Publication Cultures and the Citation Impact of Open versus Closed Access

By
Thomas Eger[1], Armin Mertens[2], and Marc Scheufen[3]

Abstract
Does open access to journal articles foster citations to these articles? We compare the citation impact of gold and green open access in two disciplines: Biology, and Economics & Management. The empirical analysis covers all articles of these disciplines included in the Web of Science "Journal Citation Reports" between 2000 and 2019. We show that, controlling for confounding variables pertaining to the journals and articles, gold OA increases citations across all articles. However, the individual disciplines feature starkly different effects: a 18.3% increase in biology, compared to a decrease by 30.9% in economics & management. Green OA increases citations to all articles by 82.2%. These results are confirmed by a number of robustness checks.

JEL Codes: L17, O33

Keywords: Open access, citations, academic journals, publication cultures

Content

[1] Thomas Eger, University of Hamburg, thomas.eger@uni-hamburg.de.
[2] Armin Mertens, German Economic Institute (IW, Cologne), armin.mertens@iwkoeln.de.
[3] Marc Scheufen, German Economic Institute (IW, Cologne) and Ruhr-University Bochum, scheufen@iwkoeln.de / marc.scheufen@rub.de.

1. Introduction

With the advent of electronic publishing and the internet, the traditional business model of academic publishing, based on subscription fees paid by the readers/libraries (Closed Access or CA journals), has to some extent been replaced and to some extent complemented by different types of open access (OA). Regarding scholarly articles, two broad models of OA can be distinguished. *Gold OA* refers to scholarly journals that are financed not by subscription fees but by fees paid by the authors or their sponsors (so-called article-processing charges – APC). Readers have free online access to such articles. A special case are *hybrid OA journals*, where authors can either pay the APC and make their articles freely available, or not pay the APC, which means that readers will have to pay to access the full text. Issues of such journals typically comprise articles that are freely available as well as articles with a paywall.

The second broad model, *Green OA*, refers to the posting of papers with a repository prior to or after their publication in a traditional journal. We may distinguish between institutional repositories, e.g. at university websites, and subject-based repositories, such as PubMed Central for biomedical and life sciences, arXiv with a focus on physics, RePEc for economics, and SSRN for the social sciences. Moreover, some authors post versions of their articles on their individual websites. The latter option, however, entails considerably larger search costs for potential readers compared to postings on institutional or subject-based websites.

The rise of the different types of OA triggered a discussion as to whether OA will foster citation counts, or more precisely: Do OA articles receive systematically more citations than comparable CA articles? The answer is important to the authors of scholarly articles, especially in the natural and social sciences, as their reputation strongly hinges on citation counts – to some extent on their individual citation counts, but in particular on the average citation counts of the journals they publish in, as a proxy of the reputation of those journals within the corresponding discipline.[4]

In the following, we begin with a brief survey of the most important literature on the above question. Thereafter, we present the results of our own empirical research, focusing on systematic differences between two publishing cultures: a predominantly *gold culture*, such as Biology, which is characterized by strong OA journal (gold OA) usage, and a predominantly *green culture*, such as Economics & Management, which features few OA journals but extensive use of OA repositories (green OA). The empirical results confirm our main hypothesis that the differences in publishing cultures and the resulting differences in citation counts to articles published in OA journals are due to the self-reinforcing perceptions of the reputation of OA journals in different disciplines.

2. Survey of the literature

In the last twenty years, many contributions have measured the impact of open access on citation counts.[5] Using cross-sectional data, early studies found considerable OA citation advantages of up to several hundred percent. Here are a few examples: *Lawrence* (2001) analyzed 119,924 peer-reviewed conference articles in computer science and related areas published between 1989 and 1999, part of which were openly accessible online, while the others were available only in print. Focusing on 1,494 venues that contained at least five offline and five online articles, he found an average OA citation advantage of 336% (median of 158%).

[4] One should also keep in mind that while citation is important for academic authors, many academic articles are also read and used by a broader public, without easy access to academic libraries. OA may thus increase the dissemination of content beyond what we may see in the citation counts.

[5] See the excellent survey by *Craig et al* (2007). More recent surveys are provided by *Swan* (2010), *Davis/Walters* (2011) and *Lewis* (2018).

Harnad/Brody (2004) were the first to study the effects of Green OA.[6] They compared 95,012 journal articles in physics and mathematics that were deposited as pre-prints in arXiv, an important natural science repository, to all the other 14 million articles that were published between 1992 and 2001 in the same journal and the same year as each of the Green OA papers. They found average OA citation advantages between 250% and 580% across the years of publication. *Antelmann* (2004) analyzed 2,017 randomly selected articles published between 1999 and 2002 in high-impact journals across four disciplines (mathematics, engineering, political science, philosophy). Within this sample, she compared the citation counts of those articles whose full text is freely available online at a location other than the publisher's website to the CA articles, i.e. all others. The mean OA citation advantage ranges from 45% in philosophy to 91% in mathematics. *Hajjem et al* (2005) automatically analyzed more than 1.3 million articles across ten disciplines published between 1992 and 2003. Between 5% and 16% of the articles were OA, depending on the discipline, year and country. The authors compared the citation counts of the OA articles, i.e. those whose full text was freely available online to the CA articles, i.e. articles that were only available in the subscription journal. They found average OA citation advantages between 36% and 172%, depending on the discipline.
All of these studies assume that the differences in citation counts are primarily due to the articles' access status (see also *Swan* 2010, 1). However, this assumption is far from plausible (*Craig* et al 2007, 17). As a matter of fact, most citing authors are academics, who can access most of the journals offered by their institutions' libraries from their home computers. Consequently, a potential citing author does not care much whether her library has paid a subscription fee for a certain article or whether this article is freely available on the internet, perhaps financed by an APC. All that matters is that the article is available somehow. However, with about 33,100 peer-reviewed English-language journals plus a further 9,400 non-English-language journals collectively publishing over 3 million articles a year as of mid-2018, even the richest universities cannot afford to subscribe to all journals.[7] With this in mind, rather than a general citation advantage of OA, we would expect a selective citation advantage, for two reasons. First, an OA citation advantage may persist with respect to articles published in less prominent journals that are not available in many libraries, and with respect to poorer universities with tight budgets and limited subscription potential.[8] Second, if the full text is not available, authors sometimes cite an article purely on the basis of the abstract. Facilitating access to the full text may lead to more informed citations, i.e. citations to lower-quality articles may be replaced with citations to articles of higher quality.[9] We would expect that article citation counts are in particular driven by factors such as the relevance of the topic (how many other authors deal with similar research questions?) and by the quality of the article (author reputation, journal ranking).
Measuring the impact of OA on citation counts is anything but trivial. Ideally, one would compare otherwise identical articles that only differ in their access status. Some studies therefore compare OA and CA articles in the same (issue of the same) journal, such as OA and CA articles in hybrid journals. This approach presumes that the articles in a given journal are sufficiently homogeneous. However, this is unlikely to be the case. Even in the top journals

---

[6] See also the summary in *Craig et al* (2007, 9).
[7] See https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf. For 2015, Ulrich's journal database Ulrichsweb even listed 111,770 active scholarly journals: 66,734 peer-reviewed, 47,826 available online, 15,025 open access, and 10,916 included in Thomson-Reuters Journal Citation Reports. See *Moed* (2017, 194).
[8] The main driver of the OA movement was the so-called serials crisis, i.e. the drastic increase in the subscription fees for academic journals, which increased the financial pressure on the libraries (see e.g. *Eger & Scheufen*, 2018). This development induced many libraries to forego some journals and, in particular, many books. The OA citation advantage could therefore also apply especially to citations to book chapters that are posted in OA repositories.
[9] See the contribution by *McCabe &Snyder* to this special issue.

with high average citation counts, only a few articles are cited very often, while most receive only a few citations, and many are never cited at all.[10] Thus, the distribution of citation counts is typically extremely skewed, which impairs the reliability of citation count averages across OA and CA articles within the same journal. Other studies refer to Green OA and compare articles in traditional journals with and without pre- or post-prints posted in repositories. However, these articles are typically not homogeneous either as they are made public at different points in time and may differ in terms of content. Many publishers as the copyright holders condition their consent to Green OA after publication on whether the archived article refers to the published, the accepted or the submitted version, and they may require embargo periods (https://v2.sherpa.ac.uk/romeo).

In more recent studies of the determinants of citation counts, the OA status is only one of several explanatory variables. As a result, the OA citation advantage is typically much lower than in the early studies. In particular, two causes of this discrepancy are discussed.[11] The first is *selection bias*: Some studies assume that more prominent authors are more likely to publish their article OA, and/or authors generally prefer to publish their better works open access. Selection bias implies reverse causality: Instead of the OA status causing citation counts, the perceived quality of the article and thereby the expected citation counts cause the probability of an article being published OA. To remedy this problem, the analysis should also control for author prominence, which could for example be proxied by prior citation history, publications in highly ranked journals, and the funding organization. To measure the quality of an article independently of its citation count is a challenge. Moreover, what is the rationale for this bias? This is by no means a trivial question. When an article that is (to be) published in a traditional journal can be made OA as a post-print or pre-print without additional costs to the author, why should they decide against OA? A quality bias could arise only if the repository employs some kind of quality filter, but it would not be based on the authors' self-selection. When an author is wondering whether to submit an article to a traditional journal or an OA journal, they will typically first of all consider the reputation of the journals. OA journals enjoy high reputation in some disciplines (e.g. Biology) but not in others (e.g. Economics).[12] In disciplines where OA is highly regarded, if the author's or the sponsor's APC budget is restricted, authors will tend to reserve their higher quality papers for OA journals – so we have a selection bias. The same holds when an article is accepted by a hybrid OA journal and the author must decide whether to pay the APC and publish OA. By contrast, when the author's institution automatically pays APCs for any article accepted by specific journals – as is for example the case with the members of the DEAL project in Germany, two big deals between a consortium of German research institutions and Wiley and Springer Nature – then this potential cause of selection bias vanishes.

A second potential source of discrepancy refers to the time when an article is first published. In some disciplines with long delays between submission and final publication of an article, it is common practice to post a pre-print in a repository. If the pre-print is sufficiently similar to the published paper, the citation counts are subject to an *early view bias*, which must be distinguished from any effect of the OA status: Articles whose pre-prints are freely available long before their publication in a journal may receive more citations not because of their OA status per se but rather because of the early availability of the pre-print – in other words, due to simple *timing*. Thus, for a given window of time during which citations are counted, the longer the period between the upload of the manuscript and the publication of the final paper, the stronger the early view bias, which will also depend on the specific citation culture in each discipline. In some disciplines, such as traditionally Health Science and Biology, many journals

[10] *Haustein & Larivière* (2015, 6) state that "as a rule of thumb, 80% of citations are received by 20% of documents and many are never cited, especially in the humanities".

[11] See in particular *Kurtz et al* (2005).

[12] See *Eger & Scheufen* 2018. As an extreme case, so-called predatory OA journals try to exploit inexperienced authors by charging APC without providing adequate publisher services, such as quality control ("Beall's list").

used to follow the 'Ingelfinger rule', which bans the publication of articles that have previously been published elsewhere, so that pre-prints were typically not available. However, the application of the Ingelfinger rule appears to be eroding.[13] Journal publishers as the copyright-holders restrict the self-archiving of post-prints in several ways (see the Sherpa/Romeo database[14]). Publishers often impose an embargo period of between 6 months and several years, which implies a *late view bias*, and not an early view bias, for the OA version – provided no self-archived pre-print is available. When comparing OA and CA articles in the same issue of a hybrid OA journal, there is obviously no early or late view bias.

These methodological limitations apply not so much to more recent studies of the OA citation advantage. *Eysenbach* (2006) conducted a longitudinal bibliometric analysis of 1,492 articles published in the second half of 2004 in the "Proceedings of the National Academy of Sciences", a multi-disciplinary hybrid OA journal. 14.2% of these articles were published OA. Eysenbach's logistic regression model controls for potentially confounding variables, such as the number of authors, the authors' lifetime publication counts and impact, submission track (with different levels of rigor in peer review), the country of the corresponding author, funding organization, and discipline. He found that OA articles are 1.7 to 2.9 times more likely to be cited, depending on the period over which the citations were counted (0-6, 4-10, or 10-16 months after publication). Furthermore, the number of authors and funding from competitive grants significantly predict the likelihood that an article is cited at least once within the applicable time window. A secondary analysis took into account that CA articles may also be self-archived in OA repositories, with the result that higher levels of openness are associated with higher citation counts.
Moed (2007) analyzed 74,521 articles on condensed matter physics deposited in arXiv between 1992 and 2005. Most of these articles were published in one of 24 condensed matter physics journals. He compared the citation counts *with* OA (10.2% of the sample), i.e. to articles that were available both as the arXiv version and as the final journal version, to those *without* OA, i.e. all other articles in these journals. Controlling for the *early view effect* by imposing specified windows of time for counting citations and for the strong *selection effect*, i.e. prominent authors are over-represented in articles deposited in arXiv, he found a small median OA citation differential 4-6 years after publication: 7% for more productive and 14% for less productive authors (24% and 29%, respectively, 1-3 years after publication).
Evans & Reimer (2009) investigated more than 26 million articles and the associated citations across 14 disciplines from the 8,253 most highly cited journals that came online since 1998. The citation data ends 2005. The authors distinguish between commercial online availability and free online availability (i.e. OA). Controlling for commercial online availability, the average citation advantage of OA journals remains modest (about 8%). The influence of OA was more than twice as strong in the developing world than in the wealthy Northern and Western Hemisphere. However, it was weaker in the very poorest countries with limited electronic access.
Gargouri et al. (2010) analyzed 27,197 articles from 1,984 CA journals in several disciplines over the period 2002 to 2006. They compared citation counts (within the same journal/year) of CA articles to those of articles that were made OA as refereed post-prints, either by voluntary self-archiving or by mandatory self-archiving in the institutional repositories of four institutions. 6,215 of the articles originate from institutions with mandatory OA policies and 20,982 from institutions without such mandates. About 15% of the articles from institutions without mandate were voluntarily self-archived, whereas about 60% were self-archived when it was mandated. The logistic regression analysis controlled for article age, journal impact

[13] https://science.sciencemag.org/content/357/6358/1344; https://polecopub.hypotheses.org/2017#identifier_1_2017.
[14] https://v2.sherpa.ac.uk/romeo.

factor, number of co-authors, references, pages, field, article type and country. It found that OA articles are cited significantly more than CA articles. Furthermore, there is no evidence that the citation advantage is smaller for mandated OA than for self-selected OA. The authors conclude that it is "highly unlikely that the OA advantage is either entirely or mostly the result of an author bias toward selectively self-archiving higher quality – hence higher citability – articles" (p. 9).

Gaulé & Maystre (2011) analyzed a sample of 4,388 biology papers published between May 2004 and March 2006 by the "Proceedings of the National Academy of Sciences", a hybrid OA journal. They compared the citation counts over the two years for those articles that were made OA in return for an APC of 1.000 USD (17% of the sample) to the remaining articles (CA).[15] The analysis controls for quality differences (selection effect) by taking into account characteristics of the last authors (typically the principal investigator in natural science) and article quality,[16] as well as for the availability of funds for the APC (proxied by publication in the last quarter of the financial year) and several other factors. The authors failed to find a statistically significant difference in the citation counts.

Mueller-Langer & Watt (2014) also studied OA citation advantages with reference to hybrid OA journals. They analyzed 1,329 articles published between January 2000 and December 2011 in 15 hybrid OA economics journals. 208 of the articles were published OA under specific HOA pilot agreements between two commercial publishers and several research institutions. Performing Poisson quasi-maximum likelihood regressions and controlling for a range of factors, they found an OA citation advantage of 22% to 26%. However, when institution quality and citations to RePEc pre-prints were additionally controlled for, the significance vanished.

Davis (2011) used a randomized controlled trial of 3,245 research articles and reviews published between January 2007 and February 2008 in 36 journals in various disciplines to explore the effects of free access on article downloads and citations. 712 articles (22%) were randomly assigned to the open access treatment group. The control group comprised the remaining 2,533 articles, which were only available to subscribing individuals and institutions.[17] The effects of the open access treatment on article downloads and citations was estimated by multivariate linear regression models that controlled for journal, article type and length, number of authors, and self-archiving of the articles. While making an article open access doubled the number of full-text downloads in the first year after publication, there was no effect on citation frequency within the first three years. Davis explains these results by social stratification: While the typical citing authors are affiliated with a small number of elite research universities with excellent access to the scientific literature and therefore do not depend on open access, many readers do not belong to the core research community. The latter are the real beneficiaries of open access to scientific literature.

McCabe & Snyder (2014) analyzed a sample of 100 top journals in Ecology, Botany, Multidisciplinary Science and Biology between 1996 and 2005. They distinguished between *open access* and *online-availability*, on the journal's own website or one of the major digital aggregators, such as JSTOR, EBSCO, ProQuest, Ingenta, Gale, and OCLC. The dataset includes over 200,000 cited articles with 4.8 million cites contained in the 8,000+ ISI-indexed journals. The authors focus on the volume level, i.e. all articles that a journal publishes in a given year. While for about 55% of the observations, the full volume was available online for the full year, only 6% of the observations were freely available for the full volume for the full year. Citations to the average journal volume peak in the second year after publication, and there is a significant secular upward trend in the sample citations. While the online availability of the sample grew quickly, reaching about 80% in 2005, open access grew only slowly (10%

[15] The CA articles are made open after 6 months.

[16] As a crude proxy for article quality, the authors used the evaluation given by biology professors on the website F1000 Biology (81% not evaluated, 12% recommended, 6% must read, and 1% exceptional).

[17] Many participating journals made all articles freely available after some embargo period.

in 2005). To control for quality differences, the authors used a Poisson estimator with volume fixed effects and a journal-specific quadratic age profile of citation counts. They obtained an OA citation advantage of 8.1%, the benefit being concentrated among the top-ranked journals, whereas the bottom-ranked journals suffer a statistically significant reduction in cites. The authors explain this result by intensified competition for the readers' attention, which benefits high-quality articles at the detriment of low-quality articles.

In another similar study, but with a focus on the impact of *online availability* on citations, *McCabe & Snyder* (2015) analyzed 260,000 articles from 100 journals in Economics (63) and Business (37) that were published between 1956 and 2005 and cited between 1980 and 2005 in ISI-indexed journals. Unlike in the previous study on science journals, here citations to the average journal volume peak only in the fifth year after publication, while the secular upward trend in the frequency of citations is confirmed. From 1995 onwards, when full-text articles started to be posted online, online availability continuously rose. By 2005, 88% of the journal volumes were available online. Controlling for quality effects by means of fixed effects for journal volumes, as well as for age and time effects, the authors no longer found a statistically significant effect of online availability on citation counts. However, they did find substantial heterogeneity across platforms; for instance, a doubling of JSTOR subscriptions causes a 10% increase in citations. Finally, they observed a larger impact of JSTOR availability on citations from developing countries and no effect on the citations from authors in non-English speaking Europe. Regarding the impact of OA, they conclude that "the modest size of these effects, and the current lack of evidence that free online access performs better, implies that the citation benefits of open-access publishing have been exaggerated by its proponents. Even if publishing in an open access journal were generally associated with a 10% boost in citations, it is not clear that authors in economics and business would be willing to pay several thousand dollars for this benefit, at least in lieu of subsidies" (p. 163).

Ottoviani (2016) used the University of Michigan's institutional repository service "Deep Blue" to randomly select 3,850 papers that were made OA from 2006 onwards via blanket licensing agreements between the publishers and the library. Thus, any self-selection bias is avoided. However, since the articles in the sample were only opened after some embargo period, there is a "late view bias" to the detriment of the OA papers. These OA articles were matched with 89,895 CA articles from the corresponding journal issue. The original publication dates range from 1990 to 2013. "By comparing citations to subscriber-only/now-open (opened) articles with the corresponding subscriber-only/still-subscriber-only (closed) articles in that journal issue before and after availability in Deep Blue, we can determine what effect opening them may have had, i.e. a post-embargo [open access citation advantage]" (p. 2). The sample is dominated by physical science, health science, and engineering articles. Ottaviani focused on article-by-article differences[18] in citation counts and found an OA citation advantage with a lower bound of approximately 20%. He also found that better (above median) articles gain more from being OA.

Archambault et al (2017) examine more than 3 million papers indexed in the Web of Science (WoS) from 22 disciplines published between 2007 and 2009, and the related citations (almost 35 million) in more than 12,000 WoS-indexed journals from 2007 to mid-2016. They identified gold and green OA papers that were published in peer reviewed journals via the aggregate "1scienceOAIndx", finding that across all disciplines, OA paper of either type received 50% more citations than strictly paywalled papers. Green OA is more effective than gold OA in 20 of the 22 fields. Gold OA is more effective only in Biology and Biomedical Research; in Clinical Medicine, the difference is very small. According to the authors, these results are due to the NIH and Wellcome Trust OA mandates. In 15 of the 22 disciplines, the impact of gold

[18] This is calculated in three ways: "each opened article is compared to the mean and the median article in the same journal which it appeared in and also to equivalent articles in that issue" (p. 3), where "equivalence" refers to citation counts.

OA on citation counts of course exceeds that of strictly paywalled papers; however, it is lower than the overall average impact on citation counts to all types of papers in those disciplines. In six disciplines, gold OA papers even receive fewer citations than strictly paywalled papers (Agriculture, Fisheries & Forestry; Built Environment and Design; Chemistry; Enabling & Strategic Technologies; General Science & Technology; Historical Studies), which the authors do not discuss any further. The authors see no indication of an early view bias of OA papers, due to embargoes for many green OA papers. However, a quality selection bias may still persist. *Piwowar et al.* (2018) analyzed a sample of 100,000 WoS-indexed articles and reviews from several disciplines published between 2009 and 2015. They found an OA citation advantage of 18%, with differences regarding the way in which the papers are made OA.[19] While the citation advantage amounted to 33% for self-archiving (Green OA) and 31% for hybrid OA, Gold OA yielded an average *negative* citation impact of 17% below the global average and 9% below that of strictly paywalled papers. The authors suggest that this negative effect of Gold OA may be due to the increase in the number of newer and smaller OA journals that are considered less prestigious and not always published in English, and the continued growth of so-called 'mega journals' such as PLOS ONE.

3 Data exploration

To analyze the impact of Open Access on citations in different publishing cultures, we rely on a variety of data sources. These sources and our approaches to them are described in the following subsection, followed by some descriptive statistics on the journal and article level.

3.1 Data

Our analysis is based on comparing the citation impact of OA between Biology and Economics & Management. In this regard, we build upon work that identifies the publishing culture as an important determinant of researcher's attitudes towards OA publishing. In the following, we distinguish between a predominantly *gold culture*, such as Biology, which is characterized by strong OA journal (gold OA) usage, and a predominantly *green culture*, such as Economics & Management, which features few OA journals but extensive use of OA repositories (green OA).[20]

The data contain article information from all journals included in the WoS "Journal Citation Reports" (JCR) database in the categories Economics & Management (green culture) and Biology (gold culture). We cover the period 2000 to 2019, including all 83 journals in Biology and 563 journals in Economics & Management listed in the WoS. While the two disciplines are represented by a rather uneven number of journals, this is to some extent compensated by the number of articles and authors in each category: Biology comprises 267,340 articles written by

[19] Apart from gold, hybrid and green OA, the authors introduced a fourth form, 'bronze OA': "free to read on the publisher page, but without any clearly identifiable license" (p. 5).

[20] *Eger & Scheufen* (2018) conducted a survey on OA in 25 countries with more than 10,000 completed questionnaires that captured the perception of academic authors regarding gold and green OA. The authors categorized academic disciplines along three publishing cultures: (1) *Gold culture*: strong use of OA journals (gold road) but little use of OA repositories (green road), for example Biology; (2) *Green culture*: extensive use of OA repositories but little use of OA journals, for example Economics & Management; (3) *Grey culture*: intermediate use of both OA journals and OA repositories, for example Philosophy. The data of the present paper by contrast indicate that Biology features strong use of both gold OA and green OA. This difference may be due to the fact that many of the repositories included in the Directory of Open Access Repositories (openDOAR) post only abstracts. However, in the present paper, we cannot distinguish between repositories with and without full text.

a total of 1,248,992 authors, whereas Economics & Management accounts for 517,184 articles by 1,281,181 authors. The two fields and associated publication cultures thus differ with respect to the number of authors per publication and the number of articles published in each journal.

Besides the journal name and category, we also collected information on the total number of cites per journal and year, the Journal Impact Factor (JIF), and various other indicators of journal quality (e.g. Immediacy score or Eigenfactor score) from the JCR database. The JCR data were matched with article-level data and information on journal OA availability using the ISSN identifiers. Additionally, data on the OA status of the journals were acquired via the Directory of Open Access Journals (DOAJ) API. The two sources were merged using either the ISSN or journal titles.

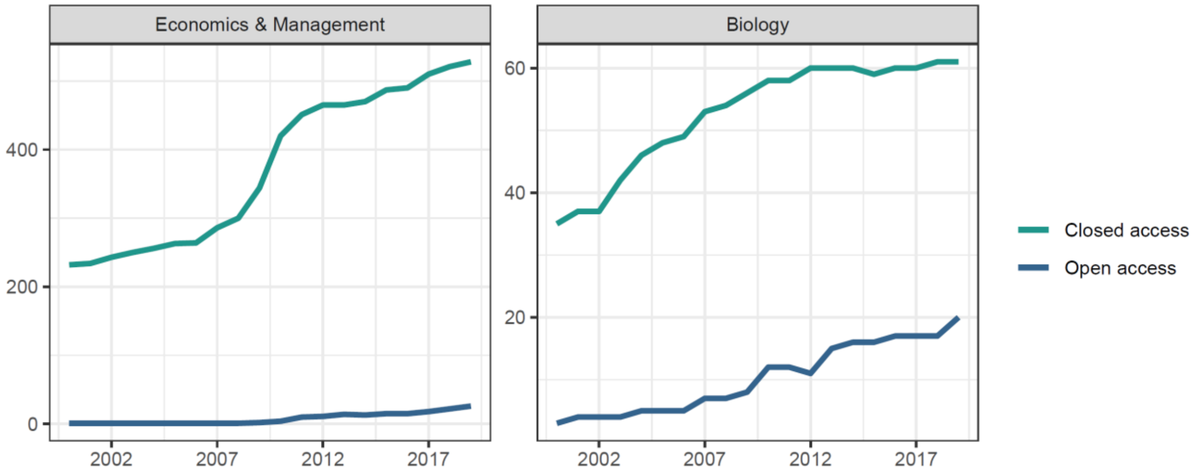
The article and author information was gathered using Microsoft's Project Academic Knowledge API accessing Microsoft Academic Graph (MAG) data. We thus retrieved information on all the articles published in the 646 journals over the sample period. This includes the number of citations per article, the number of references and authors, and the date of publication, as well as information about the authors of each article (institutional affiliation, number of publications and citations of the institution, number of publications and citations of the author). The resulting data set comprises 2,530,173 author-article observations. Since the unit of analysis is not the author but the article, the information on all authors was summarized for each article, resulting in 784,529 observations on individual articles.

The MAG data also contain information about the URL of any online repository that provides access to each article. To construct the variable on potential green OA venues, using the Directory of Open Access Repositories (openDOAR) API, we checked whether any of the unique core URLs per article were listed in the openDOAR register. For an overview of the variables featured in the analysis, see Table 3.1. MAG provides yearly data on the number of citations received by every research institution and every author, but not for individual research articles. Hence, the data is essentially of cross-sectional nature since we only have information about the number of citations at the time of extraction.

3.2 Exploration

By comparing two extreme cases, Biology (gold culture) and Economics & Management (green culture), we investigate in more detail whether any citation effect differs across opposing publishing cultures. Looking at the number of open access (OA) versus closed access (CA) journals listed in WoS – i.e. only journals that meet a certain quality standard – we find considerable differences between the two fields. While in Biology, the number of both OA and CA journals has increased steadily since 2000, with a share of around 25 percent OA journals, in Economics & Management, only the number of CA journals increased, whereas the number of OA journals remained at an extremely low level for most of the observation period (Figure 3.1). In the last few years, however, a slow growth emerged in the number of OA journals.

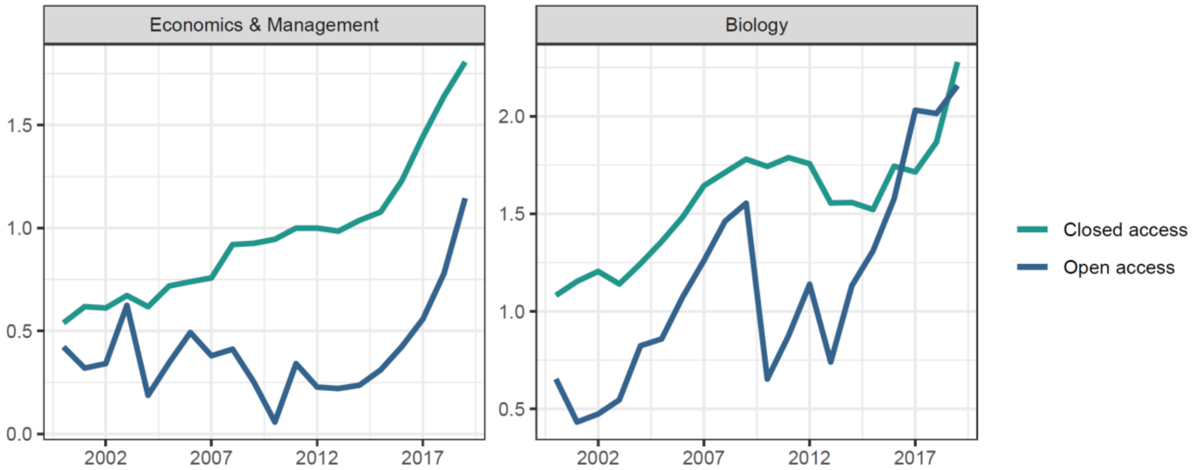
Figure 3.1: Number of OA vs. CA journals in Biology and Economics & Management authors's



Source: Authors' calculations based on WoS (2000-2019)

Most interestingly, we find that the median impact factor of the OA journals has exceeded that of the CA journals in Biology since 2016.[21] The reputation of OA journals thus equals that of traditional journals in this discipline. By contrast, in Economics & Management, we find a substantial reputation gap between OA and CA journals. Bear in mind, however, that only very few OA journals are listed in WoS for Economics & Management.[22] That is, the divergence in terms of journal reputation may be even more extreme in Economics & Management than Figure 3.2 suggests. The low esteem of OA journals in Economics & Management likely explains their little usage – an important point to bear in mind especially for young scholars seeking tenure.

Figure 3.2: Median Impact Factor of OA vs. CA journals in Biology and Economics & Management



Source: Authors' calculations based on WoS (2000-2019)

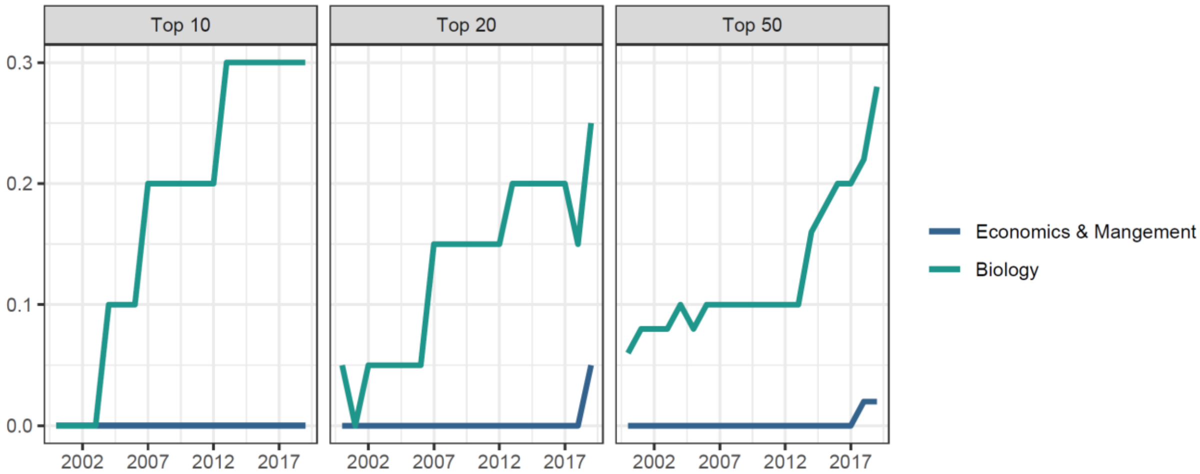
If we reduce the sample to the top journals in terms of yearly impact factors, this picture of a reputation disadvantage of OA journals in Economics & Management as compared to the

[21] Note that comparing journal impact factors is only valid within but not between disciplines, due to different publishing cultures.

[22] The number of OA (CA) journals in Economics & Management increased steadily from 1 (232) in 2000 to 26 (528) in 2019.

situation in Biology becomes even starker (Figure 3.3). The top 10 journals in Economics & Management contain not a single OA journal, whereas in Biology, three OA journals made it into the top 10 over time (eLife, PLoS Biology and BMC Biology). In Economics & Management, the Journal of Innovation & Management has been the only OA journal in the top 20 journals since 2018. In Biology, the share of OA journals increased steadily to 25 (30) percent among the top 20 (top 50) journals by 2019.

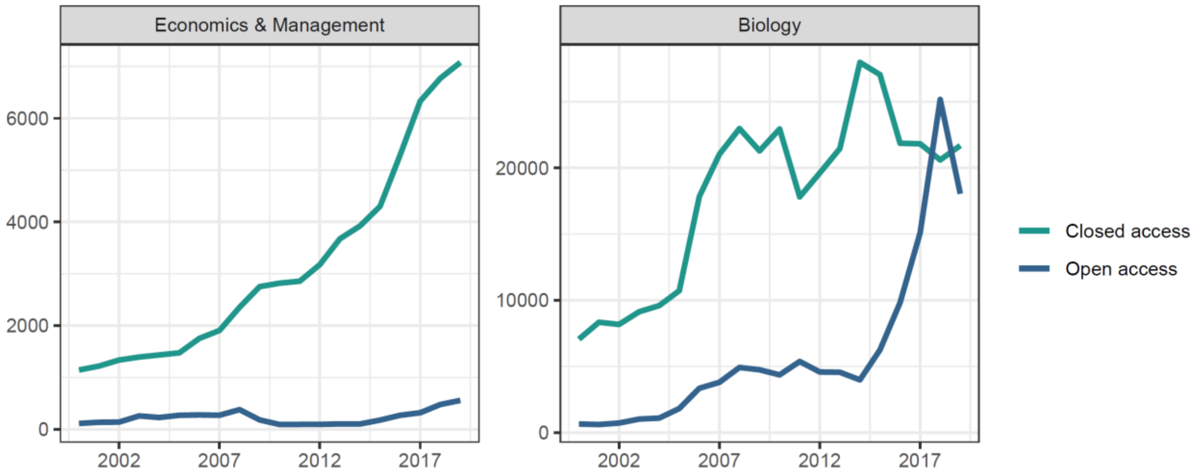
Figure 3.3: Share of top OA vs. CA journals in Biology versus Economics & Management



Source: Authors' calculations based on WoS (2000-2019)

The number of citations, our dependent variable, clearly diverges between CA and OA journals (Figure 3.4). While the number of citations for OA journals in Economics & Management remained extremely low over the entire period, citations to CA journals increased throughout. By contrast, in Biology the number of citations to OA journals increased steadily, catching up with the number of citations to A journals by the end of the period.

Figure 3.4: Number of citations per OA vs. CA journal in Biology and Economics & Management



Source: Authors' calculations based on WoS (2000-2019)

How come that OA journals enjoy so much more esteem in Biology than in Economics & Management? We think that the differences may be due to the timing of the establishment of OA journals in each discipline, combined with changes in the respective environment. Biology already featured 3 OA journals in 2000, which increased to 20 in 2019 – about one quarter of all journals in Biology. In October 2003, the Public Library of Science (PLOS) launched the

OA journal PLOS Biology, which only two years later achieved an impact factor of 13.9, the highest in general biology (*Suber* 2006, 151). There are several reasons why this OA journal succeeded so quickly and has been achieving excellent impact factors ever since (*Eckdahl*, 2004): (1) It was launched by outstanding experts in the field, allowing it to attract many excellent scholars as authors and reviewers; (2) this high level of quality in an area of great public interest has helped to attract large amounts of funding; (3) as a full-service journal, PLOS Biology also includes sections that communicate the results of high-profile original research to non-academic audiences, such as physicians, patients and policy makers. Thus, the early start of a very successful OA journal in Biology facilitated the establishment of subsequent OA journals in the field that were able to build on the strong reputation of PLOS Biology.
In Economics & Management, WoS listed the first OA journal in 2000 and, following a period of slow growth, the number stood at 26 OA journals in 2019 – 4.69% of all journals in the field. In 2010, when the discipline featured just 4 OA journals (compared to 420 CA journals), *Jeffrey Beall*, a librarian at the University of Colorado in Denver, compiled a list of so-called *predatory publishers* ('Beall's List'). This term refers to publishers whose business model aims to either exploit inexperienced academic authors or allow experienced authors to publish low-quality articles in OA journals. The strategies employed by predatory publishers include charging APCs for articles with little or no peer review, listing academics as members of editorial boards without their permission, appointing fake academics to editorial boards, mimicking the name or website style of more established journals, and so on. Beall's List was attacked by the publishers concerned, as well as by OA advocates, who argued that Beall exaggerated the problems due to his ideological objection to OA (*Pinfield* 2015, 619; *Anderson* 2015), and it went offline in January 2017.[23] In any event, the publication of Beall's List likely impaired economists' perception of the reputation of OA journals. Consequently, the top authors in Economics & Management were deterred from submitting their papers to OA journals, so the negative expectations became self-fulfilling. In Biology, the rise of OA journals, with excellent quality, began already before the publication of Beall's list and led to a reinforcement of positive expectations.

Turning from the journal level to the article level as the unit of our analysis, Table 3.1 provides an overview on the summary statistics of all variables used in our regressions.

Dependent Variable. Our dependent variable is the number of citations a paper has received since its publication from other articles published in WoS journals. The average paper in our sample receives around 29 citations. With a median of 6 citations and a maximum of 30,770 citations, however, we find considerable variation. Papers in Economics & Management (32.98) on average receive more citations than Biology papers (20.63).[24] Anecdotal evidence indicates that papers in Biology have a shorter citation half-life, whereas findings in Economics & Management tend to maintain their relevance over time.

[23] See also *Anderson* (2018, 238-47), and *Ritchie* (2020, 184-5).

[24] See Tables A.1 and A.2 in the appendix for discipline-specific summary statistics.

Table 3.1: Summary Statistics

	N	mean	std	min	median	max
<u>**Dependent Variable**</u>						
Number of citations	784,529	28.77	106.36	0	6	30,770
<u>**Independent Variables**</u>						
Regime						
Gold Open access	784,529	0.08	0.27	0	0	1
Green Open Access	784,529	0.47	0.50	0	0	1
Discipline/culture						
Economics & Management	784,529	0.66	0.47	0	1	1
Biology	784,529	0.34	0.47	0	0	1
Journal						
JIF - Bottom 25	784,529	0.26	0.44	0	0	1
JIF - Bottom 50	784,529	0.29	0.45	0	0	1
JIF - Top 50	784,529	0.24	0.43	0	0	1
JIF - Top 25	784,529	0.21	0.41	0	0	1
Article						
Institution score	784,529	28.74	28.79	0	24.36	2,099
Author score	784,529	11.99	47.28	0	3	13,673
Number of authors	784,529	3.23	3.3	1	2	500
Number of references	784,529	28.62	31.87	0	22	500
Publication year	784,529	2011.64	5.31	2000	2012	2019

Source: Authors' calculations based on WoS (2000-2019), MAG, DOAJ and openDOAR.

Independent Variables. The independent variables can be classified into four categories: (1) Regime – gold versus green open access; (2) Discipline/culture – Biology = gold culture; Economics & Management = green culture; (3) Journal – four journal impact factor groups as proxies of a journal's reputation in the year of publication; (4) Article – additional information on the article and its authors.

Regime: **'**Gold open access' is a dummy variable that indicates whether an article is published in a pure[25] OA journal. This applies to 8 percent of all articles in our sample – 20 percent in Biology versus only 1 percent in Economics & Management.[26] 'Green open access' similarly indicates whether an article has been deposited with an OA repository.[27] 47 percent of all articles make use of at least one OA repository.

Discipline/culture: We use a dummy variable for each of the two disciplines, Economics & Management (green culture, 66 percent of all articles) and Biology (gold culture, 34 percent).

[25] We follow the definition of the DOAJ to identify pure OA journals. In contrast to a hybrid OA journal, a pure OA journal offers free online access to all journal content, rather than just to a selection of papers.
[26] See Tables A.1 and A.2 in the appendix for discipline-specific summary statistics.
[27] We follow the definition of openDOAR to identify OA repositories.

Journal: The JIF-dummies indicate the group of journals that an article belongs to in term of its yearly Impact Factor (WoS).[28] In each of the two disciplines and for each year of publication, the journals were ranked according to their impact factor. The dummy variable JIF-Q1, for example, refers to the first quartiles, or top 25%, of the journals in either discipline in terms of journal impact factor.

Article: The 'institution score' denotes the number of citations per publication for a given institution in a given year. Similarly, the 'author score' indicates the sum of citations per publication for a given author in a given year. Other article information refers to the number of authors, the number of references, and the year of publication. On average, a research article is written by 2.48 researchers (maximum of 228) in Economics & Management and 4.67 (maximum of more than 500[29]) in Biology.[30] The number of papers increases over time: half of the articles in the sample were published in the first 13 years (2000 to 2012), whereas the other half took only 7 years to publish (2012 to 2019).

4 Empirical analysis

To estimate the citation effect of OA publishing in two distinct publishing cultures (Biology versus Economics & Management), we use a Poisson quasi-maximum-likelihood (PQML) model, as explained in the next subsection. Section 4.2 presents the results of our baseline model and a discipline/culture-specific view. Section 4.3 provides robustness checks by applying other methodological approaches (Negative-Binominal Hurdle Model) and additionally controlling for author fame and other measurements.

4.1 Methodology

As the number of citations is a classic example of count data, we use a Poisson model to estimate the effect of OA on the number of citations to each research article. However, the distribution of citations is heavily skewed, with an excess of zeroes. This is usually the case for citation data as many research articles are never cited at all (see for example *Baccini et al.* 2014; *McCabe & Snyder* 2015), which typically leads to over-dispersion, implying a larger variance compared to the mean and too many zeroes relative to a Poisson distribution. While the Negative Binomial model would constitute an adequate cure to over-dispersion, it too does not provide accurate results when over-dispersion is mainly caused by an excess of zeroes (*Baccini et al.* 2014). Therefore, in line with *McCabe & Snyder* (2015), we use a Poisson quasi-maximum-likelihood (PQML) model to estimate the conditional mean. It uses the mean regression and variance function from the regular Poisson model but leaves the dispersion parameter unrestricted, which, in turn, is not fixed but estimated from the data. We also estimate the more flexible hurdle model with different distributions (Poisson and Negative Binomial) as robustness checks.

4.2 Citation effects

In analyzing the citation impact of OA publishing, we compare gold/green OA in Biology versus in Economics & Management. We will start with our baseline model before turning to a

---

[28] The JIF-dummies must be calculated anew in each year of publication using the impact factors at the time of publication to avoid circular arguments. Only thus can we ensure that our category formation does not interfere with the number of citations in subsequent periods, which is the basis of the independent variable.
[29] Note that the number of authors and references is capped at 500 by Microsoft Academic. Given the extremely low number of articles reaching the cap of 500 authors, this cap should not impose a problem for the analysis.
[30] See Tables A.1 and A.2 in the appendix for discipline-specific summary statistics.

discipline/culture-specific view on the citation effects of gold and green OA while focusing on the former.

First of all, we estimate the impact of gold OA on the number of citations by three different PQML specifications, whose results are reported in Table 4.1. Column (1) refers to the basic model, which includes the regime and discipline/culture indicators, as well as journal and article information. Like models (2) and (3), it also incorporates quarter fixed effects (FE) to better account for time trends. Specification (2) adds interaction effects between the discipline and gold OA. Finally, we add quarter-discipline FE in specification (3) to control for any discipline-specific citation trends over time.

Table 4.1: The OA Citation Effect

	Dependent variable: citation_count		
	(1) Base	(2) + gold OA Interaction	(3) + quarter-discipline FE
Gold OA	0.292***	0.369***	0.183***
	(0.270, 0.314)	(0.346, 0.392)	(0.160, 0.207)
Green OA	0.833***	0.836***	0.822***
	(0.823, 0.842)	(0.827, 0.845)	(0.813, 0.832)
Economics	0.854***	0.879***	1.575***
	(0.843, 0.864)	(0.868, 0.890)	(1.500, 1.651)
JIF Bottom 50	0.616***	0.604***	0.608***
	(0.600, 0.631)	(0.589, 0.620)	(0.592, 0.623)
JIF Top 50	1.301***	1.280***	1.314***
	(1.286, 1.316)	(1.264, 1.295)	(1.298, 1.329)
JIF Top 25	2.453***	2.410***	2.443***
	(2.439, 2.468)	(2.396, 2.424)	(2.429, 2.458)
Number of References	0.010***	0.010***	0.010***
	(0.010, 0.010)	(0.010, 0.010)	(0.010, 0.010)
Institution Score	0.004***	0.004***	0.004***
	(0.004, 0.004)	(0.004, 0.004)	(0.004, 0.004)
Author Score	0.001***	0.001***	0.0005***
	(0.0005, 0.001)	(0.0005, 0.001)	(0.0005, 0.001)
Number of authors	0.011***	0.011***	0.011***
	(0.011, 0.012)	(0.011, 0.012)	(0.010, 0.012)
Gold OA* Economics		-0.580***	-0.492***
		(-0.686, -0.473)	(-0.597, -0.386)
Quarter FE	Yes	Yes	Yes
Quarter-Discipline FE	No	No	Yes
Observations	784,529	784,529	784,529

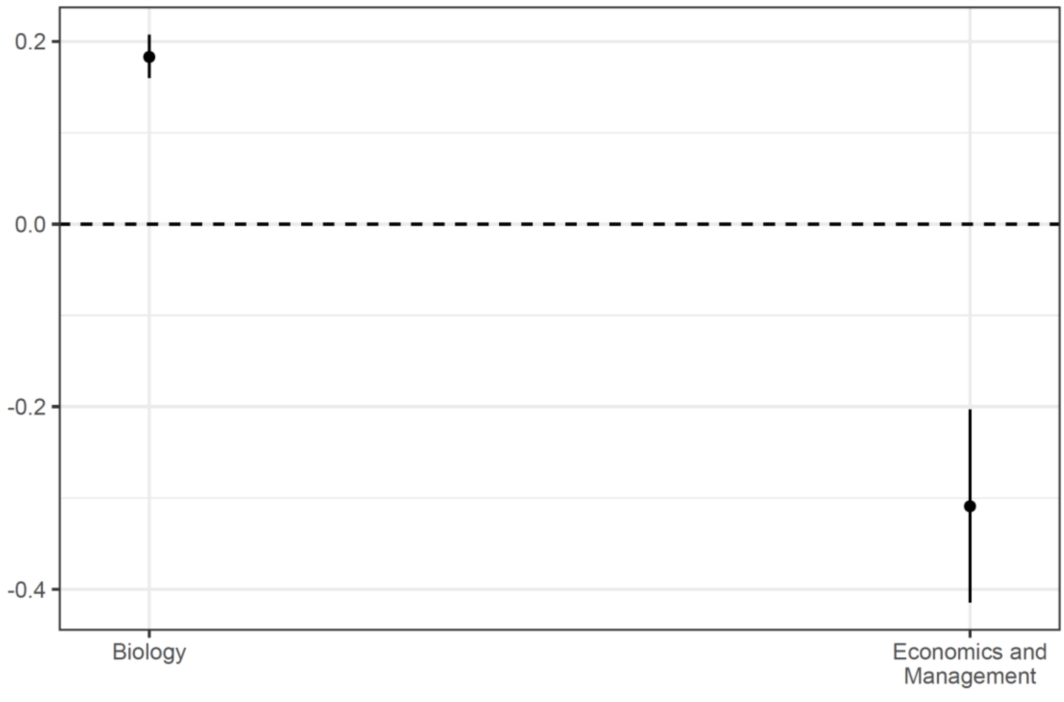
**** indicates a significance level of 0.1%. All coefficients refer to marginal effects. 95%-confidence intervals in brackets. Constant not reported.*

All regression coefficients in Table 4.1 were converted into marginal effects, which can be interpreted as proportionate changes. For instance, a marginal effect of +0.292 signifies that gold OA raises the number of citations by 29.2 percent. The number in brackets refer to the 95% confidence intervals of the marginal effects.

We find a positive and robust citation effect that is statistically significant at the 0.1% level across all specifications.[31] Moving from model (1) to model (2), the additional inclusion of the interaction effect between gold OA and Economics raises the gold OA citation effect from 0.292 to 0.369. However, the gold OA coefficient has to be interpreted differently as it indicates the effect of gold OA when the field variable (Economics) is zero. As such, the gold OA coefficient shows the increase in citations for the field of Biology. Interestingly, however, the coefficient on the interaction term itself is strongly significant and negative, which means that gold OA does less to promote citations in Economics & Management than it does in Biology. In fact, gold OA even reduces citations in Economics & Management when looking at the interaction of gold OA with the field of Economics & Management.

Our preferred specification is model (3), which also accounts for any discipline-specific citation trends by controlling for quarter-discipline FE. We find that gold OA increases the number of citations by 18.3% in Biology. In contrast, we can calculate the citation effect in Economics & Management by adding the gold OA effect with the interaction term, finding a negative effect of gold OA in Economics & Management of 30.9%, i.e. publishing in gold OA reduces citations in Economics & Management by 30.9%.[32] Figure 4.1 highlights these findings of a contradicting citation effect when comparing Biology versus Economics & Management. While gold OA boosts citations in Biology, it leads to a decrease of citations in Economics & Management.

Figure 4.1: Marginal Effects by Disciplines



[31] Given the very large number of observations, tests for statistical significance provide little merit since all coefficients become statistically significant. Hence, we report confidence intervals of the coefficients and focus more on the effects' substantial rather than their statistical significance.

[32] We also run the regressions for both publication cultures separately in Table A.3 of the appendix.

Hence, again, the research field matters to the magnitude of the Gold OA effect on citations. This emphasizes the importance of controlling not only for quarter but also for quarter-discipline FE to account for any discipline-specific trends over time. In this regard we run additional robustness checks by including the age (i.e. the number of days since publication) in Table A.7. Here, the interaction effect of age with the field of Economics & Management is strongly significant and positive, which suggests that the impact in Economics & Management increases at a higher pace than in Biology. That is, we may expect that the field of Economics & Management is slowly catching up in terms of gold OA publishing.

Moreover, green OA has a strongly significant and positive impact on citations (82.2%). We also find that the greater the reputation of a journal, the greater the effect on citations. The other journal characteristics have only slight effects on the number of citations. Significant differences between the two fields also exist with respect to green OA (Table A.3). While green OA has a positive and strongly significant effect on the number of citations in both disciplines, the effect is much larger in Biology (571.3%) than in Economics & Management (59.9%).[33] All journal and article indicators exhibit similar tendencies to those found in Table 4.1. Again, the higher the impact factor of a journal, the greater the effect on citations in both disciplines.

[33] The available data do not permit a convincing explanation of this citation boost in Biology since the repositories listed in openDOAR are very heterogeneous: some include full texts, others only abstracts; some are published before the original journal article, others afterwards; some host content that is very similar to the original journal article while on others, the content is quite different.

4.3 Robustness checks

4.3.1 Hurdle model

To check the robustness of the Quasi-Poisson models of our main analysis, we also estimate the more flexible hurdle models, as introduced by Mullahy (1986). Hurdle models are two-component models that comprise, first, a hurdle component that models zero against larger counts and, second, a truncated count component (e.g. Poisson or Negative Binomial) for positive counts. For the hurdle (zero) component, we use a binomial logistic regression for its intuitive interpretation. We report both models using the Poisson as well as the Negative Binomial distribution to estimate the truncated component (see Table A.4 in the appendix). This estimation strategy, too, yields a strongly significant citation effect of gold OA on the number of citations, with effects running in different directions depending on the field of study. The coefficient of gold OA in the zero component of the hurdle model shows an 25.3% higher chance to receive the first citation if the article was published in a gold OA journal in the field of Biology. The interaction term between gold OA and the field of Economics is also statistically as well as substantially negative. In line with our main analysis, both truncated component models also report a citation advantage of gold OA in the field of Biology of +7.9% (Negative Binomial distribution) and +11.8% (Poisson distribution), and a large negative impact of gold OA on citations in Economics & Management. Given its much higher log likelihood score, we prefer the Negative Binomial to the Poisson model.

4.3.2 Alternative variables

In our main analysis, following *Brogaard et al.* (2020), we controlled for author reputation, as proxied by an author's total number of citations to their published papers. We then averaged these sums across all the authors of a given article to obtain its author score. However, as *Brogaard et al.* (2020) argue, the citation count could be more closely related to the score of the first author or to the maximum score of all authors of a particular article. We therefore recreate model (3) of Table 4.1 for all three author score measures (see Table A.5 in the appendix), finding essentially the same outcomes. For example, raising the score of the mean (first, maximum) author by one point increases citations by 0.05% (0.03%, 0.03%). The effects of our main variable of interest (gold OA) remain similar.

Another important finding of our main analysis is the strong positive impact of a journal's reputation on citations. Clearly, journal reputation can also be proxied by measures other than the impact factor. As robustness checks of the gold OA citations effect obtained in the main analysis, we alternatively use the 5-year impact factor, the Eigenfactor, and the Immediacy Index.[34] The results of all four measures are reported in Table A.6 in the appendix. Again, the citation effect remains highly significant and robust, regardless of which measure of reputation we employ. The marginal effect ranges from +10.5% (5-year impact factor) to +41.5% (Eigenfactor).

Thirdly, in our main analysis, we controlled for time trends by using quarter FE and quarter-discipline FE (see Table 4.1). As another robustness check, we instead use the age of each journal article as a continuous variable (measured in days measured in days between publication and our acquisition of the data). The results are again quite similar (see Table A.7 in the appendix).

[34] The Eigenfactor measurement estimates the total importance of a journal by weighing citations from highly ranked journals higher than those of poorly ranked journal contributions. The Immediacy index indicates how quickly contributions of a journal are cited. For more information see the glossary of the Journal Citation Reports at http://help.incites.clarivate.com/incitesLiveJCR/.

Finally, the Microsoft Academic Graph reports an *expected* number of citations, according to some algorithm. While it remains unclear how these values come about, we retrieved the estimated citation count from the Microsoft Academic website and used it as an alternative dependent variable in lieu of the actual number of citations (see Table A.8 in the appendix). Again, the results are very similar to those reported in Table 4.1. We find a highly significant impact of Gold OA on the number of expected citations, ranging from 15% in (our preferred) model (3) to 33.1% in model (2). As before, the field of research matters: Gold OA is much less beneficial, even harmful, in Economics compared to Biology.

5 Conclusion

This paper has estimated and compared the citation impact of gold and green open access in two disciplines: Biology and Economics & Management. The empirical analysis covers all articles of the two disciplines that were included in the WoS "Journal Citation Reports" between 2000 and 2019: 267,340 articles written by a total of 1,248,992 authors in 83 Biology journals and 517,184 articles by 1,281,181 authors in 563 Economics & Management journals. Using a Poisson quasi-maximum likelihood model, we show that, controlling for quarter fixed effects, journal ranking, number of references, institutional quality, and the quality and number of authors, gold OA raises the number of citations to the articles of both disciplines. However, this positive overall effect masks crucial differences between the two publishing cultures: While citations in Biology are increased by 18.3%, citations in Economics & Management actually decline by 30.9%. This at first glance surprising result is due to the fact that in Biology, many OA journals are among the top journals in this discipline, whereas in Economics & Management, OA journals still have a poor reputation among the scholars. While in Biology 30% of the top 10 journals are OA journals, in Economics & Management, the first OA journal has yet to make its way into the top 20, and only in 2018 did one OA journal make the top 50. The late start of OA journals in Economics & Management in conjunction with the emergence of predatory journals has likely prevented OA journals from achieving a good reputation in this discipline, while the early start of some excellent OA journals in Biology facilitated the rise of OA journals with high reputation. While our analysis of gold OA focused on pure OA journals, future analysis should pay more attention to the increasing number of hybrid OA journals.
.
Green OA increases overall citations for both disciplines, with an increase of 571.3% in biology and 59.9% in Economics & Management. The very heterogeneous data on green OA make it difficult to find a satisfactory explanation for this enormous difference, which is why we focus on the citation impact of gold OA. Future research should devote more attention to the heterogeneity of green OA and its citation impact. Our general results of the PQML model are confirmed by several robustness checks.

References

Anderson, R. (2018), *Scholarly Communication. What Everyone Needs to Know*, Oxford: Oxford University Press.
Anderson, R. (2015), 'Should We Retire the Term "Predatory Publishing"?', *The Scholarly Kitchen*, May 11, accessed 28 February 2021 at https://scholarlykitchen.sspnet.org/2015/05/11/should-we-retire-the-term-predatory-publishing.
Antelmann, K. (2004). Do Open-Access Articles Have a Greater Research Impact? *College & Research Libraries*, 65(5), 372-82.
Archambault, E., Côté, G., Struck, B., & Voorons, M. (2017). Research Impact of Paywalled Versus Open Access Papers. Retrieved from https://www.1science.com/1numbr.
Baccini, A., Barabesi, L., Cioni, M., & Pisani, C. (2014). Crossing the hurdle: the determinants of individual scientific performance. *Scientometrics*, 101, 2035-2062.
Brogaard, J., Engelberg, J., Eswar, S., & Van Wesep, E. (2020). On the causal effect of fame on citations. Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3565487.
Craig, I. D., Plume, A., 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. (2011). Open Access, Readership, Citations: A Randomized Controlled Trial of Scientific Journal Publishing. *FASEB Journal*, 25(7), 2129-34.
Davis, P. M. & Walters, W. H. (2011). The Impact of Free Access to the Scientific Literature: A Review of Recent Research. *Journal of the Medical Library Association*, 99(3), 208-217.
Eckdahl, T. (2004), Review of: PLoS Biology – A Freely Available, Open Access Online Journal. *Cell Biology Education*, 3 (1), 15–17.
Eger, T., & Scheufen, M. (2018). *The Economics of Open Access. On the Future of Academic Publishing*. Cheltenham: Elgar.
Evans, J. A., & Reimer, J. (2009). Open Access and Global Participation in Science. *Science*, 323(5917), 1025.
Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLOS Biology*, 4(5), 692-698.
Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLOS ONE*, 5(10), 1-12.
Gaulé, P., & Maystre, M. (2011). Getting Cited: Does Open Access Help? *Research Policy*, 40, 1332-38.
Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data Engineering Bulletin*, 28(4), 39-47.
Harnad, S., & Brody, T. (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, 10(6),
Haustein, S., & Larivière, V. (2015). The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects.
Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. S. (2005). The Effect of Use and Access on Citations. *Information Processing & Management*, 41, 1395-402.
Lawrence, S. (2001). Free Online Availability Substantially Increases a Paper's Impact. *Nature*, 411, 2001, 521.
Lewis, C. (2018). The Open Access Citation Advantage: Does It Exist and What Does It Mean for Libraries? *Information Technology and Libraries*, 37(3), 50-65.
McCabe, M. J., & Snyder, C. M. (2015). Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals. *Review of Economics and Statistics*, 97(1), 144-165.

McCabe, M. J., & Snyder, C. M. (2014). Identifying the Effect of Open Access on Citations Using a Panel of Science Journals. *Economic Inquiry*, 52(4), 1284-1300.
Moed, H. F. (2007). The Effect of “Open Access” upon Citation Impact: An Analysis of ArXiv’s Condensed Matter Section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047-54.
Mueller-Langer, F., & Watt, R. (2014). *The Hybrid Open Access Citation Advantage: How Many More Cites is a $3,000 Fee Buying You?* Max Planck Institute for Innovation and Competition, Research Paper No. 14-02.
Mullahy (1986), Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341–365.
Ottoviani, J. (2016). The Post-Embargo Open Access Citation Advantage; It Exists (Probably), It’s Modest (Usually), and the Rich Get Richer (of Course). *PLOS ONE*, 11(8), 1-11.
Pinfield, S. (2015), Making Open Access Work. The ‘state-of-the-art’ in providing Open Access to scholarly literature, *Online Information Review*, 39(5), 604-636, accessed 28 February 2021 at http://dx.doi.org/10.1108/OIR-05-2015-0167.
Piwowar, H., Priem, J., Lativière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles. *PeerJ*. Retrieved from https://doi.org/10.7717/peerj.4375.
Ritchie, S. (2020), *Science Fictions. Exposing Fraud, Bias, Negligence and Hype in Science*, London: The Bodley Head.
Suber, P. (2006), ‘Open access in the USA’, in Jacobs, N. (ed.), *Open Access: Key Strategic, Technical and Economic Aspects*, Oxford: Chandos, 149–60.
Swan, A. (2010). *The Open Access Citation Advantage. Studies and Results to Date*. Retrieved from https://eprints.soton.ac.uk/268516.

Appendix

Table A.1: Summary Statistics for "Economics & Management"

	N	mean	std	min	median	max
Dependent Variable						
Number of citations	517,189	32.98	123.07	0	7	30,770
Independent Variable						
Regime						
Open access	517,189	0.01	0.12	0	0	1
Green open access	517,189	0.4	0.49	0	0	1
Journal						
JIF - Bottom 25	517,189	0.25	0.43	0	0	1
JIF - Bottom 50	517,189	0.25	0.43	0	0	1
JIF - Top 50	517,189	0.26	0.44	0	0	1
JIF - Top 25	517,189	0.24	0.43	0	0	1
Article						
Institution rank	517,189	28.8	27.6	0	25.81	2,099
Author score	517,189	11.05	46.6	0	2	7,856.67
Number of authors	517,189	2.48	1.68	1	2	228
Number of references	517,189	28.91	30.06	0	23	500

Source: Authors' calculations based on WoS (2000-2019), MAG, DOAJ and openDOAR.

Table A.2: Summary Statistics for "Biology"

	N	mean	std	min	median	max
Dependent Variable						
Number of citations	267,340	20.63	61.61	0	4	5,912
Independent Variables						
Regime						
Gold OA	267,340	0.2	0.4	0	0	1
Green OA	267,340	0.61	0.49	0	1	1
Journal						
JIF - Bottom 25	267,340	0.26	0.44	0	0	1
JIF - Bottom 50	267,340	0.36	0.48	0	0	1
JIF - Top 50	267,340	0.21	0.41	0	0	1
JIF - Top 25	267,340	0.16	0.37	0	0	1
Article						
Institution Score	267,340	28.63	30.98	0	20.41	1,503.48
Author score	267,340	13.82	48.52	0	5.17	13,673
Number of authors	267,340	4.67	4.83	1	4	500
Number of references	267,340	28.07	35.09	0	20	500

Source: Authors' calculations based on WoS (2000-2019), MAG, DOAJ and openDOAR.

Table A.3: Citation Effect by Disciplines

	Dependent variable: citation_count	
	(1) Biology	(2) Economics
Gold OA	0.133***	-0.364***
	(0.111, 0.155)	(-0.475, -0.253)
Green OA	5.713***	0.599***
	(5.677, 5.749)	(0.588, 0.609)
JIF Bottom 50	0.726***	0.581***
	(0.700, 0.751)	(0.560, 0.601)
JIF Top 50	1.524***	1.347***
	(1.495, 1.553)	(1.328, 1.366)
JIF Top 25	2.334***	2.456***
	(2.309, 2.359)	(2.438, 2.475)
Number of References	0.008***	0.011***
	(0.008, 0.008)	(0.011, 0.011)
Institution Score	0.003***	0.004***
	(0.003, 0.003)	(0.004, 0.004)
Author Score	0.0004***	0.001***
	(0.0004, 0.0004)	(0.001, 0.001)
Number of Authors	0.010***	0.022***
	(0.010, 0.011)	(0.020, 0.023)
Quarter FE	Yes	Yes
Observations	267,340	517,189

**** indicates a significance level of 0.1%. All coefficients refer to marginal effects. 95%-confidence intervals in brackets. Constant not reported.*

Table A.4: Hurdle Model Using Negative Binominal and Poisson Regression

	Zero component (1st stage)		Count component (2nd stage)	
	Negative Binomial	Poisson	Negative Binomial	Poisson
Gold OA	0.253***	0.253***	0.079***	0.118***
	(0.219, 0.287)	(0.219, 0.287)	(0.064, 0.094)	(0.116, 0.120)
Green OA	2.071***	2.071***	0.356***	0.343***
	(2.054, 2.088)	(2.054, 2.088)	(0.349, 0.363)	(0.342, 0.344)
Economics	3.384***	3.384***	0.887***	0.463***
	(3.288, 3.481)	(3.288, 3.481)	(0.831, 0.943)	(0.455, 0.470)
JIF Bottom 50	-0.218***	-0.218***	0.553***	0.659***
	(-0.238, -0.198)	(-0.238, -0.198)	(0.544, 0.561)	(0.657, 0.660)
JIF Top 50	-0.264***	-0.264***	1.208***	1.439***
	(-0.285, -0.243)	(-0.285, -0.243)	(1.199, 1.218)	(1.437, 1.440)
JIF Top 25	-0.421***	-0.421***	1.998***	2.817***
	(-0.442, -0.400)	(-0.442, -0.400)	(1.988, 2.008)	(2.816, 2.819)
Number of References	0.090***	0.090***	0.015***	0.009***
	(0.089, 0.090)	(0.089, 0.090)	(0.014, 0.015)	(0.009, 0.009)
Institution Score	0.012***	0.012***	0.005***	0.004***
	(0.011, 0.012)	(0.011, 0.012)	(0.005, 0.005)	(0.004, 0.004)
Number of Authors	0.026***	0.026***	0.045***	0.011***
	(0.023, 0.029)	(0.023, 0.029)	(0.043, 0.046)	(0.011, 0.011)
Author Score	0.006***	0.006***	0.005***	0.005***
	(0.006, 0.007)	(0.006, 0.007)	(0.005, 0.005)	(0.005, 0.005)
Gold OA*Economics	-0.383***	-0.383***	-0.232***	-0.368***
	(-0.452, -0.315)	(-0.452, -0.315)	(-0.273, -0.192)	(-0.378, -0.357)
Quarter FE	Yes	Yes	Yes	Yes
Quarter-Discipline FE	Yes	Yes	Yes	Yes
Observations	784,529	784,529	784,529	784,529
Log Likelihood	-2,580,085	-15,431,682	-2,580,085	-15,431,682

*** indicates a significance level of 0.1%. All coefficients refer to marginal effects. 95%-confidence intervals in brackets. Constant not reported.

Table A.5: Baseline Model by Author Reputation

	Dependent variable: citation_count		
	(1) mean author	(2) first author	(3) max author
Gold OA	0.183***	0.181***	0.169***
	(0.160, 0.207)	(0.157, 0.204)	(0.145, 0.192)
Green OA	0.822***	0.827***	0.825***
	(0.813, 0.832)	(0.818, 0.837)	(0.815, 0.834)
Economics	1.575***	1.573***	1.574***
	(1.500, 1.651)	(1.497, 1.648)	(1.498, 1.650)
JIF Bottom 50	0.608***	0.604***	0.604***
	(0.592, 0.623)	(0.589, 0.620)	(0.588, 0.619)
JIF Top 50	1.314***	1.313***	1.305***
	(1.298, 1.329)	(1.298, 1.328)	(1.290, 1.320)
JIF Top 25	2.443***	2.450***	2.428***
	(2.429, 2.458)	(2.436, 2.465)	(2.414, 2.442)
Number of References	0.010***	0.010***	0.010***
	(0.010, 0.010)	(0.010, 0.010)	(0.010, 0.010)
Institution Score	0.004***	0.004***	0.004***
	(0.004, 0.004)	(0.004, 0.004)	(0.004, 0.004)
Author Score (mean)	0.0005***		
	(0.0005, 0.001)		
Author Score (first)		0.0003***	
		(0.0003, 0.0003)	
Author Score (max)			0.0003***
			(0.0003, 0.0003)
Number of Authors	0.011***	0.011***	0.010***
	(0.011, 0.012)	(0.011, 0.012)	(0.009, 0.010)
Gold OA*Economics	-0.492***	-0.491***	-0.487***
	(-0.597, -0.386)	(-0.597, -0.385)	(-0.593, -0.381)
Quarter-Year FE	Yes	Yes	Yes
Quarter-Year-Discipline FE	Yes	Yes	Yes
Observations	784,529	784,529	784,529

*** indicates a significance level of 0.1%. All coefficients refer to marginal effects. 95%-confidence intervals in brackets. Constant not reported.

Table A.6: Citation Effect Using Alternative Measures of Reputation

	Dependent variable: citation_count			
	(1) JIF	(2) 5-Year IF	(3) Eigenfactor	(4) Immediacy
Gold OA	0.144***	0.105***	0.415***	0.263***
	(0.122, 0.167)	(0.082, 0.127)	(0.393, 0.438)	(0.240, 0.287)
Green OA	0.812***	0.999***	0.809***	0.833***
	(0.803, 0.822)	(0.989, 1.010)	(0.799, 0.820)	(0.823, 0.843)
Economics	1.473***	1.534***	1.454***	1.567***
	(1.398, 1.547)	(1.473, 1.595)	(1.392, 1.516)	(1.490, 1.645)
Journal Impact Score	4.076***	3.756***	2.696***	2.404***
	(4.060, 4.092)	(3.739, 3.773)	(2.679, 2.713)	(2.388, 2.419)
Number of References	0.010***	0.009***	0.010***	0.010***
	(0.010, 0.010)	(0.009, 0.009)	(0.010, 0.010)	(0.010, 0.010)
Institution Score	0.004***	0.004***	0.004***	0.004***
	(0.004, 0.004)	(0.004, 0.004)	(0.004, 0.004)	(0.004, 0.004)
Author Score (mean)	0.0005***	0.0004***	0.0004***	0.0005***
	(0.0005, 0.001)	(0.0004, 0.0004)	(0.0004, 0.0004)	(0.0005, 0.001)
Number of Authors	0.011***	0.010***	0.010***	0.012***
	(0.011, 0.011)	(0.010, 0.011)	(0.010, 0.011)	(0.011, 0.012)
Gold OA*Economics	-0.440***	-0.354***	-0.595***	-0.605***
	(-0.544, -0.336)	(-0.462, -0.246)	(-0.693, -0.496)	(-0.716, -0.493)
Quarter-Year FE	Yes	Yes	Yes	Yes
Quarter-Year-Discipline FE	Yes	Yes	Yes	Yes
Observations	784,529	596,085	632,865	783,089

*** indicates a significance level of 0.1%. All coefficients refer to marginal effects. 95%-confidence intervals in brackets. Constant not reported.

Table A.7: Base Model Using Age

	Dependent variable: citation_count	
	(1) Age	(2) Age: Economics
Gold OA	0.287***	0.149***
	(0.264, 0.311)	(0.125, 0.173)
Green OA	0.883***	0.897***
	(0.873, 0.892)	(0.887, 0.906)
Economics	0.896***	0.220***
	(0.884, 0.907)	(0.194, 0.246)
JIF Bottom 50	0.593***	0.583***
	(0.577, 0.609)	(0.567, 0.599)
JIF Top 50	1.239***	1.235***
	(1.223, 1.254)	(1.220, 1.250)
JIF Top 25	2.284***	2.278***
	(2.269, 2.299)	(2.263, 2.292)
Number of References	0.010***	0.010***
	(0.010, 0.010)	(0.010, 0.010)
Age	0.0004***	0.0003***
	(0.0004, 0.0004)	(0.0003, 0.0003)
Institution Score	0.004***	0.004***
	(0.004, 0.004)	(0.004, 0.004)
Author Score (mean)	0.001***	0.001***
	(0.001, 0.001)	(0.001, 0.001)
Number of Authors	0.011***	0.010***
	(0.010, 0.011)	(0.010, 0.011)
Gold OA*Economics	-0.584***	-0.519***
	(-0.693, -0.476)	(-0.628, -0.411)
Age*Economics		0.0001***
		(0.0001, 0.0001)
Observations	784,529	784,529

*** indicates a significance level of 0.1%. All coefficients refer to marginal effects.
95%-confidence intervals in brackets. Constant not reported

Table A.8: Citation Effect Using Expected Citations

	Dependent Variable: citation count		
	(1) Base	(2) + Gold OA Interaction	(3) + quarter-discipline FE
Gold OA	0.242***	0.331***	0.150***
	(0.211, 0.274)	(0.299, 0.364)	(0.117, 0.184)
Green OA	0.848***	0.851***	0.839***
	(0.836, 0.860)	(0.839, 0.863)	(0.827, 0.851)
Economics	1.363***	1.397***	2.125***
	(1.348, 1.378)	(1.382, 1.412)	(2.022, 2.228)
JIF Bottom 50	0.724***	0.711***	0.715***
	(0.703, 0.745)	(0.690, 0.732)	(0.694, 0.736)
JIF Top 50	1.575***	1.549***	1.585***
	(1.555, 1.595)	(1.529, 1.569)	(1.565, 1.605)
JIF Top 25	3.066***	3.013***	3.055***
	(3.047, 3.085)	(2.994, 3.032)	(3.036, 3.074)
Number of References	0.010***	0.010***	0.010***
	(0.010, 0.011)	(0.010, 0.011)	(0.010, 0.011)
Institution Score	0.004***	0.004***	0.004***
	(0.004, 0.004)	(0.004, 0.004)	(0.004, 0.004)
Author Score	0.001***	0.001***	0.001***
	(0.001, 0.001)	(0.001, 0.001)	(0.001, 0.001)
Number of Authors	0.012***	0.011***	0.011***
	(0.011, 0.012)	(0.011, 0.012)	(0.011, 0.012)
Gold OA*Economics		-0.617***	-0.542***
		(-0.767, -0.467)	(-0.692, -0.392)
Quarter FE	Yes	Yes	Yes
Quarter-Discipline-FE	No	No	Yes
Observations	784,529	784,529	784,529

*** indicates a significance level of 0.1%. All coefficients refer to marginal effects. 95%-confidence intervals in brackets. Constant not reported