), 'A unified framework for the Pareto Law and Matthew Effect using scale-free networks', *European Physical Journal*, **B53**, 273-277.
- Hunt, J.G. and J.D.Blair (1987), 'Content, Process, and the Matthew Effect Among Management Academics', *Journal of Management*, **13**, (2), 191-210.
- Ijiri, Y. and H.A.Simon (1974), 'Interpretations of Departures form the Pareto Curve Firm-Size Distributions', *Journal of Political Economy*, **82**, (2), 315-331.
- Katz, J.S. (1999), 'The Self-Similar Science System', *Research Policy*, **28**, 510-517.
- Keith, B. and N.Babchuk (1998), 'The Quest for Institutional Recognition: A Longitudinal Analysis of Scholarly Productivity and Academic Prestige among Sociology Departments', *Social Forces*, **76**, (4), 1495-1533.
- Kenna, R. and B.Berche (2011), 'Normalization of peer-evaluation measures of group research quality across academic disciplines', *Research Evaluation*, **20**, (2), pp. 107-116.
- Kim, J.-Y., I.Min, and C.Zimmermann (2007), *The Economics of Citation*, Department of Economics Working Paper **2007-31**, University of Connecticut, Storrs.
- Krichel, T. and C.Zimmermann (2005), 'The economics of open bibliographic data provision', *Economic Analysis and Policy*, **39**, (1), 143-152.

- Lange, L.L. and P.A.Frensch (1999), 'Gaining Scientific Recognition by Position: Does Editorship Increase Citation Rates?', *Scientometrics*, **44**, (3), 459-486.
- Lariviere, V. and Y.Gingras (2010), 'Brief communication: The impact factor's Matthew Effect: A natural experiment in bibliometrics', *Journal of the American Society for Information Science and Technology*, **61**, (2), pp. 424-427.
- Laudel, G. (2006), 'The 'quality myth': Promoting and hindering conditions for acquiring research funds', *Higher Education*, **52**, 375-403.
- Lesueur, J.Y. (2012), 'The scientific productivity of French economists: Some econometric results from the PES tournament experience', *Revue Economique*, **63**, (4), pp. 743-778.
- Maddison, D.J. (2004), 'Increasing Returns to Information and the Survival of Broadway Theatre Productions', *Applied Economics Letters*, **11**, 639-643.
- Medoff, M.H. (2006), 'Evidence of a Harvard and Chicago Matthew Effect', *Journal of Economic Methodology*, **13**, (4), 485-506.
- Merton, R.K. (1968), 'The Matthew Effect in Science', *Science*, **159**, (3810), 56-63.
- Nieri, M., D.Saletta, L.Guidi, J.Buti, D.Franceschi, S.Mauro, and G.P.Pini-Prato (2007), 'Citation Classics in Periodontology: A Controlled Study', *Journal of Clinical Periodontology*, **34**, 349-358.
- Schott, T. (1998), 'Ties between Center and Periphery in the Scientific World System: Accumulation of Rewards, Dominance and Self-Reliance in the Center', *Journal of World-Systems Research*, **4**, 112-144.
- Seiler, C. and K.Wohlrabe (2012), 'Ranking economists on the basis of many indicators: An alternative approach using RePEc data', *Journal of Informetrics*, **6**, (3), pp. 389-402.
- Serenko, A., R.A.K.Cox, N.Bontis, and L.D.Booker (2011), 'The superstar phenomenon in the knowledge management and intellectual capital academic discipline', *Journal of Informetrics*, **5**, (3), pp. 333-345.
- Simkin, M.V. and V.P.Roychowdhury (2005), 'Stochastic modeling of citation slips', *Scientometrics*, **62**, (3), 367-384.
- Simon, H.A. (1955), 'On a Class of Skew Distribution Functions', *Biometrika*, **42**, (3/4), 425-440.
- Squazzoni, F. and C.Gandelli (2012), 'Saint Matthew strikes again: An agent-based model of peer review and the scientific community structure', *Journal of Informetrics*, **6**, (2), pp. 265-275.
- Strevens, M. (2006), 'The Role of the Matthew Effect in Science', *Studies in History and Philosophy of Science*, **37**, 159-170.
- Tol, R.S.J. (2009), 'The Matthew Effect Defined and Tested for the 100 Most Prolific Economists', *Journal of the American Society for Information Science and Technology*, **60**, (2), 420-426.

van Dalen, H.P. and K.Henkens (2001), 'What makes a scientific article influential? The case of demographers', *Scientometrics*, **50**, (3), 455-482.

van Looy, B., M.Ranga, J.Callaert, K.Debackere, and E.Zimmermann (2004), 'Combining Entrepreneurial and Scientific Performance in Academia: Towards a Compounded and Reciprocal Matthew-Effect?', *Research Policy*, **33**, 425-441.

Weakliem, D.L., G.Gauchat, and B.R.E.Wright (2012), 'Sociological Stratification: Change and Continuity in the Distribution of Departmental Prestige, 1965-2007', *American Sociologist*, **43**, (3), pp. 310-327.

Zimmermann, C. (2007), *Academic Rankings with RePEc*, Department of Economics Working Paper Series **2007-36**, University of Connecticut, Storrs.



Table 1. Results for regressions, by cohort, of  $\ln(\text{number of citations})$  on  $\ln(\text{rank})$  and  $\ln^2(\text{rank})$ .

Cohort	N	Rank			Rank squared		
		$\beta$	$\sigma_\beta$	$t$	$\gamma$	$\sigma_\gamma$	$T$
1	2471	-0.48	0.02	-22.5	0.01	0.00	3.0
2	2179	0.14	0.03	5.5	-0.06	0.00	-27.2
3	1905	0.24	0.03	9.4	-0.08	0.00	-33.1
4	1688	0.36	0.03	12.9	-0.09	0.00	-36.7
5	1509	0.27	0.03	10.5	-0.09	0.00	-36.9
6	1355	0.26	0.03	10.1	-0.09	0.00	-38.1
7	1278	0.49	0.03	19.1	-0.11	0.00	-47.9
8	1210	0.37	0.03	13.1	-0.11	0.00	-41.9
9	1239	0.41	0.02	17.8	-0.11	0.00	-51.8
10	1096	0.54	0.03	18.1	-0.13	0.00	-44.7
11	1015	0.60	0.03	18.7	-0.14	0.00	-45.5
12	938	0.68	0.04	19.4	-0.16	0.00	-44.5
13	933	0.66	0.04	17.3	-0.16	0.00	-42.4
14	754	0.60	0.04	14.3	-0.17	0.00	-38.0
15	732	0.65	0.04	15.3	-0.17	0.00	-37.3
16	644	0.77	0.04	17.4	-0.18	0.00	-38.5
17	535	0.40	0.04	9.4	-0.16	0.00	-33.0
18	516	0.43	0.05	9.4	-0.15	0.01	-29.5
19	410	0.38	0.06	6.5	-0.16	0.01	-23.5
20	365	0.90	0.07	13.3	-0.23	0.01	-27.7
21	314	0.18	0.08	2.4	-0.15	0.01	-16.0
22	299	0.51	0.09	5.6	-0.19	0.01	-16.3
23	234	0.94	0.11	8.4	-0.26	0.01	-17.5
24	208	0.82	0.08	10.0	-0.25	0.01	-22.6
25	173	0.47	0.10	4.5	-0.22	0.02	-14.8
26	140	0.38	0.13	3.0	-0.23	0.02	-12.2
27	111	0.09	0.16	0.6	-0.22	0.03	-8.6
28	92	0.58	0.14	4.2	-0.29	0.02	-12.2
29	88	0.24	0.10	2.5	-0.22	0.02	-13.3
30	72	0.28	0.17	1.6	-0.26	0.03	-8.3
31	60	0.44	0.16	2.8	-0.33	0.03	-10.6
32	69	0.37	0.13	2.8	-0.32	0.02	-13.2
33	54	0.38	0.14	2.7	-0.28	0.03	-9.9
34	45	0.62	0.23	2.8	-0.38	0.05	-7.9
35	36	0.10	0.17	0.6	-0.24	0.04	-5.9
36	33	0.37	0.16	2.3	-0.32	0.04	-8.4
37	24	-0.32	0.28	-1.1	-0.16	0.08	-2.1
38	23	0.28	0.15	1.9	-0.41	0.04	-10.1
39	30	0.02	0.28	0.1	-0.32	0.07	-4.8
40	12	0.33	0.30	1.1	-0.74	0.11	-6.8
41	9	-0.44	0.43	-1.0	-0.40	0.18	-2.2

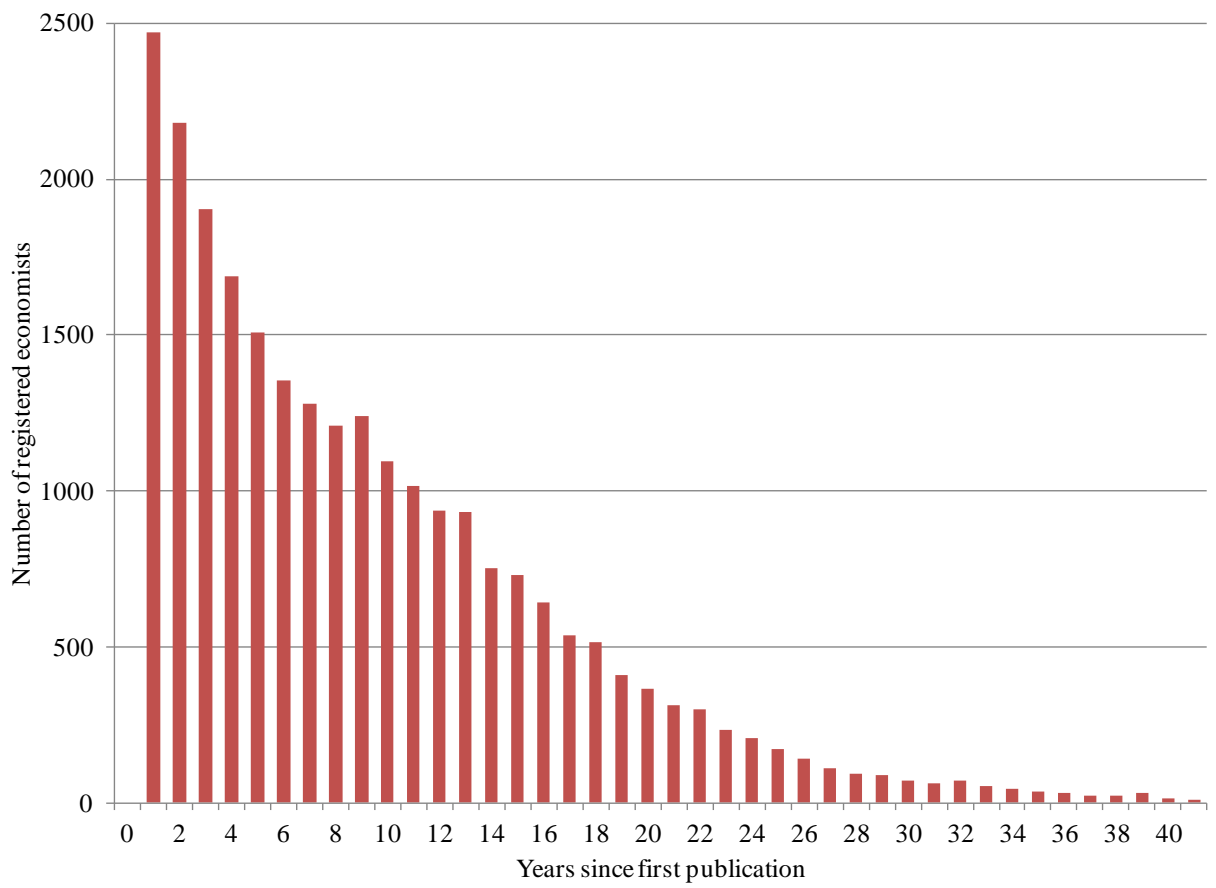


Figure 1. Number of registered economists by cohort.

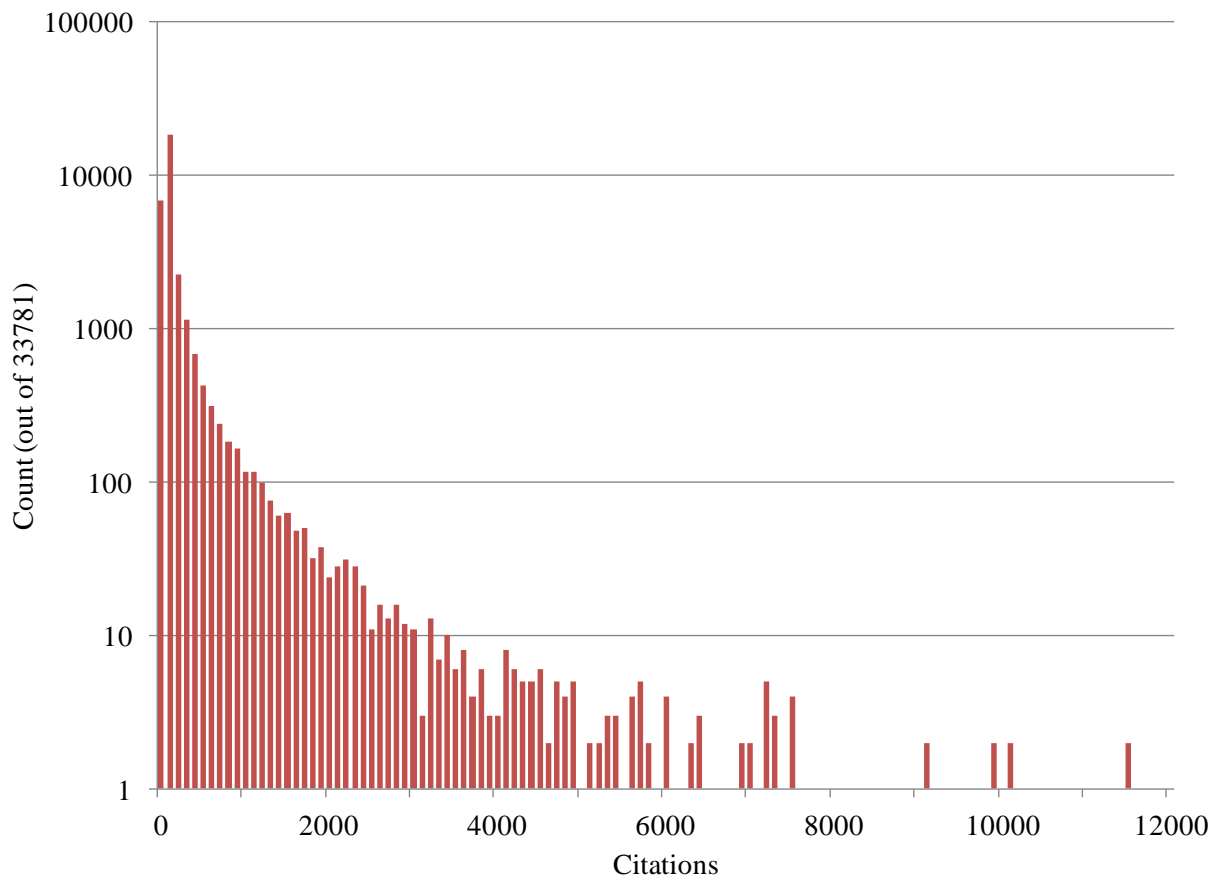


Figure 2. Histogram of the number of citations.

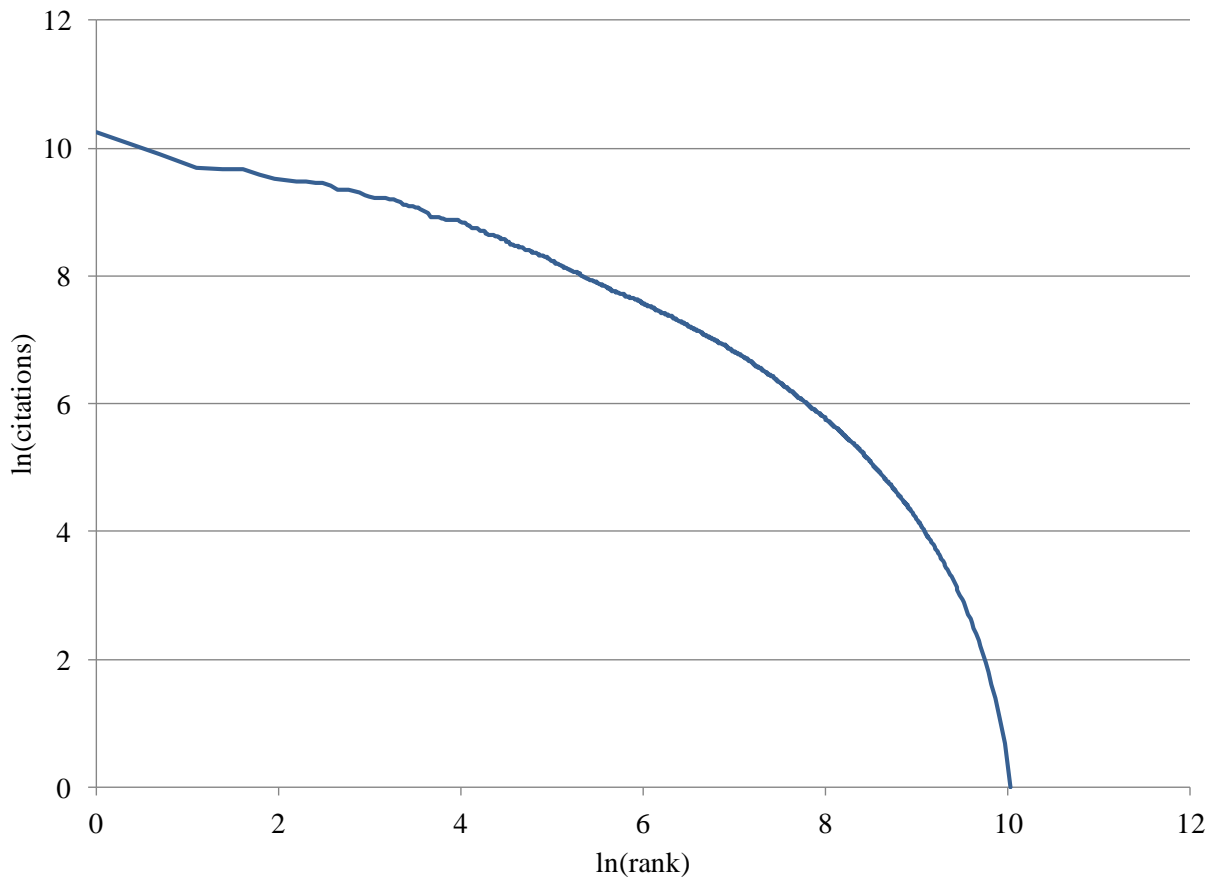


Figure 3. The natural logarithm of the number of citations as a function of the natural logarithm of its rank.

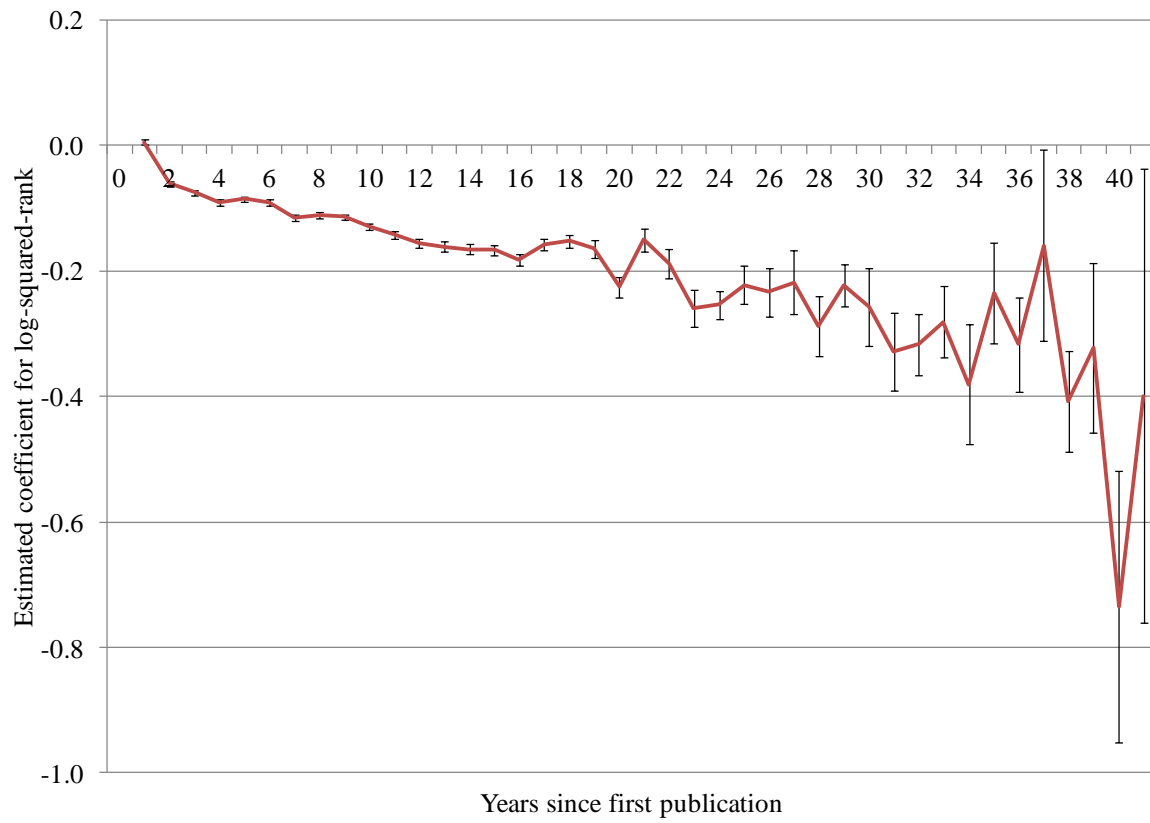


Figure 4. Cohort-specific estimates of the strength of the Matthew Effect.

## Appendix A: Matlab scripts for scraping data

```
%processperson
%this procedure scrapes data from citec.repec.org and stores the data
%
%This version: 7 November 2012 by Richard Tol

letter = ['a'; 'b'; 'c'; 'd'; 'e'; 'f'; 'g'; 'h'; 'i'; 'j'; 'k'; 'l';
'm'; 'n'; 'o'; 'p'; 'q'; 'r'; 's'; 't'; 'u'; 'v'; 'w'; 'x'; 'y'; 'z'];

address1 = 'http://citec.repec.org/p/';
address2 = '/p';
address3 = '.html';

for i=1:26,
    k=0;
    for j=1:26,
        for l=1:1000, %1000 people is enough for all names bar those
starting with ch, sa; ba goes to 1000 but not further
            number = num2str(l);

address=strcat(address1,letter(i),address2,letter(i),letter(j),number,add
ress3)

            [s, error] = urlread(address);
            if error == 1, %page exists
                k=k+1;
                vid1(k) = letter(i);
                vid2(k) = letter(j);
                vid3(k) = l;

                columnsep = strfind(s,'Articles');
                grab = s(columnsep(1)+14:columnsep(1)+18); %constants 14
and 18 by trial
                varticles(k) = convert(grab);

                columnsep = strfind(s,'Papers');
                grab = s(columnsep(2)+12:columnsep(2)+16);
                vpapers(k) = convert(grab);

                columnsep = strfind(s,'Books');
                grab = s(columnsep(1)+11:columnsep(1)+15);
                vbooks(k) = convert(grab);

                columnsep = strfind(s,'Chapters');
                grab = s(columnsep(1)+14:columnsep(1)+18);
                vchapters(k) = convert(grab);

                columnsep = strfind(s,'Co-Authors');
                grab = s(columnsep(1)+16:columnsep(1)+20);
                vcoauthors(k) = convert(grab);

                columnsep = strfind(s,'Total times cited');
                grab = s(columnsep(1)+19:columnsep(1)+24);
                vcitations(k) = convert(grab);

                columnsep = strfind(s,'Total self citations');
                grab = s(columnsep(1)+22:columnsep(1)+26);
                vselfcites(k) = convert(grab);

                columnsep = strfind(s,'years');
                grab = s(columnsep(3)-3:columnsep(3)-2);
                vyears(k) = convert2(grab);

                columnsep = strfind(s,'h-index');
                grab = s(columnsep(3)+9:columnsep(3)+13);
                vhirsch(k) = convert(grab);
```

```

        end
    end
    end
    file = strcat(letter(i),'.mat'); %store results by first letter to
    reduce chance of failure
    save(file);
    clear v*
end

%processperson
%this procedure scrapes data from citec.repec.org and stores the data
%
%RUN THIS MANUALLY FOR BA, CH, SA
%
%This version: 7 November 2012 by Richard Tol

letter = ['a'; 'b'; 'c'; 'd'; 'e'; 'f'; 'g'; 'h'; 'i'; 'j'; 'k'; 'l';
'm'; 'n'; 'o'; 'p'; 'q'; 'r'; 's'; 't'; 'u'; 'v'; 'w'; 'x'; 'y'; 'z'];

address1 = 'http://citec.repec.org/p/';
address2 = '/p';
address3 = '.html';

i= 19; %b=2 c=3 s=19
j= 1;  %a=1 h=8 a=1
k= 0;
for l=1001:2000,
    number = num2str(l);

address=strcat(address1,letter(i),address2,letter(i),letter(j),number,add
ress3)
    [s, error] = urlread(address);
    if error == 1,
        k=k+1;
        vid1(k) = letter(i);
        vid2(k) = letter(j);
        vid3(k) = 1;

        columnsep = strfind(s,'Articles');
        grab = s(columnsep(1)+14:columnsep(1)+18);
        varticles(k) = convert(grab);

        columnsep = strfind(s,'Papers');
        grab = s(columnsep(2)+12:columnsep(2)+16);
        vpapers(k) = convert(grab);

        columnsep = strfind(s,'Books');
        grab = s(columnsep(1)+11:columnsep(1)+15);
        vbooks(k) = convert(grab);

        columnsep = strfind(s,'Chapters');
        grab = s(columnsep(1)+14:columnsep(1)+18);
        vchapters(k) = convert(grab);

        columnsep = strfind(s,'Co-Authors');
        grab = s(columnsep(1)+16:columnsep(1)+20);
        vcoauthors(k) = convert(grab);

        columnsep = strfind(s,'Total times cited');
        grab = s(columnsep(1)+19:columnsep(1)+24);
        vcitations(k) = convert(grab);

        columnsep = strfind(s,'Total self citations');
        grab = s(columnsep(1)+22:columnsep(1)+26);
        vselfcites(k) = convert(grab);

        columnsep = strfind(s,'years');
        grab = s(columnsep(3)-3:columnsep(3)-2);
        vyears(k) = convert2(grab);
    end
end

```

```

        columnsep = strfind(s,'h-index');
        grab = s(columnsep(3)+9:columnsep(3)+13);
        vhirsch(k) = convert(grab);
    end
end
file = strcat(letter(i),'plus','.mat');
save(file);
clear v*

%postprocessperson
%this procedure reads the data from processperson and reorganizes
%
%This version: 7 November 2012 by Richard Tol

clear all

letter = ['a'; 'b'; 'c'; 'd'; 'e'; 'f'; 'g'; 'h'; 'i'; 'j'; 'k'; 'l';
'm'; 'n'; 'o'; 'p'; 'q'; 'r'; 's'; 't'; 'u'; 'v'; 'w'; 'x'; 'y'; 'z'];

file = strcat(letter(1),'.mat'); %initialize with a
load(file);
articles = varticles;
papers = vpapers;
books = vbooks;
chapters = vchapters;
coauthors = vcoauthors;
citations = vcitations;
selfcites = vselfcites;
years = vyears;
hirsch = vhirsch;
id1 = vid1;
id2 = vid2;
id3 = vid3;

clear v*

file = strcat('cplus.mat'); %names starting with ch
load(file);
articles = [articles varticles];
papers = [papers vpapers];
books = [books vbooks];
chapters = [chapters vchapters];
coauthors = [coauthors vcoauthors];
citations = [citations vcitations];
selfcites = [selfcites vselfcites];
years = [years vyears];
hirsch = [hirsch vhirsch];
id1 = [id1 vid1];
id2 = [id2 vid2];
id3 = [id3 vid3];

clear v*

file = strcat('splus.mat'); %names starting with sa
load(file);
articles = [articles varticles];
papers = [papers vpapers];
books = [books vbooks];
chapters = [chapters vchapters];
coauthors = [coauthors vcoauthors];
citations = [citations vcitations];
selfcites = [selfcites vselfcites];
years = [years vyears];
hirsch = [hirsch vhirsch];
id1 = [id1 vid1];
id2 = [id2 vid2];
id3 = [id3 vid3];

```



```

clear v*

for i=2:26, %b to z
    file = strcat(letter(i),'.mat');
    load(file);
    articles = [articles varticles];
    papers = [papers vpapers];
    books = [books vbooks];
    chapters = [chapters vchapters];
    coauthors = [coauthors vcoauthors];
    citations = [citations vcitations];
    selfcites = [selfcites vselfcites];
    years = [years vyears];
    hirsch = [hirsch vhirsch];
    id1 = [id1 vid1];
    id2 = [id2 vid2];
    id3 = [id3 vid3];

    clear v*
end

articles = real(ReplaceNaN(articles,0)); %replace not-a-number by 0 and
remove imaginary numbers
papers = real(ReplaceNaN(papers,0));
books = real(ReplaceNaN(books,0));
chapters = real(ReplaceNaN(chapters,0));
coauthors = real(ReplaceNaN(coauthors,0));
citations = real(ReplaceNaN(citations,0));
selfcites = real(ReplaceNaN(selfcites,0));
years = real(ReplaceNaN(years,0));
hirsch = real(ReplaceNaN(hirsch,0));
production = articles + papers + books + chapters;
productivity = ReplaceNaN(production./years,0);
othercites = citations - selfcites;
impact = ReplaceNaN(othercites./production,0);

results = [articles' papers' books' chapters' coauthors' citations'
selfcites' othercites' years' hirsch' production' productivity' impact'];
label = ['art'; 'pap'; 'bks'; 'chp'; 'coa'; 'cit'; 'slf'; 'oth'; 'yrs';
'hir'; 'tot'; 'prd'; 'imp'];
save('results.mat', 'results', 'label', 'id1', 'id2', 'id3');

```

## Appendix B: Matlab scripts to analyze data

```
%analyzeperson2
%this procedure estimates the Matthew Effect by cohort
%
%This version: 14 November 2012 by Richard Tol

clear all
load results

other = results(:,8);
filter = (other > 0);
pother = other(filter); %remove zeros
years = results(:,9);
pyears = years(filter);
maxyr = max(pyears);
for i=1:maxyr,
    filter = (pyears == i); %restrict to cohort
    foth = pother(filter);
    if length(foth) > 3,
        foth = sort(foth,'descend');
        X = ones(length(foth),2);
        for j= 2:length(foth)
            X(j,2) = X(j-1,2)+1;
        end
        foth = log(foth);
        X(:,2) = log(X(:,2)); %standard log-rank regression

        vb = inv(X'*X)*X'*foth; %coefficient
        vres = foth - X*vb;
        vsig = vres'*vres/(length(vres)-2);
        vvar = vsig*inv(X'*X); %var-covar

        X(:,3)=X(:,2).*X(:,2); %rank-squared added

        vb2 = inv(X'*X)*X'*foth;
        vres = foth - X*vb2;
        vsig = vres'*vres/(length(vres)-2);
        vvar2 = vsig*inv(X'*X);

        beta(i,:) = [i; length(vres); vb(2); sqrt(vvar(2,2)); vb2(2);
sqrt(vvar2(2,2)); vb2(3); sqrt(vvar2(3,3))];
    end
end
```