Wohlrabe, Klaus; Bürgi, Constantin

Working Paper

Do Working Papers Increase Journal Citations? Evidence from the Top 5 Journals in Economics

CESifo Working Paper, No. 8643

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Wohlrabe, Klaus; Bürgi, Constantin (2020) : Do Working Papers Increase Journal Citations? Evidence from the Top 5 Journals in Economics, CESifo Working Paper, No. 8643, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/229461

CESifo WORKING PAPERS

8643
2020

October 2020

Do Working Papers Increase Journal Citations? Evidence from the Top 5 Journals in Economics

Klaus Wohlrabe, Constantin Bürgi

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
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: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8643*

Do Working Papers Increase Journal Citations? Evidence from the Top 5 Journals in Economics

Abstract

Does it pay off in terms of citations to issue an article as a working paper before it is published in a refereed journal? We show empirically that the answer is yes, using 3167 articles published in five of the top journals in economics between 2000 and 2010. The effect is an around 25% higher number of citations on average across the investigated journals.

JEL-Codes: A100, A120.

Keywords: working paper, peer review, citations, RePEc, preprint.

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

Constantin Bürgi
St Mary's College of Maryland
47645 College Drive
USA – St. Mary's City, MD 20686
crburgi@smcm.edu

October 2020

1 Introduction

Most of the communication in science is done via scientific journals which includes the peer review process. In many fields including economics, it is quite common to release an article as a preprint or working paper as well. One of the first WP series in economics is the Cowles Foundation Discussion Papers founded in 1955. Since then, the number of working paper series has increased substantially. More than 5,000 series with almost one million working papers are listed on the RePEc website (Research Papers in Economics) as of June 2020. Working papers are usually not subject to a formal peer review process. There are several reasons for making articles working papers: Authors may want to gather input from other scientists, the article might become circulated more broadly, or it can be a way to create a time stamp signalling to be the first with an idea, especially when potentially many scientists work on a similar topic.[1]

In this paper we ask the question whether an article gets cited more often if it is also available as a working paper or not. There are multiple reasons, why one would expect a higher citation count. These can include the access to a free version of the paper, reaching a wider audience because the working paper is the result of a conference presentation. Also, the timeliness of working papers might make follow up research easier and less competitive.

Bürgi and Wohlrabe (2020) show the reverse case and working papers that get the formal stamp of approval of a refereed journal gain a boost in citations. Using more than 3,000 articles from the five top economics journals (Card and DellaVigna, 2013; Bornmann et al., 2018) published between 2000 and 2010 we show that articles which are also available as working papers have more citations.

This issue has been investigated in areas other than economics before. For example, Sarabipour et al. (2019), Fraser et al. (2019), or Fu and Hughey (2019) show for biology that journal articles which are available as preprints or working papers are cited more often.

[1] Brown and Zimmermann (2017) provide a detailed discussion on this issue.

2 Data

For our analysis we collected all articles published between 2000 and 2010 from the following five top journals: American Economic Review (AER), Econometrica, Journal of Political Economy (JPE), Quarterly Journal of Economics (QJE), and Review of Economic Studies (ReStud). We excluded the Papers and Proceedings from the AER and we do not consider articles published after 2010 as it takes time to gather sufficient citations. In total, our data set comprises 3,167 articles. For each article, we searched the RePEc website whether it is available as a working paper or not. The citation data was obtained from Citec[2] which is part of the RePEc network. In Table 1 we provide descriptive statistics for each journal. About 55% of all articles are available as a working paper as well.[3][4] This share is similar across journals with the exception of the QJE where 65% of the articles are also a working paper. On average, an article in these five journals was cited 154 times. The highest average citation count is found for the QJE with 256. The two lower panels compare the citation patterns between articles previously published as working paper with those which are not. We see a clear and statistically significant difference. Across all journals it pays off to publish an article as WP beforehand.

3 Empirical approach and results

3.1 Empirical approach

In order to test whether working paper versions of an article have a statistically significant effect on citations, we estimate the following model

$$cit_i = \alpha + \beta WP_i + \gamma X_i + \delta_{jt} + \varepsilon_i \tag{1}$$

[2]http://citec.repec.org/

[3]Baumann and Wohlrabe (2020) found that around 65% of working papers get published in a journal.

[4]This number can be considered as a lower bound. We are not able to control for papers which are posted freely as a working paper on private web pages.

Table 1: Descriptive statistics for citation count

	N	Mean	Standard deviation	Minimum	Maximum	t-Test p-value
Overall						
Total	3,167	154	254	1	5,424	
AER	1,091	154	245	1	3,550	
Econometrica	678	131	278	1	5,424	
JPE	464	142	209	1	3,077	
QJE	456	256	327	2	3,032	
ReStud	478	101	156	1	2,027	
Published as a WP before						
Total	1,761	186	298	1	5,424	
AER	568	185	280	1	3,550	
Econometrica	368	162	352	1	5,424	
JPE	257	181	256	3	3,077	
QJE	299	287	365	4	3,032	
ReStud	269	114	140	1	1,118	
Not published as WP before						
Total	1,406	113	177	1	2,192	0.000
AER	523	119	195	1	2,192	0.000
Econometrica	310	94	138	1	1,006	0.001
JPE	207	95	111	1	688	0.000
QJE	157	196	229	2	1,739	0.002
ReStud	209	84	172	1	2,027	0.019

where cit_i is the number of citations article i received. WP is a dummy variable which takes value 1 if the article is available as a working paper and 0 otherwise. We include several control variables summarized in X_i: the number of pages, pages squared and the number of authors. These variables have been frequently identified as factors that may influence citations in various studies (see the overview in Tahamtan and Bornmann (2018)). We also include journal-year dummies (δ_{jt}) that take value one if the journal is j and the year is t and zero otherwise with one dummy for each j,t combination.[5] They capture the journal quality in each year as well as the paper age and the citation practices over time. We estimate four different models: A basic OLS regression model, an OLS regression using the natural log of citations as the dependent variable, which accounts for skewness in the citation distribution, a negative binomial regression (NBR) model, where citations are interpreted as counts and an OLS regression using the inverse hyperpolic sine (IHS or $asinh$) transformation of citations, similar to log transformation, as proposed by Burbidge et al. (1988) and put forward recently by Card and DellaVigna (2020) in a citation analysis. The formal definition is $asinh(z) = \ln(z + \sqrt{1 + z^2})$. For $z \geq 2$, $asinh(z) = \ln(z) + \ln(2)$, but $asinh(0) = 0$.[6] The four approaches should yield robust results with respect to both the estimation approach and the handling of the dependent variable, the citations. For all regressions, we use standard errors that are robust against heteroskedasticity.

3.2 Results

We present our regression results in Table 2. We find statistically significant increases in the citation count for papers that are available as working papers for all four specifications. While the OLS regression suggests around 50 additional citations, the three other specifications imply an around 45% increase in the number of citations for papers where a working paper is available.[7] The coefficients of the controls are in line with for example Tahamtan and

[5] These dummies also cover what separate year and journal dummies would cover.
[6] The least cited article in our sample has 1 citation, so zeros are not an issue.
[7] In order to interpret the coefficients, it is necessary to take their exponential for those three cases.

Bornmann (2018) as longer papers and more co-authors lead to more citations as well.

Table 2: Regression results for the complete sample

	(1)	(2)	(3)	(4)
Estimation approach	OLS	OLS	NBR	OLS
Dependent variable	citations	log citations	citations	*asinh*
WP	47.716***	0.352***	0.361***	0.364***
	(8.565)	(0.040)	(0.048)	(0.041)
Pages	6.583***	0.094***	0.069***	0.100***
	(1.226)	(0.007)	(0.008)	(0.007)
Pages squared	-0.049**	-0.001***	-0.001***	-0.001***
	(0.021)	(0.000)	(0.000)	(0.000)
# of Authors	17.068***	0.150***	0.106***	0.155***
	(6.024)	(0.024)	(0.029)	(0.025)
Observations	3,167	3,167	3,167	3,167
R-squared	0.105	0.269		0.272
Journal-Year FE	YES	YES	YES	YES

The coefficients in the first column show the number of additional citations, while taking the exponential of the coefficients in the last three columns corresponds to the multiplicative increase in citations (e.g. $e^{0.352}$-1=42.19% more citations). Heteroskedasticity robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The results in Table 2 might be biased because it does not control for the influence or reputation of the authors. It could be the case that papers by prominent authors get more citations just because they are famous. This also known as the Matthew effect (Birkmaier and Wohlrabe, 2014). Papers by established authors might also be different with respect to the quality. In order to account for this issue we tried to match all authors with their corresponding Citec profiles where we retrieved the overall citation count. A prerequisite for the matching is that authors are registered in RePec. This does not hold for all authors in our sample. We were able to match at least one author for 2,434 articles in our sample. There are some cases where, i.e, for two authors only one could be matched. We handled this issue by taking the average over all available citations counts. We subtracted the individual paper

citations from the total citation counts of the authors under investigation. In Table 3 we present the results of regression after controlling for the average reputation of the authors. It shows that the results are qualitatively the same compared to Table 2. The reputation of authors is statistically significant, implying that status matters for the citation count of an article. The size of the working paper dummies is smaller compared to the regressions where reputation was not included. Converted into percentages, the last three columns imply an around 25% increase in the citation count for papers where a working paper is available.

Table 3: Regression results for the complete sample controlling for reputation of authors

	(1)	(2)	(3)	(4)
Estimation approach	OLS	OLS	NBR	OLS
Dependent variable	citations	log citations	citations	*asinh*
WP	31.011***	0.211***	0.220***	0.219***
	(8.754)	(0.044)	(0.049)	(0.046)
Pages	6.608***	0.077***	0.063***	0.081***
	(1.305)	(0.008)	(0.008)	(0.008)
Pages squared	-0.052**	-0.001***	-0.001***	-0.001***
	(0.022)	(0.000)	(0.000)	(0.000)
# of Authors	21.417***	0.107***	0.100***	0.109***
	(6.274)	(0.026)	(0.030)	(0.027)
Reputation	0.009***	0.232***	0.000***	0.234***
	(0.002)	(0.017)	(0.000)	(0.017)
Observations	2,434	2,434	2,434	2,434
R-squared	0.155	0.306		0.305
Journal-Year FE	YES	YES	YES	YES

The coefficients in the first column show the number of additional citations, while taking the exponential of the coefficients in the last three columns corresponds to the multiplicative increase in citations (e.g. $e^{0.211}$-1=23.49% more citations). Heteroskedasticity robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

We now run the four regression models controlling for reputation of the authors for each journal separately. In Table 4, we report the working paper dummy for each journal-model combination. Aside from the OLS regressions in the first column and Econometrica, the table

Table 4: Regression results for each journal separately

	(1)	(2)	(3)	(4)
Estimation approach	OLS	OLS	NBR	OLS
Dependent variable	citations	log citation	citations	*asinh*
AER	28.052	0.213***	0.255***	0.220***
	(17.073)	(0.080)	(0.091)	(0.083)
Econmetrica	2.676	0.051	0.019	0.057
	(14.031)	(0.091)	(0.096)	(0.094)
JPE	57.497***	0.318***	0.355***	0.326***
	(18.122)	(0.113)	(0.119)	(0.116)
QJE	73.154**	0.293**	0.254**	0.299**
	(30.973)	(0.124)	(0.124)	(0.126)
ReStud	8.612	0.232**	0.234**	0.243**
	(18.154)	(0.107)	(0.114)	(0.112)

The coefficients in the first column show the number of additional citations, while taking the exponential of the coefficients in the last three columns corresponds to the multiplicative increase in citations (e.g. $e^{0.213}$-1=23.74% more citations). Heteroskedasticity robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

shows similar coefficients and statistical significance as the results in table 3. The significant percentage increases across journals are between 20-40% with Econometrica benefiting the least by having a working paper and the JPE benefiting the most.

3.3 Robustness checks

We conducted several robustness checks. First, in order to ensure that highly cited papers are not driving our results, we repeat the regressions in Table 3 but exclude papers that were cited more than 100 times. The coefficients decline as Table 5 shows but remain highly significant. This result implies that papers that tend to be cited more often benefit more from having a working paper available.

Second, instead of employing a dummy variable we included in the regressions the number of working papers an article appeared in. The corresponding coefficients reported in Table

Table 5: Regression results for papers cited less than 100 times

	(1)	(2)	(3)	(4)
Estimation approach	OLS	OLS	NBR	OLS
Dependent variable	citations	log citation	citations	*asinh*
WP	5.216***	0.177***	0.130***	0.188***
	(1.291)	(0.038)	(0.030)	(0.040)
Pages	1.950***	0.078***	0.061***	0.085***
	(0.155)	(0.006)	(0.006)	(0.007)
Pages squared	-0.024***	-0.001***	-0.001***	-0.001***
	(0.003)	(0.000)	(0.000)	(0.000)
# of Authors	3.364***	0.103***	0.082***	0.109***
	(0.813)	(0.024)	(0.019)	(0.026)
Observations	1,818	1,818	1,818	1,818
R-squared	0.169	0.233		0.236
Journal-Year FE	YES	YES	YES	YES

The coefficients in the first column show the number of additional citations, while taking the exponential of the coefficients in the last three columns corresponds to the multiplicative increase in citations (e.g. $e^{0.177}$-1=19.36% more citations). Heteroskedasticity robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 6: Regression results based on the number of working papers

	(1)	(2)	(3)	(4)
Estimation approach	OLS	OLS	NBR	OLS
Dependent variable	citations	log citation	citations	*asinh*
WP	14.518***	0.091***	0.092***	0.093***
	(3.726)	(0.015)	(0.017)	(0.015)
Pages	6.567***	0.077***	0.063***	0.081***
	(1.311)	(0.008)	(0.008)	(0.008)
Pages squared	-0.051**	-0.001***	-0.001***	-0.001***
	(0.003)	(0.000)	(0.000)	(0.000)
# of Authors	18.493***	0.091***	0.084***	0.092***
	(6.390)	(0.027)	(0.030)	(0.028)
Reputation	0.009***	0.230***	0.000***	0.232***
	(0.002)	(0.017)	(0.000)	(0.017)
Observations	2,434	2,434	2,434	2,434
R-squared	0.157	0.310		0.309
Journal-Year FE	YES	YES	YES	YES

The coefficients in the first column show the number of additional citations, while taking the exponential of the coefficients in the last three columns corresponds to the multiplicative increase in citations (e.g. $e^{0.091}$-1=9.53% more citations). Heteroskedasticity robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

6 are statistically significant across specifications. The regressions imply that one additional appearance in a working paper series increases the citation count by 15 for OLS and around 10% for the other specifications.

Third, for 733 articles in our sample we were not able to recover the overall citation counts of the authors in RePEc. In order to rule out any selection effect we run the regressions only for those articles without the reputation variable. The results (Table 7) are qualitatively and quantitatively similar to those reported in Table 2.

Table 7: Regression results where no author reputation is available

	(1)	(2)	(3)	(4)
Estimation approach	OLS	OLS	NBR	OLS
Dependent variable	citations	log citation	citations	*asinh*
WP	54.175**	0.331***	0.396***	0.342***
	(24.388)	(0.090)	(0.091)	(0.092)
Pages	5.030	0.113***	0.095***	0.122***
	(3.253)	(0.013)	(0.014)	(0.014)
Pages squared	-0.038	-0.001***	-0.001***	-0.001***
	(0.066)	(0.000)	(0.000)	(0.000)
# of Authors	-6.325	0.077	0.028	0.083
	(12.537)	(0.055)	(0.058)	(0.057)
Observations	733	733	733	733
Journal-Year FE	YES	YES	YES	YES

The coefficients in the first column show the number of additional citations, while taking the exponential of the coefficients in the last three columns corresponds to the multiplicative increase in citations (e.g. $e^{0.331}$-1=39.24% more citations). Heteroskedasticity robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

4 Conclusion

This paper addressed how many more citations a journal article receives depending on whether a working paper is also available. We showed that there is a roughly 25% increase in the number of citations for articles which also appeared in working paper series. There is some heterogeneity across journals, but the impact is positive and significant for all journals. The exception is Econometrica, where do not find a significant effect.

Determining the exact channel through which the additional citations is beyond the scope of the paper, but could potentially be an interesting avenue for future research. Specifically, it would be interesting to test whether the increase is due to having a version of the paper freely available or due to other factors.

References

Baumann, A. and Wohlrabe, K. (2020). Where have all the working papers gone? evidence from four major economics working paper series. *Scientometrics*, 124:2433–2441.

Birkmaier, D. and Wohlrabe, K. (2014). The matthew effect in economics reconsidered. *Journal of Informetrics*, 8(4):880–889.

Bornmann, L., Butz, A., and Wohlrabe, K. (2018). What are the top five journals in economics? a new meta-ranking. *Applied Economics*, 50(6):659–675.

Brown, A. J. and Zimmermann, K. F. (2017). Three decades of publishing research in population economics. *Journal of Population Economics*, 30(1):11–27.

Burbidge, J. B., Magee, L., and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83(401):123–127.

Bürgi, C. and Wohlrabe, K. (2020). What is the benefit from publishing a working paper in a journal? evidence from economics. Technical report.

Card, D. and DellaVigna, S. (2013). Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1):144–61.

Card, D. and DellaVigna, S. (2020). What do editors maximize? evidence from four leading economics journals. *Review of Economics and Statistics*, 102(1):195–217.

Fraser, N., Momeni, F., Mayr, P., and Peters, I. (2019). The effect of biorxiv preprints on citations and altmetrics. bioRxiv 673665.

Fu, D. Y. and Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *Elife*, 8:e52646.

Sarabipour, S., Debat, H. J., Emmott, E., Burgess, S. J., Schwessinger, B., and Hensel, Z. (2019). On the value of preprints: An early career researcher perspective. *PLoS Biology*, 17(2):e3000151.

Tahamtan, I. and Bornmann, L. (2018). Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics*, 12(1):203–216.