



Working Papers

www.cesifo.org/wp

The Matthew Effect in Economics Reconsidered

Daniel Birkmaier
Klaus Wohlrabe

CESIFO WORKING PAPER NO. 4966
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
SEPTEMBER 2014

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

The Matthew Effect in Economics Reconsidered

Abstract

We apply the test of Ijiri and Simon (1974) to a large data set of authors in economics. This test has been used by Tol (2009, 2013a) to identify a (within-author) Matthew effect for authors based on citations. We show that the test is quite sensitive to its underlying assumptions and identifies too often a potential Matthew effect. We propose an alternative test based on the pure form of Gibrat's law. It states that stochastic proportionate citation growth, i.e. independent of its size, leads to a lognormal distribution. By using a one-sided Kolmogorov-Smirnov test we test for deviations from the lognormal distribution which we interpret as an indication of the Matthew effect. Using our large data set we also explore potential empirical characteristics of economists with a Matthew effect.

JEL-Code: A120, A140.

Keywords: Matthew Effect, Gibrat's Law, Kolmogorov-Smirnov.

Daniel Birkmaier
University of Munich
Ludwigstrasse 28
Germany – 80539 Munich
d.birkmaier@gmx.de

*Klaus Wohlrabe**
Ifo Institute – Leibniz Institute for
Economic Research
at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
wohlrabe@ifo.de

*corresponding author

We thank Jens Ruhose, Marc Piopiunik and Christian Seiler for intensive discussions. The usual disclaimer applies.

1 Introduction

The *Matthew Effect* (Merton (1968)) has recently gained more and more attention in bibliometrics.¹ It states that papers are not only cited due to their quality but due to the fame of authors. Tol (2009) suggested using an approach by Ijiri and Simon (1974) to test this effect empirically. He showed that for 99 out of 100 economists a within-paper Matthew effect was present. Tol (2013a) applied the test to different cohorts of economists using citation data from the RePEc website. Again, he found that the effect is statistically significant for all cohorts, being larger for older economists. In spirit of Tol (2009) we apply the test to a much larger data set of more than 10,000 economists using citations from RePEc. It turns out that the test is quite sensitive and detects a potential Matthew effect in more than 90% of the cases. Although the idea behind the test by Tol is intuitive and convincing, this might be an unrealistic scenario. Based on this result we discuss the robustness of the test and its underlying assumptions. The pure form of Gibrat's law, on which the test relies, leads to a lognormal distribution instead of a Pareto distribution as stated in Tol (2009). As an alternative test we propose to employ a Kolmogorov-Smirnov test to check whether the individual citation pattern of authors follows a lognormal distribution. We interpret deviations from the lognormal distribution as an indication of the Matthew effect. Our test detects considerably less authors with a Matthew effect. Furthermore, we show that such authors are not concentrated at the top, i.e. authors with the most citations. The paper is organised as follows: We first describe our data set which uses citations from RePEc. Then we apply the test by Tol (2009) to our large data set and discuss its underlying assumptions and empirical regularities. After that, we outline some theoretical considerations and propose our new test which is followed by an exposition the results of the alternative test. Finally, we investigate in a regression analysis if we can explain why some authors have a Matthew effect and some don't.

¹For a detailed review of the literature see Tol (2009) and Wang (2014).

2 Data

As Tol (2013a), we use citation data from RePEc. In economics, RePEc (Research Papers in Economics, www.repec.org) has become an essential source for the spread of knowledge and ranking of individual authors and academic institutions. RePEc is based on the 'active participation principle', i.e. that authors, institutions and publishers have to register and to provide information to the network. This approach has the main advantage that a clear assignment of works and citations to authors and articles is possible.² Indeed, the RePEc story has become a success, with more than 36,000 registered authors with listed works and 12,000 institutions in economic sciences worldwide as of August 2013. RePEc has three main sources for extracting citations: First, it reads out all publicly available documents within the network. Due to missing (open) access to the article or technical problems it is not always possible to extract all citations. Second, archive maintainers may provide meta-information on citations for their journals. Third, registered authors can upload citations for every article. See Seiler and Wohlrabe (2012) for further details on RePEc.

Data were collected on 11 July 2013 using Citec.³ We adapted and extended the Matlab code provided by Tol (2013b). We read out the publicly available citation profiles. These profiles contain, besides the aggregated citations, also citation numbers for each listed bibliographic item (books, working papers and journal articles). Furthermore we have the number of aggregated self-citations.⁴ Based on the citations we calculate the h-index.⁵ We assessed 35,633 citation pages.⁶ For the upcoming analysis we only include authors who have more than ten papers with nonzero citations.⁷ Furthermore, the sum of all citations should be at least 20.⁸ This leaves us with 10,564 authors. In Table 1 we report some descriptive statistics

²For instance, Google Scholar as a source for citation analysis potentially suffers from the problem of clear identification of citations which can lead to overestimation of citations, see Harzing and van der Wal (2009).

³See <http://citec.repec.org/index.html> for further details.

⁴Unfortunately, we were not able to extract the self-citations for each individual item.

⁵In Citec, citations are only consolidated across paper versions, not across document types. This may result in a bigger h-index than in the RePEc rankings.

⁶The full data set, including the individual citation profiles and summary statistics for each author, are available upon request from the authors.

⁷We assume that a within-paper Matthew effect for authors with only a few cited papers does not make sense. However, this does not rule out the possibility of a Matthew effect for one or two single papers with many citations.

⁸We want to exclude cases where 10 papers have one citation each.

for our sample of authors. We report all indicators which we were able to extract automatically from each individual citation page. The average number of works is about 20 and ranges between 11 and 1,063. “Number of works” includes journal articles as well as books, chapters and working papers. As it is well known from the bibliometric literature, the distribution of citations is highly skewed across authors as the mean (517) and the median (197) differ substantially. “Years in the profession” denotes the time from the first publication until now. It shows that in our sample at least two years have proceeded since the first publication and that the oldest author has published his first article 45 years ago. The Gini coefficient measures the citation concentration across papers for each author. If the coefficient equals 0 then each cited paper has the same citation count. In case of a Gini coefficient of 1 all citations are concentrated on one paper.

Table 1: Descriptive Statistics

	Full Sample ($n = 10564$)					KS Sample ($n = 633$)				
	mean	median	SD	min	max	mean	median	SD	min	max
Number of Works	50	36	42	11	1063	168	152	72	91	1063
Works/ Year	3.2	2.7	2.0	0.4	35.5	6.8	6.1	2.9	2.5	29.5
Cited papers	32	23	28	11	494	118	105	43	81	494
Fraction non-cited	0.33	0.32	0.15	0.00	0.91	0.27	0.26	0.12	0.05	0.67
Citations	517	198	1162	21	32143	2952	1828	3163	195	32143
Citations/Year	26.1	14.1	42.5	1.5	1148.0	110.0	76.7	105.5	5.7	1148.0
Citations/Works	9.1	5.5	11.9	0.1	230.4	18.9	12.4	18.6	0.8	147.1
Self-Citation Rate	0.11	0.07	0.12	0.00	1.00	0.10	0.07	0.11	0.00	0.89
h-Index	9	7	6	2	83	24	22	11	6	83
Years in the Profession	16	15	7	2	45	26	25	7	8	45
Gini coefficient	0.56	0.56	0.11	0.14	0.94	0.64	0.64	0.08	0.36	0.85

This table reports descriptive statistics for various bibliometric indicators. *Number of Works* includes working papers, journal articles, chapters in books and monographs. The *Full Sample* refers to the sample employed in the next section. *KS Sample* represents the data used for the Kolmogorov-Smirnov test.

3 The test by Tol (2009) reconsidered

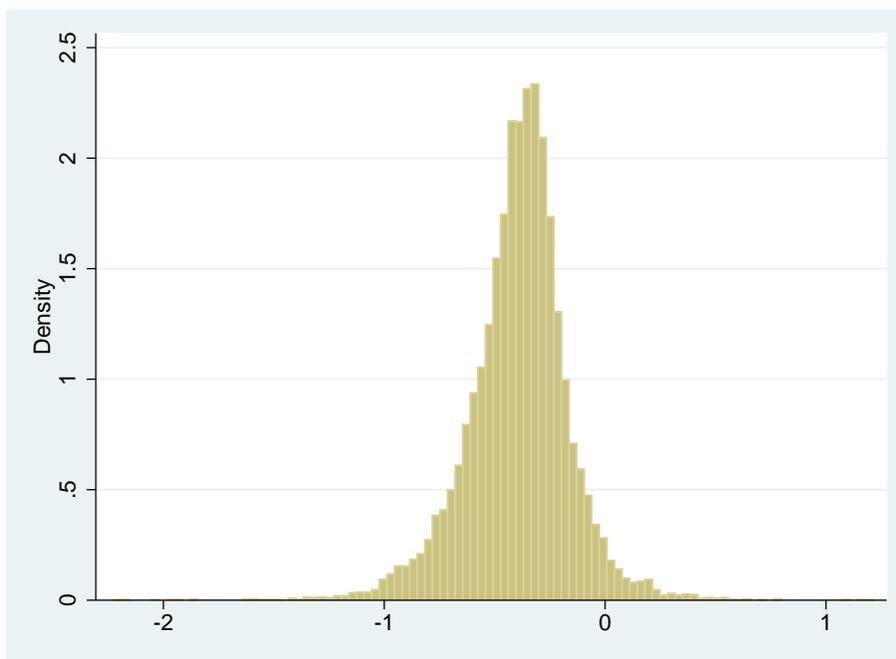
3.1 Empirical results for authors

The test of Tol (2009) for author i (within-paper) is given by

$$\ln Citations_i = \alpha_i + \beta_i \ln Rank_i + \gamma \ln^2 Rank_i. \quad (1)$$

A (potential) Matthew effect is present in case of $\gamma < 0$. We run this regression for a sample of 10,564 economists. It is important to note that both for this test and for our new proposed test only works with at least one citation are considered. In Figure 1 we plot the histogram of γ for all economists. For a large majority we have $\gamma < 0$. The mean is -0.393. In 9,755 cases (92.4%) it is also significant at the 10% level. What are reasons for so many economists to have a Matthew effect based on this intuitive test? In the following, we offer two explanations for these results, an empirical and a theoretical one.

Figure 1: Histogram of γ for 10,564 authors



3.2 The empirical validity of the test

In order to understand the behavior of the test given in equation (1) we conduct a small Monte Carlo study. Consider the probability distribution function of the Pareto distribution

$$f(x) = \alpha x^\beta \tag{2}$$

where α and β are constants, and β is called the exponent and is typically negative (i.e., $\beta < 0$). We draw N random numbers (citations) from (2), where N denotes the number of papers with citations. We start with 50 papers and increase this number in intervals of 50 up to 500. The exponential parameter β ranges from -0.2 to -2 with intervals of 0.2. For each combination of N and β we run the regression (1) and test whether γ is statistically different from zero. We conduct 1000 replications each. In Table 2 we show the results. We report both the percentage number of how many times the quadratic term was significant at the 10% level and the relative shares in case of significant negative gammas. The results are clear cut. Independently both of the shape parameter of the Pareto distribution (β) and the number of observations (N), in at least 80% of the cases the parameter γ is significantly different from zero. Looking at negative γ 's, which indicate a Matthew effect in equation (1), we find significant values in about one third of all Monte Carlo replications. These results point to the fact that the test by Tol seems to be quite sensitive with respect to a detection of a Matthew effect.

3.3 Some theoretical considerations

The approach suggested by Tol (2009, 2013a) relies on the assumption that, in the absence of anything like a 'fame' effect, the citation pattern of authors/papers has to be scale-free, meaning that citations of all authors/papers grow at the same rate, regardless of their rank (w.r.t. their relative citation count). The relationship is represented in Tol (2009, 2013a) by a Pareto distribution. In natural logarithm exposition

$$\ln(S_i) = \alpha + \beta \ln Rank_i \tag{3}$$

Table 2: Monte Carlo Results concerning the test by Tol (2009)

		β from equation (2)									
N		-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0
50	$\gamma < 0$ and sign.	0.34	0.33	0.29	0.34	0.34	0.33	0.32	0.33	0.35	0.34
	total sign.	0.82	0.87	0.88	0.89	0.91	0.93	0.91	0.92	0.94	0.93
100	$\gamma < 0$ and sign.	0.31	0.32	0.35	0.33	0.32	0.31	0.34	0.33	0.34	0.34
	total sign.	0.81	0.88	0.91	0.91	0.92	0.92	0.93	0.92	0.94	0.94
150	$\gamma < 0$ and sign.	0.27	0.30	0.32	0.34	0.34	0.35	0.34	0.36	0.33	0.30
	total sign.	0.82	0.87	0.88	0.89	0.91	0.92	0.93	0.92	0.93	0.92
200	$\gamma < 0$ and sign.	0.31	0.33	0.32	0.30	0.33	0.33	0.35	0.33	0.32	0.33
	total sign.	0.81	0.85	0.89	0.90	0.92	0.92	0.93	0.93	0.92	0.93
250	$\gamma < 0$ and sign.	0.31	0.32	0.31	0.35	0.33	0.34	0.31	0.35	0.31	0.33
	total sign.	0.80	0.87	0.88	0.90	0.90	0.91	0.92	0.92	0.95	0.94
300	$\gamma < 0$ and sign.	0.32	0.35	0.30	0.33	0.33	0.35	0.32	0.36	0.35	0.32
	total sign.	0.80	0.87	0.88	0.89	0.92	0.93	0.92	0.94	0.94	0.93
350	$\gamma < 0$ and sign.	0.32	0.33	0.33	0.34	0.35	0.32	0.35	0.33	0.33	0.32
	total sign.	0.84	0.87	0.89	0.90	0.92	0.91	0.91	0.94	0.94	0.93
400	$\gamma < 0$ and sign.	0.30	0.32	0.32	0.33	0.32	0.33	0.34	0.36	0.32	0.33
	total sign.	0.80	0.86	0.90	0.90	0.91	0.93	0.93	0.93	0.92	0.93
450	$\gamma < 0$ and sign.	0.31	0.31	0.32	0.34	0.31	0.35	0.32	0.33	0.34	0.34
	total sign.	0.80	0.87	0.85	0.91	0.92	0.91	0.93	0.94	0.93	0.92
500	$\gamma < 0$ and sign.	0.31	0.34	0.34	0.32	0.34	0.33	0.31	0.31	0.33	0.32
	total sign.	0.81	0.87	0.88	0.89	0.91	0.93	0.91	0.91	0.95	0.93

This table reports the results of the Monte Carlo study described in subsection 3.2. N denotes the number of papers and β the Pareto exponent from equation (2). It shows relative shares of significant values of γ at the 10% level both for all 1000 replications (*total sign.*) and γ being negative (*$\gamma < 0$ and sign.*).

where S_i is a bibliometric measurement, like citations or the h-index, and R_i is its corresponding rank. In case of $\beta = -1$ this is also referred to as Zipf's Law. Tol (2009, 2013a) suggests that the proportionality of rank and size would be violated if larger objects grew faster than smaller ones, a property which could be captured by a quadratic term, as suggested by Ijiri and Simon (1974) for firm sizes. This leads to equation (1). In this estimation framework, if $\gamma < 0 (> 0)$, large objects would grow systemically faster (slower) than small ones, while for $\gamma = 0$, the growth rate would not depend on their relative size.

Zipf's law has been extensively investigated in the regional and urban science literature (see Gabaix and Ioannides (2004)). When studying the size distribution of cities, for example, a relationship like (3) with a coefficient of β equal to -1 seems to be a robust finding, at least for the largest cities, even if there appears to be some variation across regions. It implies a Pareto distribution in the upper part of the distribution. Another empirical regularity for city sizes is that of scale-independent proportionate growth - a stochastic process described by Gibrat's Law. But, as Eeckhout (2004) indicates, the pure form of Gibrat's Law implies a lognormal distribution which makes both findings hard to reconcile. According to Eeckhout (2004), the reason why Zipf's Law seems to hold for larger cities might be due to the fact that the Pareto and the lognormal distribution have similar upper-tail properties, but they can differ quite a lot in the lower tail. More precisely, Gabaix (1999) shows that, under certain conditions, the stochastic growth process will converge to a Pareto distribution with exponent equal to -1: Zipf's law. However, it is unknown how these conditions could look like for the topic considered here. Therefore, we draw two conclusions. First, the Matthew effect builds upon the theoretical assumption of proportionate growth independently of its rank. Therefore, one should look rather for deviations from the lognormal instead of the Pareto distribution. Second, as shown in the previous subsection, the γ coefficient has a tendency to become significant too often as one might expect in our context. In the log-log notation, a lognormal distribution would not necessarily imply linearity (see Eeckhout (2004)), but possibly lead to a curved line. Thus, significant γ 's could be obtained without violating the proportionate growth assumption as found by the test by Tol and our small Monte Carlo study.

4 An alternative test

Based on the results from the previous section we propose an alternative test which relies on the idea whether the citation distribution of an author, a cohort or a journal follows a lognormal distribution. The following two steps are necessary:

1. Fit a lognormal distribution to the empirical data at hand in order to obtain the corresponding expectation value and the variance.⁹
2. Run a Kolmogorov-Smirnov (KS) test whether the empirical distribution follows a lognormal distribution with the parameters obtained from the previous step. The null hypothesis is that the empirical distribution and the theoretically predicted distribution are identical.¹⁰

In order to test for a Matthew effect we conduct a one-sided test, i.e. whether the empirical distribution is larger than the theoretical one. We do this by assuming that the Matthew effect generates *more* citations than theoretical citation distribution without a Matthew effect. Before we apply our test we have to ask what the power of the KS test is, i.e. how many observations are necessary. Therefore, we conduct again a Monte Carlo study. As the Pareto distribution is the natural competitor to the lognormal one, we test how well the KS test can discriminate between these two distributions. We use the same framework as described in subsection 3.2 by drawing Pareto random numbers for different distribution parameters (β). We start with 10 observations and increase them by ten up to 500. For each of the 1000 replications and (N, β) -combinations we test the null hypotheses whether the N observations follow a lognormal distribution. Table 3 reports the power of the KS test. It shows that independently of the shape parameter β , about 80 observations are necessary for a reliable test result.

Given the previous results we included only authors which have at least 80 papers with at least one citation. This leaves us with 633 economists. In Table 1 we report the corresponding descriptive statistics (*KS sample*). Compared to the full sample it is obvious that we have only

⁹We employ the Matlab routine LOGNFIT from the Statistics toolbox.

¹⁰The corresponding Matlab routine is KSTEST from the Statistics toolbox.

Table 3: Power of the KS test against the Pareto distribution

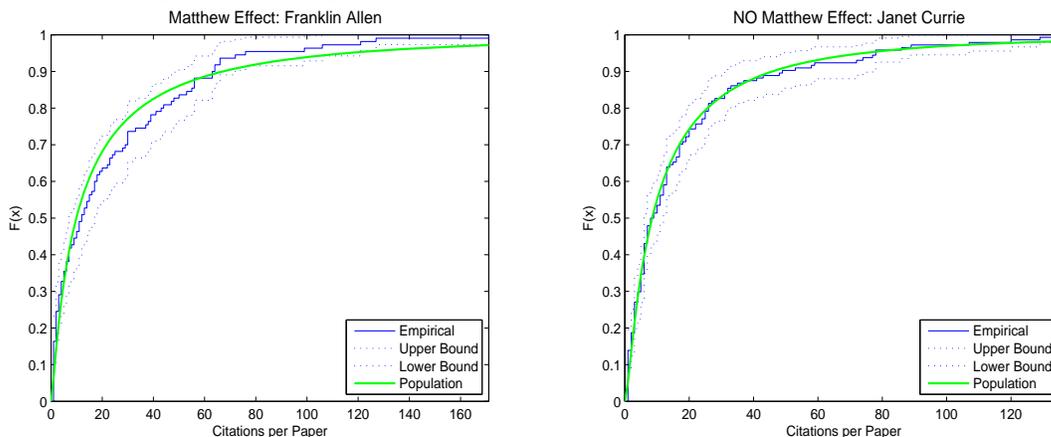
N	β from equation (2)									
	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0
10	3.4	1.2	2.6	2.0	3.2	1.5	1.8	1.8	2.1	2.1
20	11.9	12.4	9.6	10.8	13.1	11.1	10.3	10.6	13.6	12.6
30	26.8	27.5	26.8	24.6	22.2	24.2	27.7	25.1	25.2	27.7
40	40.3	42.4	45.2	42.9	42.8	42.7	40.0	43.4	41.8	43.7
50	59.6	58.2	62.7	59.6	57.7	62.3	64.3	59.0	60.7	58.7
60	75.3	73.2	75.3	73.9	78.3	73.1	76.7	73.6	74.6	73.5
70	88.0	86.0	81.9	84.1	85.2	85.8	86.0	84.0	85.1	86.4
80	92.3	93.1	91.7	92.5	91.7	92.0	92.4	91.9	93.0	91.5
90	95.9	96.1	96.9	96.2	96.2	95.3	96.3	95.8	95.5	95.5
100	98.4	98.5	98.3	97.1	98.3	98.8	97.7	98.4	98.9	98.1
110	99.7	99.4	99.6	99.5	99.2	99.6	99.4	99.3	99.4	99.9
120	100.0	99.8	99.8	100.0	99.8	99.8	99.7	99.8	99.7	99.8
130	99.9	100.0	99.9	99.9	99.8	99.8	99.9	100.0	100.0	99.8
140	100.0	100.0	100.0	99.9	100.0	99.9	100.0	100.0	100.0	100.0

This table reports the relative share of rejection of the null hypothesis that the N observations (random draws from a Pareto distribution) follow a lognormal distribution. The results are based on 1,000 replications and a significance level of 10%.

top authors included in our sample. Furthermore, each author is in the profession for at least eight years. We assume that this is sufficient to converge to a log-normal distribution as stated above. Before we turn to the aggregated results we illustrate the test with an example from our data set. In Figure 2 we plot both the empirical and theoretical (“Population”) cumulative distribution function (CDF) both for an economist with a Matthew effect (left graph) and no Matthew effect (right graph). The dashed lines denote the 95% confidence intervals of the empirical CDF. The population CDF is obtained by fitting a lognormal distribution and estimating the corresponding mean and variance. Given these two parameters the CDF is simulated. For Franklin Allen (left panel) it is obvious that beyond 60 citations the empirical CDF is clearly located above the theoretical CDF. In contrast, for Janet Currie (right panel) the empirical and theoretical CDF lie close together. The KS test cannot reject the null that the citation data is lognormally distributed, i.e. there is no Matthew effect.

The test by Tol suggests a Matthew effect for the whole sample of 633 economists, i.e. $\gamma < 0$ and being statistically significant. In contrast, our KS test indicates a Matthew effect for only 276 economists (44%). Given our theoretical discussion above concerning the underlying

Figure 2: Examples of cumulative distribution functions for citations



Notes: This figure plots both the empirical and theoretical (“Population”) cumulative distribution function (CDF) both for an economist with a Matthew effect (left graph) and no Matthew effect (right graph). The dashed lines denote the 95% confidence intervals of the empirical CDF. The population CDF is obtained by fitting a lognormal distribution and estimating the corresponding mean and variance. Given these two parameters the CDF is simulated.

distribution, we compared the data both to a fitted lognormal and a Pareto distribution. In 630 out of 633 cases the Pareto distribution has a larger mean-squared error compared to the lognormal one, i.e. the latter one fits the empirical data better.

In Table 6 in the Appendix we report the top 100 economists according to the number of overall citations. We display some individual bibliometric statistics for each economist. Furthermore, we report the estimated value of γ from the test by Tol together with its p-value. As all p-values are almost zero, the test suggests that all of these reported economists profit from a Matthew effect as stated above. In the last column, we report the one-sided p-value of the KS test. A value smaller than 0.1 points to a Matthew effect which we indicate by a \checkmark . Using our test, we find only for 13 economists in our top 100 a Matthew effect. The large citation counts for the other authors are solely due to the quality of their papers. In the top 10, only Joseph Stiglitz benefits from his fame and gathers additional citations. Otherwise, for the best economists according to the citation count, like Andrei Shleifer, and for Nobel laureates like James Heckman, Robert Lucas or Paul Krugman we find no Matthew effect. The rather small number of top economists with a Matthew effect might be explained by the findings of Wang (2014) who finds only weak evidence of prestige effect.

5 Can we characterize authors with a Matthew effect?

In this section we ask the question, whether economists with a Matthew effect exhibit specific characteristics. In Table 4 we compare the means between the two groups, economists with ($N = 276$) and without ($N = 374$) a Matthew effect (at the 10% level) for certain bibliometric indicators. For reasons of comparability, we also report the means of the full sample from Table 1 for the KS test. We use a two-sided t-test with the null hypotheses that the means of the two groups are equal. This is only true for the years in the profession. For all other bibliometric indicators the means are statistically different from each other. It stands out, that economists with no Matthew effect have on average better indices of bibliometric success. This mean, they have written more works, gathered more citations, have a higher h-index and a lower self-citation rate. This might be surprising, as one would expect it the other way round.¹¹

Table 4: Mean comparison across economists with and without a Matthew effect

	Full Sample	Matthew Effect		p-value two-sided t-test
		No	Yes	
N	633	357	276	
Number of Works	168	200	143	0.00
Papers/ Year	1.4	1.3	1.6	0.00
Cited papers	118	112	127	0.00
Fraction non-cited	0.27	0.21	0.35	0.00
Citations	2952	3885	1746	0.00
Citations/Year	110.0	145.4	66.0	0.00
Citations/Paper	18.9	26.8	8.0	0.00
Self-Citation Rate	0.10	0.07	0.15	0.00
h-Index	26	28	19	0.00
Years in the Profession	26	26	26	0.35
Gini coefficient	0.64	0.66	0.63	0.00

This table reports mean comparison across economists with and without a Matthew effect for different bibliometric indicators. The last column reports the p-value of a two-sided t-test with the null hypothesis that the means of the two groups are equal.

We want to investigate this issue in a regression. As the cutoff point for the p-values at which we define a potential Matthew effect is arbitrary, we decided to use the one-sided

¹¹The conclusions remain qualitatively the same if we define the Matthew effect at the 5% instead of the 10% level.

p-values obtained from our KS-test as the dependent variable. This allows us, on the one hand, to account for the continuous nature of the p-value and, on the other hand, to prevent the use of a dummy variable that discards potentially valuable information. As the p-value is bounded between zero and one, we employ a Tobit regression which accounts for bounded dependent variables. As independent variables we use unique bibliometric indicators, which are not derived from other indicators such as cites per year etc. Furthermore, they should be related to our KS test, i.e. we employ the number of cited papers and not the overall number of works. Therefore we include in our regression: *cited papers*, *citations*, *years in the profession*, the *Gini coefficient*, and the *self-citation rate*.¹² A priori, we expect for all variables, except the last one, a negative sign. For example, a higher citation count might increase the probability that the author-specific citation distribution deviates from a lognormal distribution, which results in a Matthew effect. The expected sign for the self-citation rate is ambiguous. Your degree of popularity in the science community might be raised both with self- and “external” citations. However, we believe that the latter is more plausible.

In the first column of Table 5 we report the regression coefficients. All variables are significant and negative. There is one exception for the citation count. We obtain a positive and significant coefficient which stands in contrast to prior expectations but is in line with Table 4, where, on average, economists with a Matthew effect have fewer citations. In a second step we want to control for potential nonlinear effects. Therefore, we test for each variable whether the corresponding squared term is significant. This turns out to be the case for all variables but the years in the profession. Taking the non-linear effect into account, we calculate the marginal effects evaluated at the mean, i.e. all variables are held constant at their corresponding mean. In order to get an idea about the direction of the influence for different intervals we plot the marginal effects at the means for the four variables with significant squared effects in Figure 3. The upper left graph shows that only very large citation counts (>16,000) increase the probability of a Matthew effect. This result might be driven by the economist Joseph Stiglitz, which is the only one who has an Matthew effect beyond this

¹²Other explanatory variables might be useful in this context, like number of coauthors, affiliations both of the authors and the coauthors, or the type of publication. These information were not available for automatic extraction from the citation websites.

threshold. Furthermore, only five economists have a citation count larger than 16,000. This implies that, in general, a larger citation count lowers your probability of a Matthew effect. A possible conclusion might be that the Matthew effect is a temporary phenomenon. This is confirmed by the scatter plot of the citation count against the p-value of the KS test in Figure 4. The larger the citation count the higher the probability that no Matthew effect occurs. The temporary phenomenon is not necessary in terms of time but in citation counts. Concerning the number of cited works it shows that beyond some threshold (>300) there is no positive effect.¹³ This means that, in general, every cited paper increases your likelihood to benefit from a Matthew effect. The results for the Gini coefficient are intuitive. The more unequal your citation distribution the higher the probability that this does not follow a lognormal distribution. The results for the self-citation rate show that the higher the self-citation rate the lower the probability of a Matthew effect. To put it differently, more citations from other authors increase your degree of popularity in the science community. In order to summarize, an economist has the highest probability of a Matthew effect if he has many cited papers, a low self-citation rate, has been long in the profession and the citation distribution across its own papers should be concentrated on a few papers.

Table 5: Results from the Tobit regression

	(1)	(2)	Marginal Effects
Cited papers	-0.001495***	-0.003735***	-0.002485***
Cited papers squared		0.000005***	
Citations	0.000028***	0.000070***	0.000058***
Citations squared		-0.000000***	
Self-Citation Rate	-0.391365***	-0.395177***	-0.304135***
Self-citation rate squared		0.437747**	
Years in the Profession	-0.013553*	-0.001381*	-0.001381*
Gini coefficient	-0.255752***	3.98185***	-0.476757***
Gini coefficient squared		-3.46237***	
Oberservations	633	633	

This table reports coefficients and marginal effects (evaluated at the mean) for a Tobit regression. Marginal effects refer to column (2) taking the non-linear squared terms into account. Dependent variable: p-value from the one-sided KS test, see Section 4. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

¹³But this is rarely the case in our sample: only 6 economists have more than 300 cited works.

Figure 3: Marginal effects of the Tobit regression

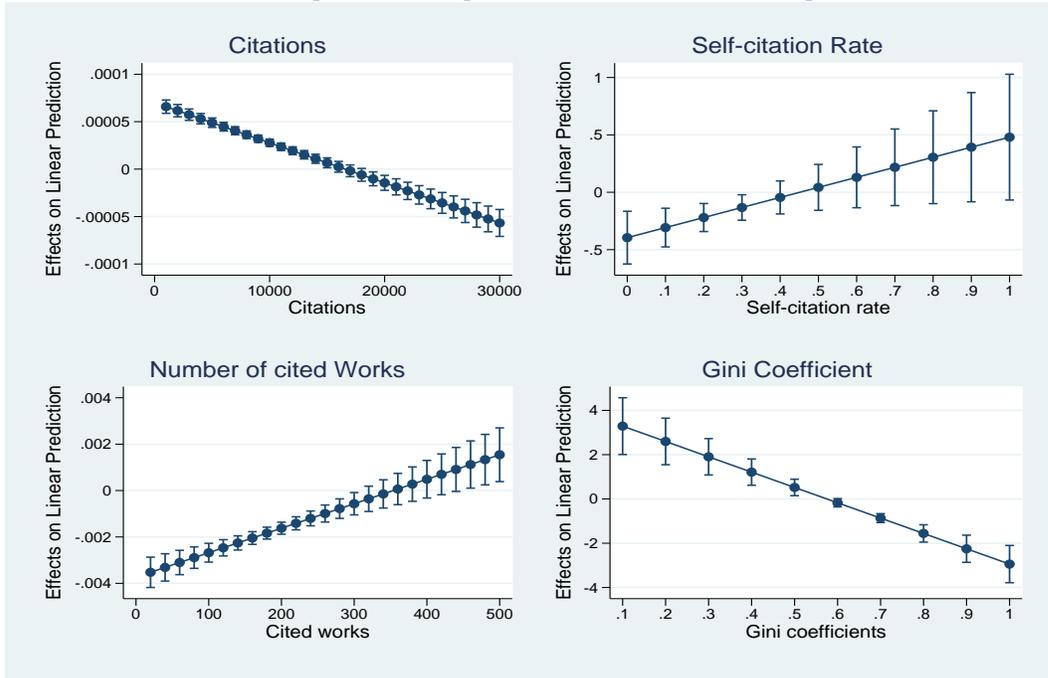
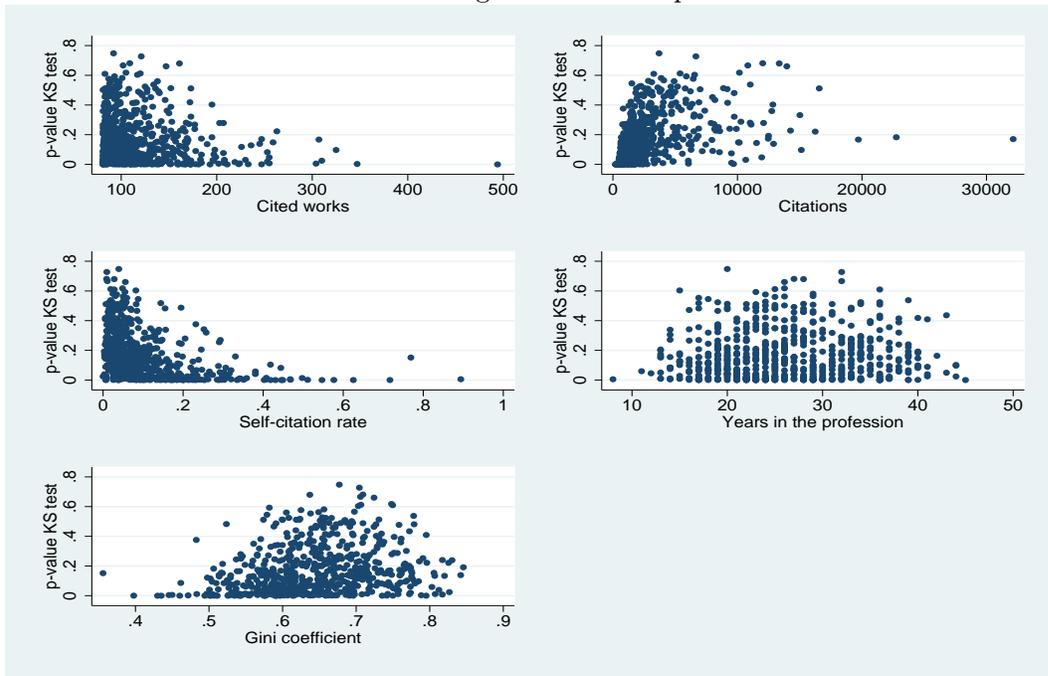


Figure 4: Scatter plots



Notes: This figure show scatter plots of the p-value of the one-sided KS test against the independent variables from the Tobit regression.

6 Conclusion

We show that the approach by Tol (2009, 2013a) to test for the Matthew effect, at least in economics, might lead to a too large number of authors with a Matthew effect. This might be due to both the sensitivity of the estimation approach and the assumption that the underlying distribution is a Pareto one. We propose an alternative test based on the idea that Gibrat’s law in its pure form states that the citation distribution of an individual author follows a lognormal distribution. Therefore, one could test for deviations from the lognormal distribution using a Kolmogorov-Smirnov test. We demonstrate that our approach leads to considerably fewer economists with a Matthew effect. In a regression analysis we show that with an increasing citation count the probability of a Matthew effects shrinks. Our interpretation of this finding is that the Matthew effect might be a temporary phenomenon, not in terms of time but in terms of the citation count.

References

- ECKHOUT, J. (2004): “Gibrat’s law for (all) cities,” *American Economic Review*, 94(5), 1429–1451.
- GABAIX, X. (1999): “Zipf’s law for cities: an explanation,” *The Quarterly Journal of Economics*, 114(3), 739–767.
- GABAIX, X., AND Y. M. IOANNIDES (2004): *Handbook of regional and urban economics* vol. 4, chap. The evolution of city size distributions, pp. 2341–2378. Elsevier.
- HARZING, A., AND R. VAN DER WAL (2009): “A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business,” *Journal of the American Society for Information Science and Technology*, 60(1), 41–46.
- IJIRI, Y., AND H. A. SIMON (1974): “Interpretations of departures from the Pareto curve firm-size distributions,” *The Journal of Political Economy*, 82(2), 315–331.
- MERTON, R. K. (1968): “The Matthew effect in science,” *Science*, 159(3810), 56–63.

- 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), 389–402.
- 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.
- (2013a): “The Matthew effect for cohorts of economists,” *Journal of Informetrics*, 7(2), 522–527.
- (2013b): “The Matthew Effect for Cohorts of Economists,” Working Paper Series 5513, Department of Economics, University of Sussex.
- WANG, J. (2014): “Unpacking the Matthew effect in citations,” *Journal of Informetrics*, 8(2), 329–339.

Appendix

Table 6: Top 100 Economists

Rank	Name	Citations	Works	C/W	Fraction		Years	h-index	Gini	Test Tol (2009)		KS-Test	
					non-cited	non-cited				γ	p-value	p-value	Matthew?
1	Andrei Shleifer	32143	278	116	0.11	0.11	28	83	0.75	-0.46	0.00	0.17	
2	Robert J. Barro	22757	225	101	0.13	0.13	40	66	0.77	-0.49	0.00	0.18	
3	James J. Heckman	19707	351	57	0.13	0.13	39	71	0.71	-0.35	0.00	0.17	
4	John Y. Campbell	16570	188	89	0.08	0.08	31	67	0.62	-0.48	0.00	0.51	
5	Kenneth S Rogoff	16247	210	78	0.20	0.20	33	60	0.73	-0.53	0.00	0.22	
6	Joseph Stiglitz	15128	438	35	0.26	0.26	44	53	0.78	-0.25	0.00	0.10	✓
7	Mark Gertler	15013	113	147	0.17	0.17	27	49	0.73	-0.58	0.00	0.33	
8	N. Gregory Mankiw	14259	166	86	0.08	0.08	31	55	0.73	-0.41	0.00	0.23	
9	Ross Levine	13964	164	90	0.10	0.10	26	54	0.72	-0.41	0.00	0.66	
10	Alberto Alesina	13354	176	76	0.09	0.09	28	59	0.64	-0.37	0.00	0.68	
11	Jean Tirole	12900	282	46	0.13	0.13	33	64	0.69	-0.43	0.00	0.14	
12	Carmen Reinhart	12843	251	51	0.22	0.22	26	53	0.73	-0.37	0.00	0.40	
13	Jordi Gali	12756	117	111	0.11	0.11	23	45	0.76	-0.53	0.00	0.36	
14	Gary S. Becker	12496	136	92	0.22	0.22	40	46	0.70	-0.70	0.00	0.19	
15	Olivier Blanchard	12474	209	60	0.12	0.12	34	55	0.71	-0.43	0.00	0.17	
16	Richard Blundell	12098	252	48	0.18	0.18	30	52	0.75	-0.37	0.00	0.28	
17	James H. Stock	12030	117	103	0.07	0.07	27	47	0.71	-0.53	0.00	0.68	
18	Daron Acemoglu	11949	328	36	0.22	0.22	21	56	0.72	-0.40	0.00	0.05	✓
19	Robert F. Engle	11531	173	67	0.17	0.17	34	47	0.76	-0.34	0.00	0.14	
20	Maurice Obstfeld	11099	201	55	0.12	0.12	33	49	0.71	-0.36	0.00	0.27	
21	Edward C. Prescott	11029	159	69	0.16	0.16	39	42	0.78	-0.39	0.00	0.54	
22	Elhanan Helpman	11019	195	57	0.14	0.14	35	53	0.70	-0.46	0.00	0.28	
23	Robert King	10831	114	95	0.11	0.11	32	44	0.71	-0.50	0.00	0.67	
24	Lawrence H. Summers	10664	238	45	0.13	0.13	36	48	0.70	-0.41	0.00	0.03	✓
25	Paul Krugman	10279	183	56	0.16	0.16	34	43	0.76	-0.41	0.00	0.29	
26	Tim Bollerslev	10157	122	83	0.14	0.14	26	43	0.75	-0.39	0.00	0.62	
27	Edward Ludwig Glaeser	10148	192	53	0.12	0.12	23	48	0.72	-0.42	0.00	0.15	
28	M Hashem Pesaran	10079	323	31	0.20	0.20	34	45	0.76	-0.25	0.00	0.15	
29	Robert Hall	9914	183	54	0.32	0.32	38	34	0.83	-0.36	0.00	0.24	
30	Mark W. Watson	9742	137	71	0.19	0.19	29	50	0.65	-0.55	0.00	0.32	
31	Lawrence Christiano	9734	166	59	0.16	0.16	28	50	0.67	-0.40	0.00	0.48	
32	Peter C. B. Phillips	9688	513	19	0.32	0.32	34	40	0.79	-0.21	0.00	0.00	✓

Continued on next page.

Table 6 – cont. from previous page.

Rank	Name	Citations	Works	C/W	Fraction non-cited	Years	h-index	Gini	Test Tol (2009)		KS-Test	
									γ	p-value	p-value	Matthew?
33	Ben S. Bernanke	9528	374	29	0.66	32	43	0.77	-0.65	0.00	0.01	✓
34	Andrew J. Oswald	9517	189	50	0.14	32	43	0.76	-0.41	0.00	0.07	✓
35	Clive W. J. Granger	9381	185	51	0.22	39	37	0.82	-0.29	0.00	0.24	
36	Raghuram G. Rajan	9211	144	64	0.21	21	46	0.67	-0.47	0.00	0.41	
37	Martin Eichenbaum	9178	138	67	0.08	29	48	0.68	-0.39	0.00	0.51	
38	Ernst Fehr	8874	184	48	0.17	20	43	0.73	-0.34	0.00	0.51	
39	Allen N. Berger	8578	170	50	0.15	26	46	0.67	-0.49	0.00	0.22	
40	Jeffrey Alexander Frankel	8556	336	25	0.25	34	42	0.73	-0.30	0.00	0.08	✓
41	Michael Woodford	8470	171	52	0.25	26	47	0.65	-0.54	0.00	0.25	
42	Francis Diebold	8272	185	45	0.12	26	46	0.70	-0.41	0.00	0.23	
43	Guillermo Calvo	8188	188	44	0.22	36	37	0.77	-0.29	0.00	0.43	
44	Lars E.O. Svensson	8167	216	38	0.18	33	46	0.66	-0.39	0.00	0.24	
45	Alan B. Krueger	8110	188	43	0.09	26	44	0.69	-0.39	0.00	0.27	
46	Robert J. Shiller	8012	184	44	0.14	40	38	0.74	-0.41	0.00	0.09	✓
47	Lawrence F. Katz	7993	142	56	0.08	29	48	0.64	-0.48	0.00	0.45	
48	Andrew Rose	7950	208	38	0.18	25	47	0.71	-0.46	0.00	0.16	
49	David E. Card	7879	238	33	0.15	33	46	0.66	-0.36	0.00	0.28	
50	Pierre Perron	7451	154	48	0.24	29	28	0.82	-0.33	0.00	0.13	
51	Dani Rodrik	7374	189	39	0.22	27	40	0.73	-0.36	0.00	0.36	
52	Christopher Sims	7172	110	65	0.25	41	34	0.76	-0.59	0.00	0.09	✓
53	Sergio Rebelo	7004	105	67	0.10	25	41	0.65	-0.47	0.00	0.51	
54	Guido Tabellini	6985	152	46	0.15	28	44	0.64	-0.46	0.00	0.32	
55	Angus Deaton	6908	199	35	0.14	40	42	0.69	-0.34	0.00	0.42	
56	Luigi Zingales	6887	112	61	0.13	19	39	0.67	-0.52	0.00	0.24	
57	John Haltiwanger	6841	195	35	0.18	30	41	0.68	-0.42	0.00	0.19	
58	Kevin M. Murphy	6691	94	71	0.13	31	33	0.73	-0.59	0.00	0.17	
59	Gene Grossman	6665	137	49	0.12	32	37	0.70	-0.37	0.00	0.73	
60	Thorsten Beck	6557	152	48	0.23	15	40	0.70	-0.37	0.00	0.60	
61	Paul Milgrom	6528	122	54	0.23	36	37	0.69	-0.58	0.00	0.34	
62	Martin Ravallion	6498	304	21	0.13	25	44	0.62	-0.36	0.00	0.22	
63	Asli Demirguc-Kunt	6487	208	31	0.17	24	45	0.62	-0.37	0.00	0.32	
64	Campbell Harvey	6450	95	68	0.08	24	37	0.63	-0.44	0.00	0.58	
65	Oded Galor	6371	103	62	0.05	27	38	0.70	-0.40	0.00	0.41	
66	James Poterba	6281	310	20	0.24	33	42	0.63	-0.31	0.00	0.13	
67	Thomas Sargent	6273	244	26	0.19	40	44	0.65	-0.35	0.00	0.10	✓
68	John B. Taylor	6269	162	39	0.33	37	34	0.81	-0.42	0.00	0.15	

Continued on next page.

Table 6 – cont. from previous page.

Rank	Name	Citations	Works	C/W	Fraction non-cited	Years	h-index	Gini	Test Tol (2009)		KS-Test	
									γ	p-value	p-value	Matthew?
69	Patrick J. Kehoe	6220	153	41	0.09	27	39	0.69	-0.36	0.00	0.49	
70	Lars Peter Hansen	6189	112	55	0.18	35	34	0.76	-0.31	0.00	0.38	
71	Peter Schmidt	6111	135	45	0.27	39	28	0.85	-0.32	0.00	0.19	
72	Stephen Nickell	6089	147	42	0.19	34	34	0.76	-0.31	0.00	0.14	
73	Barry Eichengreen	6063	437	14	0.30	29	42	0.69	-0.31	0.00	0.01	✓
74	Martin S. Feldstein	6052	427	14	0.27	41	36	0.69	-0.25	0.00	0.02	✓
75	Geert Bekaert	5892	98	60	0.05	23	40	0.58	-0.44	0.00	0.59	
76	Jeremy Stein	5867	118	50	0.05	25	39	0.63	-0.48	0.00	0.41	
77	Zvi Griliches	5858	143	41	0.13	43	33	0.72	-0.36	0.00	0.44	
78	Richard H. Thaler	5799	94	62	0.12	34	40	0.57	-0.44	0.00	0.51	
79	Drew Fudenberg	5769	198	29	0.17	33	40	0.66	-0.38	0.00	0.24	
80	Avinash Kamalakar Dixit	5634	138	41	0.17	34	33	0.77	-0.31	0.00	0.17	
81	Anthony J. Venables	5572	189	30	0.19	27	36	0.75	-0.32	0.00	0.07	✓
82	James R. Markusen	5489	132	42	0.11	33	39	0.68	-0.47	0.00	0.37	
83	William Easterly	5392	140	39	0.14	25	34	0.71	-0.30	0.00	0.61	
84	John List	5294	285	19	0.21	16	41	0.67	-0.34	0.00	0.12	
85	George Borjas	5280	129	41	0.10	37	37	0.67	-0.40	0.00	0.25	
86	Timothy Besley	5279	193	27	0.23	26	41	0.65	-0.41	0.00	0.19	
87	Rene M. Stulz	5244	175	30	0.13	29	37	0.65	-0.41	0.00	0.39	
88	Eric S. Maskin	5191	153	34	0.18	34	37	0.66	-0.48	0.00	0.52	
89	Donald W. K. Andrews	5161	157	33	0.17	29	30	0.81	-0.27	0.00	0.13	
90	Robert C. Feenstra	5134	178	29	0.13	31	33	0.72	-0.30	0.00	0.20	
91	Mark P. Taylor	5130	216	24	0.24	24	35	0.70	-0.38	0.00	0.12	
92	Halbert White	5128	136	38	0.24	35	29	0.83	-0.30	0.00	0.22	
93	Boyan Jovanovic	5115	107	57	0.13	34	31	0.77	-0.37	0.00	0.19	
94	Frederic Mishkin	5084	242	22	0.30	33	42	0.61	-0.42	0.00	0.18	
95	Glenn Rudebusch	5052	146	35	0.21	27	39	0.64	-0.49	0.00	0.42	
96	Jose Scheinkman	4999	99	51	0.09	37	28	0.77	-0.43	0.00	0.22	
97	Joshua Angrist	4986	132	38	0.11	25	38	0.66	-0.38	0.00	0.49	
98	Julio Rotemberg	4974	139	36	0.19	32	29	0.76	-0.44	0.00	0.17	
99	Ricardo J Caballero	4910	152	34	0.09	25	37	0.61	-0.36	0.00	0.56	
100	David Blanchflower	4888	136	36	0.12	25	35	0.73	-0.34	0.00	0.48	

Notes: This table reports the top 100 economists based on the number of citations in RePEc (July 2013). C/W denotes the number of cites per work. From the test by Tol (2009) - equation (2) - we report the parameter γ and its corresponding p-value. The p-value of the Kolmogorov-Smirnov test whether the citation distribution is lognormal is reported in column *KS-Test*. The ✓ denotes whether the null hypothesis of the KS test is rejected on the 10% level, indicating a *Matthew effect*.