Check for updates

# Does open data boost journal impact: evidence from Chinese economics

Liwei Zhang[1] · Liang Ma[2]

## Abstract

To encourage research transparency and replication, more and more journals have been requiring authors to share original datasets and analytic procedures supporting their publications. Does open data boost journal impact? In this article, we report one of the first empirical studies to assess the effects of open data on journal impact. China Industrial Economics (CIE) mandated authors to open their research data in the end of 2016, which is the first to embrace open data among Chinese journals and provides a natural experiment for policy evaluation. We use the data of 37 Chinese economics journals from 2001 to 2019 and apply synthetic control method to causally estimate the effects of open data, and our results show that open data has significantly increased the citations of journal articles. On average, the current- and second-year citations of articles published with CIE have increased by 1~4 times, and articles published before the open data policy also benefited from the spillover effect. Our findings suggest that journals can leverage compulsory open data to develop reputation and amplify academic impacts.

**Keywords** Open science · Open data · Economics · Impact factor · Research transparency

## Introduction

In the past decades, with the development of web tools, innovations in science publishing are taking place as an "open science movement" at multiple levels: open access to publications, open data in an available manner for scientists, and open peer review for manuscripts (Ule 2020). Open data, as a key component of the open science movement, requires scientists to share their research datasets to encourage research transparency and replication, which is the most essential components of scientific publications (Whitlock, et al.

✉ Liang Ma
liangma@ruc.edu.cn

Liwei Zhang
liwei_zhang@sdu.edu.cn

1   School of Innovation and Entrepreneurship, Shandong University, Jimo District, Qingdao, Shandong province, China

2   School of Public Administration and Policy, Renmin University of China, Haidian District, Beijing, China

2010; Peng 2011). With the development of web-based tools, the data sharing and informal debate is changing the traditional scholarly communication model that scientists must embrace the culture of sharing and rethink their vision of databases (Anonymous 2005). Recently, more and more scientists, research societies, universities, journals, publishers and funders have taken attention to advocate "Open Data" which was first developed in the fields of medical, pandemic, and clinical research (McCain 1995; Ross et al. 2009; Krumholz 2012; Ross et al. 2012a, b; Reardon 2014; Anonymous 2015) and the fields of genomic, biological, and ecological research (Campbell and Bendavid 2002; Kauffmann and Cambon-Thomsen 2008; Cambon-Thomsen et al. 2011; Reichman et al. 2011; Whitlock 2011; Borgman 2012). The global pandemic of COVID-19 in 2020 has drawn renewed attention in data sharing among the global scientific community (Zastrow 2020).

Given the appealing of open data, more and more funding agencies and societies have announced their open data policies. The National Institutes of Health released a statement of sharing research data in 2002, which expected investigators supported by NIH funding to make their research data available to the scientific community for subsequent analyses (NIH 2002). The key members of Open Knowledge Foundation (e.g., the Working Group on Open Data in Science, which is known as the Open Science Working Group (OKF)) collaborated with John Wilbanks of Creative Commons to spend two years to develop a set of principles for open scientific data, which is named "Panton Principles" to promote open data in science that is gathering momentum (Murray-Rust et al. 2010). The Royal Society released "science as an open enterprise" aiming to identify the principles, opportunities, and problems of sharing and disclosing scientific information (The Royal Society 2012). Public Library of Science (PLoS) developed a "free the data" policy, requesting all data underlying the findings of published articles fully available (Bloom et al. 2014). The Yale University launched "Yale Open Data Access Project" (YODA) (Krumholz et al. 2013)—one of several pioneering data-sharing platforms. And Clarivate Analytics launched Data Citation Index to provide a single point of access to quality research data from global repositories across disciplines (Clarivate Analytics 2012).

In recent years, an increasing number of journals and publishers, such as Springer Nature,[1] Elsevier,[2] *Science*,[3] *PNAS*,[4] *Molecular Biology and Evolution*,[5] have developed their data availability policies. The International Association of Scientific, Technical and Medical Publishers (STM) is fully committed to supporting research data sharing, also launched the 'STM 2020 Research Data Year' in January 2020 to increase the number of journals with data policies and articles with Data Availability Statements (DAS), to increase the number of journals that deposit the data links to the SCHOLIX framework, and to increase the citations to datasets along the FORCE 11 data citation guidelines (International Association of Scientific, Technical and Medical Publishers 2020). In addition, aiming to provide researchers with free and open valid data and biological discovery resources and revolutionize publishing by promoting reproducibility of analyses and data dissemination, organization, understanding, the journal *GigaScience*[6] was established by Beijing Genomics Institute (BGI) and BioMed Central on July 2012. Also, the Dryad Digital Repository was established with the support of NSF, which is one of the
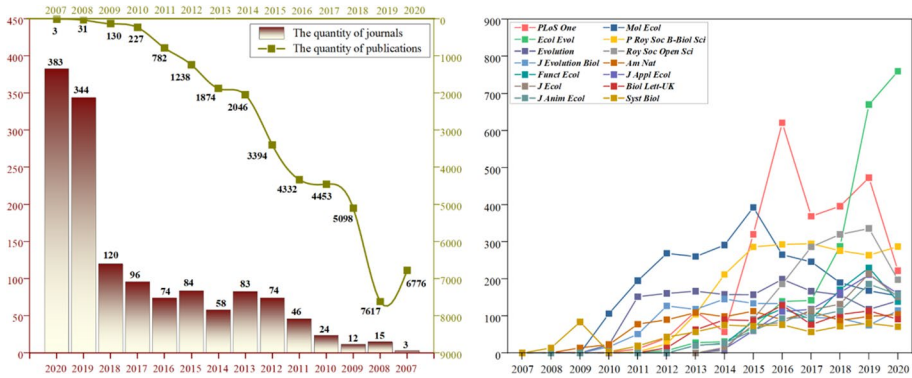
---

[1] https://www.springernature.com/gp/authors/research-data-policy/data-policy-faqs.

[2] https://www.elsevier.com/authors/tools-and-resources/research-data/open-data.

[3] https://www.sciencemag.org/authors/science-journals-editorial-policies.

[4] https://www.pnas.org/authors/editorial-and-journal-policies.

[5] https://academic.oup.com/mbe/pages/availability-of-data-and-materials.

[6] https://academic.oup.com/gigascience/pages/About.

**Fig. 1** The increasing number of journals and open-data publications on Dryad (left), and the number of open-data publications from top journals on Dryad (right)

most commonly-used research data services to adopt a joint data archiving policy among a group of leading journals and scientific societies (Dryad 2011). Dryad provides a general-purpose home for a wide diversity of data types to make research data discoverable, freely reusable, and citable. Until December 28, 2020, 1,416 journals have joint Dryad and published 38,001 publications, and both the number of joining journals and publications has been increasing (shown in Fig. 1).

Nowadays, China is becoming the epicenter of academic article retractions. There have been many high-profile cases of faked peer reviews and academic article sales, and some of them even involved with prominent scholars (Xin 2009; Hvistendahl 2013; Tang 2019; Tang et al. 2020; Mallapaty 2020). For example, a research article written by Chunyu Han who was then an associate professor at Hebei University of Science and Technology was published by *Nature Biotechnology* in 2016. Many scholars tried to replicate the findings of this article but all failed, suggesting its experimental method is not repeatable (Javidi-Parsijani et al. 2017; Khin et al. 2017; Lee et al. 2016; Burgess et al 2016). *Nature Biotechnology* was appealed to request Han to publish the source data, while he withdrew this article in 2017 (Gao et al. 2017).

Given the growing problems of academic article retraction and other misconducts, government agencies, universities, scientific community, and the society at large pay close attention to academic integrity, and several government policies have been implemented to strengthen academic integrity (Yang 2013; Mallapaty 2020). Open data is one of the instruments advocated by the scientific community to promote academic integrity and impact, but its effects have not been well assessed. China Industrial Economics (CIE), a Chinese journal hosted by the Institute of Industrial Economics, the Chinese Academy of Social Sciences, has mandated its authors to share their research data (e.g., source data, programs, processed data, case-study materials, and attachments) since November 2016. This is the first attempt among Chinese social sciences journals in embracing open data, and it has been the only one by the time of writing this article.[7] Chinese Journal of Sociology also

---

[7] Fu and Chen (2014) examined the 2013 version of the highest international impact academic journals of China, and revealed that 22.9% of science/technology/medicine journals and 8.9% of humanities and social sciences journals have data publish policy. Peng and Han (2019) examined the 65 English journals published by the Chinese Academy of Sciences and found that 37 (or 57%) have data policy. We applied the same approaches and went through each of the 2020 version of Chinese academic journals of China with the highest international impact (see: http://hii.cnki.net/cajz/), and we found that 5 of 135 science/

published data processing programs of some published articles, but it is not a mandatory requirement for its authors.

The adoption of open data by CIE provides a natural experiment for policy evaluation, and our research question is whether and to what extent open data improved the impacts of journal articles. Specifically, we used the case of CIE to explore the impact of the open data policy on citation performance of journal publications. Our results show that open data has significantly increased the citations of journal articles, and articles published before the open data policy also benefited from the spillover effect. Our findings suggest that journals can leverage compulsory open data to develop reputation and amplify academic impacts.

In the rest of this article, we firstly review the literature and develop the hypotheses to be tested in this study. We then introduce the data and methods used and report the key empirical findings. We finally discuss the research and policy implications of our findings.

## Literature review and research hypotheses

### A review of the literature on open data

For the case of open data, there is one viewpoint in scientific community that data provides the evidence for the published body of scientific knowledge, which is the foundation for all scientific progress. The more data is made openly available in a useful manner, the greater the level of transparency and reproducibility, hence the more efficient the scientific process becomes to the benefit of society, such as in responding to COVID-19 outbreaks (Hanson et al. 2011; UNESCO 2020; Homolak et al. 2020). Despite its merit and benefits, there are still many obstacles for open data in the scientific community (Molloy 2011).

Regarding the merit and benefits of open data, some advocated open data to push the development of science from the perspective of research transparency and replication. In the report of "Science as an open enterprise", there is a view to advocate open availability of research data, because it could allow validation, replication, reanalysis, new analysis, reinterpretation, or inclusion into meta-analyses, facilitate reproducibility of research and improve the reproducibility of science. From the analysis of existing shared neuroimaging data sets, Poldrackh and Gorgolewsk (2014) pointed that data sharing generated lots of benefits, including: maximizing the contribution of research subjects, enabling new questions, enhancing reproducibility, improving research practices, test bed for new analysis methods, reducing the cost of doing science, and protecting valuable scientific resources. Sa and Grieco (2016) pointed that open research data not only contributed to scientific development but also enabled science-policy dialogue helpful to scientific decision-making.

The merit of open data has been examined in various research fields, and the results are generally supportive of open data. Ioannidis et al. (2009) found that data availability improved the published results which are reproducible by independent scientists from the analysis of gene expression profiling based on the microarray published in Nature Genetics from 2005 to 2006, and he suggested that stricter publishing rules will enforce data

---

Footnote 7 (continued)

technology/medicine journals and 4 of 67 humanities and social sciences journals have open data policy. In the case of Chinese humanities and social sciences journals, only CIE compulsorily mandates open data and really publishes data together with articles. Part of the reason that their results (Fu and Chen 2014) are overstated is that they might use a rather broad definition of data publishing policy.

availability. Brown (2003) argued that the increase of received data and applications in Genomic and Proteomic Databases (GPD, one of the open data databases) improved the research explosion in the field of molecular biology. Ross et al. (2012a, b) pointed that making individual patient data available to the whole research community enabled scientists to derive full benefit from the enormous resources devoted to human clinical trial research. Gotzsche (2012) suggested that sharing raw data helped to strengthen and open up health research.

As mentioned above, it is widely believed that open data will greatly stimulate the development of science. However, the development and application of open data is not optimistic. The open data movement is facing challenges, and a complete cultural shift needs to be further down the line (Van Noorden, 2014). For instance, a survey by Vines et al. (2013) took a subset of evolutionary biology articles and found only 6 of 51 *PLoS ONE* articles had shared their research data. There are big obstacles to widely advocate open data, especially because many scientists and journals are unwilling to open data. Van Noorden (2014) indicated that not everyone complied with the open-data rules, and even some scientists are worried about being required to share their data. Piwowar and Vision (2013) pointed that sharing data will bring many benefits, such as helping to identify errors, discouraging fraud, useful for training new researchers, and increasing efficient use of funding and patient population resources by avoiding duplicate data collection. Meanwhile, Piwowar and Vision also pointed there are also some risks and cost, such as being afraid that other researchers would find errors in their results, or "scoop" additional analyses they have planned for the future. In addition, some scholars pointed that the lack of tools for datasets repositories, the underdevelopment of computational methods, and inadequate financial support postpone open data to be an effective research tool (Gymrek and Farjoun 2016; Leonelli 2016).

From specific research disciplines, previous studies also pointed out the open data movement is facing some challenges. Reichman et al. (2011) argued that the development of ecology which is a synthetic discipline will be benefited from open access to data from the earth, life, and social sciences, but there are series of technological, social, and cultural challenges. For example, ecologists have had few incentives for sharing information because sharing data was not viewed as a valuable scholarly endeavor or as an essential part of science. The research by Zipper et al. (2019) showed that open data and programs are transforming water science by enabling synthesis and enhancing reproducibility, but there might also risks. For example, researchers unintentionally violate the privacy and security of individuals or communities by sharing some sensitive information. Murray-Rust (2008) suggested that many publishers claimed copyright for open data and do not allow its reuse without authentication, which is one of the barriers for open data. He also argued that there are large variations of open data traditions among different disciplines. For instance, bioscience has a long tradition of requiring data to be published and then aggregated in publicly funded databanks. Telescopes, satellites, particle accelerators, and neutron sources as "large science" are adept at making universally available for re-use. For "small science", however, scientists typically publish many independent publications which report individual experiments.

To realize the expectation of open data, some scholars examined the intentions of open data, such as the initial open data research especially in the fields of medicine and biomedical. For example, Campbell et al. (2002) took a survey of 100 geneticists and life scientists from 100 US universities, and found that there are several reasons for unwilling to share data, such as costing too much effort (mentioned by 80% scientists), protecting the graduate students, postdoctoral fellows or junior faculty (mentioned

by 64% scientists), and protecting the ability to publish of themselves (mentioned by 53% scientists). Vogeli et al. (2006) implemented a survey for the data withholding by second-year doctoral students and postdoctoral fellows in the fields of life science, computer science, and chemical engineering from 50 universities, and found that data withholding demonstrated negative effects on trainees, while trainees of life science have to address this issue more. Piwowar (2011) found that the situation of open data is different in different field. For example, the authors from the field of cancer and human subjects are relatively reluctant to share data.

Meanwhile, Piwowar and Chapman (2010) pointed that the journal impact and policies of open science could affect the researchers' willingness to share data. For example, researchers are more likely to share the source data when their studies were published in high-impact journals, or when the first or last authors had high influence. Piwowar (2011) also analyzed 11,603 academic articles published from 2000 to 2009 in the field of gene expression microarray creation, and the results showed that the authors were more likely to share their data if the articles got published on open access journals, journals with relatively strong data sharing policy, or journals funded by a large number of NIH grants.

In addition, during the forum of open data for better science organized by *National Science Review*, many scholars believed that open data is able to represent better science, but there are some privacy and intellectual property rights issues impending its wide acceptance. At the same time, data sharing is not a part of the tenure evaluation system or any other evaluation systems, and we need to act more to advance open data (Zhao, 2018). In the report of "Science as an open enterprise" from the Royal Society, it demonstrated that scientists need to be more open among the public and the social network to improve their self-cognition to improve data open. Bolukbasi et al. (2013) indicated open data needs a crediting and cooperation culture. Consequently, the key point to promote open data is to improve personal incentives of open data, while receiving higher citations is an important motivator for archiving their data publicly (Tenopir et al.2011).

Whether and how open data helps to increase research efficiency and quality is the heart of the matter about which scientists and journals concern, as a well-known adage: you cannot manage what you do not measure. Meanwhile, open data is considered as the key to resolve or mitigate academic misconduct. Although there have been more and more studies on open data, few of them examined whether open data will increase the citations of articles and improve the impact of journals.

The above review suggests that the impact of open data has not been well examined, and the existing studies are limited in at least three aspects. First, most studies focused on open access (Craig et al. 2007; van Vlokhoven 2019), while the relevance of open data has not been investigated. For the influence of open science on citations, scholars analyzed the relationship between open-access articles and their citations, and different processing methods generated different conclusions (Craig et al. 2007; Ingwersen and Elleby 2011; Wang, et al. 2015).

Second, the existing studies are subject to endogeneity concerns and cannot generate causal inference. For the influence of open data on citations, Piwowar Day, and Fridsma (2007) examined 85 cancer microarray clinical trial publications with respect to the availability of their data. Piwowar and Vision (2013) analyzed the influence of data availability of the articles in field of created gene expression microarray on citations, and found that open data helped to increase the citations to some extent. Piwowar and colleagues (2013) examined the effects of open data on article citations, but they only included small samples and did not take time and other factors into account. The authors are voluntary to share their research data in Harvard Dataverse, Mendeley Data Repository and other data sharing

platforms, and the estimates of open data effect would be biased if we cannot control for their intentions and other heterogeneities.

Lastly, our understanding of open data in natural sciences is much more progressed than that of social sciences, such as "psychology and other social and behavioral sciences have invested in relatively few open domain-specific repositories" (Martone, Garcia-Castro and VandenBos, 2019), which merit more research to further the acceptance of open data. It is thus imperative to examine the effects of open data on academic and social impact in the various fields of social sciences.

## 2.2 Research hypotheses

In this article, we discuss whether and how open data will help to develop the reputation and amplify the impact of journals. Previous studies indicated that research data policies of funding agencies and journals could influence researchers' willingness to share research data (Schmidt, Gemeinholzer and Treloar 2016; Giofre et al. 2017). So, we expect that if a journal adopts the open data policy to require all its authors to share their research data, then its reputation could be improved in at least three ways. Open data policy can be either "mandatory" (required) or "voluntary" (suggested), and we expect that the former is stronger and would generate higher impacts than the latter.

Firstly, journal articles sharing research data are more likely to be trusted and favored by the scientific community because they can be replicated and checked by rerunning their data analytic processes (Piwowar and Vision 2013). People would trust journal articles with open data because data sharing means integrity and reciprocity, and they would more likely to cite them than those without open data (Piwowar and 2007; Piwowar and Vision 2013; McKiernan et al. 2016; Christensen et al. 2019; Colavizza et al. 2020).

Secondly, a journal mandating open data may deter unqualified and even dishonest authors from submitting their research, and only research articles standing the tests of any kinds dare to submit and be published. As long as authors expect that their data and analytic procedures would be shared among peers and subject to countless checks, they would submit their manuscripts only when they are confident in their data and findings (van Vlokhoven 2019). By setting a high standard in research integrity, journals of open data can elicit quality manuscripts and deter unqualified studies, which helps to build their reputation among the research community (Piwowar 2011).

Thirdly, research data shared by authors might be used for other purposes than follow-up research, and we argue that the learning function of open data helps to boost journal impact (Ule 2020). Journal articles sharing the original data and analytic procedures would be valuable teaching and learning resources for research methods (particularly for quantitative methods, despite qualitative and mixed methods also benefit), and their replication in classes and by students helps to increase their coverage and impact. Given the above discussions, we propose the first hypothesis:

**H1** The impact of articles published in the journals which implement the open data policy is higher than that of other journals.

We argue that there is a spillover effect of open data and journal articles which did not share their research data would also benefit from journals' open data policy. Open data helps to develop journal reputation, and the impact of articles published before the introduction of the open data policy would also increase. For instance, the open data policy was

implemented in the end of 2016, and manuscripts submitted after 2016 would be required to share their original data. Articles published in 2017 were mandated to share research data, and their impacts would be amplified due to data sharing (H1). The 2015 and 2016 articles were not required to do so, but we expect that their impacts would also increase due to the reputational effect of open data. Such rippling effect is expectable because open data not only increase the exposure of articles sharing research data, articles published before the introduction of open data would also benefit from the increasing reputation of the journals as a whole. We thus develop the second hypothesis:

**H2** Articles published before the implementation of the open data policy are more likely to be cited than those published with other journals.

## Data and methodology

### Data

In this study we collect the data of journal articles and their citations from the Chinese Social Sciences Citation Index (CSSCI) database, the Chinese version of SSCI. CSSCI updates its journal list every two years, and there are about 70 economics journals indexed by CSSCI. We believe economics journals included in CSSCI are comparable with CIE, and we collected the data of 37 journals which have been indexed from 2001 to 2019 (see Table 1). These journals cover all key economics research areas such as industrial economics, government economics, financial economics, accounting economics, rural economics, international trade and taxation research, and economic theory (see Table 5).

We collected the variables of articles published with the 37 journals from 2001 to 2019, the period of time when the data are available, and we totally covered 703 observations. We only included full-length research articles, and editorials, conference reports, and leaders' speeches were excluded. Similar to the rule of SSCI, CSSCI only counts the citations of the citing articles indexed in its journal list.

The number of articles published per year and the number of citations per article vary substantially across journals. ERJ has received the highest number of citations per article (45.16 times), followed by CIE (15.1), Journal of Financial Research (13.57), Accounting Research (9.22), Chinese Rural Economy (8.92), and China Rural Survey (7.97). The results show that CIE has a relatively higher research impact among the 37 journals.

### Synthetic control method

We use the Synthetic Control Method (SCM) to estimate the effect of open data on journal impact. SCM is a counterfactual analysis, which is a statistical analytic method to evaluate the effects of policies commonly used in the fields of economics, environmental science, and political science. SCM was originally developed by Abadie and Gardeazabal in 2003, and it is a new policy effect assessment method expanding the traditional difference-in-differences (DID) method (Abadie and Gardeazabal 2003; Abadie et al. 2010, 2015). The basic idea of SCM is to construct a synthetic 'counterfactual' unit which is a weighted average of multiple control groups, and then to identify the difference of the results between the treated group and the control group to estimate policy effects before and after policy implementation.

**Table 1** The average number of papers published by each journal and the average number of citations by each article

| Journals | N of papers (annual) | N of citations (per paper) |
|---|---|---|
| Finance & Economics | 171.11 | 2.36 |
| Research on Financial and Economic Issues | 227.89 | 1.94 |
| Journal of Finance and Economics | 144.63 | 6.13 |
| Finance & Trade Economics | 200.84 | 5.08 |
| Contemporary Finance & Economics | 230.00 | 2.30 |
| Modern Economic Science | 94.21 | 3.16 |
| Contemporary Economic Research | 181.11 | 1.35 |
| Reform | 214.32 | 2.76 |
| Studies of International Finance | 137.26 | 4.24 |
| International Economic Review | 83.05 | 3.88 |
| Intertrade | 149.16 | 1.36 |
| Journal of International Trade | 205.63 | 4.07 |
| Macroeconomics | 184.11 | 2.04 |
| Accounting Research | 157.74 | 9.22 |
| Journal of Financial Research | 180.42 | 13.57 |
| Economic Science | 69.42 | 7.08 |
| Economic Theory and Business Management | 146.63 | 3.49 |
| Economic Review | 117.26 | 4.02 |
| Comparative Economic and Social Systems | 128.47 | 3.87 |
| Economic Perspectives | 273.00 | 2.98 |
| Economist | 149.63 | 4.15 |
| Economic Research Journal | 164.63 | 45.16 |
| Research on Economics and Management | 180.05 | 1.65 |
| Economic Review Journal | 295.84 | 1.24 |
| Nankai Economic Studies | 71.84 | 5.83 |
| Journal of Agrotechnical Economics | 132.11 | 4.17 |
| Issues in Agricultural Economy | 197.95 | 5.19 |
| Shanghai Journal of Economics | 163.16 | 2.21 |
| Auditing Research | 94.26 | 4.73 |
| The Journal of World Economy | 110.37 | 16.48 |
| Word Economy Studies | 155.11 | 3.91 |
| Taxation Research | 314.11 | 1.55 |
| Asia–pacific Economic Review | 149.68 | 1.12 |
| China Industrial Economics | 162.32 | 15.10 |
| China Rural Survey | 56.16 | 7.97 |
| Chinese Rural Economy | 119.05 | 8.92 |
| China Land Science | 124.37 | 4.25 |

SCM is appropriate to estimate policy effects when there is only one member in the treatment group. As CIE is the only journal which has adopted the open data policy among economics journals in China, it is not reliably to directly observe the influence of open data policy on citations because journals in China would developed without open

data (the counterfactual). Also, it is not easy to find an equivalent journal comparable with CIE due to their heterogeneities. In this regard, SCM is a suitable method to estimate the influence of open data on academic journals.

As mentioned above, the research purpose of this article is to examine the influence of open data on journal impact. We used the number of citations for articles published with CIE and after the open data policy as the core research variable. The observation journal is defined as $J + 1$ journals ($J + 1 = 37$), while the first journal is CIE which is the treated journal. Journals from the second to the $J + 1$st are defined as control journals. We can observe the number of citations for the $J + 1$st journal in the corresponding year $t$ ($2001 \leq t \leq 2019$). $P_{1t}$ is defined as the number of citations of CIE articles, and $P_{1t}^N$ is defined as the number of citations of the counterfactual outcome that would be realized if there were no implementation of the open data policy. $P_{jt}$ is defined as the number of citations of the J journal in the control group. $D_t$ is a treatment indicator which satisfies:

$$D_t = \begin{cases} 1, 2017 \leq t \leq 2019; \\ 0, 2001 \leq t \leq 2016 \end{cases} \qquad (1)$$

$a_t$ is defined as the number of citations increased after the implementation of the open data policy, and the estimation model for the treated journal CIE is as following:

$$P_{1t} = P_{1t}^N + \alpha_t D_t, \qquad (2)$$

So, $a_t$ as the increment of the number of citations for the published articles on CIE after policy implication is as following:

$$\alpha_t = P_{1t} - P_{1t}^N, for D_t = 1 \qquad (3)$$

To get the value of $at$, $P_{1t}^N$ needs to be evaluated in formula (3) as $P1t$ can be observed directly. As Abadie and colleagues (2015) suggested, the synthetic control unit was taken as a weighted average of the control units in the pool, and these weights were used to construct $P_{1t}^N$. This means that a synthetic control can be represented by a ($Jx1$) vector of weights $W = (w2,..., wJ+1)$, for journals $j = 2,..., J + 1$, which satisfy $0 \leq wj \leq 1$ and $w2 +...+ wJ+1 = 1$. Using a vector of some optimal weights $w* = (w_2^*,..., w_{J+1}^*)'$, $P_{1t}^N$ is taken as the weighted average of $Pjt$ which is the indicator of the number of citations for journals in the control group. So, $at$ which is the incremental number of citations for CIE after 2016 is as following:

$$\hat{\alpha}_t = P_{1t} - \sum_{j=2}^{J+1} w_j^* P_{jt}, for D_t = 1 \qquad (4)$$

CIE kicked off the open data policy in November 2016, the articles published in 2017 are still in the early stage for the observation points after the implementation of this policy. We thus used the current- and the second-year citations for the articles published after the implementation of the open data policy, and both are dependent variables used to reflect the short policy effect. The current-year citations include the data of articles published from 2017 to 2019, while the second-year citations only includes the data of articles published in 2017 and 2018 because articles of 2019 would be cited in 2020. The data from 2001 to 2019 show that the number of articles published by CIE has decreased, but the current-year citation has quickly increased (see Fig. 2).
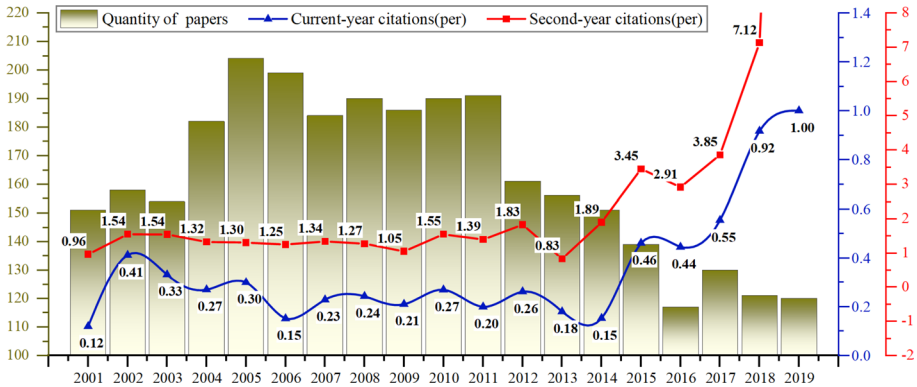
**Fig. 2** The number of papers, current-year citations per paper, and second-year citations per paper of CIE

Firstly, we selected the commonly used indicators which influence the citations of journal articles as control variables for the estimation model, including the number of authors, the number of funding sources, the number of keywords, the number of references, and the number of pages. We downloaded the raw data of the articles published with the 37 journals from 2001 to 2019 and calculated the statistics included in models.

Secondly, we selected the number of articles published each year, the founding year, and the CSSCI rank of the journals as control variables for model estimations. The CSSCI rank is calculated by aggregating journal impact factor and peer review among prominent scholars. Moreover, as the time span of the data is very long, we included current-year citations and second-year citations of 2001, 2005, 2010, and 2015 as control variables. The descriptive statistics for all the variables are summarized in Table 2. The Stata 15.1 with the Synth package is utilized to estimate the models.

## Results

### The match between real CIE and synthetic CIE

Table 3 shows the composite results of the journals in control group. In the case of current-year citations, the synthetic CIE is combined by Journal of Finance and Economics, International Economic Review, Economic Research Journal (ERJ), the Journal of World Economy, China Land Science, and their weights are 0.032, 0.337, 0.2, 0.348, and 0.083 respectively. When we used the second-year citations as the indicator, the synthetic CIE is combined by International Economic Review, ERJ, the Journal of World Economy, and China Land Science, and their weights are 0.337, 0.22, 0.383, and 0.06 respectively. In the fitting result of both synthetic CIEs, International Economic Review, ERJ, and the Journal of World Economy are the journals with highest weights, suggesting these journals are very comparable with CIE.

Table 4 lists the real observed value of each control variable in real CIE, the two versions of synthetic CIE (Model 1 shows the fitted value of current-year citations, and Model 2 reports the fitted value of second-year citations), and the composition of average value of the 37 control journals. The results show that the value of both versions of

**Table 2** The descriptive statistics of the control variables

| Variables | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Current-year citations (per) | 703 | 0.13 | 0.13 | 0.00 | 1.00 |
| Second-year citations (per) | 703 | 0.72 | 0.72 | 0.02 | 7.12 |
| The number of authors | 703 | 1.82 | 0.38 | 1.12 | 3.51 |
| The number of funds | 703 | 0.57 | 0.56 | 0.00 | 2.10 |
| The number of keywords | 703 | 3.92 | 0.39 | 3.00 | 5.94 |
| The number of references | 703 | 14.44 | 9.17 | 0.00 | 42.14 |
| The number of pages | 703 | 8.16 | 3.66 | 3.07 | 23.71 |
| The started year of publication | 703 | 1979.95 | 10.08 | 1955.00 | 1998.00 |
| The quantity of papers | 703 | 160.46 | 68.58 | 47.00 | 449.00 |
| The journals' rankings of CSSCI | 666 | 25.12 | 16.89 | 1.00 | 68.00 |
| Current-year citations (per, 2001) | 37 | 0.08 | 0.11 | 0.00 | 0.60 |
| Current-year citations (per, 2005) | 37 | 0.10 | 0.10 | 0.01 | 0.58 |
| Current-year citations (per, 2010) | 37 | 0.12 | 0.11 | 0.02 | 0.63 |
| Current-year citations (per, 2015) | 37 | 0.15 | 0.11 | 0.03 | 0.60 |
| Second-year citations (per, 2001) | 37 | 0.42 | 0.50 | 0.05 | 3.05 |
| Second-year citations (per, 2005) | 37 | 0.52 | 0.54 | 0.13 | 3.27 |
| Second-year citations (per, 2010) | 37 | 0.74 | 0.64 | 0.22 | 3.94 |
| Second-year citations (per, 2015) | 37 | 1.05 | 0.73 | 0.29 | 3.64 |

The annual statistics of the papers published with the 37 journals from 2001 to 2019 were reported. The CSSCI ranking of the 2017–2018 Journals has been greatly questioned, which was suspended from then on

**Table 3** The weights for each fitting journal

| Journals | Weight | |
|---|---|---|
|  | Current-year citations | Second-year citations |
| Journal of Finance and Economics | 0.032 | – |
| International Economic Review | 0.337 | 0.337 |
| Economic Research Journal | 0.200 | 0.220 |
| The Journal of World Economy | 0.348 | 0.383 |
| China Land Science | 0.083 | 0.060 |

synthetic CIE are close to the real value of CIE, except for the quantity of articles and current-year citations (per, 2001). The similarities of the two control variables in synthetic CIE and the real variables in CIE are much higher than that of the control variables in the average of 36 control journals, suggesting the two versions of synthetic CIE are very comparable with real CIE and can be used as its counterfactual.

The difference in the number of articles between synthetic CIE and real CIE is mainly due to the fact that the number of the article published with CIE has declined from 180 (from 2004 to 2011) to 120 (from 2016 to 2019) (see Fig. 2). For the case of current-year citations (per, 2001), the values of CIE fluctuated significantly from 2001

**Table 4** The comparison of observation value and fitted value for each control variable

| Variables | Real CIE | Synthetic CIE | | Average of 36 control journals |
|---|---|---|---|---|
| | | Model 1 | Model 2 | |
| The number of authors | 1.95 | 1.86 | 1.85 | 1.82 |
| The number of funds | 0.96 | 0.74 | 0.75 | 0.56 |
| The number of keywords | 4.04 | 3.95 | 3.95 | 3.91 |
| The number of references | 16.85 | 17.27 | 17.84 | 14.28 |
| The number of pages | 9.70 | 10.25 | 10.48 | 8.08 |
| The started year of publication | 1984 | 1979.51 | 1979.55 | 179.83 |
| The quantity of papers | 169.56 | 115.81 | 115.47 | 160.41 |
| The journals' rankings of CSSCI | 7.69 | 10.22 | 9.08 | 25.61 |
| Current-year citations (per) | 0.26 | – | 0.26 | 0.12 |
| Second-year citations (per) | 1.59 | 1.46 | – | 0.69 |
| Current-year citations (per, 2001) | 0.12 | 0.22 | – | 0.08 |
| Current-year citations (per, 2005) | 0.30 | 0.26 | – | 0.09 |
| Current-year citations (per, 2010) | 0.27 | 0.33 | – | 0.11 |
| Current-year citations (per, 2015) | 0.46 | 0.37 | – | 0.14 |
| Second-year citations (per, 2001) | 0.96 | – | 1.05 | 0.41 |
| Second-year citations (per, 2005) | 1.30 | – | 1.36 | 0.50 |
| Second-year citations (per, 2010) | 1.55 | – | 2.00 | 0.71 |
| Second-year citations (per, 2015) | 3.45 | – | 2.25 | 0.98 |

to 2002. Despite these differences, the fitted CIE before the implementation of the open data policy can well mirror the features of real CIE.
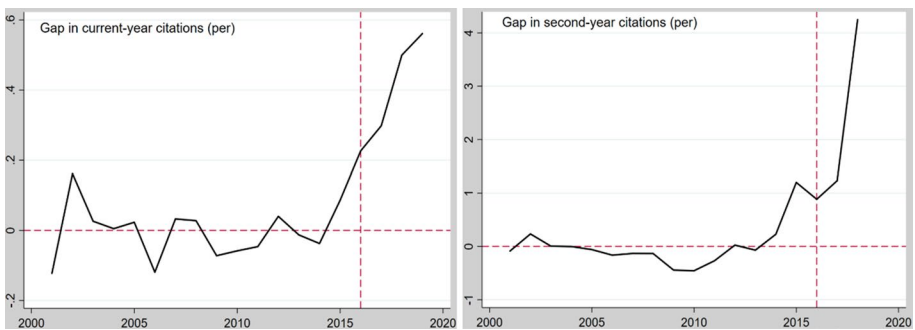
Both the current- and second-year citations of CIE quickly increased in 2015 but declined in 2016, which made real CIE and synthetic CIE deviated during this period of time. One of the reasons is that the "Made in China 2025" initiative was proposed by the State Council (the Chinese central government) on 25th March 2015 to drive the promotion of Chinese manufacturing industry (State Council of the People's Republic of China 2015). As a result, the articles related to Chinese manufactures were highly cited in 2015. There are nine articles with the "Made in China 2025" initiative in their titles published with CIE in 2015, and they received totally 339 citations. In contrast, the articles with similar titles published in 2013, 2014, 2016, and 2017 only received 11, 37, 15, and 121 citations respectively. For example, two of the nine articles (Wang and Li 2015; Huang and He 2015) received 82 and 76 citations respectively, but there are fewer citations on this topic after 2016.

In addition, the editorial department of CIE has speeded up the review process from 2015, replying to authors with initial reviews within two weeks, finishing the reviews and publishing articles in two to three months. The faster review process helped to improve the citations by attracting high-quality manuscripts and responding to hotspots.

At the same time, both real CIE and synthetic CIE got apparent negative value (real value is smaller than the synthetic value) from 2006 to 2013 in both current- and second-year citations, and the reason is that the number of articles published with CIE has quickly increased from 2004 to 2011 (from 150~160 to 180~210). The CSSCI journal rank also fluctuated sharply during this period of time. The delaying effect of citation led to the decline of current- and second-year citations between 2006 and 2013. However, after CIE
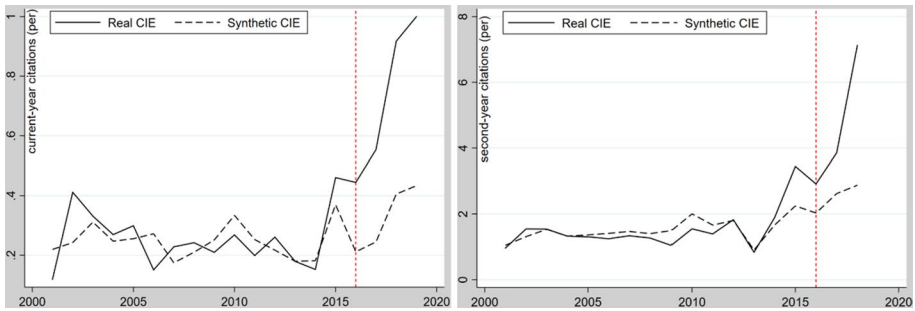
**Fig. 3** The trends in current-year citations (left) and second-year citations (right) of real CIE and synthetic CIE
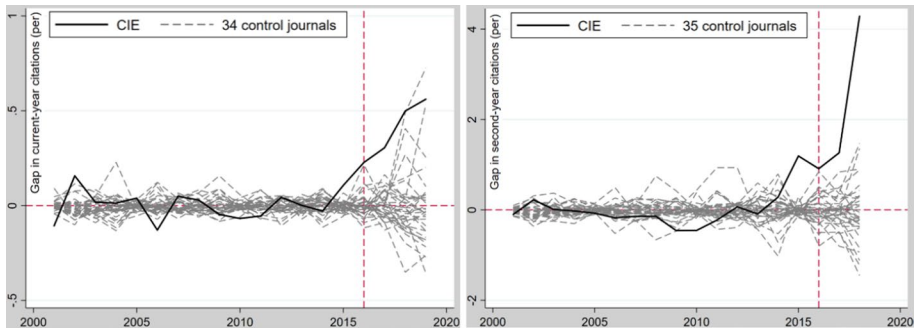


**Fig. 4** Current-year citations gap (left) and second-year citations gap (right) between real CIE and synthetic CIE

adopted policies to improve citations (e.g., to reduce the number of articles), the deviation has been substantially narrowed. These differences may noise our estimates, but generally we believe the results are robust.

## The SCM estimates

Figure 3 shows the comparison of current- and second-year citations between real CIE and synthetic CIE. Firstly, from the perspective of current-year citations, there have been bigger differences between real CIE and synthetic CIE after the open data policy from 2017, and the value of real CIE is apparently higher than the synthetic CIE. For example, the current-year citations of real CIE was 1.0 in 2019, which is 0.6 points higher than that of synthetic CIE (0.4) (Fig. 4).

Second, in the case of second-year citations, there is also a big difference between real CIE and synthetic CIE after the implementation of the open data policy. For example, the value of second-year citations for real CIE is 7.1 in 2018, which is 3.0 points higher than that of synthetic CIE (4.1). The results indicate that open data led to significant differences of both current- and second-year citations between real CIE and synthetic CIE, suggesting open data has significant policy effects on citations and H1 is supported.

**Fig. 5** Current-year citations gap in CIE and placebo gaps in 34 control journals (left) and second-year citations gap in CIE and placebo gaps in 35 control journals (right). *Note*: Journals with values of RMSPE higher than that of CIE were dropped

The open data policy has not only improved the citations of articles published after its implementation, but also generated a spillover effect by improving the whole citations of CIE articles. For example, the 2019 data for each journal show that the articles of CIE with data available increased citations (the average citations reached to 6 times for each article published in 2017 and 2018), and the citations of articles published prior to the open data policy also increased (the average citations are 4 and 6 times for each article published in 2014 and 2015/2016 respectively). The increments of citations for the articles published after 2014 were similar, while the citations of articles published before 2015 were not significantly affected. The results suggest that H2 is supported.

## Robustness checks

In order to verify the validity of the results (i.e., to verify the difference of the predictor variables in the empirical analysis is indeed caused by the open data policy and not by unobservable factors, and to verify the significant policy effect on citations), this article utilizes placebo test proposed by Abadie et al. (2010) which is similar to the Permutation Test in Rank Test to judge if there is other journal with same results of CIE and the probability of such possibility. To run this test, we first supposed all the other 36 control journals started with the open data policy in November 2016. We then constructed the corresponding composite control group with SEM to estimate the policy effect in the supposed conditions. Finally, we compared the real policy effect of CIE with the policy effect in the supposed conditions of 36 control journals. If there is a significant difference between the two policy effects, we can then believe that the policy effect of open data is significant.

Root Mean Square Prediction Error (RMSPE) is used to measure the difference between values predicted by a model and the values actually observed from the environment that is being modeled (Abadie et al. 2010). We used RMSPE to exclude the journals with unsatisfactory fitting effect before November 2016, and the bigger the value of RMSPE, the worse the fitting effect. We excluded the journals with the value of RMSPE higher than that of CIE, including *International Economic Review* and ERJ (for the case of current-year citations); ERJ (for the case of second-year citations). As shown in Fig. 5, CIE has the biggest treatment effect in both current- and second-year citations, and their probabilities are 0.0286 (1/35) and 0.0278 (1/36) respectively. Both values are smaller than 0.05, the level of significance. The results show that our estimates of CIE is reliable.
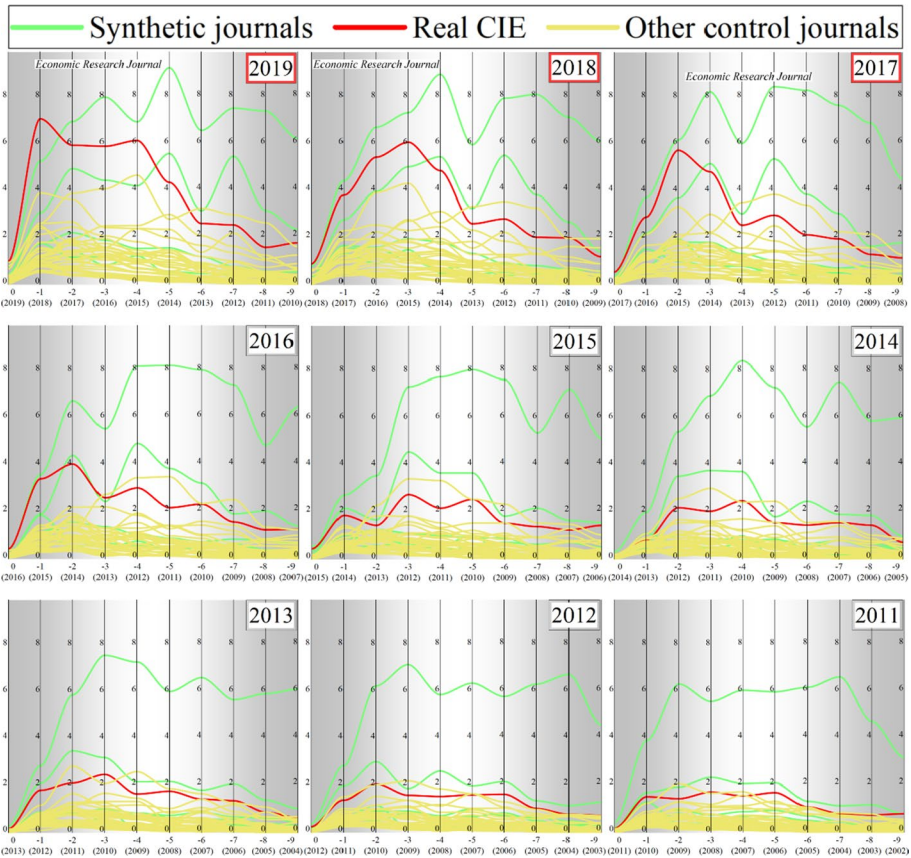
**Fig. 6** The annual citations per paper of synthetic journals, real CIE, and the control journals

At the same time, to further verify the results, we used the data of the 5 synthetic journals including ERJ, real CIE, and other control journals to compare their estimates (Fig. 6). The nine charts display the citations per article in each year for the corresponding journals. The longitudinal axis is the number of citations per article, and the horizontal axis is the year when the articles were published. The results show that the number of citations per article of CIE articles increased quickly after the implementation of the policy, and the values for the current- and second-year citations are close to or exceed that of ERJ, the best economics journal in China. For example, taking the result of 2019 as example, the value of CIE is higher than that of ERJ. In contrast, this pattern did not happen to the citation statistics before 2015, suggesting the results of this article hold.

## Discussion and conclusion

### 5.1 Research and policy implications

Open science is a general trend of the scientific enterprise, and open data is among its core components (Ule 2020). Our results in this article show that the open data policy significantly

improves the number of citations of articles in Chinese journals. Specifically, the current-year citations of CIE increased from 0.2~0.4 to 0.9~1.0 after the open data policy, which means it increased by around 1.5~4.0 times. Meanwhile, the second-year citations increased from about 1.0 to 4.0, and the magnitude of policy effect is similar to the case of current-year citations. The current- and second-year citations are the immediate response of journal impact, which suggests the journal impact of CIE has been quickly increasing and the policy effect of open data is remarkable. With the increasing impact, CIE will lead and speed up the open science of Chinese research in a predictable time frame, and academic dishonest would also be mitigated with the promotion of open science in China.

The results suggest that the effects of open data have spilled over to the articles published before the implantation of the policy, mainly due to the improved overall reputation of CIE. As a matter of fact, it is well-known that the quality of CIE has increased quickly in the field of economics, and prominent universities such as Renmin University of China have already taken CIE as one of the best economics journals in their faculty research evaluation. Given the current trend, CIE might surpass ERJ which is the best journal of Chinese economics. The success of CIE in open data may help to facilitate more journals to embrace open data, which will contribute to scientific integrity and development in China.

To encourage research transparency and replication, more and more social science journals have been requiring authors to share original datasets and analytic procedures supporting their publications. Despite the advocacy of open data, few studies have used causal inferences such as DID and SCM to rigorously examine its effects on journal impact. In this article, we report one of the first empirical studies to assess the effects of open data on journal impact. We not only revealed the direct effects of open data on current- and second-year citations of journal articles, but also probed its spillover effects due to the enhancement of journal reputation as a whole. Our findings thus contributed to the ongoing debates about the merits and challenges of open data by providing solid evidences and novel perspectives.

The findings reported in this article show that open data can boost journal impact, and journals can embrace open data to improve their reputation and generate broader impacts. Our results suggest that journals can leverage open data to improve research integrity and impact, which is a doable strategy for journals catching up with top journals. Prior studies reveal that open data might not be widely accepted by the voluntary approach, and few scientists are willing to share research data of their own free will (Eschenfelder and Johnson 2014). In contrast, mandatory requirements of journals in political science, economics, and psychology have substantially facilitated the acceptance of open data in the past decade. It is thus promising to persuade journals, publishers, funding agencies, and universities to mandate open data among their authors, grant receivers, and tenure-applicants. With the increasing advocacy and engagement of top-tier journals in these disciplines, open data would sooner or later becomes the gold standard or leading paradigm of publishing practices.

There is a gap in scientific research between China and the developed countries such as United States and European countries, and there are less high-impact articles and more academic fraud in the Chinese science. However, China has been catching up by embracing open science to improve scientific research quality and integrity, and the open data policy of CIE is one of the examples (Yang 2013). CIE is a pioneering journal in China to introduce open data, and its success revealed that open data is an effective instrument to improve journal reputation and impact. It is thus meaningful to generalize the successful story of CIE to other journals in China and other countries.

Despite that open data improves the academic impacts of journal articles, it is still challenging to persuade journals to mandate open data. It is thus difficult to generalize the

practices of CIE among other Chinese journals, despite our evidence suggests it works in boosting academic impacts of published articles. Given that authors have lots of choices in selecting target journals, editors are reluctant to adopt open data policies as first movers, otherwise they would lose authors who disagree with data sharing mandates. Funding agencies and societies can help mitigate journal editors' concerns by mandating their grant recipients and members to share their research data. For instance, the National Social Science Foundation of China has been funding most prestigious social science journals, and open data could be a precondition for receiving its grants.

Open data is found to improve citation performance of published articles, but its unintended consequences should also be considered. In the case of social sciences, original data accompanying published articles might include individual-level information, and people's privacy would be leaked if the data were linked with other sources of data. It is thus imperative for journals to develop formal procedures and capacities to scrutiny the data submitted by the authors to avoid data security concerns. Also, given the plague of misconducts in the Chinese academic community, authors might worry about the risks of cheating and abuse of their shared research data. Journals and other institutions should work together to protect authors' data rights, otherwise open data would not sustain in the long term.

### Limitations and future research avenues

Our research is limited in three aspects, and we call for future studies to extend our understanding of open data. First, our findings can be generalized to other social sciences, but to what extent they hold in natural sciences is up to future tests. Disciplines differ in their traditions and practices of open data, and the citation impacts of open data policies would differ and deserve future tests. Also, Chinese journals are different from international ones in many ways and we hope future replication and extension of our findings in other contexts.

Second, given the short window of the experiment, we can only assess the short-term effects of open data, and long-term effects could be examined in future studies. We believe that open data would generate more profound impacts in the long term, and our current effect estimates might be underestimated. Also, we only examined the citation advantages of open-data studies, while social and educational impacts of open data can be examined in future studies. Because the research data shared by authors can be used by various users (e.g., course instructors, students, policy consultants, journalists), open data might generate profound social and learning benefits.

Lastly, we call for future studies to explore the underlying mechanisms through which open data affects journals' impacts. We theoretically discussed the mechanisms of open data effect, but how they work can be empirically explored further. Also, we only examined the impacts of mandatory open data policy, and whether and how voluntary data sharing policies affect citation performance is an open question. Despite these limitations, as one of the first studies to use SCM to causally assess open data effects, our findings contribute to the literature by providing new evidence to promote open data among the academia.

### Appendix

See Table 5

**Table 5** The description of the 37 sample journals included in this study

| Journals | Affiliated institution | Research field |
|---|---|---|
| Accounting Research | Accounting society of China | Financial accounting, financial management, management accounting, corporate governance, auditing, etc |
| Asia–pacific Economic Review | Fujian Academy of Social Sciences | Asia–Pacific Aspects, Asia–Pacific Economic Cooperation, National Dedicated Lines, Taiwan-Hong Kong-Macao Economy, Opening to the Outside World, WTO Economics, Special Zone and Development Zone Economy, Development and Environment, Comparison Window, Financial Focus, Asia–Pacific Enterprises |
| Auditing Research | China Auditing Society | Study and implement the spirit of the 19th National Congress of the CPC, research topics, audit research, low-carbon evaluation system research, large-scale audit project organization method research, state-owned enterprise audit comparative research, etc |
| China Industrial Economics | Institute of Industrial Economics of CASS | Situation and outlook, national economic operation, industrial economy, economic system reform, enterprise operation and management |
| China Land Science | China Society and Technology Association, China Land Science Society, Chinese land surveying and Planning Institute | Land management, land planning, land use, land consolidation, land economy, land information, land law, real estate prices, land evaluation |
| China Rural Survey | Rural Development Institute, Chinese Academy of Social Sciences | Rural land; rural organizations and systems; rural labor; agricultural product production and circulation; rural construction and development; resources, environment and sustainable development; urbanization; finance and finance; rural poverty and peasant welfare; rural science and technology system; rural society; rural Governance; rule of law in rural areas; rural education; rural culture, etc |

**Table 5** (continued)

| Journals | Affiliated institution | Research field |
|---|---|---|
| Chinese Rural Economy | Rural Development Institute, Chinese Academy of Social Sciences | Rural situation and policy analysis, rural property rights system, rural land issues, rural human capital and labor transfer, agricultural input and growth, food issues, food quality and safety, sustainable development, rural finance and insurance, sectoral economy, regional economy, poverty issues, Farmers' income and consumption, rural urbanization, agricultural products circulation and trade, rural finance and public goods supply, rural organization system, agricultural technological progress, foreign agriculture and world agriculture |
| Comparative Economic and Social Systems | Central Compilation and Translation Bureau | Frontiers of reform theory, comparison of transition economies, comparison of regional development, enterprise reform, economic literature, research on political system reform issues, interviews with famous economists and entrepreneurs, leadership forums and corporate clubs, reform policy and countermeasure research, forward-looking Theoretical research and contention, reform case studies, etc |
| Contemporary Economic Research | Jilin University of Finance and Economics | "Capital", Political Economics, History of Economic Thought, Western Economics, etc |
| Contemporary Finance & Economics | Jiangxi University of Finance and Economics | Theoretical economy, public economy and management, modern finance, business administration, industrial economy and regional economy, finance and accounting, international trade and investment |
| Economic Perspectives | Institute of Economic Chinese Academy of Social Sciences | New Theory of Economic Science, Economic Hotspot Analysis, Macroeconomic Discussion, Sectoral Economy, Regional Economy, Financial Research, Conference Summary, Academic Materials, Economic System Reform, Enterprise Management, Surveys and Recommendations, Chinese and Foreign Academic Exchanges, Foreign Economic Theory, Overseas Economy Academic Expo, World Economy, Book Review, etc |

**Table 5** (continued)

| Journals | Affiliated institution | Research field |
|---|---|---|
| Economic Research Journal | Institute of Economic Chinese Academy of Social Sciences | Expert forum, hot topics, economic theory, finance and taxation, enterprise and development, new rural construction, northeast old industrial base, research on agriculture, rural areas and farmers, financial securities, marketing, accounting research, labor economy, enterprise management, regional economy, management World, economic and trade exploration, economy and law, income and distribution, tourism economy, social security research, industrial economy, group economy, urban economy, circular economy, ecological economy, legal analysis, humanistic exploration, educational perspective, etc |
| Economic Review | Wuhan University | Economic theory research, Chinese economic research, industrial economics research, western economics research, etc |
| Economic Review Journal | Jilin Academy of Social Sciences | Comparison and innovation, theoretical discussion, macroeconomics, reform and opening-up forum, regional economy, etc |
| Economic Science | Peking University | Reform aspect, economics and economic theory, western development, securities and investment, financial research, corporate economy, business and trade research, fiscal and tax research, etc |
| Economic Theory and Business Management | Renmin University of China | Theoretical exploration, academics, economic hotspots, public economy, financial research, industrial economy, regional economy, new rural construction, business management, international economy, development and overview, etc |
| Economist | Southwestern University of Finance and Economics; Sichuan Social Science Academic Foundation (Xinzhi Research Institute) | China Economics Forum, theoretical longitude and latitude, reform exploration, comparison and reference, research reports, financial research, fiscal and tax research, agricultural research, scholar style, new observations, etc |
| Finance and Economics | Southwestern University of Finance and Economics | Financial forums, corporate economy, theoretical economy, macroeconomics, economic latitude, and agriculture |

**Table 5** (continued)

| Journals | Affiliated institution | Research field |
| --- | --- | --- |
| Finance and Trade Economics | National Academy of Economic Strategy | Finance and taxation, finance and currency, international economy and trade economy, city and real estate, cost and price, circulation economy, industrial economy, service economy, tourism economy, information and e-commerce, etc |
| International Economic Review | Institute of World Economics and Politics Chinese Academy of Social Sciences | Time comment, contention, monograph, finance, international politics, Chinese economy, region, peers |
| Intertrade | China Commerce and Trade Press | China's economy and trade, international business, regional cooperation, foreign investment, finance, service trade, laws, materials, etc |
| Issues in Agricultural Economy | Chinese Association of Agricultural Economics; Institute of Agricultural Economics and Development, Chinese Academy of Agricultural Sciences | Urbanization and industrialization, farmers' income and structural adjustment, agricultural development, rural society, factor allocation, market and trade, organization and system, industrial economy, contention and discussion, sustainable development, foreign agricultural economy |
| Journal of Agrotechnical Economics | Chinese Association of Agricultural Economics; Institute of Agricultural Economics and Development, Chinese Academy of Agricultural Sciences | Economic theory and its application, scientific and technological progress, benefit evaluation, industrial economy, regional economy, doctoral forum, etc |
| Journal of Finance and Economics | Shanghai University of Finance and Economics | Public economy and management, finance and accounting research, economic history and economic thought history research, financial research, regional economic research, industrial economic research, international economic research, etc |
| Journal of Financial Research | China Society for Financeand Banking | Financial banking system, policies and explanations, banking system and business, financial shopping malls, credit and investment, insurance theory, commercial banks, rural credit cooperatives, financial banking theory, etc |

**Table 5** (continued)

| Journals | Affiliated institution | Research field |
| --- | --- | --- |
| Journal of International Trade | University of International Business and Economics | Economic and trade forums, business strategies, global value chains, "Belt and Road" topics, international investment and transnational operations, trade and environment, service trade, trade barriers, international finance, international economic law, academic frontiers, etc |
| Macroeconomics | Academy of Macroeconomic Research | Situation analysis, theoretical research, industrial development, response to WTO entry, policy recommendations, world economy, etc |
| Modern Economic Science | Xi'an Jiaotong University | Financial research, macroeconomic research, industrial economic research, business management research, business enterprise management research, trade economic research, agricultural economic research, rural economic research, business enterprise management, commercial economic research, fiscal and tax research, socialist research with Chinese characteristics, etc |
| Nankai Economic Studies | School of Economics, Nankai University | Finance, international economy, transformation, Chinese economy, macroeconomics, basic economic theory, political economy research, etc |
| Reform | Chongqing Academy of Social Sciences | Timely review of reform, capital market discussion, international economic review, labor market research, SME reform, private enterprise research, comparison and reference |
| Research on Economics and Management | Capital University of Economics and Business | Macroeconomics, industrial economy, finance and insurance, public economy and management, enterprise management, corporate governance, labor economy and human resources, rural economy, etc |
| Research on Financial and Economic Issues | Dongbei University of Finance and Economics | Economic observation, theoretical research (name column of the Ministry of Education), industrial organization, finance and investment |

**Table 5** (continued)

| Journals | Affiliated institution | Research field |
|---|---|---|
| Shanghai Journal of Economics | Institute of Economics, Shanghai Academy of Social Sciences | New Trends in China's Economy, Economic Development Report, New Trends in Shanghai's Economy, Global Observatory, Legends of Overseas Chinese Businessmen, Trend Theory, etc |
| Studies of International Finance | Bank Of China Ltd., China Institute of International Finance | Financial theory and policy, global finance, banking research, financial markets, etc |
| Taxation Research | China Tax Magazine | Fiscal and tax hotspots, difficulties, focus issues and international tax comparisons |
| The Journal of World Economy | China Society of World Economics; Institute of World Economics and Politics Chinese Academy of Social Sciences | Macroeconomics, microeconomics, international trade, international finance, regional and country studies, Chinese economy, capital markets, etc |
| Word Economy Studies | Institute of World Economy, Shanghai Academy of Social Sciences | Exploring the Road to a Powerful Country, Research on Powerful Companies, Regional and Country-specific Studies, International Finance, International Trade, International Investment, Analysis of International Relations, International Enterprise Forum, etc |

The data are from the scope of theme of each journal. The journals are listed alphabetically

# References

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review, 93*(1), 113–132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science, 59*(2), 495–510.

Anonymous. (2005). Let data speak to data. *Nature, 438*(7068), 531.

Anonymous. (2015). Data access and research transparency (DA-RT): A joint statement by political science journal editors. *European Union Politics, 16*(3), 323–324.

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

Bolukbasi, B., Berente, N., Cutcher-Gershenfeld, J., Dechurch, L., Flint, C., Haberman, M., et al. (2013). Open Data: Crediting a Culture of Cooperation. *Science, 342*(6162), 1041–1042.

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

Brown, C. (2003). The changing face of scientific discourse: Analysis of genomic and proteomic database usage and acceptance. *Journal of the American Society for Information Science and Technology, 54*(10), 926–938.

Burgess, S., Cheng, L. Z., Gu, F., Huang, J. J., Huang, Z. W., Lin, S., et al. (2016). Questions about ngago. *Protn and Cell, 7*(12), 913–915.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., et al. (2002). Data withholding in academic genetics- Evidence from a national survey. *JAMA-The Journal of the American Medical Association, 287*(4), 473–480.

Campbell, E. G., & Bendavid, E. (2002). Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. *Journal of Health Care Law & Policy, 6*(2), 241–255.

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.

Christensen, G., Dafoe, A., Miguel, E., Moore, D. A., & Rose, A. K. (2019). A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLoS ONE, 14*(12), e0225883.

Clarivate Analytics. (2012). Data Citation Index. Retrieved Dec 30, 2020, from https://clarivate.com/webof sciencegroup/solutions/webofscience-data-citation-index/

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS ONE, 15*(4), e0230416.

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.

Dryad. (2011). Joint Data Archiving Policy (JDAP). Retrieved Dec 30, 2020, from https://datadryad.org/docs/ JointDataArchivingPolicy.pdf

Eschenfelder, K. R., & Johnson, A. (2014). Managing the data commons: Controlled sharing of scholarly data. *Journal of the Association Society for Information Science and Technology, 65*(9), 1757–1774.

Fu, T., & Chen, M. (2014). The analyses of and suggestions for data publishing policies of academic journals in China (in Chinese). *China Publishing Journal, 12,* 31–34.

Gao, F., Shen, X. Z., Jiang, F., Wu, Y. Q., & Han, C. Y. (2017). DNA-guided genome editing using the natronobacterium gregoryi Argonaute (Retraction of Vol 34, Pg 768, 2016). *Nature Biotechnology, 35*(8), 979.

Giofre, D., Cumming, G., Fresc, L., Boedker, I., & Tressoldi, P. (2017). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. *PLoS ONE, 12*(4), e0175583.

Gotzsche, P. C. (2012). Strengthening and opening up health research by sharing our raw data. *Circulation Cardiovascular Quality & Outcomes, 5*(2), 236–237.

Gymrek, M., & Farjoun, Y. (2016). Recommendations for open data science. *Gigaence, 5,* 22.

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

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.

Huang, Q. H., & He, J. (2015). The core capability, function and strategy of Chinese manufacturing industry——comment on 'Chinese manufacturing 2025.' *China Industrial Economics, 6,* 5–17.

Hvistendahl, M. (2013). China's publication bazaar. *Science, 342*(6162), 1035–1039.

Ingwersen, P., & Elleby, A. (2011). Do Open Access Working Papers Attract more Citations Compared to Printed Journal Articles from the same Research Unit? In E. Noyons, P. Ngulube, & J. Leta (Eds.), *Proceedings of ISSI 2011 Durban: 13th International Conference of the International Society of Scientometrics and Informetrics* (pp. 327–332), Leuven: Int Soc Scientometr & Informetr (ISSI).

International Association of Scientific, Technical and Medical Publishers (STM). (Jan, 2020). STM 2020 Research Data Year. Retrieved Dec 30, 2020, from https://www.stm-assoc.org/standards-technology/2020-stm-research-data-year/

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

Javidi-Parsijani, P., Niu, G. G., Davis, M., Lu, P., Atala, A., & Lu, B. S. (2017). No evidence of genome editing activity from Natronobacterium gregoryi Argonaute (NgAgo) in human cells. *PLoS ONE, 12*(5), e0177444.

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.

Khin, N. C., Lowe, J. L., Jensen, L. M., & Burgio, G. (2017). No evidence for genome editing in mouse zygotes and HEK293T human cell line using the DNA-guided Natronobacterium gregoryi Argonaute (NgAgo). *PLoS ONE, 12*(6), e0178768.

Krumholz, H. M. (2012). Open science and data sharing in clinical research basing informed decisions on the totality of the evidence. *Circulation-Cardiovascular Quality and Outcomes, 5*(2), 141–142.

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.

Lee, S. H., Turchiano, G., Ata, H., Nowsheen, S., Romito, M., Lou, Z., et al. (2016). Failure to detect DNA-guided genome editing using Natronobacterium gregoryi Argonaute. *Nature Biotechnology, 35*(1), 17–18.

Leonelli, S. (2016). Open data: Curation is under-resourced. *Nature, 538*(7623), 41.

Mallapaty, S. (Aug 21, 2020). China's research-misconduct rules target 'paper mills' that churn out fake studies. Retrieved Dec 30, 2020, from https://www.nature.com/articles/d41586-020-02445-8

Martone, M. E., Garcia-Castro, A., & VandenBos, G. R. (2019). 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.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., et al. (2016). How open science helps researchers succeed. *Elife, 5,* e16800.

Molloy, J. C. (2011). The open knowledge foundation: open data means better science. *Plos Biology, 9*(12), e1001195.

Murray-Rust, P. (2008). Open data in science. *Serials Review, 34*(1), 52–64.

Murray-Rust, P., Neylon, C., Pollock, R., & Wilbanks, J. (Feb 19, 2010). Panton Principles, Principles for open data in science. Retrieved July 20, 2020, from https://www.pantonprinciples.org/

National Institutes of Health (NIH). (Mar 1, 2002). NIH Announces Draft Statement on Sharing Research Data. Retrieved July 20, 2020, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-035.html

Peng, L., & Han, Y. (2019). The analyses and implications of data policies of China's scientific journals–- The case of English journals published by the Chinese Academy of Sciences (in Chinese). *Chinese Journal of Scientific and Technical Periodicals, 30*(8), 870–877.

Peng, R. D. (2011). Reproducible research in computational science. *Science, 334*(6060), 1226–1227.

Piwowar, H. A. (2011). Who shares? who doesn't? factors associated with openly archiving raw research data. *PLoS ONE, 6*(7), e18657.

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.

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: Data sharing in neuroimaging. *Nature Neuroscience, 17*(11), 1510–1517.

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

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., Lehman, R., & Gross, C. P. (2012a). The importance of clinical trial data sharing: Toward more open science. *Circulation Cardiovascular Quality and Outcomes, 5*(2), 238–240.

Ross, J. S., Lehman, R., & Gross, C. P. (2012b). The importance of clinical trial data sharing: Toward more open science. *Circulation-Cardiovascular Quality and Outcomes, 5*(2), 238–240.

Ross, J. S., Mulvey, G. K., Hines, E. M., Nissen, S. E., & Krumholz, H. M. (2009). Trial publication after registration in clinicaltrials gov: A cross-sectional analysis. *PLoS Medcine, 6*(9), e1000144.

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

Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). Open data in global environmental research: The Belmont Forum's open data survey. *PLoS ONE, 11*(1), e0146695.

State Council of the People's Republic of China. (2015). The announcement of "Made in China 2025". Retrieved Dec 30, 2020, from http://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm.

Tang, L. (2019). Five ways china must cultivate research integrity. *Nature, 575*(7784), 589–591.

Tang, L., Hu, G. Y., Sui, Y., Yang, Y. H., & Cao, C. (2020). Retraction: The "Other Face" of research collaboration? *Science and Engineering Ethics, 26*(3), 1681–1708.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE, 6*(6), e21101.

The Royal Society. (Jun 21, 2012). Final report - Science as an open enterprise. Retrieved July 20, 2020, from https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/

Ule, J. (2020). Open access, open data and peer review. *Genome biology, 21*(1), 86.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (Apr 20, 2020). Big Data Platforms for a Global Pandemic. Retrieved Dec 20, 2020, from http://www.unesco-hist.org/index.php?r=en/article/info&id=1551

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

van Vlokhoven, H. (2019). The effect of open access on research quality. *Journal of Informetrics, 13*(2), 751–756.

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., et al. (2013). Mandated data archiving greatly improves access to research data. *Faseb Journal, 27*(4), 1304–1308.

Vogeli, C., Yucel, R., Bendavid, E., Jones, L. M., Anderson, M. S., & Louis, K. S. (2006). Data withholding and the next generation of scientists: Results of a national survey. *Academic Medicine, 81*(2), 128–136.

Wang, L., & Li, H. Y. (2015). Research on GVCs intergrating routes of China's manufacturing industry——perspectives of embedding position and value-adding capacity. *China Industrial Economics, 2,* 76–88.

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

Whitlock, M. C., McPeek, M. A., Rausher, M. D., Rieseberg, L., & Moore, A. J. (2010). Data archiving. *American Naturalist, 175*(2), 145–146.

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

Xin, H. (2009). Retractions put spotlight on China's part-time professor system. *Science, 323*(5919), 1280–1281.

Yang, W. (2013). Research integrity in China. *Science, 342*(6162), 1019.

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

Zhao, W. J. (2018). Open data for better science. *National Science Review, 5*(4), 593–597.

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.