CESifo WORKING PAPERS

8328
2020

May 2020

Where Have All the Working Papers Gone? Evidence from Four Major Economics Working Paper Series

Alexandra Baumann, Klaus Wohlrabe

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. 8328*

Where Have All the Working Papers Gone? Evidence from Four Major Economics Working Paper Series

Abstract

Working papers or preprints have become an important part in the scientific landscape. Such papers present research before (potentially) being published in refereed journals. But is every working paper finally published in a journal? We answer this question for four major working paper series in economics. Based on linked data in RePEc and a random sample we provide an estimate of 66.5% of more than 28,000 investigated working papers that are published in a journal. About 8% are released as a book chapter. For the remaining 25.5% we find no evidence for what happened to the article.

JEL-Codes: A120, A140.

Keywords: working paper, journals, scholary communication, publication analysis, RePEc.

Alexandra Baumann
TU Munich / Germany
Alexandra.C.Baumann@gmx.de

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

1 Introduction

Articles published in printed academic journals are the backbone of scientific communication. With the emergence of the internet a second channel has become more and more important, where so called preprints, working papers (WP) or any other kind of preliminary articles published (mostly) as a pdf-file are deposited. These kind of articles or papers have in almost all cases not been subject to a formal peer review process. There are several reasons for this type of publishing. Before submitting to a journal, authors want to gather input from other scientists. It can be a way of signaling in the scientific competition to be the first, especially when potentially many scientists work on the same or a similar topic.[1] Preprints are either published in a WP series, on a corresponding preprint server or in a private repository. WP are quite common in the social sciences – especially in economics. One of the first WP series is the Cowles Foundation Discussion Papers founded in 1955. Three well-known preprint repositories are arXiV, SSRN (Social Science Research Network) and RePEc (Research Papers in Economics), where the number of listed articles is huge. As of December 2019 on arXiv were more than 1.6 million articles listed, on SSRN about 770,000 and on RePEc over 800,000 working papers. Li et al. (2015) provide an overview of various repositories and their role in scholarly communication.

A natural question that arises is, if an article published as a preprint or WP has been finally published in a scientific journal, i.e. passing the peer review process. There are only a few articles that have dealt with this issue. Brown and Zimmermann (2017) check which articles from the *Journal of Population Economics* (275 in total) were published previously as a WP. McCabe and Snyder (2015) investigated for about 900 economics articles whether they have been deposited on a public available archive as

[1] Brown and Zimmermann (2017) provide a detailed discussion on this issue. Sarabipour et al. (2019) outline the value of preprints for early stage researchers.

RePEc or university web page. Larivière et al. (2014) analyzed almost 750,000 papers published on arXiV whether they have been finally published in a journal listed in Web of Science. Abdill and Blekhman (2019) did the same for 37,600 articles from bio_RXiv. Tsunoda et al. (2019) also analyzed from bio_RXiv, using a smaller sample of 17,800 articles. Finally, Anderson (2020) shows that over a period of five years 30% of preprints uploaded on bio_RXiv never get published in a journal.

We contribute to this literature using data covering more than 28.000 articles published in four major economics working paper series. This this the largest data set in the social sciences so far. Our analysis is based on RePEc data. This article investigates where all these WP have gone: in a journal or a book chapter. Besides the pure counting we also matched the impact factors to these journals in order to assess the selection in to high and low-quality journals. Furthermore, evidence is provided that a growing number of articles are published in several working paper series. Finally, we answer the question of how wide the time gap is between the date of the WP and journal publication.

2 Data

In socio-economic sciences RePEc has become an essential source for the spread of knowledge and ranking of individual authors and academic institutions. The RePEc network is growing continuously, as of December 2019 there were 2.8 million pieces of research from 3,200 journals and 5,000 working paper series. Additionally, more than 55,000 authors and 14,000 institutions from 101 countries are listed on the website.[2] Our study is based on articles from four major working paper series in economics

[2]RePec data has been often used in bibliometric research. See, for example, Zimmermann (2013), Seiler and Wohlrabe (2013), Rath and Wohlrabe (2016), Bornmann and Wohlrabe (2019), García-Suaza et al. (2020) or Wohlrabe and Gralka (2020) among others.

- NBER (National Bureau of Economic Research) Working Papers
- CEPR (Centre for European Policy Research) Discussion Papers
- IZA (Institut für die Zukunft der Arbeit) Discussion Papers
- CESifo (Center for Economic Studies) Working Papers.

This choice is driven by the number of yearly published papers, reputation and influence. These four belong to the most cited[3] and downloaded[4] working paper series on RePEc. The four series are published by networks of economists. Table 1 states the approximate number of network members retrieved from their websites as of December 2019. The largest network is NBER with about 1600 members and the smallest CEPR with 1,300. Submitting papers to the series is only allowed to members of the corresponding networks and joining the network is only possible by invitation. Once an author is a member of a specific network they are free to submit any working paper. With this procedure the networks want to assure a specific level of quality of the submitted papers as invitations are only issued to established or promising researchers.

Many working papers series are associated with an organisation, faculty or a university often indicated by the name of the series. Thus, at least the submitter of a working paper should be affiliated with the issuing organisation. There are only a few series where everyone can submit a paper. A prominent example is the Munich Personal Archive (https://mpra.ub.uni-muenchen.de/).

We extracted meta-information for these four major working paper series in Summer 2019 for the time period 2000 to 2012. WPs published later than 2012 are not being considered as there can be a substantial delay in the process to the final journal publication. This can be attributed to three potential reasons. First, the economics

[3]https://ideas.repec.org/top/top.wpseries.simple.html
[4]https://ideas.repec.org/top/top.wpseries.download10.html

publishing process has slowed down (Ellison (2002)). Second, if a paper was accepted by a journal, it takes sometimes a long period before finally being published in an issue. And third, the provision of the meta data in RePEc is often delayed.

In total we collected information for 28,877 WPs (Table 1). The majority of WPs (∼11,000) were published by the NBER network. Although the CESifo network is older (founded in 1991) it has published fewer WPs (3,820) than the younger (established 1998) IZA network (7018 WPs). Figure 1 plots the annual number of WPs. For all years, the NBER publishes the most WPs with more than 1,000 in 2011 and CESifo the least with about 400 in 2012. There seems to be general trend in publishing more WPs in every year, especially in the early 2000s, with CEPR being an exception here. The institution seems to have established a plateau of about 500 WPs per year.[5]

RePEc links papers with corresponding title automatically. On the website it is stated "As long as two works have the same titles and are both listed in an author's profile, we will link them automatically. Just give us some time. However, if the titles differ, you can create the links yourself by using this online form". So, if RePEc is not able to link an article automatically a registered author can do this. We checked randomly the automated linkages for various working papers and found that RePEc does not only match exact recordings but also those with a similar title.

Every RePEc page for each WP was manually assessed and checked whether it was published either as journal article or a book chapter. In case it was published in a journal the corresponding meta data was collected. Additionally, we also checked whether a paper has been also published in other working paper series.

In order to evaluate the quality of journal publications we matched the corresponding annual Journal Impact Factors (JIF) published by Web of Science if available.

[5]CEPR has currently increased this number over 800 due to their website www.cepr.org.

Table 1: Summary statistics for the working paper series

	Members as of Dec. 2019	Number of Working Papers	Average per Year
NBER	1,600	11,208	862
CEPR	1,300	6,831	525
IZA	1,500	7,018	540
CESifo	1,500	3,820	294
Total		28,877	555

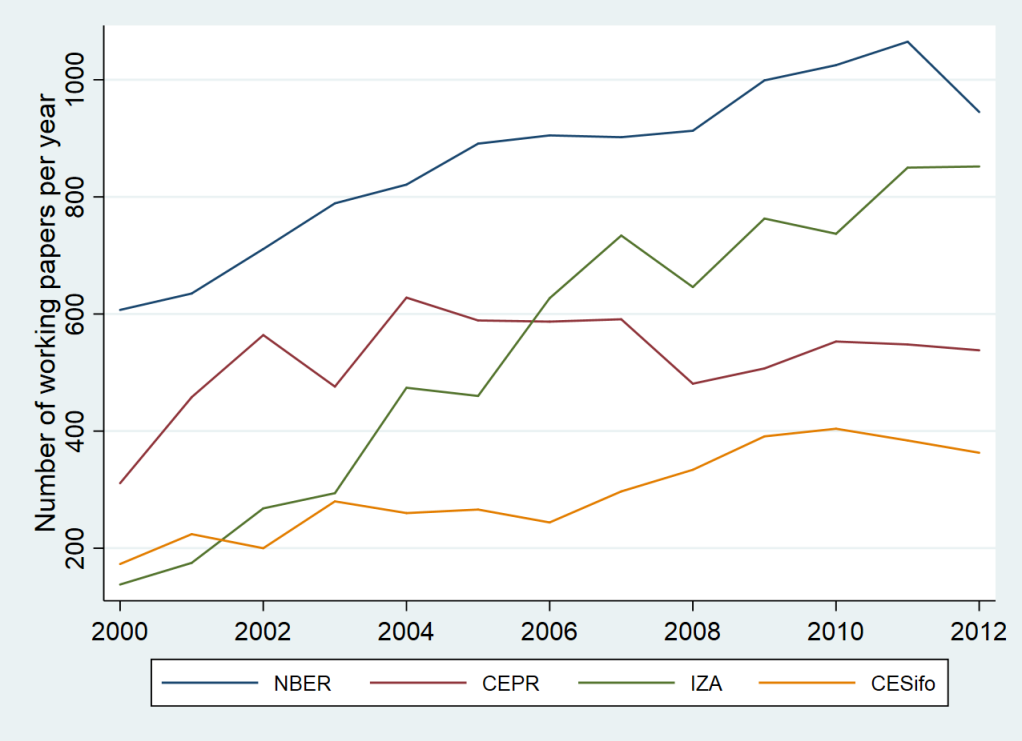
3 Results

3.1 Overlap of publications in working paper series

In how many series has a WP been published? We state the answer in Figure 2. In the left panel we plot a distribution across WP series.[6] The majority of WP is released exclusively in one series. This especially pronounced for the NBER series with about 60%. For the CEPR series an article is published either in one or two series to the same amount (31%). The right panel of Figure 2 shows the average number of series an article has been issued in. There is a clear upward tendency, i.e., papers are published in more and more series simultaneously. The papers released in the NBER series have the lowest rate of additional publications in other working paper series and the papers published in CEPR have the highest rate. In Table 2 we show the overlap between the series. For example, 17% of all CEPR discussion papers have been simultaneously published as a NBER WP, which is the highest overlap among all series. From the NBER perspective these are 10.5%. The smaller percentage share is due to higher number of NBER WPs compared to the CEPR series (see Table 1). The smallest overlap (<3%) is between the CESifo and the NBER working paper series.

[6] In order to increase readability we excluded some articles which have been published in more than six WP series.

Figure 1: Quantitative development of working papers over time

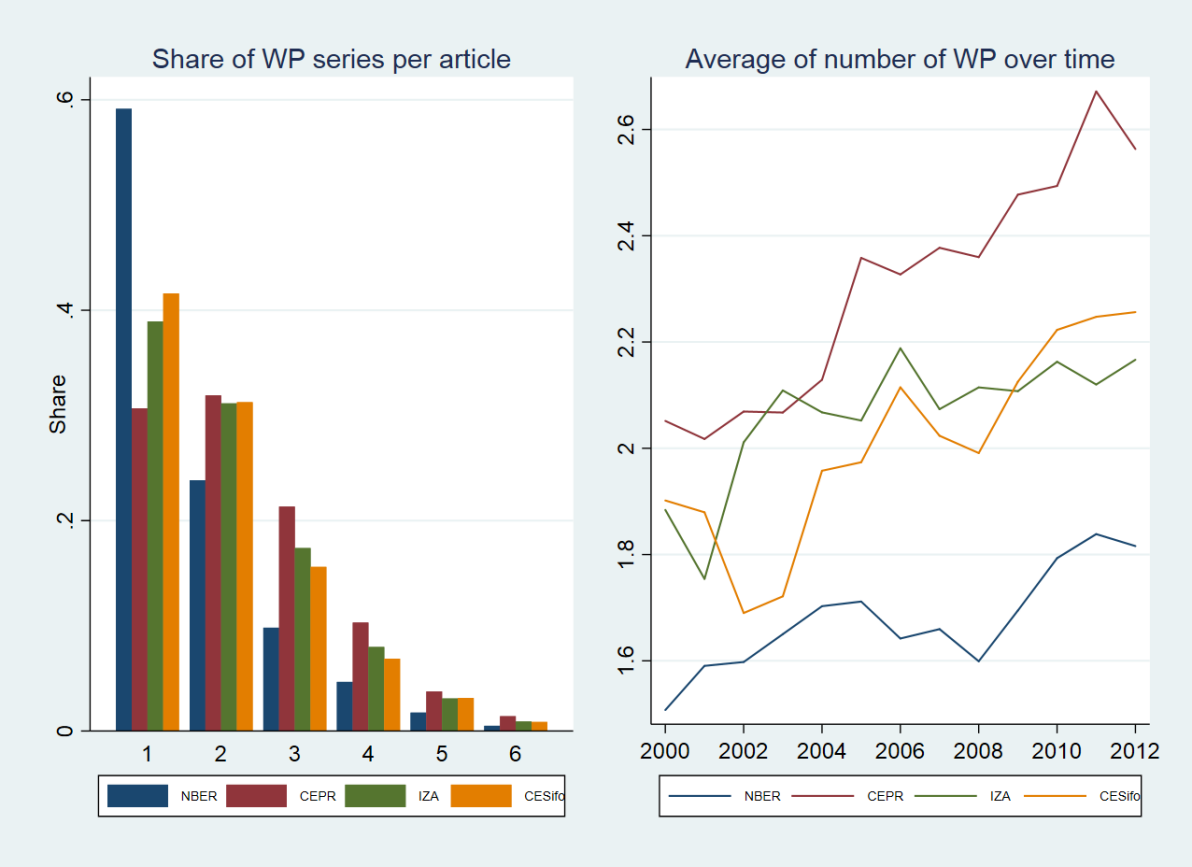


3.2 Where have the working papers gone?

3.2.1 Basic estimates

We start answering the question where the working papers have gone using matching provided on the RePEc website. The results are stated in Table 3. The share of WP that have been published as a book chapter ranges between 0.8% for CESifo and 7.5% for NBER papers. In total we have a share of about 4%. With respect to journal publications we find that almost 50% of all WPs in our sample have been issued in a scientific journal. This number varies only marginally across WP series. For the IZA series we find a slightly lower value with about 47%. The WPs were published in 622 different journals. The most articles (862) were finally published in the *American Economic Review*, followed by the *Journal of Public Economics* (456) and the *Review*

Figure 2: Number of Publications as Working Paper



of Economics and Statistics. More than 100 WPs in total were published in 38 journals. In contrast, for 189 journals we found only one corresponding WP.

For the moment, we do not find evidence where the remaining 46.5% of WPs have finally been published. Figure 3 plots the development over time. It shows that the publication rate in journals is quite stable and there seems to be no obvious trend.

How do our results compare to the existing literature? In Brown and Zimmermann (2017) 55% of the articles in the *Journal of Population Economics* were published as a WP. McCabe and Snyder (2015) detected the same number for articles published in 2005. For those before this date the share is somewhat lower. Larivière et al. (2014)

Table 2: Overlap between the four WP series in percent

	NBER	CEPR	IZA	CESifo
NBER		10.53%	3.98%	0.90%
CEPR	17.07%		10.95%	3.34%
IZA	6.54%	11.16%		5.19%
CESifo	2.59%	5.97%	9.37%	

Note: This table reports the relative share of papers in a WP series (row) which are also published in another series (column).

found that 64% of articles published on arXiv were finally issued in a journal list in Web of Science. In Tsunoda et al. (2019) this amounts to about 40%. To summarize, our results are similar.

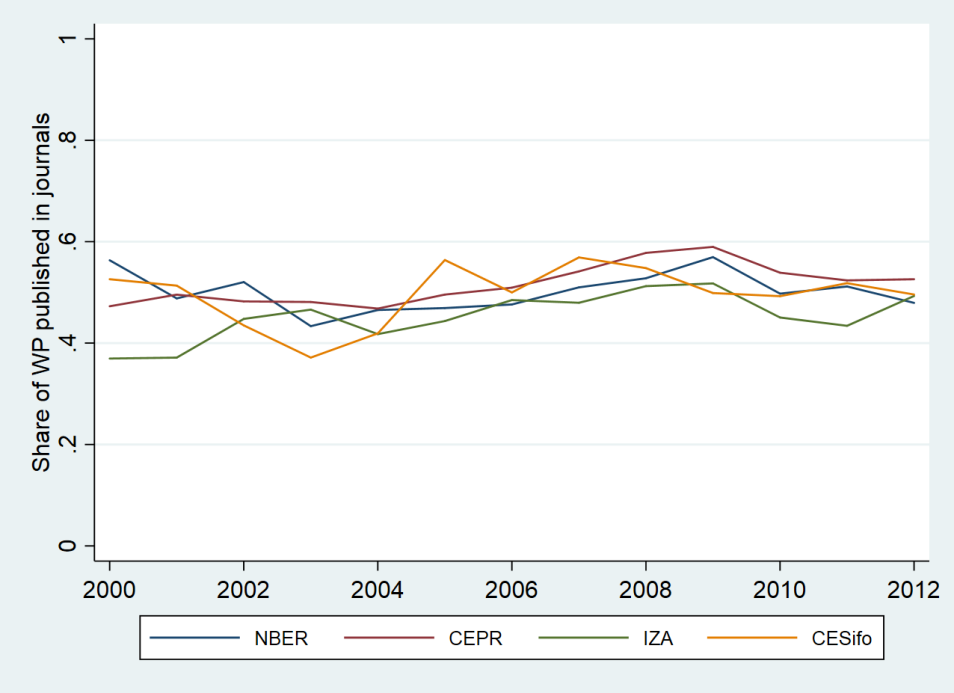
Table 3: Working Papers published in journals or as chapters across series

	WP	WP published as book chapters		WP published in journals		No evidence	
	N	N	relative	N	relative	N	relative
NBER	11,208	843	7.52%	5,614	50.09%	4,751	42.39%
CEPR	6,831	132	1.93%	3,525	51.60%	3,174	46.46%
IZA	7,018	114	1.62%	3,278	46.71%	3,626	51.67%
CESifo	3,820	30	0.79%	1,903	49.82%	1,887	49.40%
	28,877	1119	3.88%	14,320	49.59%	13,438	46.54%

3.2.2 Additional evidence based on a random sample

How robust or reliable are our results? Although our results are comparable to others in the literature the share seems rather moderate. We already mentioned that RePEc automatically matches working paper titles with journal article titles. Furthermore, authors can link working papers to journal publications via their author account in RePEc. However, there can be several reasons why our figures underestimate the true value. First, and in our opinion the most important one, the title might have been changed during the revision process. Second, it has been published in a journal or book

Figure 3: Publication of WP in journals over time



chapter that is not listed in RePEc. This might apply especially to journals outside economics or statistics. In order to investigate these issues, we draw a random sample from the non-matched articles, specifically 100 per WP series. For each of these 400 articles we searched the authors webpages and CVs (if available), looking for papers with similar titles or themes. We also read abstracts in order to identify a link between a working paper and a corresponding journal article.

Table 4 shows the results of our efforts. In our random sample we were able to match 36% of the working papers to a journal article. Across series the share is somewhat similar. Looking at the journals we find that almost all articles were published in (economics) journals that are also listed in RePEc and that many article titles (substantially) changed. Therefore, RePEc was not able to match them automatically. We also find that 9% were published as a book chapter. Here the NBER working paper series stands out, because 21% of the 100 investigated papers were published in a book.

In case of book chapters our investigations show that often the corresponding books are not listed in RePEc. For a total of 55% in our random sample no evidence whether it was finally published in a journal or book was found.

Combining the results from the RePEc matching (Table 3) and the random sample with individual matching (4) we provide an estimate both for journal publications and book chapters, for the former arriving at a value of 66.46% (49.56%+36%·46.53%) and 7.83% (3.88%+9%·46.53%) for the latter. It follows that for approximately 25% we find no record or evidence of an additional publication outlet besides the original WP.

Table 4: Further matching evidence based on a random sample

	N	Journal		Chapter		No Reference	
		N	relative	N	relative	N	relative
NBER	100	39	39%	21	21%	40	40%
CEPR	100	30	30%	2	2%	68	68%
IZA	100	40	40%	3	3%	57	57%
CESifo	100	36	36%	8	8%	56	56%
Total	400	145	36%	34	9%	221	55%

3.2.3 Adjusting for quality

In order to draw a proper picture not just the quantity but as well the quality of the journal publications has to be taken into account. Therefore, the JIF impact factor in the specific year is matched to each article, if available. Finally, we average over these JIFs. The corresponding time series can be seen in Figure 4. WPs published by the NBER are usually issued in the best journals according to JIF. CEPR is ranked second. The IZA and CESifo WP series are evenly off. This ranking follows the RePEc IF ranking for the WP series. Thus, there seems to be a relationship between the reputation of a WP series and the publication in prestigious journals. This claim is supported by looking at publications in the so-called top-5 journals in economics (Card

and DellaVigna (2013)). About 10% of NBER papers are issued in these elite journals. This share drops to 1.7% for the CESifo series.

Figure 4: Quality-weighted journal output

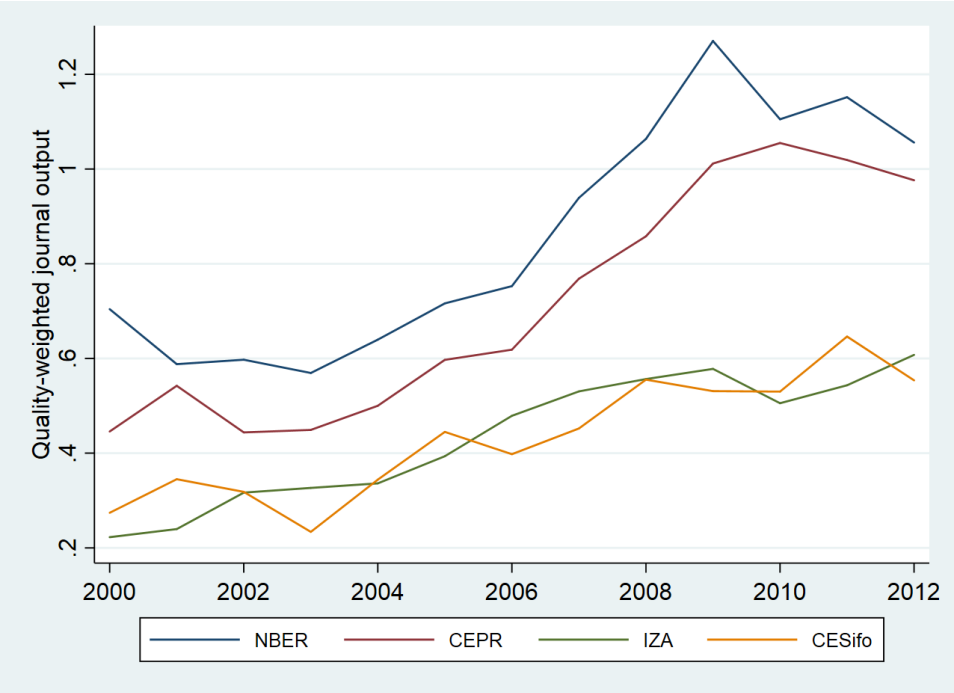


Table 5: Publications in top-5 journals

	NBER		CEPR		IZA		CESifo	
	N	relative	N	relative	N	relative	N	relative
American Economic Review	563	5.02%	191	2.80%	82	1.17%	26	0.68%
Econometrica	100	0.89%	41	0.60%	20	0.28%	9	0.24%
Journal of Political Economy	167	1.49%	44	0.64%	14	0.20%	7	0.18%
Quarterly Journal of Economics	253	2.26%	67	0.98%	23	0.33%	6	0.16%
Review of Economic Studies	108	0.96%	86	1.26%	28	0.40%	17	0.45%
Sum	1191	10.63%	429	6.28%	167	2.38%	65	1.70%

3.3 Time period up to journal publication

As already mentioned before, it takes time until a WP is released as an article. We recorded for every WP that has been published in a journal the time elapsed between

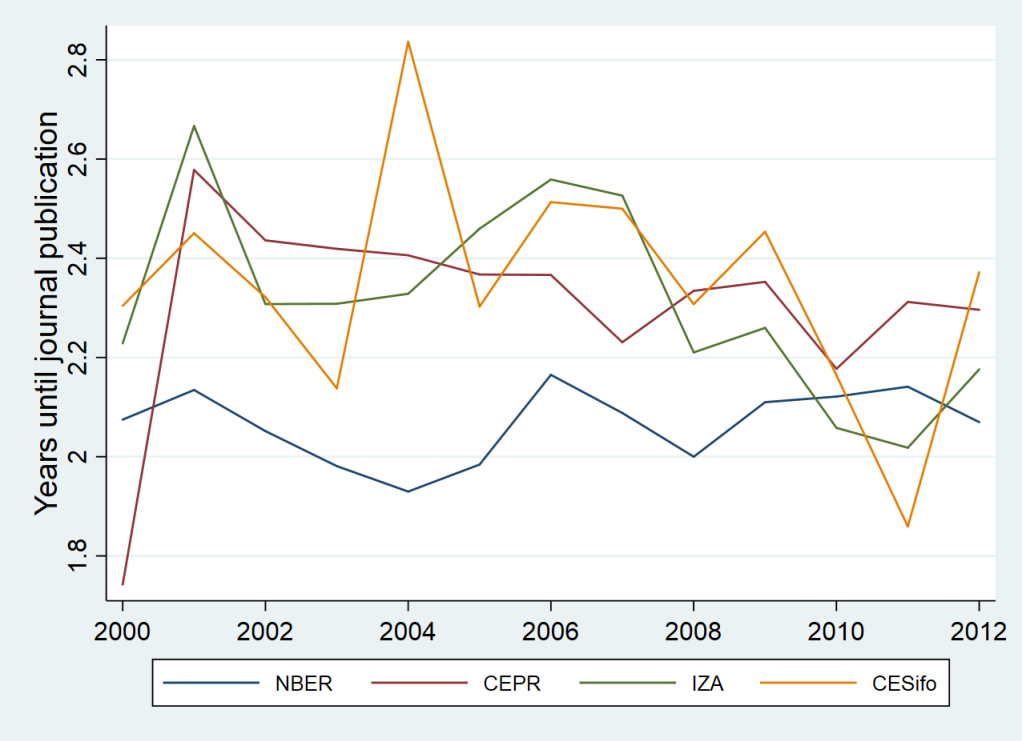
the publication years. In Table 6 the corresponding periods in years are reported. The negative numbers are WP that have been published *after* the journal release. The corresponding share of WP is very small (<1%). About 15% of WP are released in the same year as the journal article. Almost half of the journal articles are either published one or two years after the respective WP. For almost 20% the publication process took four years or more. The average time to publication is 2.2 years. Comparing the different series we find quite similar numbers with the tendency that the waiting period for NBER papers is somewhat lower (2.1 years). The longest period is found for the CESifo and CEPR series with 2.3 years. In Figure 5 we plot the corresponding development over time. There seems to be no obvious time trend across series.

There are several issues that influence the reported delay in Table 6 and Figure 6 which makes the interpretation difficult and it should be handled with caution. First, every author has different strategies when to publish an article as a WP. Some do it just after the paper is finished and before the formal submission to a journal. Extra time can be added when the paper is presented at various conferences and seminars. Other authors wait until after the journal revision or the article has finally being accepted. Additionally, there are different handling times at journals from the submission to the final acceptance. This time span can depend on the length of the article, its difficulty, number of referees, speed of referees, handling times of the editor, or the time for doing the revision by the authors. This list might be expanded and surely varies across authors and papers. The time between acceptance and publication in an issue also varies across journals. Finally, one has also keep in mind that the time between the WP and the journal release can be artificially over- or understated as it is based on annual figures. Suppose a WP is released in December in year t. The time to publication in a journal in $t+1$ is overstated compared to WP that was published in January in t.

Table 6: How long does it take a WP to be published in a journal?

Years	All WP		NBER		CEPR		IZA		CESifo	
	N	relative	N	relative	N	relative	N	relative	N	relative
-5	1	0.0%	0	0.0%	1	0.0%	0	0.0%	0	0.0%
-2	6	0.0%	4	0.1%	1	0.0%	0	0.0%	1	0.1%
-1	67	0.5%	35	0.6%	19	0.5%	10	0.3%	3	0.2%
0	2,083	14.5%	956	17.0%	503	14.3%	408	12.4%	216	11.4%
1	3,618	25.3%	1420	25.3%	860	24.4%	859	26.2%	479	25.2%
2	3,295	23.0%	1312	23.4%	736	20.9%	767	23.4%	480	25.2%
3	2,424	16.9%	898	16.0%	601	17.0%	595	18.2%	330	17.3%
4	1,431	10.0%	503	9.0%	411	11.7%	320	9.8%	197	10.4%
5	698	4.9%	248	4.4%	178	5.0%	168	5.1%	104	5.5%
6	354	2.5%	116	2.1%	104	3.0%	87	2.7%	47	2.5%
7	174	1.2%	52	0.9%	61	1.7%	36	1.1%	25	1.3%
8	92	0.6%	37	0.7%	27	0.8%	16	0.5%	12	0.6%
9	27	0.2%	9	0.2%	9	0.3%	6	0.2%	3	0.2%
10	23	0.2%	12	0.2%	4	0.1%	5	0.2%	2	0.1%
11	11	0.1%	6	0.1%	4	0.1%	1	0.0%	0	0.0%
12	8	0.1%	3	0.1%	3	0.1%	0	0.0%	2	0.1%
13	2	0.0%	1	0.0%	0	0.0%	0	0.0%	1	0.1%
14	3	0.0%	1	0.0%	2	0.1%	0	0.0%	0	0.0%
15	2	0.0%	1	0.0%	1	0.0%	0	0.0%	0	0.0%
16	1	0.0%	0	0.0%	0	0.0%	0	0.0%	1	0.1%

Figure 5: Average time to publication in journals



4 Conclusion

This article analyzed whether, when and where a working paper has been published either in a journal or as a book chapter. Based on RePEc matching and a random sample we found that approximately 66% of about 28,000 investigated working papers where released in a journal. Additionally, about 9% were issued as a book chapter. We have no record of what happened to the remaining 25% of WPs. Some caveats of our analysis should be mentioned:

1. The title of a WP could have changed and finally been published under a different name in a journal.

2. A WP might be published in a journal not covered by RePEc.

3. A WP has not been connected to the journal article in RePEc.

4. A WP has been completely revised, both with respect to the title as well the content, and therefore it was not possible to match it.

We are not able to quantify how large (or small) these effects are in our example. However, a certain share of WPs seem to be given up, i.e. has not passed formal peer review of a journal.

References

Abdill, R. J. and Blekhman, R. (2019). Meta-research: Tracking the popularity and outcomes of all biorxiv preprints. *eLIFE*, 8:e45133.

Anderson, K. R. (2020). biorxiv: Trends and analysis of five years of preprints. *Learned Publishing*, Forthcoming.

Bornmann, L. and Wohlrabe, K. (2019). Normalisation of citation impact in economics. *Scientometrics*, 120(2):841–884.

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

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

Ellison, G. (2002). The slowdown of the economics publishing process. *Journal of Political Economy*, 110(5):947–993.

García-Suaza, A., Otero, J., and Winkelmann, R. (2020). Predicting early career productivity of phd economists: Does advisor-match matter? *Scientometrics*, 122(1):429–449.

Larivière, V., Sugimoto, C. R., Macaluso, B., Milojević, S., Cronin, B., and Thelwall, M. (2014). arxiv e-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology*, 65(6):1157–1169.

Li, X., Thelwall, M., and Kousha, K. (2015). The role of arxiv, repec, ssrn and pmc in formal scholarly communication. *Aslib journal of information management*, 67(6):614–635.

McCabe, M. J. and Snyder, C. M. (2015). Does online availability increase citations? theory and evidence from a panel of economics and business journals. *Review of Economics and Statistics*, 97(1):144–165.

Rath, K. and Wohlrabe, K. (2016). Recent trends in co-authorship in economics: Evidence from RePEc. *Applied Economics Letters*, 23(12):897–902.

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.

Seiler, C. and Wohlrabe, K. (2013). Archetypal scientists. *Journal of Informetrics*, 7(2):345–356.

Tsunoda, H., Sun, Y., Nishizawa, M., Liu, X., and Amano, K. (2019). An analysis of published journals for papers posted on biorxiv. *Proceedings of the Association for Information Science and Technology*, 56(1):783–784.

Wohlrabe, K. and Gralka, S. (2020). Using archetypoid analysis to classify institutions and faculties of economics. *Scientometrics*, page Forthcoming.

Zimmermann, C. (2013). Academic rankings with repec. *Econometrics*, 1(3):249–280.