Do open access articles in economics have a citation advantage?

Klaus Wohlrabe and Daniel Birkmeier

Ifo Institute for Economic Research

24. June 2014

Do open access articles in economics have a citation advantage?

Klaus Wohlrabe
Ifo Institute for Economic Research
Poschingerstr. 5
81679 Munich, Germany
Phone: +498992241229
wohlrabe@ifo.de
- *Corresponding Author*

Daniel Birmaier
Ifo Institute for Economic Research
Poschingerstr. 5
81679 Munich, Germany
Phone: +498992241229
birkmaier@ifo.de

Abstract

We investigate whether articles in economics that are freely available on the web have a citation advantage over articles with a gated access. Our sample consists of articles from 2005 from 13 economic journals (including the top five journals). In addition to standard mean comparisons we also use a negative-binomial regression model with several covariates to control for potential selection effects and quality bias. Using citation data from three different databases (Web of Science, RePEc and Google Scholar) we show that articles that are freely available on the internet have indeed a significantly higher citation count.

JEL Code: A12, A14
Keywords: Open Access, Citations, Web of Science, RePEc, Google Scholar

1 Introduction

We ask the question whether journal articles in economics which are freely available on the internet[1], either as full text or as a working paper, gather on average significantly more citations than articles that are not. We are the first that use three different bibliometric databases, Web of Science, Repec and Google Scholar, to investigate this issue.

Previous findings across different fields of science claim that there is a citation advantage of open access publications (see e.g. Lawrence (2001), Antelman (2004), Harnad and Brody (2004), Norris, Oppenheim, and Rowland (2008) or Eysenbach (2006)). However, the channels through which this correlation might work are unclear. There are doubts - as expressed

[1] We label this in the following as 'open access' (OA).

for example by Kurtz, Eichhorn, Accomazzi, Grant, Demleitner, Henneken, and Murray (2005), Moed (2007) or Craig, Plume, McVeigh, Pringle, and Amin (2007) - that the pure fact of a paper being OA has an effect per se. The literature proposes a couple of potential candidates for mechanisms that might be responsible alone or, maybe more plausibly, together with other factors.

First, while neither of both articles finds an explicit OA advantage, Kurtz, Eichhorn, Accomazzi, Grant, Demleitner, Henneken, and Murray (2005) or Davis and Fromerth (2007) suggest a quality bias, i.e. the assumption that authors (self-)archiving their publications on freely accessible sources select their better works which provides a relationship between citations and OA status, or that more prominent authors are more inclined to deposit their work at archives like ArXiv (Kurtz, Eichhorn, Accomazzi, Grant, Demleitner, Henneken, and Murray (2005)) or in working paper series as it is common in economics. The results of Gaulé and Maystre (2011) affirm that channel.

On the contrary, Metcalfe (2006) assigns this coincidence to the higher visibility of the papers rather than to such a kind of self-selection. Furthermore, Kurtz, Eichhorn, Accomazzi, Grant, Demleitner, Henneken, and Murray (2005) and Moed (2007) find evidence that there might be an "early-view" channel, i.e. the positive citation effect that stems from the mere possibility of reading an article before it is officially published in an electronic or printed journal (distinct from "open (or free)" versus "not open" access, as defined by Moed (2007)). Some other suggestions are available. Wren (2005) finds that - at least in the field of biomedicine - articles from high-impact journals are more likely to be presented at non-journal websites and thus are signaling a higher quality. Frandsen (2009) emphasizes the fact that the advantage in question could also be due to scholars obeying the principle of least effort, citing OA articles just because they are easily available.

While the major part of the previous literature focuses on other fields of research, mainly natural sciences, relatively few studies are available for economics so far. Norris, Oppenheim, and Rowland (2008) conduct a simple mean comparison for 22 economic journals and find a positive OA effect on citations. McCabe and Snyder (2011) investigate how online access to articles from 100 top economics and business journal changed the citation patterns but do not investigate a potential OA effect, whereas in a most recent paper using journals from ecology, botany and general science, McCabe and Snyder (2013) empirically and theoretically evaluate the issue, arriving at differentiated estimates. They discover a significant OA advantage on average, but when testing by different journal quality levels, they find what they call a ’superstar effect’. That means, while the OA effect is positive for high quality articles, it has got a

negative sign for lower quality articles. Supported by a theoretical model, their explanation is that "converting content from closed to open access broadens the set of articles over which citing authors search. This broader search process encourages substitution by readers towards higher quality articles". In our paper, we will also look at different quality levels of journals but arrive at qualitatively different results.

We investigate in the following the issue, whether articles from economic journals that are freely available on the web have a higher citation count than those that are not. We collect citations for 13 economic journals from three bibliometric databases: Web of Science, Repec and Google Scholar. Simple mean comparisons, as they are standard in the literature, might be biased due to (self-)selections and quality differences of articles and journals. Therefore we use a negative-binomial count data model to account for dispersion in citations and covariates that might influence article citations. We find that articles that are OA significantly gather more citations than those that are not.

2 Empirical Approach

2.1 Data

We take all articles from the American Economic Review[2], Quarterly Journal of Economics, Journal of Political Economy, Econometrica, and Review of Economic Studies published in 2005[3]. This corresponds to the top five journals in the comprehensive journal ranking by Combes and Linnemer (2010)[4]. In order to account for different quality levels of journals we furthermore include the top two journals from the categories: 'AA' (Journal of Financial Economics, Journal of Monetary Economics) , 'A' (Review of Financial Studies, Journal of Business and Economics Statistics), 'B' (Agricultural Economics, China Economic Review) and 'C' (Finanzarchiv, Geneva Risk and Insurance Review). Finally, we have 639 articles.

We check whether an article is freely available online by using Google Scholar. We inserted the title of the article plus the extension "filetype:pdf" into the search field. If a pdf-file was available on the first two pages than we labeled the article as OA. An OA article can be either a working paper or a (self-) archived original article. A potential caveat is that we can-

[2] We exclude the Papers and Proceedings.
[3] Johnston, Piatti, and Torgler (2012) find for the AER that the citation peak for individual articles is about five to seven years after publication.
[4] This category is labeled with 'AAA'.

not identify for how long an article has been freely available. We collected citations from three different bibliometric databases: Web of Science, RePEc and Google Scholar[5]. Previous studies focused mainly on citations from Web of Science. Our approach allows accounting for different coverage rates both of journals and citations in these different databases. We collected all of our data in October 2012.

In Table 1 we tabulate descriptive statistics for the full sample and each journal separately. In column two we report the share of open access articles for the full sample and each journal. Almost 80 % of our our articles in our sample are available freely on the web. This share is higher than the 65% found by Norris, Oppenheim, and Rowland (2008) for 22 Economics journals with 1141 articles. The largest share is given for Econometrica with 96% and lowest with 33% for Agricultural Economics. It seems that articles published in better ranked journals have a higher probability that they are open access.

The third column denotes the share of articles that have been published in a working paper series. Almost 50% of the articles appeared in a working paper series. Again, there is a tendency that articles from high–impact journals are also published as a working paper.

Columns four to six report the average citations per article for the different bibliometric databases. The number of citations for Web of Science and RePEc are quantitatively similar, whereas the ones from Google Scholar are considerably higher. This might be due to a larger coverage of bibliometric items.

2.2 Mean comparisons

The standard approach to test whether the citation count between open and restricted access articles is a mean comparison. We conduct a two-sided t-test with unequal variances. The results are given in Table 2 both for the full sample and each journal separately. Then for each database we report the average citations and the corresponding p-value of the t-test. We reject the null of an equal mean for all three databases, i.e., that OA articles have on average a higher citation count compared to its restricted counterparts. This does not hold for each journal. In case that the null cannot be rejected, i.e., the average citations are significantly not different, this holds almost for every bibliometric database. These results seem to point to the fact that articles published as OA receive on average more citations than those that are not.

[5] See Seiler and Wohlrabe (2012) and references therein for comprehensive comparisons between these databases.

Table 1: Descriptive Statistics

	OA		WP Share			WOS					REPEC	GS	Authors	Pages	Empirical vs. Theoretical	Top10 University	Top 4 WP
Full Sample	0,79	0,49	31	27	151	2,03	25	0,68	0,22	0,21							
Econometrica	0,96	0,67	35	28	156	2,02	34	0,29	0,31	0,13							
American Economic Review	0,90	0,66	42	39	200	1,96	18	0,63	0,35	0,33							
Journal of Political Economy	0,93	0,57	59	65	288	1,90	33	0,57	0,48	0,29							
Quarterly Journal of Economics	0,85	0,80	54	51	260	2,13	39	0,75	0,59	0,60							
Review of Economics and Statistics	0,91	0,65	34	32	169	1,96	26	0,46	0,25	0,24							
Journal of Financial Economics	0,95	0,34	44	25	229	2,24	35	0,89	0,14	0,19							
Journal of Monetary Economics	0,80	0,69	25	37	132	1,93	24	0,67	0,13	0,33							
Review of Financial Studies	0,88	0,25	37	21	183	2,15	37	0,65	0,21	0,13							
Journal of Business and Economic Statistics	0,82	0,54	15	13	54	1,92	12	1,00	0,08	0,08							
Agricultural Economics	0,33	0,15	6	4	28	2,19	12	0,92	0,03	0,00							
China Economic Review	0,73	0,23	14	6	47	1,95	22	0,86	0,11	0,05							
Geneva Risk and Insurance Review	0,82	0,18	3	1	8	1,73	15	0,18	0,00	0,09							
FinanzArchiv	0,46	0,36	3	4	15	1,79	22	0,50	0,01	0,14							

Notes: OA = share of open access articles, WP = share of articles also published as working papers, WoS = average citations based on WoS database, RePEc = average citations based on RePEc database, GS = average citations based on Google Scholar database, Authors = average number auther per article, Type = share of empirical articles, TOP 10 = share of authors affiliated with top university (see text for details), TOP4 = share of articles published in a top working paper series (see text for details).

Table 2: Mean citation comparison

	N		Web of Science			RePEc			Google Scholar		
	OA=1	OA=0	OA=1	OA=0	p-value	OA=1	OA=0	p-value	OA=1	OA=0	p-value
Full Sample	508	131	36	12	**0.00**	32	9	**0.00**	177	50	**0.00**
Econometrica	53	2	35	26	0.71	29	12	0.32	159	79	0.38
American Economic Review	83	9	44	23	**0.02**	40	26	0.32	210	106	**0.05**
Journal of Political Economy	39	3	64	5	**0.00**	70	3	**0.01**	309	15	**0.00**
Quarterly Journal of Economics	34	6	55	50	0.81	54	36	0.27	268	214	0.54
Review of Economics and Statistics	42	4	35	19	**0.05**	35	9	**0.01**	180	53	**0.00**
Journal of Financial Economics	75	4	44	34	0.58	25	16	0.30	232	172	0.52
Journal of Monetary Economics	49	12	27	16	0.13	40	27	0.39	142	89	0.30
Review of Financial Studies	35	5	40	16	**0.01**	23	7	**0.05**	198	77	**0.01**
Journal of Business and Economic Statistics	32	7	17	9	0.14	14	7	0.16	59	30	0.15
Agricultural Economics	28	56	7	6	0.39	6	2	**0.02**	40	22	**0.06**
China Economic Review	16	6	18	6	**0.01**	7	1	**0.01**	58	15	**0.00**
Geneva Risk and Insurance Review	9	2	3	0	**0.02**	2	0	**0.02**	9	1	**0.00**
FinanzArchiv	13	15	2	3	0.63	3	4	0.76	15	15	0.96

Notes: N = Number of Articles, OA=1 and WP=1, articles that are freely available on the internet or as working paper, respectively. The p-values are obtained from a two--sided t--test with unequal variances. The null is that the average citations are equal. P-values smaller than 0.1 are in bold.

2.3 Regression approach

In a second step, we estimate the OA effect on citations by modeling citation count of a paper by using a negative-binomial regression model. In comparison to a Poisson regression model, this approach is able to handle explicitly over-dispersion, a feature often found for citation data[6]. As controls we follow mainly Johnston, Piatti, and Torgler (2012): proportion of authors that were affiliated with a top 10 university[7] (Top 10); number of authors (Authors); the number an article appeared in a working paper series (nWP); years since first publication until 2012 (Years)[8]; a dummy whether an article was primarily an empirical one or not (Empirical)[9]; publication in top working paper series (TOP WP)[10] and the number of pages (Pages). In Table 1 we report some descriptive statistics for our controls. Furthermore, we include journal dummies which control for unobserved effects and journal quality.

Table 3 shows the regression results for different specifications and for each bibliometric database separately. Instead of coefficients we report incidence-rate ratios which can be interpreted ceteris paribus as relative percentage changes. We start with the interpretation of the Web of Science database. The first column denotes the result for the OA effects without any controls. The significant incidence-rate ratio of 3.079 denotes that OA articles have on average a 307.9% higher citation count compared to articles that are not freely available on the web. This corresponds clearly to the mean comparison results in Table 1. Column 2 reports the same results with control variables. The OA citation advantage reduces to 35% and remains highly significant. The control variables have the expected sign and are almost all significant.

These significant results hold also mainly for the other two bibliometric databases. The open access advantages is higher and still significant. In case of RePEc this advantage rises to about 55% and for Google Scholar to 63.5%. This also corresponds to the results in Table 1, where it can be seen that the citations counts for both databases are on average higher compared to the Web of Science database.

In columns 3 and 4 of Table 3 we report the results only for the top five journals. The citation advantage effect shrinks and remains significant with the exception of the Web of Science database. In the last two columns we show the results for the remaining journals. The

[6] See for example Ketzler and Zimmermann (2012) or Johnston, Piatti, and Torgler (2012).
[7] The ranking is based on the analysis in Amir and Knauff (2008).
[8] Number of years since publication, either as a working paper or as a journal article.
[9] For a definition see Johnston, Piatti, and Torgler (2012).
[10] Our top working paper series are from the following institutions: NBER, CEPR, IZA and CESifo.

effect is larger and significant across all three databases. Therefore we can conclude that our results are robust.

3 Conclusions

Using mean comparisons and regression models we test the hypothesis whether articles in economics that are freely available on the internet have a higher citation count than those that are not. Our sample consists of 639 articles published in 2005 from 13 economic journals, including the so-called top five journals. We find a significant OA effect on citations counts for economic articles. This effect is robust across three different bibliometric databases. One may conclude that researchers should make their articles freely available on the web to increase their citation count. Future research can investigate the issue whether our finding may change over time.

Table 3: Regression results

	Full Sample		Top 5 Journals		Non Top 5 Journals	
Citations from Web of Science						
OA	3.079 ***	1.349 ***	1.661 *	1.320	3.320 ***	1.377 **
nWP		1.064		1.087		0.986
Years		1.060 *		1.075		1.063 **
Authors		1.109 **		1.184		1.055
TOP 10		1.466 ***		1.569 ***		1.341
Pages		1.012 **		1.008		1.017 ***
Empirical		1.858 ***		1.857 ***		1.860 ***
TOP WP		1.407 ***		1.317 **		1.616 **
Citations from RePEc						
OA	3.523 ***	1.546 ***	3.342 ***	1.505 ***	3.342 ***	1.601 ***
nWP		1.215 **		1.192 *		1.198
Years		1.141 ***		1.137 **		1.156 **
Authors		1.070		1.221		0.975
TOP 10		1.460 ***		1.470 ***		1.531 **
Pages		1.009		1.009		1.011
Empirical		1.756 ***		1.578 ***		1.982 ***
TOP WP		1.335 **		1.309 *		1.392
Citations from Google Scholar						
OA	3.553 ***	1.635 ***	3.813 ***	1.521 **	3.813 ***	1.701 ***
nWP		1.101 *		1.115		1.038
Years		1.087 **		1.111 *		1.077 *
Authors		1.048		1.169		0.971
TOP 10		1.631 ***		1.720 ***		1.570 **
Pages		1.013 **		1.008		1.019 **
Empirical		1.871 ***		1.870 ***		1.895 ***
TOP WP		1.529 ***		1.490 ***		1.620 **
Journal Dummies		Yes		Yes		Yes
N	639	639	275	275	364	364

Notes: This table reports the results of the negative-binomial regression model using clustered standard errors. The dependent variables are citations. We report incidence-rate ratios. The dispersion coefficient (α) is not reported but is in all cases significantly different from zero, i.e. the data is overdispersed. ***, **, * denote significance at the 1%, 5% and 10% level.

References

Amir, R., and M. Knauff (2008). Ranking Economics Departments Worldwide on the Basis of PhD Placement. The Review of Economics and Statistics, 90(1), 185–190.

Antelman, K. (2004). Do open-access articles have a greater research impact? College & Research Libraries, 65(5), 372–382.

Combes, P.-P., and L. Linnemer (2010). Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of all Journals in Economics. Working papers, HAL.

Craig, I. D., A. M. Plume, M. E. McVeigh, J. Pringle, and M. Amin (2007). Do open access articles have greater citation impact? A critical review of the literature. Journal of Informetrics, 1(3), 239–248.

Davis, P. M., and M. J. Fromerth (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? Scientometrics, 71(2), 203–215.

Eysenbach, G. (2006). Citation advantage of open access articles. PLoS Biology, 4(5), e157.

Frandsen, T. F. (2009). Attracted to open access journals: a bibliometric author analysis in the field of biology. Journal of Documentation, 65(1), 58–82.

Gaulé, P., and N. Maystre (2011). Getting cited: Does open access help? Research Policy, 40(10), 1332–1338.

Harnad, S., and T. Brody (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-lib Magazine, 10(6).

Johnston, D. W., M. Piatti, and B. Torgler (2012). Citation Success Over Time: Theory or Empirics?, School of Economics and Finance Discussion Papers and Working Papers Series 284, School of Economics and Finance, Queensland University of Technology.

Ketzler, R., and K. F. Zimmermann (2012). A Citation-Analysis of Economic Research Institutes. IZA Discussion Papers 6780, Institute for the Study of Labor (IZA).

Kurtz, M. J., G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, E. Henneken, and S. S. Murray (2005). The effect of use and access on citations. Information Processing & Management, 41(6), 1395–1402.

Lawrence, S. (2001). Online or invisible. Nature, 411(6837), 521.

McCabe, M., and C. Snyder (2011). Did Online Access to Journals Change the Economics Literature? SSRN Working Paper 1746243.

McCabe, M. J., and C. M. Snyder (2013). The Rich Get Richer and the Poor Get Poorer: The Effect of Open Access on Cites to Science Journals Across the Quality Spectrum. SSRN Working Paper 2269040.

Metcalfe, T. S. (2006). The citation impact of digital preprint archives for solar physics papers, Solar Physics. 239(1), 549–553.

Moed, H. F. (2007). The effect of open access on citation impact: An analysis of ArXiv's condensed matter section, Journal of the American Society for Information Science and Technology. 58(13), 2047–2054.

Norris, M., C. Oppenheim, and F. Rowland (2008). The citation advantage of open-access articles, Journal of the American Society for Information Science and Technology. 59(12), 1963–1972.

Seiler, C., and K. Wohlrabe (2012). Ranking Economists on the Basis of Many Indicators: An Alternative Approach Using RePEc Data. Journal of Informetrics, 6(2), 389–402.

Wren, J. D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. BMJ, 330(7500), 1128.