Is open science a double-edged sword?: data sharing and the changing citation pattern of Chinese economics articles

Liwei Zhang[1] · Liang Ma[2]

© Akadémiai Kiadó, Budapest, Hungary 2023

Abstract

Data sharing is an important part of open science (OS), and more and more institutions and journals have been enforcing open data (OD) policies. OD is advocated to help increase academic influences and promote scientific discovery and development, but such a proposition has not been elaborated on well. This study explores the nuanced effects of the OD policies on the citation pattern of articles by using the case of Chinese economics journals. *China Industrial Economics* (CIE) is the first and only Chinese social science journal so far to adopt a compulsory OD policy, requiring all published articles to share original data and processing codes. We use the article-level data and difference-in-differences (DID) approach to compare the citation performance of articles published in CIE and 36 comparable journals. Firstly, we find that the OD policy quickly increased the number of citations, and each article on average received 0.25, 1.19, 0.86, and 0.44 more citations in the first four years after publication respectively. Furthermore, we also found that the citation benefit of the OD policy rapidly decreased over time, and even became negative in the fifth year after publication. In conclusion, this changing citation pattern suggests that an OD policy can be double edged sword, which can quickly increase citation performance but simultaneously accelerate the aging of articles.

✉ Liang Ma
  liangma@ruc.edu.cn

  Liwei Zhang
  liwei_zhang@sdu.edu.cn

1 School of Innovation and Entrepreneurship, Shandong University, 72 Binhai Road, Jimo District, Qingdao, Shandong Province 266237, China

2 School of Public Administration and Policy, Renmin University of China, 59 Zhongguancun Avenue, Haidian District, Beijing 100872, China

Introduction

At present, scientific research and publishing is facing new challenges and opportunities from open science (OS), which mainly includes open access to publications, open data for scientists, and open peer review for manuscripts. OS has been developing rapidly since the twenty-first century. Taking open access (OA) as example, OA journals indexed in Directory of Open Access Journals (DOAJ) have been witnessed an exponential growth trend since the launch of the Budapest Open Access Initiative (2002) (Pandita & Singh, 2022). There was a 20-fold increase of journals adopting hybrid OA from 2007 to 2013 by the five largest publishers (Elsevier, Sage, Springer, Taylor & Francis, and Wiley) (Laakso & Björk, 2016). Consequently, OA has progressively become a key issue for policy makers, and more and more countries, funding agencies, and institutions have formulated policies to promote its development (Maddi et al., 2021; Zhang et al., 2022).

Meanwhile, as a crucial part of OS, more and more scientists, research societies, universities, funders have also taken attention to open data (OD), and more and more journals and their publishers have been introducing OD policies. Given the increasing coverage and influences, OD also become a hot topic. Similarly, as OA is associated with the increased academic influences of articles (Vincent-Lamarre et al., 2016) which usually refer to "OA citation advantage" (OACA) (Lawrence, 2001), scientists also focus on the academic influences brought by OD, particularly the increase of citations. However, the citation advantage is still a controversial proposition with inconclusive findings, and few studies have examined the varying effects of OD on the citation performance of scientific articles over time. Given the salience of OD for scientific research, it is relevant to explore its nuanced effects on the citation performance and patterns of articles. Therefore, in this study we explore the changing effects of the OD policy over time to fill the gap in the literature.

We specifically use the case of Chinese economics journals to explore the citation advantage of OD articles. *China Industrial Economics* (CIE) is the first and only Chinese social science journal so far to adopt a compulsory OD policy, requiring all published articles to share original data and processing codes. We use the article-level data and difference-in-differences (DID) approach to compare the citation performance of articles published in CIE and 36 comparable journals. We find that the OD policy quickly increased the number of citations in the first four years after publication, but the citation benefit of the OD policy rapidly decreased over time, and even became negative in the fifth year after publication. Our findings suggest that the OD policy is a double-edged sword, which can quickly increase citation performance but simultaneously accelerate the aging of articles. We report one of the first empirical studies to document the changing citation pattern of OD articles, and contribute to the ongoing literature on OD and OS.

In the rest of this article, we first review the prior OS literature and develop theoretical hypotheses. We then introduce the data and methods used in this study, and report the empirical findings. We lastly discuss the theoretical and policy implications of our findings.

Literature review and theoretical hypotheses

Enforcement of OD

With the development of OS, OD policies have been increasingly implemented globally, for which funding agencies, research institutions, journals, and publishers require scientists

to share their research data. To promote the movement of data sharing, pioneering societies and organizations have taken steps starting about a decade ago. For example, scientists seeking National Science Foundation (NSF, 2010) funding were required to make primary data available to others within a reasonable length of time. The National Institutes of Health (NIH, 2003)has required a data-sharing plan for all large funding grants since 2003. The Royal Society (2012) released the "science as an open enterprise" report aiming to identify the principles, opportunities and problems of sharing and disclosing scientific information. Yale University launched the "Yale Open Data Access Project (YODA)", which is a pioneering data-sharing mechanism (Krumholz et al., 2013).

For journals which lie at the heart of the shift towards OS, the past decade has witnessed an increase in the number of scientific journals adopting OD policies (Hrynaszkiewicz et al., 2020; Jones et al., 2019). For example, aiming to address the issues of complexity and lack of clarity for authors across the data policy landscape, Springer-Nature (2021) began to roll out standard research data policies in 2016, and launched a new service called "Research Data Support" for its authors. The Public Library of Science (PLoS) and its journals developed a 'free the data' policy, requesting all data underlying the findings of articles to be fully available (Bloom et al., 2014). Meanwhile, Molecular Biology and Evolution (2023), PNAS (2023), Science (2023) among other journals, also developed a series of data availability policies.

What is more, the activities of data sharing more focused on the field of natural sciences (Enwald et al., 2022; McCain, 1995), especially in medical, pandemic, and clinical research (Reardon, 2014; Ross et al., 2009; Yancy et al., 2018), and biological, genomic and ecological research (Cambon-Thomsen et al., 2011; Kauffmann & Cambon-Thomsen, 2008; Powers & Hampton, 2019; Reichman et al., 2011; Whitlock, 2011). As a consequence, the willingness to share research data as OA data was higher in the natural sciences (Ohmann et al., 2021). In contrast, there are no well-developed traditions of OD and OA in the humanities and social sciences (Enwald et al., 2022), although there have already been emerging practices in psychology (Martone, 2018)and politics (Hartzell, 2015). Even in these fields, data sharing is still in its infancy.

Benefits and challenges of OD

As a consequence of the emerging enforcement of OD policies, OD has attracted extensive attention and debates in the scientific community. Firstly, some scholars argue that the main purpose of OD is to promote the dissemination and verification of scientific research findings, and make data available for reuse and research reproducible to further develop scientific research and discovery (Borgman, 2012; Cousijn et al., 2022; Zuiderwijk et al., 2020). Specifically, the benefits generated by OD include but not limite to: enabling the analysis of large volume of data, maximizing the contribution of research subjects, enabling new questions and testing novel hypotheses, enhancing reproducibility, providing test beds for new analytical methods, protecting valuable scientific resources, reducing the cost of doing science and avoiding duplication of effort (Campbell, 2015; Molloy, 2011; Rouder, 2016). For example, the global COVID-19 pandemic makes the actual sharing of raw data critical to global recovery (Strcic et al., 2022; Zastrow, 2020), which motivated the scientific community to work together in order to gather, organize, process, and distribute data on the novel biomedical hazard (Homolak et al., 2020).

Furthermore, by publishing their findings with maximum transparency, OD can elicit great attention to global scientific collaboration, bring new opportunities to research, and

eventually increase the efficiency of the global innovation system (Gold, 2021). In addition, OD not only contributes to scientific development, but also is beneficial for the public, and the whole of society because research data have become a common resource from which business, government and the whole of society can extract value, such as open data systems which can enable relevant research and inform evidence-based governmental decisions (Sá & Grieco, 2016).

Secondly, OD strengthens research transparency and helps eliminate scientific misconducts. Because a lack of transparency in science could erode trust in science, similar to the case of used car markets (Vazire, 2017), many scholars argue that OD encourages reproducibility of research through transparency and helps to promote scientific integrity (Banks et al., 2019; Kretser et al., 2017, 2019; Kwasnicka et al., 2020; Pencina et al., 2016; Ross et al., 2018). As gatekeepers of publication, scientific journals have the responsibility to make data publicly and permanently available (Hanson et al., 2011). For example, a few major journals of biomedical sciences jointly released the principles and guidelines of data availability (as deposing the data in public repositories) for research reproducibility, rigor, and transparency (McNutt, 2014). Therefore, in terms of the value of accessibility, research scrutiny, and data transparency, OD has important ethical value that all scientists should aspire to achieve (Inkpen et al., 2021).

Thirdly, scientists and journals pay close attention to whether and how OD helps increase articles' impact, specifically the number of citations. Regarding the citation impact of data sharing, some scholars have pointed out that data availability improves the published results which are reproducible by independent scientists (Ioannidis et al., 2009); therefore, sharing research data may increase citation rates relatively (Piwowar & Chapman, 2010; Piwowar et al., 2007, 2008). In addition, Zhang & Ma (2021) suggested that journals can leverage compulsory OD to develop reputation and amplify academic impact.

Although it is widely believed that OD will greatly stimulate the development of science, researchers still worry about the crises brought about by the "openness" of research data, and there can be tricky and mountainous challenges to embrace the move to make scientific findings transparent (Gewin, 2016; Van Noorden, 2014). A survey by 1,800 UK-based academics revealed that while most academics recognized the importance of sharing research data, most of them had never shared or reused research data (Zhu, 2020). The reason is that scientists are worried that OD will create some risks, such as being afraid that other researchers would find errors in their results, or "scoop" additional analyses that they have planned for the future (Piwowar & Vision, 2013). As said by David Hogg, an astronomer of New York University, "everybody has a scary story about someone getting scooped" (Van Noorden, 2014). In addition, Zipper et al. (2019)also pointed that researchers could unintentionally violate the privacy and security of individuals or communities by sharing sensitive information.

The hypotheses

As mentioned above, the research data shared by authors could be independently replicated and reanalyzed by other researchers, which might generate lasting effects through methodological training and teaching. In addition, numerous studies have documented the citation advantage of OA, which is usually concerned as the core issue of OS. But the conclusions are inconsistent: some researchers revealed that OA could increase citations and usage (Eysenbach, 2006; Lin, 2021; Piwowar et al., 2007; Wang et al., 2015; Zhang et al., 2021), but others reject this assertion (Craig et al., 2007; Davis et al., 2008; Norris et al,

2008). Given the above discussions, we expect the citation performance of articles with OD associated with them (OD articles) to be higher than that of articles that do not have OD associated with them (non-OD articles). Therefore, we propose the first hypothesis as follows:

H_1 The academic influence (i.e., the number of citations) of the articles published in journals enforcing the OD policy is higher than that of other journals without OD policies.

Furthermore, there is no study exploring the changing pattern of citations after OD policy implementation. In other words, it is unknown whether the citation effects of the OD policy would change over time. Cited half-life (CHL) is an important indicator of aging for scientific articles, which is the median age of citations of an article received from other articles, and is the watershed between the two life periods of an article. A higher value of CHL means that an article receives attention for a longer period of time after its publication. In this study, CHL provides a fruitful indicator to examine if OD could delay or accelerate articles' aging of citation, and the changing dynamics of citation patterns.

It is pointed out that the CHL of Chinese economics articles is about 3–4 years (Wang, 2005; Zheng et al., 2014), which means that half of the citations are received from its publication up to the 3–4 years after that, and the other half are earned after 3–4 years from an article's publication date. The timeliness of data shared by authors matters in an article's citation potential, and scientists favor timely data in research. In other words, the citation advantage of OD articles declines together with the obsoleting of data. We expect that the citation performance of OD articles would be higher than that of non-OD articles in the early period of time, but such a citation benefit would decline over time. Based on the above arguments, we propose the second hypothesis as follows:

H_2 The OD policy could change the citing life of articles, and specifically it will accelerate their aging over time.

Data and methods

Data

In China, the movement of OS is still in its infant stage. A survey of OS policies enforced by high-impact Chinese journals (International Research Center for the Evaluation of Chinese Academic Documents & Tsinghua University Library, 2020a, 2020b) found that only a few journals had adopted OS policies until 2020 (Table 1). With the coming of the OS era, Chinese researchers, government, research institutions, and universities are strongly concerned about the influences of OS in China. For example, the National Natural Science Foundation of China (NSFC) conducted a survey of its researchers' attitudes towards OS in April 2020.

As shown in Table 1, CIE, a Chinese journal managed by the Institute of Industrial Economics, the Chinese Academy of Social Sciences, mandated its authors to share their research data (e.g., original data, processing code, processed data, and supplemental materials) in November 2016. CIE's pilot created a quasi-natural experiment for us to explore the impact of OS on journal publications, which helps to address endogeneity

Table 1 List of Chinese journals enforcing OS policies

Discipline	Journal	Mandatory	Moderated	None
Natural Sciences (Totally 133)	The most influential journals	0	2[a]	30
	The outstanding journals	1[b]	2[c]	98
Social Sciences (Totally 146)	The most influential journals	1[d]	3[e]	67
	The outstanding journals	0	0	75

[a]Acta Physica Sinica, Environmental Science

[b]Quaternary Sciences

[c]Natural Gas Industry, Biodiversity Science

[d]China Industrial Economics (CIE)

[e]Management World, Chinese Journal of Sociology, Economic Management Journal

concerns in causal effect estimates (Gaulé & Maystre, 2011). Therefore, we utilized the articles published in CIE as the research sample.

We collected the data from the Chinese Social Sciences Citation Index (CSSCI) database, the Chinese counterpart of Social Sciences Citation Index (SSCI) developed by the Institute of Chinese Social Sciences Research and Assessment, Nanjing University. It covers more than 500 Chinese journals from 25 disciplines, such as management, economics, law, history, and politics. The journals collected in CSSCI are usually the leading journals in each discipline, and we chose all 37 indexed economics journals.

The self-archiving policy might be a confounding effect, and it is important to consider whether journals have adopted self-archiving. As shown in Table 1, the most influential and outstanding Chinese journals of social sciences have not adopted OS policies (totally 146 journals), not to mention self-archiving policies. Meanwhile, we searched all the journals included in CSSCI and did not find any journals other than CIE adopting self-archiving policies. In addition, we also checked the top Chinese universities in social sciences and did not discover any of them adopting self-archiving platforms. So, it follows that self-archiving has not been adopted in Chinese social sciences journals, and its confounding effect would not be a big concern.

Authors might deposit and share their research data and codes through personal websites, social media (e.g., ResearchGate), and open data repositories (e.g., Harvard Dataverse), and these sporadic OD practices might also confound our model estimates. It would beyond the scope of this study to explore whether and how individual authors share their data voluntarily and personally, and such practices would further corroborate our findings because we reasonably assume that the articles published in no-CIE economics journals do not share data, and our estimates would be attenuated if they did so. We thus believe that personal OD practices would not significantly change our core results.

Because CIE initiated the OD policy in November 2016, we used articles published from 2014 to 2021 as the sample. The articles published in CIE are the treatment group, while those of the other 36 journals are the control group. The years from 2014 to 2016 are the pre-treatment period, while the years from 2017 to 2021 are the post-treatment period.

A total of 31,101 articles were collected (articles, reviews, comments, research reports are included, and editorials, meeting reports, and leader speeches are excluded),

and each article included the data about author(s), journal, year of publication, key words, funder(s), page numbers, and references. The number of citations for each article was counted, and the mean number of citations of the 37 journals were calculated.

As shown in Fig. 1, the articles published on *Journal of Financial Research*, *Economic Research Journal*, *Word Economy Studies* and CIE have received the largest numbers of citations.

Model specifications and variables

We used the DID model to address the problem of endogeneity in estimating the effect of the OD policy. The DID model is usually used to construct differential statistics reflecting policy effects, and it assumes that the treatment group and the control group meet the same trend of variation. In the DID model, the difference of the dependent variable between the pre- and post-treatment periods in each group is first estimated, and then the differences are compared between the treatment group and the control group, meaning the difference in the differences. By doing so, the preexisting differences are eliminated and the resulting estimate is the pure treatment effect.

As mentioned above, as the half-life of Chinese economics journals is about 3–4 years (Wang, 2005; Zheng et al., 2014), we used the count of cited times of each article in the 1st year to 5th year after publication as the dependent variables, respectively. We created a *Treat* variable to designate the articles published on CIE as the treatment group (coded 1), while that of the other journals the control group (coded 0). We created a *Time* variable to indicate the period pre or post the OD policy, and it was coded 1 for articles published after November 2016 and 0 otherwise. In the DID model estimates, the interaction term *Treat*Time* is the pure effect of the data-sharing policy. The DID model was developed as follows:

$$Citations_t = \beta_0 + \beta_1 Treat + \beta_2 Time + \beta_3 Treat \times Time + \beta_4 Controls_t + \varepsilon_t, t = 1, 2, 3, 4.$$

We followed prior studies (Wang & Shapira, 2015) to include conventional control variables at the article level that might affect the number of citations, including the number of authors, keywords, references, and pages of each article, as well as whether the article was sponsored by national funding agencies (i.e., the NSFC, the National Social Science
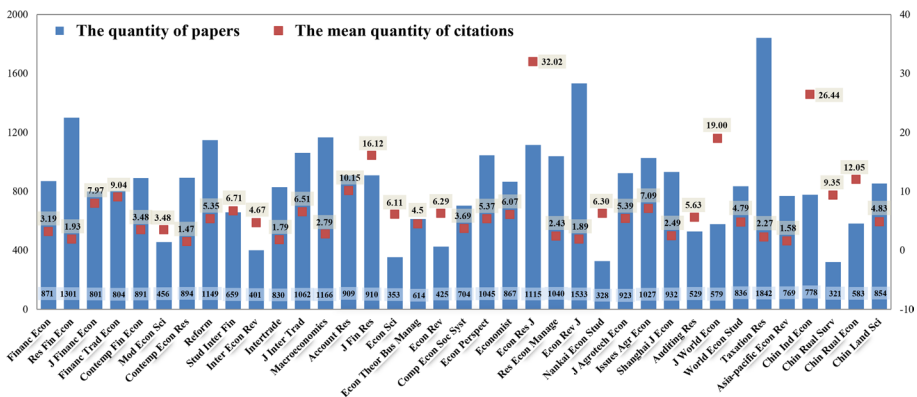

Fig. 1 The number of articles and the average citations of each article among the 37 journals (up to 2021)

Foundation of China (NSSFC), and the Ministry of Education (MoE)) and whether the article's first affiliated institution was a top university covered by the "Double World-Class" plan initiated by the MoE (Christensen & Ma, 2022). A recent study shows that articles listed first in each issue are about 30% more likely to be viewed, downloaded, and subsequently cited (Feenberg et al., 2017). We included a dummy variable to indicate whether an article was listed as the first article in each issue.

We also included two journal-level variables in the models. First, we controlled for the age of each journal, because journal age may be correlated with citation and journal reputation (Cahn & Glass, 2018; Sombatsompop et al., 2002). Second, we also included the compound impact factor (CIF, similar to IFs of SSCI-indexed journals) of each journal to control for its influence on citations (Tahamtan et al., 2016). In addition, as all 37 journals are either monthly or bi-monthly publications, we controlled for time fixed effects by including the variable of "i.Bi-month", which is a time series split by every two months. The descriptions of the dependent, independent, and control variables are listed in Table 2.

Results

The descriptive statistics

The descriptive statistics of the key variables are listed in Table 3. The count of cited times of each article vary substantially, but on average its scientific impact is weak. On average

Table 2 The descriptions of key variables

Variable	Description
Dependent variable	
1st year	The count of cited times in the 1st year of each article after publication
2nd year	The count of cited times in the 2nd year of each article after publication
3rd year	The count of cited times in the 3rd year of each article after publication
4th year	The count of cited times in the 4th year of each article after publication
5th year	The count of cited times in the 5th year of each article after publication
Independent variable	
Treat*Time	The interaction term between Treat and Time
Treat	The articles published on CIE are the treatment group (coded 1), while that of the other journals are the control group (coded 0)
Time	The period before the OD policy is coded 1, and otherwise 0
Control variable	
#Author	Number of authors of each article
#Keyword	Number of keywords of each article
#Reference	Number of references of each article
#Page	Number of pages of each article
#First	1 if the article is the leading article of each issue; 0 if not
#University	1 if the article's first affiliated institution is a top university; 0 if not
#Fund	1 if the article was funded by the NSFC, NSSFC, or MoE; 0 if not
#CIF	Compound impact factor of the article's published journal
#Age	Age of the article's published journal

Table 3 The descriptive statistics of key variables

Variable	Type	N	Mean	SD	Min	Max
1st year	Count	31,101	0.1724	0.6082	0	24
2nd year	Count	31,101	1.1803	2.2262	0	68
3rd year	Count	31,101	1.6256	3.2952	0	106
4th year	Count	26,498	1.5427	3.5136	0	109
5th year	Count	21,533	1.3813	3.5644	0	178
#Author	Count	31,101	2.0981	0.9794	0	22
#Keyword	Count	31,101	3.8964	0.8091	0	10
#Reference	Count	31,101	20.6741	14.9795	0	243
#Page	Count	31,101	10.1672	4.5354	1	57
#First	Dummy	31,101	0.0750	0.2635	0	1
#University	Dummy	31,101	0.4400	0.4964	0	1
#Fund	Dummy	31,101	0.2449	0.4300	0	1
#CIF	Interval	31,101	3.0208	2.0258	0.889	11.179
#Age	Interval	31,101	40.3445	10.5557	22	65

each article was written by 2.1 authors with 10.2 pages, including 3.9 keywords and 20.7 references. We first compared the cited times (citations) between CIE and other leading journals with highest values of CIF in the first, second, third, fourth, and fifth year after publication (Fig. 2). The articles published in CIE received more and more citations over time, substantially more than that of the other three journals since 2017. The mean number of citations of the CIE articles in the first year after publication has been higher than 1.0 since 2019, which is much more than that of the other three journals.

Regarding the mean number of citations in the second and third years after publication, the CIE articles also received more citations than those of the other three journals since 2017. However, the relative advantages are both less than that of the first year.

For the case of the fourth and fifth year after publication, there is no obvious difference of the mean number of citations between articles published on CIE and that of others. The
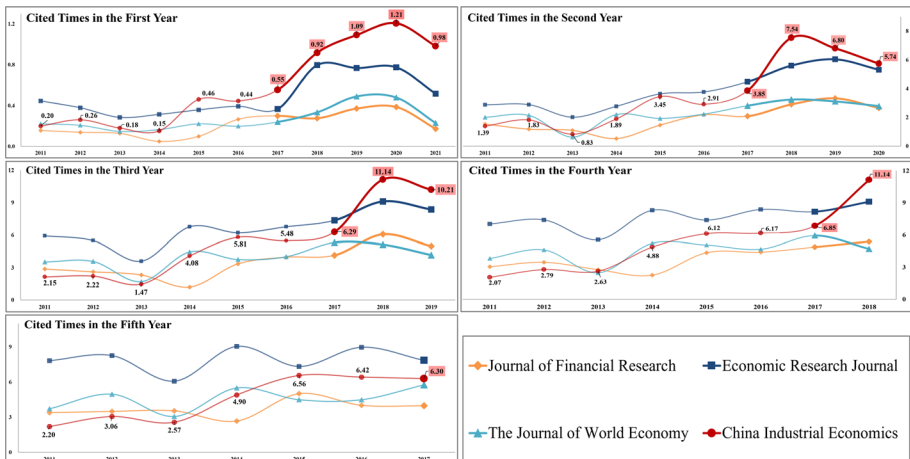

Fig. 2 The mean citations of articles published in the four top journals

data-sharing policy significantly increased the citation performance in first three years after the articles were published, but the increasing momentum rapidly decreased and finally disappeared over time.

The DID model estimates

To estimate the effect of the OD policy, we first estimated a model in which the articles published on CIE (including all 390 articles published after November 2016) formed the treatment group, while the articles published in the other 36 journals were the control group. The variable *Treat*Time* is positive and significant in the first-, second-, and third-year models, and the coefficients are 0.249, 1.071, and 0.497 respectively (Table 4).

The results reveal that the OD policy has increased the number of times the articles were cited in the first, second, and third years after publication by 0.249, 1.071, and 0.497 respectively. However, the coefficients turn negative albeit nonsignificant ($\beta = -0.093$) in the fourth year after publication, and become even significantly negative ($\beta = -1.237$, $p < 0.01$) in the fifth year (Table 4).

The coefficients of *Treat* are all positive and significant at the 1% level, meaning that the OD impact has been increasing steadily. The coefficients of *Time* are all negative, suggesting that the citation advantage of OD takes time to build up.

The results show that the OD policy increased the citation performance of articles in the short term (usually in the first three years), but the effect drops rapidly over time, even turning into a negative impact in the fifth year after publication. In other words, the data-sharing policy has changed the articles' citing life and accelerated their aging over time. Both H1 and H2 are supported.

For the results of the control variables at the article- and journal-levels, we find that articles with a greater number of pages (#page), being the leading article of each issue (#first), or published in journals with higher values of compound impact factor (#CIF) have received more citations. In contrast, the effects of the number of authors (#author), the number of keywords (#keyword), the number of references (#reference), whether the first affiliated institution is a top university (#university), whether it is nationally funded (#fund), and the age of each journal (#age) are not significantly related to citation performance.

In the online appendix, we ran a series of robustness tests to confirm our findings reported in the main text. Since not all articles published on CIE released raw data (mainly case study articles), we first compared the articles published on CIE with and without data sharing. The findings by using alternative samples are substantially consistent with our main results (see Tables A1 and A2 in Online Appendix). The parallel trend assumption should be met in estimating DID models, and the results in Fig. A1 in Online Appendix support this. Lastly, Table A3 in Online Appendix shows that the model estimates combining Propensity Score Matching (PSM) and DID are also similar to our main findings. Taken together, our findings are robust and reliable.

Discussions

The OD policies enforced by scientific journals could increase the number of cited times of articles (Inkpen et al., 2021; Zhang & Ma, 2021), but it is unknown whether and how they could change the citing life and pattern. As one of the top economics journals in China, CIE was among the first Chinese journals (and the only social science journal) to mandate

Table 4 The DID model estimates of all CIE articles

Variables	1st year	2nd year	3rd year	4th year	5th year
Treat*Time	0.2486*** (0.0426)	1.0712*** (0.1497)	0.4971** (0.2161)	− 0.0933 (0.2539)	− 1.2366*** (0.3144)
Treat	0.1500*** (0.0299)	1.1220*** (0.1052)	2.4153*** (0.1519)	2.6643*** (0.1622)	2.9304*** (0.1678)
Time	− 0.3533*** (0.0298)	− 0.5310*** (0.1048)	− 0.3672** (0.1513)	− 0.3356** (0.1512)	− 0.5108*** (0.1525)
#Author	− 0.0160*** (0.0034)	− 0.0187 (0.0121)	0.0146 (0.0175)	0.0376* (0.0213)	0.0493** (0.0247)
#Keyword	− 0.0009 (0.0041)	0.0415*** (0.0145)	0.0388* (0.0210)	0.0202 (0.0244)	0.0043 (0.0283)
#Reference	0.0001 (0.0003)	0.0064*** (0.0010)	0.0113*** (0.0014)	0.0123*** (0.0017)	0.0134*** (0.0019)
#Page	0.0077*** (0.0010)	0.0519*** (0.0034)	0.0946*** (0.0049)	0.1076*** (0.0057)	0.1126*** (0.0065)
#First	0.1922*** (0.0124)	0.7095*** (0.0435)	0.5952*** (0.0628)	0.3402*** (0.0731)	0.0851 (0.0842)
#University	− 0.0139** (0.0067)	− 0.0143 (0.0235)	0.0259 (0.0340)	0.0737* (0.0393)	0.0766* (0.0450)
#Fund	− 0.0066 (0.0097)	− 0.0023 (0.0342)	0.0648 (0.0493)	0.1656*** (0.0527)	0.1304** (0.0547)
#Age	0.0001 (0.0004)	0.0048*** (0.0012)	0.0154*** (0.0018)	0.0227*** (0.0020)	0.0259*** (0.0023)
#CIF	0.0270*** (0.0021)	0.2755*** (0.0072)	0.5074*** (0.0104)	0.5874*** (0.0125)	0.5930*** (0.0149)
Bi-month	Yes	Yes	Yes	Yes	Yes
Constant	0.1774*** (0.0280)	− 0.6078*** (0.0985)	− 2.0392*** (0.1422)	− 2.5644*** (0.1589)	− 2.7582*** (0.1748)
N	31,101	31,101	31,101	26,498	21,533
Adjust R^2	0.1192	0.1868	0.2262	0.2288	0.2037

Standard errors are in parentheses

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

the OD policy, providing an ideal quasi-natural experiment for us to causally estimate the OD effects. We used the articles published in the 36 comparable Chinese economics journals as the control group, and DID model estimates revealed the lasting effects of the OD policy on citation performance after years of publication.

First, we found that the OD policy has significantly increased the citation performance of the articles in the first four years after publication, and a series of robustness checks confirmed the findings. This research suggested that the OD policy could significantly and positively affect the academic impact of articles in the short term, which is consistent with prior studies on this topic (Inkpen et al., 2021; Molloy, 2011; Piwowar & Vision, 2013; Sá & Grieco, 2016). Journals should consider enforcing the data-sharing policy to accelerate scientific communication and to improve the development of science, especially for the case of China which is troubled by academic fraud.

Second, we found that there was first an increasing and then decreasing trend, and the effect of the OD policy on citation performance seems to have an inverted U-shaped relationship over time. What is more, the results show that there is a negative effect on citation performance in the fifth year after articles are published. Although there is a significant positive effect in the first few years, the effect of the OD policy rapidly decreases as time goes by, and even changes to be negative. On the contrary, the effects of article- and journal-specific attributes such as the number of references and pages of each article (#reference and #page), whether the article was funded (#fund), and the age and compound impact factor of the journal the article was published (#age and #CIF), which represent articles' academic quality, become increasingly salient over time. This means that the citation performance becomes increasingly dependent on articles' inherent academic quality in the long term.

Third, the OD policy changes articles' citing life. The half-life of Chinese economics journals is about 3–4 years (Wang, 2005; Zheng et al., 2014), while the citation effect of the OD policy changes to negative in the fifth year. The OD policy accelerates the aging of articles, and authors and journals should pay more attention to articles' own academic quality, which defines whether articles could have lasting and continuous effects.

Our findings contribute to the ongoing debates about the academic influences of OD, and help clarify the nuanced effects of OD on the citing life and pattern of journal articles. Zhang & Ma (2021) used journal-level data and a synthetic control method to estimate the OD policy effect, and our study is a follow-up replication and extension of their findings. We used article-level data and the DID method with a longer period of coverage, and our study generated more convincing and nuanced findings than prior studies.

The results reveal that OD is a two-edged sword, and its citation advantage declines and even turns into citation disadvantage. Part of the reason might be that articles with OD are more quickly replicated and extended by follow-up studies, making their long-term effects shrink. Also, OD articles are more likely to focus on well-developed topics, which are more prone to becoming obsolete in comparison with emerging topics without data sharing. It is thus fruitful to dig further into the mechanisms through which OD matters and opposite responses.

Our study is of course not without limitations. First, the citation advantages of OD articles reported in this article are from our analyses of articles published in Chinese economics journals, and should be replicated and extended in other disciplines and fields, as well as in journals of other languages to extend the generalizability of the findings.

Second, the DID model helps to address the endogeneity concerns in regard to causal inference, but still the experiment is not randomly assigned and controlled. It is thus beneficial to seize other more ideal contexts and cases to explore the OD citation advantages.

Third, our estimates of the OD policy effects might be biased due to several data limitations. Authors of articles which are not mandated to share their research data may also do so through personal websites or data repositories, and it would help to collect data to explore these cases. There might be additional attention brought to CIE due to the new initiative of OD, which could be explored in future studies. We could not differentiate the citations to the research data and those to the articles per se, which requires manual checks of citations one by one, and it would be relevant to consider this possibility in future studies. Also, journal self-archiving policies could confound the findings, and it is helpful to consider its effects in future studies. Finally, the mechanisms through which OD works merit further explorations, which would help us to uncover the black box of the OS ecosystem.

Declarations

Conflict of interest There is no conflict of interest.

References

Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 questions about open science practices. *Journal of Business and Psychology, 34*(3), 257–270.

Bloom, T., Ganley, E., & Winker, M. (2014). Data access for the open access literature: PLoS's data policy. *PLoS Medicine, 11*(2), e1001607.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology, 63*(6), 1059–1078.

Budapest Open Access Initiative. (2002). *Read the Budapest Open Access Initiative*. Retrieved from https://www.budapestopenaccessinitiative.org/read

Cahn, E. S., & Glass, V. (2018). The Effect of Age and Size on Reputation of Business Ethics Journals. *Business & Society, 57*(7), 1465–1480.

Cambon-Thomsen, A., Thorisson, G. A., & Mabile, L. (2011). The role of a bioresource research impact factor as an incentive to share human bioresources. *Nature Genetics, 43*(6), 503–504.

Campbell, J. (2015). Access to scientific data in the 21st century: Rationale and illustrative usage rights review. *Data Science Journal, 13*, 203–230.

Christensen, T., & Ma, L. (2022). Chinese University Administrations: Chinese Characteristics or Global Influence? *Higher Education Policy, 35*, 255–276. https://doi.org/10.1057/s41307-020-00208-8

Cousijn, H., Habermann, T., Krznarich, E., & Meadows, A. (2022). Beyond data: Sharing related research outputs to make data reusable. *Leading Publishing, 35*(1), 75–80.

Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (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., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: Randomised controlled trial. *BMJ-British Medical Journal, 337*(7665), a568.

Enwald, H., Grigas, V., Rudzioniene, J., & Kortelainen, T. (2022). Data sharing practices in open access mode: A study of the willingness to share data in different disciplines. *Information Research-an International Electronic Journal, 27*(2), 932. https://doi.org/10.47989/irpaper932

Eysenbach, G. (2006). Citation advantage of open access articles. *PloS Biology, 4*(5), 692–698.

Feenberg, D., Ganguli, I., Gaule, P., & Gruber, J. (2017). It's good to be first: order bias in reading and citing NBER working papers. *Review of Economics and Statistics*, 99(1), 32–39.

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

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature, 529*(7584), 117–119.

Gold, E. R. (2021). The fall of the innovation empire and its possible rise through open science. *Research Policy, 50*(5), 104226.

Hanson, B., Sugden, A., & Alberts, B. (2011). Making data maximally available. *Science, 331*(6018), 649.

Hartzell, C. (2015). Data access and research transparency (DA-RT): A joint statement by political science journal editors. *Conflict Management and Peace Science, 32*(4), 355–355.

Homolak, J., Kodvanj, I., & Virag, D. (2020). Preliminary analysis of COVID-19 academic information patterns: A call for open science in the times of closed borders. *Scientometrics, 124*(3), 2687–2701.

Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., & Goudie, S. (2020). Developing a research data policy framework for all journals and publishers. *Data Science Journal, 19*(5), 1–15.

Inkpen, R., Gauci, R., & Gibson, A. (2021). The values of open data. *Area, 53*(2), 240–246.

International Research Center for the Evaluation of Chinese Academic Documents, & Tsinghua University Library. (2020a). *Annual Report for International Citation of Chinese Academic Journals (Natural Science)*. China Academic Journal Electronic Publishing House.

International Research Center for the Evaluation of Chinese Academic Documents, & Tsinghua University Library. (2020b). *Annual Report for International Citation of Chinese Academic Journals (Social Science)*. China Academic Journal Electronic Publishing House.

Ioannidis, J. P. A., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X. Q., Culhane, A. C., et al. (2009). Repeatability of published microarray gene expression analyses. *Nature Genetics, 41*(2), 149–155.

Jones, L., Grant, R., & Hrynaszkiewicz, I. (2019). Implementing publisher policies that inform, support and encourage authors to share data: Two case studies. *Insights, 32*(1), 11. https://doi.org/10.1629/uksg.463

Kauffmann, F., & Cambon-Thomsen, A. (2008). Tracing biological collections: Between books and clinical trials. *JAMA-Journal of the American Medical Association, 299*(19), 2316–2318.

Kretser, A., Murphy, D., Bertuzzi, S., Abraham, T., Allison, D. B., Boor, K. J., et al. (2019). Scientific integrity principles and best practices: Recommendations from a scientific integrity consortium. *Science and Engineering Ethics, 25*(2), 327–355.

Kretser, A., Murphy, D., & Dwyer, J. (2017). Scientific integrity resource guide: Efforts by federal agencies, foundations, nonprofit organizations, professional societies, and academia in the United States. *Critical Reviews in Food Science and Nutrition, 57*(1), 163–180.

Krumholz, H. M., Ross, J. S., Gross, C. P., Emanuel, E. J., Hodshon, B., Ritchie, J. D., et al. (2013). A historic moment for open science: The Yale University open data access project and medtronic. *Annals of Internal Medicine, 158*(12), 910–911.

Kwasnicka, D., Ten Hoor, G. A., van Dongen, A., Gruszczynska, E., Hagger, M. S., Hamilton, K., et al. (2020). Promoting scientific integrity through open science in health psychology: Results of the synergy expert meeting of the European health psychology society. *Health Psychology Review, 15*(3), 333–349.

Laakso, M., & Bjork, B. C. (2016). Hybrid open access-A longitudinal study. *Journal of Informetrics, 10*(4), 919–932.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature, 411*(6837), 521–521.

Lin, W. Y. C. (2021). Effects of open access and articles-in-press mechanisms on publishing lag and first-citation speed: A case on energy and fuels journals. *Scientometrics, 126*(6), 4841–4869.

Maddi, A., Lardreau, E., & Sapinho, D. (2021). Open access in Europe: A national and regional comparison. *Scientometrics, 126*(4), 3131–3152.

Martone, M. E., Garcia-Castro, A., & VandenBos, G. R. (2018). Data sharing in psychology. *American Psychologist, 73*(2), 111–125.

McCain, K. (1995). Mandating sharing: Journal policies in the natural sciences. *Science Communication, 16*(4), 403–431.

McNutt, M. (2014). Journals unite for reproducibility. *Science, 346*(6210), 679.

Molecular Biology and Evolution. (2023). Availability of data and materials. Retrieved January 19, 2023, from https://academic.oup.com/mbe/pages/availability-of-data-and-materials?login=true

Molloy, J. C. (2011). The open knowledge foundation: open data means better science. *PLoS Biology, 9*(12), e1001195. https://doi.org/10.1371/journal.pbio.1001195

National Institutes of Health (NIH). (2003). Final NIH statement on sharing research data. Retrieved January 17, 2021, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

National Science Foundation. (2010). Scientists seeking NSF funding will soon be required to submit data management plans. Retrieved January 17, 2021, from https://www.nsf.gov/news/news_summ.jsp?cntn_id=116928

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

Ohmann, C., Moher, D., Siebert, M., Motschall, E., & Naudet, F. (2021). Status, use and impact of sharing individual participant data from clinical data: a scoping review. *BMJ Open, 11*(8), e049228. https://doi.org/10.1136/bmjopen-2021-049228

Pandita, R., & Singh, S. (2022). A Study of Distribution and Growth of Open Access Research Journals Across the World. *Publishing Research Quarterly, 38*(1), 131–149.

Pencina, M. J., Louzao, D. M., McCourt, B. J., Adams, M. R., Tayyabkhan, R. H., Ronco, P., & Peterson, E. D. (2016). Supporting open access to clinical trial data for researchers: The Duke clinical research Institute-Bristol-Myers Squibb supporting open access to researchers initiative. *American Heart Journal, 172*, 64–69.

Piwowar, H. A., Becich, M. J., Bilofsky, H., & Crowley, R. S. (2008). Towards a data sharing culture: Recommendations for leadership from academic health centers. *PLoS Medicine, 5*(9), 1315–1319.

Piwowar, H. A., & Chapman, W. W. (2010). Public sharing of research datasets: A pilot study of associations. *Journal of Informetrics, 4*(2), 148–156.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE, 2*(3), e308.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *Peer J, 1*, e175.

PNAS. Editorial and Journal Policies. (2023). Retrieved January 19, 2023, from https://www.pnas.org/author-center/editorial-and-journal-policies#Embargo%20Policy

Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological Applications, 29*(1), e01822.

Reardon, S. (2014). Clinical-trial rules to improve access to results. *Nature, 515*(7528), 477.

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science, 331*(6018), 703–705.

Ross, J. S., Mulvey, G. K., Hines, E. M., Nissen, S. E., & Krumholz, H. M. (2009). Trial publication after registration in ClinicalTrials.Gov cross-sectional analysis. *PloS Medicine, 6*(9), e1000144.

Ross, M. W., Iguchi, M. Y., & Panicker, S. (2018). Ethical aspects of data sharing and research participant protections. *American Psychologist, 73*(2), 138–145.

Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavior Research Methods, 48*(3), 1062–1069.

Sá, C., & Grieco, J. (2016). Open data for science, policy, and the public good. *Review of Policy Research, 33*(5), 526–543.

Science. (2023). Science Journals: Editorial Policies. Retrieved January 15, 2023, from https://www.science.org/content/page/science-journals-editorial-policies

Sombatsompop, N., Ratchatahirun, P., Surathanasakul, V., Premkamolnetr, N., & Markpin, T. (2002). A citation report for Thai academic journals published during 1996–2000. *Scientometrics, 55*(3), 445–462.

Spring Nature. (2021). Research data support service. Retrieved January 17, 2021, from https://www.springernature.com/gp/authors/research-data/research-data-support

Strcic, J., Civljak, A., Glozinic, T., Pacheco, R. L., Brkovic, T., & Puljak, L. (2022). Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv. *Scientometrics, 127*(5), 2791–2802.

Tahamtan, I., Afshar, A. S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics, 107*(3), 1195–1225.

The Royal Society. (2012). *Science as an open enterprise*. The Royal Society Science Policy Centre report.

Van Noorden, R. (2014). Confusion over open-data rules. *Nature, 515*(7528), 478.

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra-Psychology, 3*(1), 1.

Vincent-Lamarre, P., Boivin, J., Gargouri, Y., Larivi`ere, V., & Harnad, S. (2016). Estimating open access mandate effectiveness: The MELIBEA score. *Journal of the Association for Information Science and Technology, 67*(11), 2815–2828.

Wang, J. N. (2005). The survey on the peak citations in various subjects of social science and the requirements of citations evaluation in time-period (社会科学各学科引文的高峰值调查及引文评价的时段性要求). *Information and Documentation Services (chinese Journal), 3*, 81–82.

Wang, J., & Shapira, P. (2015). Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers. *PLoS ONE, 10*(2), e0117727.

Wang, X. W., Liu, C., Mao, W. L., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics, 103*(2), 555–564.

Whitlock, M. C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution, 26*(2), 61–65.

Yancy, C. W., Harrington, R. A., & Bonow, R. O. (2018). Data sharing-the time has (not yet?) come. *JAMA Cardiology, 3*(9), 797–798.

Zastrow, M. (2020). Open science takes on Covid-19. *Nature, 581*(7806), 109–110.

Zhang, G. Y., Wang, Y. Q., Xie, W. X., Du, H., Jiang, C. L., & Wang, X. W. (2021). The open access usage advantage: A temporal and spatial analysis. *Scientometrics, 126*(7), 6187–6199.

Zhang, L. W., & Ma, L. (2021). Does open data boost journal impact: Evidence from Chinese economics. *Scientometrics, 126*(4), 3393–3419.

Zhang, L., Wei, Y. H., Huang, Y., & Sivertsen, G. (2022). Should open access lead to closed research? The trends towards paying to perform research. *Scientometrics*. https://doi.org/10.1007/s11192-022-04407-5

Zheng, H. L., Wang, X., Guo, J. H., Zhang, Y., Li, H. H., & Gu, H. Y. (2014). Citation-based economics literature needs analysis taking doctoral dissertation citation in Renmin University of China as an example (基于引文的经济学文献需求特征分析——以中国人民大学经济学博士学位论文引文为例). *Information Science (Chinese Journal), 32*(10), 48–51.

Zhu, Y. M. (2020). Open-access policy and data-sharing practice in UK academia. *Journal of Information Science, 46*(1), 41–52.

Zipper, S. C., Whitney, K. S., Deines, J. M., Befus, K. M., Bhatia, U., Albers, S. J., et al. (2019). Balancing open science and data privacy in the water sciences. *Water Resources Research, 55*(7), 5202–5211.

Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLoS One, 15*(9), e0239283. https://doi.org/10.1371/journal.pone.0239283