

Do working papers increase journal citations? Evidence from the top 5 journals in economics

Klaus Wohlrabe & Constantin Bürgi

To cite this article: Klaus Wohlrabe & Constantin Bürgi (2020): Do working papers increase journal citations? Evidence from the top 5 journals in economics, Applied Economics Letters, DOI: 10.1080/13504851.2020.1855303

To link to this article: <https://doi.org/10.1080/13504851.2020.1855303>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 30 Nov 2020.



Submit your article to this journal [↗](#)



Article views: 172



View related articles [↗](#)




View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Do working papers increase journal citations? Evidence from the top 5 journals in economics

Klaus Wohlrabe ^a and Constantin Bürgi^b

^aCenter for Macroeconomics and Surveys, Ifo Institute for Economic Research, Munich, Germany; ^bDepartment of Economics, St Mary's College of Maryland, Mary's City, MD, United States

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.

KEYWORDS

Working paper; peer review; citations; RePEc; preprint

JEL CLASSIFICATION

A10; A12

I. 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 article, 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 (2020b) 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, Butz, and Wohlrabe 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.

II. 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 were

CONTACT Klaus Wohlrabe  wohlrabe@ifo.de

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

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

obtained from Citec² 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.³⁴ 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.

III. Empirical approach and results

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

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

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 hyperbolic sine (IHS or $asinh$) transformation of citations, similar to log transformation, as proposed by Burbidge, Magee, and Robb (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

²<http://citec.repec.org/>

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

⁴This 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.

⁵These dummies also cover what separate year and journal dummies would cover.

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.

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 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	<i>asinh</i>
WP	47.716*** (8.565)	0.352*** (0.040)	0.361*** (0.048)	0.364*** (0.041)
Pages	6.583*** (1.226)	0.094*** (0.007)	0.069*** (0.008)	0.100*** (0.007)
Pages squared	-0.049** (0.021)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
# of Authors	17.068*** (6.024)	0.150*** (0.024)	0.106*** (0.029)	0.155*** (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

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	<i>asinh</i>
WP	31.011*** (8.754)	0.211*** (0.044)	0.220*** (0.049)	0.219*** (0.046)
Pages	6.608*** (1.305)	0.077*** (0.008)	0.063*** (0.008)	0.081*** (0.008)
Pages squared	-0.052** (0.022)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
# of Authors	21.417*** (6.274)	0.107*** (0.026)	0.100*** (0.030)	0.109*** (0.027)
Reputation	0.009*** (0.002)	0.232*** (0.017)	0.000*** (0.000)	0.234*** (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

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 is 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 citation 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.

Heteroskedasticity robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4. Regression results for each journal separately.

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

⁶The least cited article in our sample has 1 citation, so zeros are not an issue.

⁷In order to interpret the coefficients, it is necessary to take their exponential for those three cases.

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

Heteroskedasticity robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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 but remain highly significant.⁸ 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 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 are qualitatively and quantitatively similar to those reported in Table 2.

IV. 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 we 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.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Klaus Wohlrabe  <http://orcid.org/0000-0002-5651-2641>

References

- Baumann, A., and K. Wohlrabe (2020). “Where Have All the Working Papers Gone? Evidence from Four Major Economics Working Paper Series.” *Scientometrics*, 124:2433–2441.
- Birkmaier, D., and K. Wohlrabe. 2014. “The Matthew Effect in Economics Reconsidered.” *Journal of Informetrics* 8 (4): 880–889. doi:10.1016/j.joi.2014.08.005.
- Bornmann, L., A. Butz, and K. Wohlrabe. 2018. “What are the Top Five Journals in Economics? A New Meta-ranking.” *Applied Economics* 50 (6): 659–675. doi:10.1080/00036846.2017.1332753.
- Brown, A. J., and K. F. Zimmermann. 2017. “Three Decades of Publishing Research in Population Economics.” *Journal of Population Economics* 30 (1): 11–27. doi:10.1007/s00148-016-0620-1.
- Burbidge, J. B., L. Magee, and A. L. Robb. 1988. “Alternative Transformations to Handle Extreme Values of the Dependent Variable.” *Journal of the American Statistical Association* 83 (401): 123–127. doi:10.1080/01621459.1988.10478575.
- Bürge, C., and K. Wohlrabe (2020a). “Do Working Papers Increase Journal Citations? Evidence from the Top 5 Journals in Economics.” CESifo Working Paper 8643.
- Bürge, C., and K. Wohlrabe (2020b). What is the benefit from publishing a working paper in a journal? Evidence from economics. Technical report.
- Card, D., and S. DellaVigna. 2013. “Nine Facts about Top Journals in Economics.” *Journal of Economic Literature* 51 (1): 144–161. doi:10.1257/jel.51.1.144.

⁸Detailed results are available in Bürge and Wohlrabe (2020a).

- Card, D., and S. DellaVigna. 2020. "What Do Editors Maximize? Evidence from Four Leading Economics Journals." *Review of Economics and Statistics* 102 (1): 195–217. doi:10.1162/rest_a_00839.
- Fraser, N., F. Momeni, P. Mayr, and I. Peters. 2019. "The Effect of Biorxiv Preprints on Citations and Altmetrics." In *bioRxiv*, 673665.
- Fu, D. Y., and J. J. Hughey. 2019. "Releasing a Preprint Is Associated with More Attention and Citations for the Peer-reviewed Article." *Elife* 8: e52646. doi:10.7554/eLife.52646.
- Sarabipour, S., H. J. Debat, E. Emmott, S. J. Burgess, B. Schwessinger, and Z. Hensel. 2019. "On the Value of Preprints: An Early Career Researcher Perspective." *PLoS Biology* 17 (2): e3000151. doi:10.1371/journal.pbio.3000151.
- Tahamtan, I., and L. Bornmann. 2018. "Core Elements in the Process of Citing Publications: Conceptual Overview of the Literature." *Journal of Informetrics* 12 (1): 203–216. doi:10.1016/j.joi.2018.01.002.