Research Policy 49 (2020) 103886

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Does online access promote research in developing countries? Empirical evidence from article-level data

Frank Mueller-Langer[a,b,c,*], Marc Scheufen[d,e], Patrick Waelbroeck[f,g,#]

[a] European Commission, Joint Research Centre, Digital Economy Unit, Seville, Spain
[b] Max Planck Institute for Innovation and Competition, Munich, Germany
[c] Bundeswehr University Munich, Department of Business Administration, Munich, Germany
[d] Ruhr University Bochum, Bochum, Germany
[e] German Economic Institute (IW), Cologne, Germany
[f] Télécom ParisTech, Paris, France
[g] CESIfo, Munich, Germany

ARTICLE INFO

Keywords:
Online access
Scientific productivity
Access to knowledge as publication input
Difference-in-difference-in-differences (DDD) estimation
Instrumental variables (IV) estimation
Bayesian Markov-Chain-Monte-Carlo (MCMC) estimation

JEL codes:
L17
O33

ABSTRACT

Universities in developing countries have rarely been able to subscribe to academic journals in the past. The "Online Access to Research in the Environment" initiative (OARE) provides institutions in developing countries with free online access to more than 11,500 environmental science journals. We analyze the effect of OARE on (1) scientific output and (2) scientific input as a measure of accessibility in five developing countries. We apply difference-in-difference-in-differences estimation using a balanced panel with 249,000 observations derived from 36,202 journal articles published by authors affiliated with 2,490 research institutions. Our approach allows us to explore effects across scientific fields, i.e. OARE vs. non-OARE fields, within institutions and before and after OARE registration. Variation in online access to scientific literature is exogenous at the level of scientific fields. We provide evidence for a positive marginal effect of online access via OARE on publication output by 29.6% with confidence interval (18.5%, 40.6%) using the most conservative specification. This adds up to 2.07 additional articles due to the OARE program for an average institution publishing 7.0 articles over the observation period. Moreover, we find that OARE membership eases the access to scientific content for researchers in developing countries, leading to an increase in the number of references by 8.4% with confidence interval (5.6%, 11.2%) and the number of OARE references by 14.5% with confidence interval (7.5%, 21.5%). Our results suggest that productive institutions benefit more from OARE and that the least productive institutions barely benefit from registration.

1. Introduction

While global online access has laid the groundwork for involving all nation-states in science, universities in developing countries have rarely been able to subscribe to academic journals in the past (Annan, 2004). For instance, most libraries in Sub-Saharan African countries had no access to any scientific journal for years (Suber and Arunachalam, 2005). The Online Access to Research in the Environment (OARE) initiative seeks to provide free or reduced-fee online access for researchers of registered institutions in the field of environmental science. It was launched by the United Nations Environment Programme (UNEP) and Yale University in October 2006. We focus our analysis on the five developing countries that are most active in terms of both publishing (number of articles in Web of Science) and registration with the OARE initiative: Kenya and Nigeria (Sub-Saharan Africa) and Bolivia, Ecuador and Peru (South America).

* Corresponding author at: European Commission, Joint Research Centre, Digital Economy Unit, Edificio Expo, Calle Inca Garcilaso 3, 41092 Seville, Spain
E-mail address: frank.mueller-langer@ip.mpg.de (F. Mueller-Langer).

We are grateful to Kimberly Parker and the World Health Organization for providing us with data on registration dates of the institutions that joined the Research4Life initiatives. We thank Luis Aguiar, Néstor Duch-Brown, Estrella Gomez-Herrera, Dietmar Harhoff, Stan Liebowitz, Bertin Martens, Fabio Montobbio, Laura Rosendahl Huber, Joel Waldfogel, Richard Watt, Michael Weber and the participants of the 2016 MPI-IC research seminar, the 2016 Workshop on "The Organization, Economics and Policy of Scientific Research", the 2015 MPI-IC guest lecture series and the 2015 SERCI Conference for comments. Research assistance was provided by Sebastian Osterrieth, Laura Sundsgaard and Christoph Winter. The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

https://doi.org/10.1016/j.respol.2019.103886
Received 9 January 2019; Received in revised form 9 October 2019; Accepted 4 November 2019
Available online 22 January 2020

We investigate the impact of the OARE initiative on both local publication output as well as publication input. As for the input analysis, we explore the effect of online access to OARE journals on the number of references in general and the number of references to articles published in OARE journals.

All developing countries are eligible, but the initiative distinguishes between so-called Group A and Group B countries. Registered research institutions in Group A countries (gross national income (GNI) per capita below $1,600) receive free online access to all journals that are available under the OARE initiative whereas institutions in Group B countries (GNI per capita below $5,000) receive access for a reduced fee of $1,000 per year. Notably, the OARE registration date varies by institution, i.e., we have different pre-and post-OARE cutoffs for different OARE member institutions.

Using bibliometric article-level data from Web of Science (WoS) and OARE registration data from January 2000 to June 2012, we analyze the impact of OARE on the publication output (i.e. number of articles) and input (i.e. number of references) of research institutions. Our identification strategy is based on the fact that OARE limits free or reduced-fee online access to environmental (OARE) journals. We use a difference-in-difference-in-differences (DDD) estimation that explores differences in publication output and input across OARE and non-OARE fields within institutions that registered with the OARE initiative and those that did not – before and after joining OARE. The underlying idea is that only researchers working on environmental issues can be impacted by free or reduced-fee online access to OARE journals after an institution has registered with OARE. In contrast, other scientific fields within the same institution (and OARE fields within the same institution but before OARE registration) will not benefit from the OARE program. Using this methodology, we compare the publication output (i.e., published journal articles) and publication input (i.e., references cited in these articles) in OARE fields in a given OARE member institution with the output and input in other fields at the same institution and with the output and input of non-member institutions – before and after joining the initiative.

The DDD method has the advantage that it can deal with concerns regarding self-selection at the institutional level that could imply that more productive universities might be more likely to register with OARE. We mitigate these concerns by looking at OARE and non-OARE disciplines in the same institution. In addition, we use instrumental variables and Bayesian methods to account for potential unobserved endogeneity. We find that OARE membership increases the overall number of publications of a research institution by +29.6% (95%-CI: 18.5% to 40.6%), i.e. OARE membership adds, in average, 2.07 (95%-CI: 1.3 to 2.8) articles per institution over the whole period of 50 quarters. This is equivalent to 0.04 extra articles per quarter for an average institution, a rather small economic effect. However, the most productive institutions benefit most from OARE membership while the least productive institutions barely benefit from registration. These findings may have important policy implications as a higher productivity level in academic research may have a positive effect on economic growth and other means of economic prosperity, e.g. environmental innovation (Romer, 1986 & 1990; Griliches, 1957 & 1992; Jaffe, 1989).[1] Moreover, our results suggest that OARE membership promotes access to academic knowledge as an input in the scientific production, measured by the number of references cited by researchers in our dataset. We find that researchers in OARE fields within OARE-registered institutions cite 8.4% (95%-CI: 5.6% to 11.2%) more journal articles than other researchers and also cite 14.5% (95%-CI: 7.5% to 21.5%) more articles from OARE journals than other researchers, respectively.

The remainder of the paper is organized as follows: Section 2 relates our work to the literature on the economics of science, innovation and economic growth. In Section 3, we describe the OARE initiative, discuss potential issues related to self-selection and endogeneity and provide an overview of the data and the variables under study. Section 4 describes the methodology. In Section 5, we present the results of our empirical analysis, provide extensions and discuss robustness checks. Section 6 concludes.

2. Related literature

Access to scientific research has recently attracted widespread interest from economics scholars (Furman and Stern, 2011; McCabe and Snyder, 2015; Sorensen, 2004) and policy-makers (European Commission, 2012). In particular, open access (OA) has been subjected to a broad discussion on whether it is a promising new business model in the digital economy (Suber, 2012; Scheufen, 2015; McCabe and Snyder, 2018).[2]

The literature on open access can broadly be structured along three lines of research: studies investigating the effects of different publishing models (Shavell, 2010[3]; Jean and Rochet, 2010); studies analyzing the impact of different publishing models on readership and citations (Gaulé and Maystre, 2011; McCabe and Snyder, 2014, 2015; Mueller-Langer and Watt, 2018); and studies directed towards a scientist's attitude and behavior regarding OA publishing (Hanauske et al., 2007; Eger et al., 2015). Our paper seeks to contribute to the first line of research. In particular, we study the effects of a change in the ability of researchers in developing countries to access academic works. We analyze the impact of this change before and after these researchers' institutions joined the OARE initiative, and we compare the results to those disciplines within institutions for which the access mode remained unchanged over time. Our research discusses the impact of free or reduced-fee online access on scientific production in developing countries, for which we find little prior literature.[4] However, the need for such research is emphasized by Annan (2004). Our DDD approach allows us to examine the effect of OARE controlling for article characteristics and institutional characteristics such as rank, city population and the distance to the largest domestic city. Evans and Reimer (2009a) emphasize the need to further assess the role of open access in developing countries. Evans and Reimer (2009b, p. 5) show that "lower-middle-income countries tend to much more frequently cite freely available journals, but the poorest countries do not." Thus, scientists in the poorest countries seem to have virtually no access to online content.

[1] We will elaborate on the link between academic research and economic growth in Section 2.

[2] Two arguments mainly drive this debate. First, with the advent of the Internet and the development of technologies to digitize information goods, scientific journal publishers have found new means to price discriminate ("big deals"), which has led to a sharp increase in journal subscription prices (Bergstrom and Bergstrom, 2004; Ramello, 2010) and hence higher costs of access to research. In contrast, OA provides free and unrestricted access to academic works (McCabe and Snyder, 2005, 2014). Second, the copyright system that is behind these pricing schemes is built on the idea that commercial exclusivity granted by copyright generates the main incentive for the creator of a copyright work. Researchers, in contrast, are primarily motivated by reputation rather than by financial gains. Especially for journal articles, authors typically do not receive any royalties, since the copyright is generally transferred to the publisher. Some authors even argue that an abolishment of copyright and hence a forced OA regime would foster scholarly esteem (Shavell, 2010).

[3] Shavell (2010) argues that (a) readership is higher under open access, (b) a higher readership increases scholarly esteem, (c) research institutions would bear the costs of a shift towards the "author pays" model and (d) there are several reasons why legal action is necessary to facilitate a change towards a universal OA regime. Several researchers have critically assessed the assumptions made in Shavell (2010). See Mueller-Langer and Scheufen (2013) for a review.

[4] Gaulé (2009), Frandsen (2009) and Davis (2011) are notable exceptions that we will discuss in this section.

Evans and Reimer (2009a) suggest that poor infrastructure and slow internet access may explain this difference in citation rates. McCabe and Snyder (2015) criticize their paper, arguing that Evans and Reimer (2009a) do not control for citation trends.[5] Our approach complements the two papers, as we analyze both input and output trends of access to academic works for researchers in the developing world.[6] We contribute to this strand of literature by investigating the role of free and reduced-fee online access in developing countries on scientific output and input. Our paper is also related to Gaulé (2009), Frandsen (2009) and Davis (2011). Gaulé (2009) and Frandsen (2009) explore the access restrictions to scientific literature of developing country researchers as compared to developed country researchers. Using a database of 43,150 scientific papers published by Swiss and Indian researchers in 2007, Gaulé (2009) compares backward citation patterns for Swiss and Indian scientists to analyze the difficulties faced by developing country researchers in accessing scientific literature. Gaulé (2009) finds that Indian authors have shorter reference lists and are more likely to cite articles that appeared in open access journals. In a similar vein, Frandsen (2009) conducts a publication and citation analysis for the use of OA journals (for 451 journals) in biology by researchers in developing countries. Frandsen's (2009) results suggest that researchers from developing countries are not attracted to open access journals to a larger extent than researchers from developed countries.

Our paper is closely related to Davis (2011) that analyzes the impact of a digital collection of journal articles in agriculture and allied subjects (TEEAL) on (1) the number of produced articles, (2) the number of references (in general) and (3) the number of TEEAL references for researchers in eleven developing countries. While Davis (2011) finds no effect on the production output of subscribed institutions, he provides evidence for a significant impact on the number of references both in general and for the TEEAL journals. However, our approach differs from Davis (2011)'s approach in several aspects. First, we investigate the impact of free and reduced-fee online access to scientific works for member versus non-member institutions, while Davis (2011) explores the effect of *offline* access to scientific works. In this regard, it is particularly noteworthy that the TEEAL journal content is shipped to subscribing institutions on a portable hard-drive copy (CD-ROM) upon payment of $ 5,000.[7] As a result, both the offline access and the payment may be seen as substantial costs in making use of TEEAL journals for researchers in the developing world. Secondly, the OARE initiative offers access to more than 11,500 journals in environmental sciences, while TEEAL offers access to over 200 journals in agricultural sciences. Thirdly, we broaden the geographical scope by looking at Sub-Saharan African and South American countries. Finally, we do not only look at the differences in impact for member versus non-member institutions, but also explore the effect within institutions by applying a triple difference approach to investigate both input and output effects. Similar to Davis (2011), we also find that free or reduced-fee access to academic journals increases the number of references both in general as well as with respect to the initiative (TEEAL in Davis (2011), OARE in our analysis).

Our paper also contributes to the literature in the broader field of economics of science and innovation investigating the role of science and scientific research in the advancement of technologies and hence in fostering economic growth (Dasgupta and David, 1994; Dosi, 1988; Merton, 1973; Murray et al., 2016).[8] In general, Romer (1986, 1990) highlights the role of academic research as a major factor for technological innovations and hence for economic growth. Before Romer, the literature especially by Solow (1956) and Swan (1956) were able to explain the role of academic research for economic growth by means of a residual as growth was exogenously determined. Romer's endogenous growth theory emphasizes the relevance of spill-overs from academia. Accordingly, when free online access increases scientific output, this eventually may have a positive effect on innovation and economic growth. Extending on Romer (1990) several authors have emphasized the importance of knowledge spillovers from science for economic growth (Griliches, 1992; Jaffe, 1989; Audretsch and Feldman, 1996; Acs et al., 1994).[9] However, these spillovers cannot be taken for granted as we find that only 5 percent of eligible institutions are OARE members which points to the unused potential of the initiative.

3. Data and variables

3.1. The OARE initiative

The OARE initiative is led by the United Nations Environment Programme (UNEP) in partnership with major publishers in environmental science.[10] OARE was launched in October 2006. Today, OARE offers access to more than 11,500 peer-reviewed scientific journal titles published by 461 OARE partners in more than 100 eligible countries. In this regard, eligibility distinguishes between Group A (free online access) and Group B (low-cost access) countries,[11] depending on the countries' GNI per capita. Institutions in countries with a GNI per capita at or below $ 1,600 receive free access to the full content of journal articles, while institutions in countries with a GNI per capita below $ 5,000 pay a fee of $ 1,000 per year. However, institutions have to register to OARE in order to receive access. In this respect, OARE offers courses and workshops for librarians and researchers in order to make the initiative known to a wider audience.

Fig. 1 illustrates the rate of adoption of OARE over time (quarters) in all countries (as given by the middle solid line) and separately in Group A countries (upper dashed line) and Group B countries (lower dashed line).[12] The rate of adoption is measured by the cumulative number of institutions that joined OARE in a given quarter divided by the total number of institutions, i.e. 2,490 institutions in all countries (Group A countries: 1,599; Group B: 891). Finally, it is worth noting that about 5% of all eligible research institutions in Group A countries and about 4% of all eligible research institutions in Group B countries had joined OARE in the last quarter under study (June 2012).[13]

To account for self-selection, we apply a triple difference-in-difference approach dealing with concerns regarding self-selection at the institutional level that could imply that more productive universities

[5] We follow McCabe and Snyder's (2015) approach to control for trends in publication output.

[6] Input is measured by the relative number of cited OARE articles in a given article, while output is measured by the total number of articles of a given institution.

[7] We obtained the information on the fee payment from the TEEAL website at https://teeal.org/purchase.

[8] See also Stephan (1996) for an overview of the economics of science literature.

[9] See Diamond (1994) for an overview on Zvi Griliche's contributions for understanding the economics of technology and growth. See also Geroski (2000), Hall (2004), Hall and Kahn (2003) and Mansfield (1961, 1963).

[10] See http://oare.research4life.org/content/en/partners.php for an overview on the major partners of the OARE initiative.

[11] Please note that countries can also convert from one group to the other if the GNI per capita changes over time. As such, Bolivia changed from group A to group B in 2017. For the time horizon under study, however, we do not find any group transitions.

[12] We used Internet Archive's Wayback Machine to explore possible group changes over time. All countries under study remained in the same group for the period under study, i.e. 2000 to 2012.

[13] Note that the total number of eligible institutions refers to institutions that have observable research output in the period under study. We exclude non-research institutions from our analysis, i.e., we drop institutions that did not publish any journal article during the period under study.

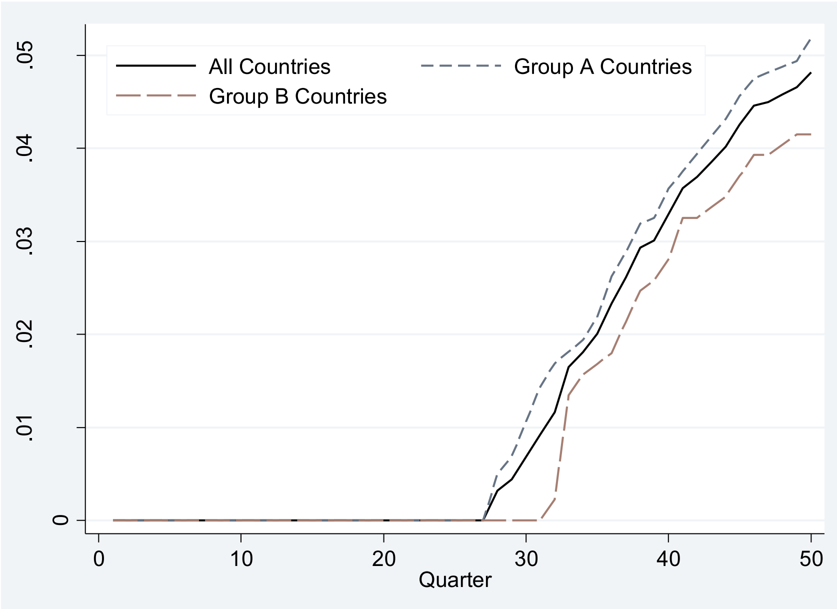


Fig. 1. OARE ADOPTION OVERALL AND BY COUNTRY GROUP Adoption patterns of OARE for all countries (as given by the middle solid line) and by country group (dashed lines). The OARE initiative started in quarter 28. Quarter 28 is thus the earliest possible quarter for OARE registration. Group A countries: Kenya, Nigeria and Bolivia. Group B countries: Ecuador and Peru. Fig. 1 suggests that different institutions registered with OARE at different points in time. This implies that we have multiple cut-offs for before and after OARE depending on the respective institution. Fig. 1 also implies that only about 5% of all eligible institutions joined OARE after a period of 5 ½ years. It points to the unused potential of the OARE initiative.

might be more likely to register with OARE. The triple difference-in-difference approach compares the treatment effect in OARE field articles with a control group of non-OARE field articles at the institution level, thus mitigating self-selection of institutions into the OARE initiative.

3.2. Instrumental variables and endogeneity

There might still be endogeneity issues if the unobserved variable is correlated with the OARE treatment effect. There are two problems that we may worry about: (1) unobserved endogenous benefits (not controlled for by other independent fixed effects, such as institutions having better digital infrastructures) which could result in self-selection into OARE; (2) unobserved endogenous information problems resulting from the fact that only well-informed institutions can join the OARE initiative.

We extend our analysis by using an instrumental variable (IV) approach that takes into account the fact that many institutions located in similar areas in a given country joined other online-access initiatives in addition to OARE (see also Section 4.2. “DDD using Instrumental Variables” and Section 5.2. “The Effect of OARE using IV Estimation”). In particular, next to the OARE initiative there are further initiatives within the “Research4Life”-program initiated to foster access to scientific content also in health science (HINARI) and agricultural science (AGORA). The Health Inter-Network Access to Research Initiative (HINARI) was launched in January 2002 (quarter 9), offering free or reduced-fee access to more than 15,000 journals in biomedical and health research. The Access to Global Online Research in Agriculture (AGORA) initiative was launched in October 2003 (quarter 16), providing with access to a collection of around 14,500 key journals in agricultural science for researchers in developing countries.

Using information on these Research4Life initiatives, we implement our IV approach using two methodologies. First, we use a two-stage least squares method that initially analyses the factors that explain self-selection into the OARE initiative using instruments based on the average number of institutions that joined OARE, HINARI and AGORA in similar areas and then uses the predicted value of the treatment effect in the productivity equation. Secondly, we simultaneously estimate the parameters of the selection and the productivity equation using Bayesian MCMC methods (see also the methodology Section 4.3. and the results Section 5.3.).

3.3. Data

Our dataset is built from three main sources. First, we collected bibliometric article-level data from WoS for the five countries under study. We focus our analysis on five countries for the following reasons. On the one hand, we choose the most productive countries in terms of the total number of research articles from January 2000 to June 2012 for both geographical regions (Sub-Saharan Africa and South America). On the other hand, we look at countries that exceed a threshold of at least 20 OARE institutions in order to have variation across institutions within countries.[14] Second, we gathered institutional data including institutions’ registration with OARE. Third, we extracted the rank of the institutions from the Ranking Web of Universities.

Regarding the first data source, we collected a panel dataset containing metadata for 36,202 research articles. The period under study starts in January 2000 (quarter 1) and ends in June 2012 (quarter 50). We obtain article metadata from WoS. The WoS data contain the institutions of the authors, the title of the paper, journal information (publication date, number of pages, volume number, issue number) and the number of citations. Overall, we have 2,490 institutions that published at least one article over the period under study.

We use article-level data for assigning different characteristics to each single article, accounting for the field of research, institutional affiliations of the authors, cooperation with authors from outside the developing world and other controls such as number of references, pages etc. Since the OARE initiative offers free or reduced-fee online access to research in environmental science, we create a dummy variable indicating whether an article falls under an OARE research area. We define an article as falling under an OARE research area if its “Research Area” provided by WoS also appears frequently in the titles of OARE journals. We proceed as follows. First, for all articles under study, we extract all terms from the WoS “Research Area” field. Second, we order these research area terms by frequency, i.e., we count how many articles in the data fall under a given single-word term (henceforth, WoS research area terms). For instance, in the case of articles of authors affiliated with Nigerian universities, the term “environmental”

[14] Please also note that the data creation process involved manual matching of institutions using different versions of search terms in Stata's string matching functions. We will further elaborate on the data creation process below in Section 3.3.

appears 2,179 times, whereas the term "architecture" appears once. Next, we extract the 200 most frequent terms that appear in the complete list of titles of OARE journals (henceforth, top 200 title terms). Matching these two lists (WoS research area terms and top 200 title terms), we obtain the top 50 OARE research areas. The top 50 OARE research areas are given by the 50 most frequent WoS research area terms that are also included in the top 200 title terms. Finally, we use a "top 50 OARE research"-dummy (which is one if an article falls under the top 50 OARE research fields, 0 otherwise). Distinguishing between OARE fields and non-OARE fields within a given institution allows us to explore the effect of online access to OARE journals on scientific output in OARE fields before and after OARE registration as well as in OARE fields as compared to non-OARE fields before and after OARE registration.[15]

Our sample contains all articles of researchers of the countries under study, including both single and multiple authored articles. However, dividing the share of a publication between different (local) authors to determine the respective contributions of authors is a challenging task for at least two reasons. First, there is no consensus within and across disciplines on how to account for multiple authorships. In particular, taking each author of a paper fully into account would overestimate the output produced. Creating a weight for multiple authored papers by dividing each publication by the number of authors, however, would also necessarily involve assumptions on the habits of co-authorship. In some disciplines (or publishing cultures), the order of authors has clear implications. Sometimes the first author or the last author is perceived as the "main author" of a research article. Other disciplines choose the order of authors alphabetically or by status. All of this makes it hard to operationalize multiple-authored papers from one country. Second, to the best of our knowledge, McCabe and Snyder (2015) is the only reference that discusses the issue of single versus multiple authors with respect to online access. They nevertheless restrict their sample to single authors (from a local country) due to the difficulties in dividing the share of a multiple authored paper between the authors. However, to consider multiple co-authored articles in addition to single authored articles has two main advantages. First, only looking at single authored articles would substantially reduce our sample by 18,955 articles, that is, more than 50% of our dataset. Second, multiple co-authored articles may have different characteristics than single authored articles.

In the light of these advantages, we address these concerns as follows. First, we account for multiple authorship by simply dividing the institutional share of each paper by the number of authors. For instance, a paper with two authors from two institutions leads to an increase in output of 0.5 for each of these institutions.[16] Second, to test the robustness of our results, we provide the regression results for single authored papers in Section 5.5.2 (Table 9).[17] The results are remarkably similar.

To construct the (balanced) panel, we gather article level information by institution, field (OARE vs. non-OARE) and quarter for each country under study. For each country, we then merge rank and city information – including population and distance data – from separate datasets. Subsequently, we merge all individual country data into one dataset.[18] We distinguish country-specific information by generating a unique country ID for all countries. In total, we obtain 249,000 institution-discipline-quarter triplets, which constitute our unit of observation.

In assigning institutions to authors of articles from the countries under study, we use Stata string-matching functions, searching for snippets of institution names and abbreviations. In particular, we manually account for different versions and spellings of institutions, as WoS does not provide with a unique number or code to unambiguously identify a particular institution of interest. Most importantly, also spelling errors, case sensitivity as well as abbreviations impede an automatic matching of articles and institutions. Last but not least, we repeat the string matching process for each country file for each author level, accounting for up to 11 levels (11 authors for each article) and distinguishing different author level IDs. We unambiguously identify 459 research institutions that are part of the Ranking Web of World Universities and/or OARE member institutions, forming the core universities for the string-matching process.[19] For each country under study, we find a large number of institutions that are neither included in the Ranking Web of Universities list nor in the list of OARE institutions. For these institutions, we generate unique institution IDs as follows. First, we order the institutions in a given country alphabetically. Second, we identify all instances of a given institution in the raw data. For instance, a given institution can have multiple versions because of abbreviations, use of different languages, or typos. Thereby, we also use the city where an institution is located to identify different versions of a given institution, manually assigning identical institution IDs in such cases.

Moreover, we assign institution IDs to track the relative position of an institution in the university ranking list. For a given country, a lower institution ID reflects a better rank. The rank variable, in addition, reflects the absolute worldwide position of the institution in the Ranking Web of World Universities. This ranking provides information on the performance of 22,123 research institutions worldwide on the basis of the web presence as well as the impact of institutions. The former aspect is particularly noteworthy as web presence provides also a proxy for the technical expertise needed to set up online access to journals.

Finally, we assign city IDs to construct distance and population variables. To give an example, we identify 74 cities in Nigeria with a population of more than 100,000 inhabitants (pop variable) using the World Population Review (2017), listing population numbers for each city in each country of the world. In addition, we identify 64 cities from our Nigeria sample with fewer than 100,000 inhabitants. We assign city IDs 1 to 138 to the Nigerian cities, where a lower number denotes a larger population. As a further control, the variable *distance_1* was created by using Google maps and by computing the distance in km from the city in which an institution is located to the largest domestic city, as suggested by the first itinerary option by car. In addition, we create a *distance_2* variable indicating the distance of an institution's location to the next domestic 1 million city.

3.4. Definition of variables

Table 1 provides an overview of the variables under study and summary statistics at the institution-discipline-quarter level.[20] Variables can be grouped into six categories: dependent variable, countries, main variable of interest, article characteristics, institutional characteristics and city characteristics.

[15] See Section 4 on the methodology. See also Section 3.4.2 on the definition of our treatment variable.

[16] Note that due to the complexity of the manual string matching process we restrict our calculations to include multiple authored papers with up to 11 authors. This restriction does not reduce our sample by too much because 94% of all papers have 11 or less authors.

[17] Appendix 1 provides summary statistics for the sample of single local authors.

[18] We take the mean for the continuous variables, the max for the binary variables and the sum for the publication variable in performing the collapse command.

[19] In total, 163 institutions in Nigeria, 96 in Peru, 82 in Kenya, 62 in Ecuador and 56 in Bolivia. Please note that for all institutions that are not listed, we still have the information that their rank is above 25,000. Since we use broadly defined rank categories, we can therefore include rank information for all institutions under study.

[20] Appendix 2 provides summary statistics by country group.

Table 1
Summary Statistics.

	mean	sd	min	max	N
Dependent variables					
# publications	0.140	1.194	0	93.82	249,000
# references	20.60	21.27	0	293	249,000
# OARE references	4.293	7.222	0	135	249,000
Countries					
Kenya	0.256	0.436	0	1	249,000
Nigeria	0.257	0.437	0	1	249,000
Bolivia	0.129	0.335	0	1	249,000
Ecuador	0.130	0.336	0	1	249,000
Peru	0.228	0.419	0	1	249,000
Main variable of interest					
OARE treated (DDD)	0.007	0.0833	0	1	249,000
Article characteristics					
# co-authors USA	0.472	1.329	0	37.25	249,000
# co-authors EUR	0.543	1.651	0	57	249,000
# pages	6.205	6.306	0	120	249,000
Institutional characteristics					
Rank1: rank ≤ 5,000	0.0201	0.140	0	1	249,000
Rank2: 5,000 < rank ≤ 10,000	0.0161	0.126	0	1	249,000
Rank3: 10,000 < rank ≤ 15,000	0.0181	0.133	0	1	249,000
Rank4: 15,000 < rank ≤ 25,000	0.0253	0.157	0	1	249,000
Rank5: rank > 25,000	0.920	0.271	0	1	249,000
Rank, in 1,000	7.185	4.855	0.749	21.79	249,000
City characteristics					
Distance from largest domestic city, in 100 km	3.254	3.613	0	20.64	249,000
Distance from closest dom. city with > 1 million inhabitants, in 100 km	1.991	3.216	0	20.64	249,000
Pop0: pop ≤ 100, in 1,000	0.220	0.414	0	1	249,000
Pop1: 100 < pop ≤ 500, in 1,000	0.106	0.308	0	1	249,000
Pop2: 500 < pop ≤ 1,000, in 1,000	0.159	0.366	0	1	249,000
Pop3: 1,000 < pop ≤ 5,000, in 1,000	0.335	0.472	0	1	249,000
Pop4: pop > 5,000, in 1,000	0.180	0.384	0	1	249,000

We use a balanced panel and take into account journal articles by both single and multiple local authors. Data is aggregated at the institution-discipline-quarter level that constitutes our unit of observation.

3.4.1. Dependent variables

Our dependent variable, $y_{s,t,r}$, indicates the number of publications of institution s in quarter t in discipline r. Recall from Section 3.3 that we consider journal articles from both single local authors as well as multiple local authors to create our dependent variable. In addition, as reported in Table 1, the maximum number of publications per institution/quarter/discipline is 93.82. Hence, our dependent variable is not a count variable (and we thus refrain from running a Poisson model).[21] Following McCabe and Snyder (2015), Fig. 2 illustrates patterns in publication output over time, i.e., from quarter 1 to quarter 50.

Publication output follows a steady upward trend, reaching a level in the last quarter under study about 8% higher than in the base quarter 1. The results presented in Fig. 2 suggest that it is important to control for secular trends in publication output.[22] As discussed in Section 4 (Methodology), we control for these secular trends in the regressions by including binary variables for quarters at the discipline level.

3.4.2. Independent variables

Countries: We study 2,490 institutions from five countries of which two are located in Sub-Saharan Africa (Kenya and Nigeria) and three in South America (Bolivia, Ecuador, Peru). At the institution-discipline-quarter level, 51.3% of our observations are from Sub-Saharan Africa.

Main variable of interest: (OARE) treated is our main variable of interest. We construct this treatment variable by interacting three dummy variables. First, *OARE* indicates whether papers are written by authors affiliated with OARE institutions. We generate the OARE dummy by using the institution IDs of all institutions that are part of UNEP's list of OARE institutions. *OARE* (not reported in the table) takes on the value 1 if the respective institution of an article under study is an OARE institution and the value 0 otherwise. Second, the *after* dummy (not reported in Table 1) accounts for the registration date (in quarters) of a certain OARE institution. Its value is 1 if the article under study was written by an author affiliated with an OARE institution after the institution joined the OARE program and 0 otherwise.[23] Recall that the *after* dummy turns 1 in different quarters for different OARE institutions as OARE institutions registered at different points in time. Third, we generate an OARE research field dummy capturing whether a particular article is within the top-50 OARE research areas or not. This allows us to compare differences within institutions, i.e. differences between disciplines that are core OARE research fields (e.g. environmental science) versus non-OARE fields of research (e.g. economics).

Article characteristics: #co-authors USA (*#co-authors EUR*) indicates the average number of co-authors from the US (Europe). *#pages* indicates the average number of pages.

Institutional characteristics: Five variables indicate the rank of an institution constructed from the Ranking Web of Universities (2014). *Rank1* represents the best institutions (rank ≤ 5,000) whereas *Rank4* corresponds to institutions with the lowest reported ranks (15,000 < rank ≤ 25,000). *Rank5* indicates that an institution is not listed, which implies that its rank is above 25,000; these institutions are

[21] The histogram of the number of publications at the institution-discipline-quarter level is shown in Appendix 3.

[22] Using a panel of citations to economics and management journals, McCabe and Snyder (2015) explore whether online availability of journals increases cites to scientific works. McCabe and Snyder's (2015) Fig. 2 illustrates that citations follow a steady upward trend. Based on this finding, McCabe and Snyder (2015) argue that it is important to account for these secular citation trends in order for the online access variable of interest to be identified. Here, we address the underlying identification problem in a similar fashion.

[23] For non-OARE members *after* is set to 1 for all quarters after quarter 28 (launch of OARE).

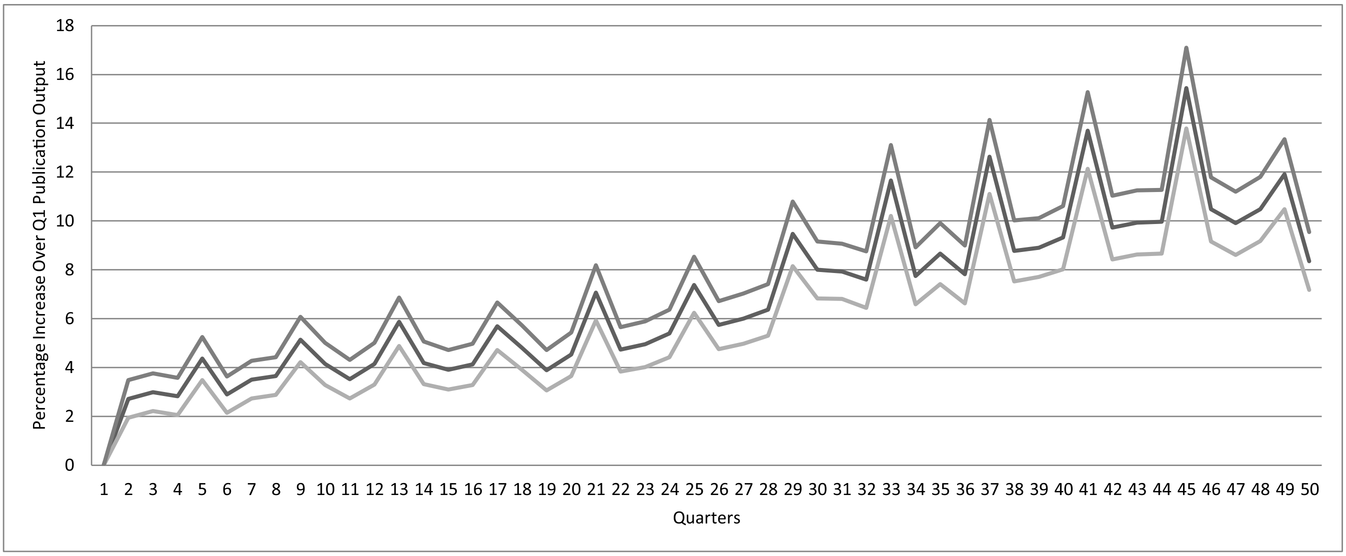


Fig. 2. SECULAR TREND IN PUBLICATION OUTPUT Fig. 2 illustrates patterns in publication output over time, i.e., from quarter 1 to quarter 50. We use a balanced panel and take into account journal articles by both single and multiple local authors. In order to be able to interpret the OLS results for the 49 quarter dummies directly in terms of percentage changes compared to the reference quarter Q1, here we use the log of the number of publications of institution *s* in quarter *t* in discipline *r* as dependent variable. Following McCabe and Snyder (2015, Fig. 2) the middle curve plots a set of quarter fixed effects from OLS regression using Q1 as reference quarter. We use the *xtreg* command implemented in STATA. The underlying regression also includes binary variables for countries and quarter-discipline pairs. Outside lines bound the 95% confidence interval based on robust standard errors clustered at the institution level. This figure illustrates that publication output follows a steady upward trend, reaching a level in the last quarter under study about 8% higher than in the reference quarter Q1.

the least productive in scientific output.[24]

City characteristics: We use two different distance measures, indicating (1) the distance in 100 km of a given city to the largest domestic city (henceforth also *distance_1*) or (2) the distance in 100 km of a given city to the next domestic 1 million city (henceforth also *distance_2*).[25] City population dummies indicate the number of inhabitants of the city where an institution is located: *Pop0* indicates cities with less than 100,000 inhabitants whereas *Pop4* indicates cities with more than 5,000,000 inhabitants.

4. Methodology

4.1. DDD using OLS regressions

In order to analyze the effect of the OARE initiative, we use a DDD method for comparing the change in research output and input for research fields in the treatment group (i.e. environmental sciences in registered institutions after OARE registration) with the change in research output for scientific fields in the control group (i.e. environmental sciences in registered institutions before OARE registration, non-environmental sciences in registered institutions and all research fields in unregistered institutions) before and after a given institution has registered (or not) with OARE. The intuition behind the DDD approach is the following. Within an OARE institution, only researchers working on environmental issues can be impacted by free or reduced-fee online access to environmental (OARE) journals after the institution has registered with OARE. In contrast, other scientific fields within the same institution (and OARE fields before OARE registration) will not benefit from the OARE program. Exploring effects of online access across scientific fields within a given institution mitigates concerns of self-selection at the institutional level, for instance, because better/more productive institutions might be more likely to register with OARE. Following McCabe and Snyder (2015), we also include quarter-discipline and institution-discipline fixed effects in the regressions to account for the effects of institutions, disciplines and time on research productivity.

The dependent variable, $y_{s,t,r}$, is the number of published articles by researchers from institution *s* in quarter *t* in research area *r* (henceforth, discipline). We use the specification outlined in Eq. (1):

$$y_{s,t,r} = \beta_0 + \beta_{1,t,r} fe_{t,r} + \beta_{2,s,r} fe_{s,r} + \beta_3 \text{treated}_{s,t,r} + \sum_k \beta_{4,k} x_{k,s,t,r} + \varepsilon_{s,t,r} \quad \text{with } k = 1, \cdots, K; \ \ t = 1, \ldots T \tag{1}$$

where $fe_{t,r}$ are quarter-discipline fixed effects (100 quarter-discipline pairs in the balanced panel using the full sample); $fe_{s,r}$ are institution-discipline fixed effects (4,890 institution-discipline pairs). Variable $treated_{s,t,r}$ is our main variable of interest. It accounts for the fact that institutions registered with the OARE initiative at different points in time and that other disciplines than environmental sciences in a given institution will not benefit from OARE. In other words, *treated* is 1 if an institution is an OARE institution and if articles of affiliated researchers are published in the OARE research discipline in a quarter after the institution registered with OARE (and 0 otherwise). $x_{k,s,t,r}$ are *k* control variables ($k = 1, \ldots, K$).[26] $\varepsilon_{s,t,r}$ are unobservable effects assumed independent across *s, t* and *r*. When we refrain from including institution-discipline fixed effects, we include institutional characteristics (such as worldwide rank and the number of publications during which an institution published at least one article) and city characteristics (such as distance to largest domestic city/next 1-million city and city population).

We also explore the effect of OARE on publication input as measured by (a) the number of cited references and (b) the number of cited

[24] Using rank categories instead of the actual rank has the advantage of being invariant to small variations in rank over time.

[25] We do not have distance information for 206 institutions, as the respective cities do not appear in Google maps. For these cities, we proxy the distance to the largest domestic city by taking the average distance in the respective country. We use the same approach for our alternative distance variable.

[26] For instance, we include article characteristics such as the number of pages, co-authors USA and co-authors EUR. We also include dummy variables indicating the country where a given institution is located.

OARE references. For this purpose, we use again the specification outlined in (1). The only difference is that $y_{s,t,r}$ now indicates (a) the number of references in articles by researchers from institution *s* in quarter *t* in research discipline *r*, and (b) the number of OARE references in articles by researchers from institution *s* in quarter *t* in research discipline *r*, respectively.

4.2. DDD using instrumental variables

A potential endogeneity problem arises if the unobservable variable in the self-selection equation is correlated with the unobserved productivity factors (see Section 3.2 above). There are two such unobservable factors that we need to deal with. First, only well-informed institutions can join the OARE initiative, especially since institutions that had registered with earlier Research4Life initiatives (i.e., HINARI and AGORA) may be better informed about the benefits of joining OARE. This could lead to unobserved endogenous information problems that could explain the low OARE adoption rate. Since registration with OARE is either costless (for the majority of the institutions of our sample that belong to group A countries) or very cheap, direct monetary cost are unlikely to be a factor for joining OARE. However, there might be hidden administrative costs that could explain OARE membership. Secondly, unobserved endogenous problems resulting from an insufficient ICT-infrastructure could also influence OARE membership. Indeed, a sufficiently developed infrastructure is a prerequisite for online access. We construct three instruments that can account for unobserved informational and infrastructure effects at the institution level. For each institution, we compute the average number of institutions that have joined the OARE, the HINARI and the AGORA initiatives in similar geographic areas of the country in which the institution is located. More specifically, we constructed five distance categories according to whether a given institution is located in a city with more than one million inhabitants, less than 50 km away from such a city, between 50 km and 250 km, between 250 km and 750 km and more than 750 km away from a city with more than one million inhabitants. We also use the five categories of cities according to the size of the population: less than 100,000, between 100,000 and 500,000, between 500,000 and 1 million, between 1 million and 5 million, more than 5 million. Then, we compute the average number of institutions that joined the OARE, the AGORA and the HINARI initiatives – in similar distance and population areas where a given institution is located – for each quarter in the sample. From this we use the one-quarter lagged values as instruments (respectively *m_OARE, m_AGORA* and *m_HINARI)*. These instruments should control for both different levels of awareness about the existence of Research4Life before institutions join the OARE program as well as infrastructure related issues that could lead to self-selection into the OARE program. They should be uncorrelated with individual unobserved productivity factors at the institution level and thus can be used as valid instruments.

4.3. DDD using Bayesian MCMC estimation

We estimate the treatment effect using Bayesian estimation techniques based on a data augmentation MCMC algorithm described in Appendix 4 that can deal with endogeneity. There are two equations. The first equation corresponds to the self-selection process and determines the outcome of the binary treatment effect within a latent variable framework. The second equation is the productivity Eq. (1). We assume that the unobserved variables of both equations follow a bivariate normal distribution with correlation coefficient ρ. The MCMC algorithm simulates the latent variable of the first equation to generate the endogenous binary treatment effect. The Bayesian approach explicitly deals with the correlation between the unobserved variables of the two equations. If there are any unobserved variables that determine whether an institution self-selects into the OARE program, the Bayesian method accounts for its potential endogeneity on the estimation of the treatment effect.

5. Empirical analysis

Our empirical analysis explores the effect of OARE on publication output, applying triple difference regressions using OLS in Section 5.1., instrumental variables estimation (Section 5.2.) and Bayesian MCMC estimation (Section 5.3.). In Section 5.4., we extend our analysis to the effect of OARE on publication input, covering both OARE references as well as total number of references. Section 5.5. provides additional robustness checks.

5.1. The effect of OARE on publication output using OLS

We estimate the impact of OARE membership on scientific output by using eight different specifications in Table 2. Specifications (1) to (8) use OLS estimation.[27] Column (1) reports the OLS regression coefficients for the basic model, including the treatment variable, country and quarter dummy variables as well as dummy variables for 100 quarter-discipline pairs. We add article characteristics in (2), institutional rank information in (3), city population in (4), distance to the largest domestic city in (5) and distance to the closest domestic city with more than 1 million inhabitants in (6). In specification (7), we include institution-discipline fixed-effects instead of country dummy variables and institutional and city characteristics (rank, population and distance). In specification (8), we add dummy variables indicating the numbers of quarters during which an institution published at least one article.

We find a positive and robust OARE effect that is statistically significant at the 1% level across all specifications.[28] The marginal OARE effect ranges from +29.6% in specification (7) to +31.3% in specification (8).[29] We also ran the regressions separately for Group A and Group B countries (Appendix 5) and for each of the five countries (Appendix 6). The OARE treatment effect is positive and statistically significant for the subgroups. It is higher for institutions in Group A countries (i.e. free access countries) than for institutions in group B countries (i.e. reduced-fee countries).

Moving from column (1) to column (2), we consider the effects of article characteristics on publication output. R-squared is similar (0.092 versus 0.085) and the OARE effect remains almost the same when we include article characteristics in (2). R-squared increases from 0.085 to 0.107 while the OARE effect decreases only slightly when we add institutional rank information in (3).[30]

We also find that lower-ranked institutions are less productive in terms of publication output, since the coefficients associated with lower ranks ($5000 < rank \leq 10000$, $10000 < rank \leq 15000$ and $rank \geq 25{,}000$) as compared to the best rank category $rank \leq 5000$ (reference category) are negative and statistically significant at least at the 5% level across columns (3) to (6).[31]

[27] We use the *xtreg* command in STATA. The institution-discipline-quarter triplets constitute the unit of observation. Random institution-discipline effects are included in columns (1) to (6).

[28] All country dummy variables are negative. Recall that Nigeria is the reference country. This suggests that Nigeria has the largest publication output.

[29] We obtain these results by dividing the OARE treatment effect by the total number of publications of an average institution over the full period. For instance, for specification (7), we obtain 2.069/7.0 = 0.296 and for specification (8) 2.191/7.0 = 0.313, respectively.

[30] Note that the Ranking Web of Universities that we use to create the rank variable is mainly based on the assessment of the web presence of institutions, e.g., it uses link analysis for quality evaluation. In this respect, an institution's web performance provides a proxy for its technical expertise to set up online access to journals.

[31] In specification (8), the rank dummies capture productivity effects that are not captured by the number of quarters with publications FE while in columns (3) to (6) the rank dummies capture all productivity effects.

Table 2
Effect of OARE on Publication Output.

Model:	(1) Base	(2) + Article info	(3) + Rank	(4) + Pop.	(5) + Dist._1	(6) + Dist._2	(7) + Inst.-Disc. FE	(8) # of quarters with publications FE
Dependent variable:	*w*	*w*	*w*	*w*	*w*	*w*	*w*	*w*
OARE treated (DDD)	2.110***	2.100***	2.093***	2.093***	2.093***	2.093***	2.069***	2.191***
	(0.401)	(0.401)	(0.399)	(0.399)	(0.399)	(0.399)	(0.394)	(0.0216)
# pages		0.00512*	0.00441	0.00431	0.00433	0.00434	0.00671*	0.000611
		(0.00307)	(0.00303)	(0.00303)	(0.00303)	(0.00303)	(0.00382)	(0.000391)
# co-authors USA		0.00402	0.00376	0.00377	0.00379	0.00375	0.00137	−0.0069***
		(0.0105)	(0.0103)	(0.0103)	(0.0103)	(0.0103)	(0.0115)	(0.00145)
# co-authors EUR		0.0431	0.0429	0.0428	0.0429	0.0429	0.0493	0.00279**
		(0.0319)	(0.0320)	(0.0320)	(0.0320)	(0.0320)	(0.0361)	(0.00116)
Rank2: 5,000 < rank ≤ 10,000			−0.819**	−0.818**	−0.822**	−0.814**		0.00778
			(0.395)	(0.396)	(0.396)	(0.396)		(0.0182)
Rank3: 10,000 < rank ≤ 15,000			−0.907***	−0.888***	−0.897***	−0.888***		0.0413**
			(0.343)	(0.342)	(0.341)	(0.342)		(0.0180)
Rank4: 15,000 < rank ≤ 25,000			−0.00279	0.00136	0.0108	0.00991		0.213***
			(0.537)	(0.533)	(0.534)	(0.533)		(0.0165)
Rank5: rank > 25,000			−0.984***	−0.999***	−0.997***	−0.994***		0.0859***
			(0.325)	(0.327)	(0.327)	(0.327)		(0.0130)
Pop1: 100 < pop ≤ 500, in 1,000				−0.0565	−0.0203	−0.0361		−0.0768***
				(0.0488)	(0.0506)	(0.0532)		(0.00965)
Pop2: 500 < pop ≤ 1,000, in 1,000				0.0579	0.0973	0.0734		−0.0485***
				(0.0667)	(0.0730)	(0.0718)		(0.00979)
Pop3: 1,000 < pop ≤ 5,000, in 1,000				0.0605	0.0356	0.0393		−0.0236***
				(0.0630)	(0.0584)	(0.0637)		(0.00913)
Pop4: pop > 5,000, in 1,000				0.00191	−0.0987	−0.0415		−0.0508***
				(0.0364)	(0.0634)	(0.0413)		(0.00931)
Distance from largest domestic city, in 100 km					−0.0173**			
					(0.00837)			
Distance from closest dom. city > 1 mill. inhab., in 100 km						−0.00966		0.00113
						(0.00619)		(0.000807)
Constant	0.333***	0.262***	1.158***	1.150***	1.237***	1.160***	0.0451	−0.0501**
	(0.0601)	(0.0623)	(0.319)	(0.321)	(0.322)	(0.320)	(0.0302)	(0.0230)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES	YES	YES
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES	NO
# quarters with publications FE	NO	NO	NO	NO	NO	NO	NO	YES
Observations	249,000	249,000	249,000	249,000	249,000	249,000	249,000	249,000
R-squared, overall	0.0919	0.0846	0.1071	0.1079	0.1089	0.1081	0.0761	0.524
Number of Inst_Discipline	4,980	4,980	4,980	4,980	4,980	4,980	4,980	4,980
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490	2,490

We use a balanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD. We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is $rank \leq 5000$. Reference population is $pop \leq 100$. Random institution-discipline effects are included in columns (1) to (6). Robust standard errors clustered at the institutional level. Note that serial correlation is not an issue in our balanced panel because the large number of periods with 0 publications breaks any time correlation for any given institution. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

In addition, the distance to the largest domestic city has a negative impact on output that is statistically significant at the 5% level. The distance from the closest domestic city with more than 1 million inhabitants has a negative impact on output that is not statistically significant. Moreover, the distance variable also proxies the importance of a sufficient ICT-infrastructure for online access to scientific content. This suggests that institutions further away from a metropolitan region may simply benefit less from free or reduced fee online access due to insufficient ICT.

To mitigate concerns about outliers and about the large number of quarters during which an institution did not publish any article, we construct an alternative, binary dependent variable *y*. It is equal to 0 when an institution does not publish any article in a given quarter and scientific field. It is equal to 1 for any positive article contribution of affiliated authors. We run the same regressions as reported in Table 2 using this alternative, binary dependent variable. Results are reported in Table 3. Coefficients can be directly interpreted as percentage effects.

The results reported in Table 3 suggest that OARE has a positive and statistically significant effect on the probability of any positive article contribution of affiliated authors. The marginal OARE effect ranges from +15.8% in column (8) to +22.4% in column (1). These results provide additional empirical support for a robust effect of OARE on

Table 3
Effect of OARE on Publication Output (Linear Probability Model).

Model:	(1) Base	(2) + Article info	(3) +Rank	(4) +Population	(5) +Distance_1	(6) +Distance_2	(7) +Inst.-Disc. FE	(8) +# quarters with publications FE
Dependent variable:	*y*	*y*	*y*	*y*	*y*	*y*	*y*	*y*
OARE treated (DDD)	0.224***	0.224***	0.216***	0.216***	0.216***	0.216***	0.205***	0.158***
	(0.0219)	(0.0219)	(0.0224)	(0.0224)	(0.0224)	(0.0224)	(0.0230)	(0.00502)
# pages		0.00107***	0.000629***	0.000587***	0.000595***	0.000604***	−0.000529***	−2.13e-05
		(0.000154)	(0.000165)	(0.000166)	(0.000166)	(0.000166)	(0.000144)	(9.08e-05)
# co-authors USA		0.00274**	0.00262**	0.00257**	0.00258**	0.00255**	−0.000864	3.50e-06
		(0.00134)	(0.00104)	(0.00103)	(0.00104)	(0.00104)	(0.000763)	(0.000337)
# co-authors EUR		−0.00163***	−0.00177***	−0.00176***	−0.00174***	−0.00174***	−0.00203***	−0.000146
		(0.000484)	(0.000546)	(0.000549)	(0.000549)	(0.000547)	(0.000739)	(0.000271)
Rank2: 5,000 < rank ≤ 10,000			−0.200***	−0.197***	−0.198***	−0.196***		0.00140
			(0.0506)	(0.0505)	(0.0506)	(0.0506)		(0.00423)
Rank3: 10,000 < rank ≤ 15,000			−0.219***	−0.213***	−0.214***	−0.213***		0.00286
			(0.0486)	(0.0484)	(0.0484)	(0.0485)		(0.00417)
Rank4: 15,000 < rank ≤ 25,000			−0.101*	−0.0983*	−0.0969*	−0.0961*		0.00228
			(0.0555)	(0.0554)	(0.0553)	(0.0554)		(0.00383)
Rank5: rank > 25,000			−0.276***	−0.277***	−0.276***	−0.275***		0.0121***
			(0.0417)	(0.0416)	(0.0417)	(0.0417)		(0.00303)
Pop1: 100 < pop ≤ 500, in 1,000				−0.000357	0.00482	0.00498		−0.000831
				(0.0100)	(0.0107)	(0.0109)		(0.00224)
Pop2: 500 < pop ≤ 1,000, in 1,000				0.0162	0.0218*	0.0202*		−0.000723
				(0.0106)	(0.0115)	(0.0113)		(0.00228)
Pop3: 1,000 < pop ≤ 5,000, in 1,000				0.0131	0.00952	0.00755		−0.000576
				(0.00858)	(0.00849)	(0.00876)		(0.00212)
Pop4: pop > 5,000, in 1,000				0.0123	−0.00205	0.000996		−0.000473
				(0.00825)	(0.0106)	(0.00951)		(0.00217)
Distance from largest domestic city, in 100 km					−0.00247**			
					(0.00117)			
Distance from closest dom. city with > 1 mill. inhab., in 100 km						−0.00252**		1.78e-05
						(0.00115)		(0.000188)
Constant	0.127***	0.117***	0.377***	0.370***	0.382***	0.373***	0.0567***	−0.0129**
	(0.00891)	(0.00878)	(0.0423)	(0.0422)	(0.0428)	(0.0423)	(0.00585)	(0.00534)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES	YES	YES
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES	NO
# quarters with publications FE	NO	NO	NO	NO	NO	NO	NO	YES
Observations	249,000	249,000	249,000	249,000	249,000	249,000	249,000	249,000
R-squared, overall	0.0558	0.0583	0.0928	0.0934	0.0939	0.0937	0.005	0.396
Number of Inst_Discipline	4,980	4,980	4,980	4,980	4,980	4,980	4,980	4,980
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490	2,490

Dependent variable *y* is a binary variable equal 1 if an institution made any article contribution in a given quarter and discipline. We use a balanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on any publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru). We explore a linear probability model. We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is $rank \leq 5000$. Reference population is $pop \leq 100$. Random institution-discipline effects are included in columns (1) to (6). Robust standard errors clustered at the institutional level. Note that serial correlation is not an issue in our balanced panel because the large number of periods with 0 publications breaks any time correlation for any given institution.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

publication output.

5.2. The effect of OARE using IV estimation

In this section, we explicitly deal with endogeneity of unobserved factors in the productivity equation. We first explain the probability to join the OARE initiative and then use the predicted value of the treatment effect in the main productivity equation. In the first stage of the IV procedure, we explain the probability that an institution joins the OARE initiative using quarter, country and institutions fixed effects as well as the three instruments *m_OARE, m_HINARI* and *m_AGORA* using a linear probability model estimated by ordinary least squares regression. In the second stage of the IV procedure, we explain the number of publications by the predicted treatment effect obtained from the first stage in addition to the other control variables used before.

Column (1) in Table 4 gives the estimation results of the IV procedure using institution-discipline fixed effects in addition to quarter and quarter-discipline fixed effects. The IV estimate (2.5) is slightly higher

Table 4
OARE Effect using IV.

Model:	(1) IV	(2) LPM IV
First Stage		
Dependent variable:	*OARE treated*	*OARE treated*
m_OARE	0.612***	0.612***
	(0.0106)	(0.0106)
m_AGORA	0.0145	0.0145
	(0.0127)	(0.0127)
m_HINARI	−0.0162	−0.0162
	(0.0124)	(0.0124)
Constant	0.000606	0.000606
	(0.000980)	(0.000980)
Observations	249,000	249,000
R-squared	0.042	0.042
Number of Inst_Discipline	4,980	4,980
Quarter FE	YES	YES
Quarter-discipline FE	YES	YES
Inst-discipline FE	YES	YES
Second Stage		
Dependent variable:	*w*	*y*
OARE treated (DDD)	2.455***	0.743***
	(0.541)	(0.139)
Constant	0.126***	0.0602***
	(0.00802)	(0.00324)
Observations	249,000	249,000
R-squared	0.016	0.018
Number of Inst_Discipline	4,980	4,980
Quarter FE	YES	YES
Quarter-discipline FE	YES	YES
Inst-discipline FE	YES	YES

We use a balanced panel and take into account journal articles by both single and multiple local authors. Regressions based on Specification (7) of Table 2. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is rank ≤ 5000. Reference population is pop ≤ 100. Robust standard errors clustered at the institutional level. Included but not reported: # pages, # co-authors USA, # co-authors EUR. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

than the OLS estimate (2.1). However, the standard error is quite large. Column (2) reports the estimation results of the linear probability model where the dependent variable is *y*. Coefficients can be directly interpreted as percentage effects. Compared to the linear probability model effect estimated by OLS (21%), the IV procedure yields a higher treatment effect of 74%. Again, the standard error is quite high.

5.3. The effect of OARE using Bayesian MCMC estimation

In this section, we explicitly model correlation between the unobserved variable in the self-selection equation and the unobserved variable in the productivity equation by using a Bayesian MCMC estimation method. Here, we simultaneously estimate parameters of a system of two equations using Bayesian inference: parameters of the self-selection equation are jointly estimated with parameters of the productivity equation assuming a normal bivariate distribution of the unobserved variables of the two equations. Inference from the posterior distribution of the parameters is carried out using data augmentation and an MCMC algorithm. Contrary to the IV approach, the Bayesian approach is an exact likelihood procedure.[32]

Results from Table 5 were obtained from a random sample of 300 institutions. Columns (1) and (2) report results from OLS and MCMC regressions with *w* as dependent variable, respectively. Columns (3) and (4) report results from a linear probability model with *y* as dependent variable. The sample was selected to match the treatment effect estimated on the full sample by OLS as reported in Tables 2 and 3. More specifically, the random sample was selected so that the treatment effect in column (1) of Table 5 (2.06 in the second stage equation using OLS) was matched with the estimated treatment effect of specification (7) of Table 2 (2.07). In addition, the random sample also matched the treatment effect using *y* as a dependent variable in column (3) of Table 5 (0.21 in the second stage of the linear probability model) with the estimated treatment effect of column (7) of Table 3 (0.205). In all specifications reported in Table 5, we include a set of binary variables corresponding to the number of quarters during which an institution published at least one article as institution fixed effects, in addition to country, rank, population and distances variables.[33] Across columns, the Bayesian MCMC estimate of the treatment effect is slightly above the OLS estimate with a small standard error, i.e., 2.099 > 2.06 and 0.224 > 0.210. The correlation between unobservable variables of the selection and observation equations is slightly negative and significant, respectively −0.056 with *w* as the dependent variable and −0.075 with *y* as the dependent variable. Unobserved variables in the self-selection equation include the hidden (administrative and informational) costs of joining the initiative, while unobserved variables in the main equation include hidden productivity factors. A negative correlation between the unobserved variables corresponds to a negative correlation between the hidden costs of joining the OARE initiative and the unobserved productivity variables at the institution level.

5.4. Extensions

5.4.1. OARE effect by number of quarters with publications

We also ran the regressions separately for institutions with three different levels of productivity, i.e., institutions that published in at least *x* quarters in at least one of the two disciplines with $x < 2$, $2 \leq x \leq 30$, and $x > 30$, respectively. Results are reported in Table 6. Specification (1) reports results for institutions with publications in less than 2 quarters. Specification (2) reports results for institutions with publications in between 2 and 30 quarters. Finally, specification (3) reports results for the most productive institutions that publish in more than 30 quarters of our sample (with a total of 50 quarters).

The marginal OARE effect ranges from +0.4% for institutions with a low level of productivity as reported in column (1) to +43.4% for institutions with a high level of productivity as reported in column (3). It is statistically significant at the 1% level for institutions with intermediate and high levels of productivity as reported in columns (2) and (3). It is statistically significant at the 5% level for low-productivity institutions as reported in column (1). Notably, these findings suggest that productive institutions benefit more from OARE registration and that the least productive institutions barely benefit from it.

5.4.2. The effect of OARE on publication input

In addition to the impact of the OARE initiative on publication output, we investigate the effect of OARE on publication input. Thereby, we address the cumulative character of the knowledge production process and revert to the fact that researchers need access to academic journals in order to be able to create their own research. Our

[32] We use a non-informative prior distribution of the parameters. The main drawback of the MCMC procedure is that it is extremely computer intensive. The MCMC could not run on the full set of 2,490 institutions with all fixed effects.

[33] Using the number of quarters as additional fixed effects allows us to run the program on a relatively large sample of institutions. The MCMC algorithm was run for 1,000 iterations after a warmup period of 100 iterations starting at the OLS estimate and an initial value of ρ of −0.15. Note that the coefficients in the first stage of Table 5 are different from those in Table 4. Indeed, the first stage of the IV procedure is a linear probability model, while the first equation of the MCMC procedure is a standard probit model.

Table 5
OARE Effect using MCMC.

Model:	(1) OLS	(2) MCMC	(3) OLS LPM	(4) MCMC LPM
First Stage				
Dependent variable:	*OARE treated*	*OARE treated*	*OARE treated*	*OARE treated*
Average # of institutions that joined OARE (t-1)	0.810*** (0.031)	6.371*** (1.054)	0.810*** (0.031)	6.8221*** (1.1009)
Average # of institutions that joined AGORA (t-1)	0.256*** (0.041)	4.011* (2.060)	0.256*** (0.041)	3.0013 (1.9066)
Average # of institutions that joined HINARI (t-1)	−0.3007*** (0.042)	−6.43*** (2.252)	−0.3007*** (0.042)	−6.247*** (1.8116)
Constant		−3.909*** (0.623)		−3.664*** (0.5529)
Second Stage				
Dependent variable:	*w*	*w*	*y*	*y*
OARE treated (DDD)	2.06*** (0.051)	2.099*** (0.054)	0.2103*** (0.0124)	0.2237*** (0.0131)
# pages	−0.002*** (0.0007)	−0.002*** (0.0007)	0.0009*** (0.0001)	0.0009*** (0.0001)
# co-authors USA	0.0077* (0.0044)	0.0080* (0.0041)	0.0017 (0.0010)	0.0016 (0.0011)
# co-authors EUR	−0.0008 (0.0032)	−0.0008 (0.0032)	−0.0004 (0.0007)	−0.0004 (0.0007)
Rank2: 5,000 < rank ≤ 10,000	−0.775*** (0.0528)	−0.777*** (0.0511)	0.0720*** (0.0127)	0.0720*** (0.0128)
Rank3: 10,000 < rank ≤ 15,000	−0.736*** (0.0466)	−0.736*** (0.0447)	0.0049 (0.0113)	0.0055 (0.0117)
Rank4: 15,000 < rank ≤ 25,000	−0.726*** (0.0447)	−0.727*** (0.0445)	0.0217** (0.0108)	0.0222** (0.0111)
Rank5: rank > 25,000	−0.718*** (0.0347)	−0.717*** (0.0336)	0.0138 (0.0085)	0.0150* (0.0088)
Pop1: 100 < pop ≤ 500, in 1,000	−0.022 (0.0233)	−0.022 (0.0218)	−0.005 (0.0056)	−0.005 (0.0057)
Pop2: 500 < pop ≤ 1,000, in 1,000	0.0124 (0.0244)	0.0123 (0.0228)	−0.012** (0.0059)	−0.012** (0.0058)
Pop3: 1,000 < pop ≤ 5,000, in 1,000	0.0685*** (0.0201)	0.0678*** (0.0193)	−0.004 (0.0049)	−0.004 (0.0049)
Pop4: pop > 5,000, in 1,000	−0.0217 (0.0222)	−0.022 (0.0213)	−0.007 (0.0053)	−0.007 (0.0053)
Distance from closest dom. city with > 1 mill. inhab., in 100 km	0.0033* (0.0020)	0.0032* (0.0019)	−0.0002 (0.0004)	−0.0002 (0.0005)
Constant	0.127*** (0.0158)	0.675*** (0.050)	0.0208 (0.0127)	0.0196 (0.0128)
ρ		−0.056* (0.0317)		−0.075** (0.0340)
Quarter dummies	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES
Quarter-discipline FE	YES	YES	YES	YES
# quarters with publications FE	YES	YES	YES	YES
Observations	30,000	30,000	30,000	30,000
Number of Inst_Discipline	600	600	600	600
Number of Inst	300	300	300	300

We use a balanced panel and take into account journal articles by both single and multiple local authors. Results obtained from a random sample of 300 institutions. The institution-discipline-quarter triplets constitute the unit of observation. Columns (1) and (2) report results from OLS and MCMC regressions with *w* as dependent variable, respectively. Columns (3) and (4) report results from a linear probability model with *y* as dependent variable. First equation of the MCMC procedure is a standard Probit model. Robust standard errors clustered at the institutional level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

a priori belief based on the results originally formulated by Davis (2011) is that free or reduced fee online access to academic journals in environmental sciences should have an effect on access and hence the number of references used to create one's own research. In this regard, we look at both the overall number of references as well the number of OARE references, i.e., references to articles published in OARE journals.

The results on the impact of OARE on the number of references are reported in Table 7. We find a positive effect of membership to OARE on the number of references across all specifications. It is statistically significant at the 1% level across all specifications. Interestingly, next to OARE membership international collaborations with researchers from the US (and to a smaller degree also from Europe) have a positive and statistically significant effect on the number of references. This result suggests that collaborations with researchers from the US (and Europe) have a positive effect on developing-country researchers in terms of access to scientific journals.

The results on the impact of OARE on the number of OARE

Table 6
OARE Effect by the Number of Quarters with Publications.

# quarters with publication:	(1) < 2 quarters	(2) 2 ≤ quarters ≤30	(3) > 30 quarters
Model:	+Inst.-Disc. FE	+Inst.-Disc. FE	+Inst.-Disc. FE
Dependent variable:	*w*	*w*	*w*
OARE treated (DDD)	0.0246**	0.445***	3.036***
	(0.0118)	(0.110)	(0.969)
# pages	[omitted]	0.00116	−0.0103
		(0.00176)	(0.00885)
# co-authors USA	[omitted]	−0.00795	0.0258
		(0.00554)	(0.0226)
# co-authors EUR	[omitted]	0.00780***	0.0939**
		(0.00302)	(0.0430)
Constant	0.00473	0.0643***	3.182***
	(0.00380)	(0.0177)	(0.393)
Quarter dummies	YES	YES	YES
Country dummies	NO	NO	NO
Quarter-discipline dummies	YES	YES	YES
Institution-discipline FE	YES	YES	YES
Observations	130,100	109,800	9,100
R-squared, overall	0.0009	0.0451	0.1775
Number of Inst_Discipline	2,602	2,196	182
Number of Inst	1,301	1,098	91

We use a balanced panel and take into account journal articles by both single and multiple local authors. OLS results on the impact of OARE membership (*treated*) on publication output of research institutions with three different levels of productivity, i.e., institutions that published in at least x quarters in at least one of the two disciplines whereas $x < 2$, $2 \le x \le 30$, and $x > 30$, respectively. We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is $rank \le 5000$. Reference population is $pop \le 100$. Robust standard errors clustered at the institutional level. Note that serial correlation is not an issue in our balanced panel because the large number of periods with 0 publications breaks any time correlation for any given institution. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

references are reported in Table 8. We find a positive effect of membership to OARE on the number of OARE references across all specifications. It is statistically significant at the 1% level across all specifications.

5.5. Robustness

5.5.1. Pre- and post-OARE trends

There may be concerns that the post-OARE effect that we obtain is confounded with a pre-OARE trend which may undermine the interpretation of the OARE effect as a treatment effect.[34] To mitigate these concerns we implement a specification similar to specification (7) of Table 2 following the approach taken by Furman and Stern (2011, Fig. 2). Therein, we replace the OARE treated dummy with the dummy variables for each quarter preceding and following the quarter when an institution registered with OARE. From this new specification, we obtain Fig. 3.

Fig. 3 plots quarter-by-quarter pre-treatment and post-treatments effects on publication output. The data represents each of the estimated marginal pre- and post-treatment quarter effects on publication output. All marginal effects are computed relative to the treatment quarter plus/minus one quarter. Outside lines bound the 95% confidence interval based on robust standard errors clustered at the institution level. Fig. 3 does not suggest that publication output follows a clear upward trend in the 28 quarters before the OARE treatment.

5.5.2. Articles with only one local author

In our basic model, we use a balanced panel including both single- and multiple-local-authored journal articles to examine the OARE effect. As discussed in Section 3.3, dividing authorship shares across different institutions is not a trivial exercise. For robustness, we therefore run the same regressions as reported in Table 2 for the subsample of single-local-author articles, i.e. articles for which we observe only one local author who may or may not be affiliated with an OARE member institution. As before, we use a balanced panel. We create the single-local-authored dataset by dropping 18,955 articles from the sample for which we have at least two local authors. Results are reported in Table 9.

The marginal OARE effect is positive and statistically significant at the 1% level across all specifications. It ranges from +23.5% in column (7) to +24.3% in column (1).[35] These results provide additional empirical evidence for a robust OARE effect.

5.5.3. Effect of outliers

We address possible concerns that OLS results without log-transformed dependent variable are potentially sensitive to extreme outliers. Following Williams (2016), we delete observations that have at least one of the following characteristics: (a) standardized residuals' values greater than 3, (b) leverage greater than $2k/n$ where k is the number of independent variables in the regression and n is the number of observations, (c) Cook's Distance measure values greater than $4/n$. In total, 4,709 observations are deleted. 1,355 observations are deleted under (a), 13 under (b) and 3,341 under (c). Results are reported in Table 10.

The OLS OARE coefficient is statistically significant at the 1% level across all columns ranging from 1.48 in column (7) to 1.53 in column (2). These results provide additional evidence for a robust OARE effect.[36]

5.5.4. Unbalanced panel

In order to address possible concerns that there are many observations with $w = 0$ in the balanced-panel analysis, we use an unbalanced panel and take into account journal articles by both single and multiple local authors. Results are reported in Table 11.

The OLS OARE coefficient is statistically significant at the 1% level across all columns ranging from 2.94 in column (7) to 4.95 in column (1). These results provide additional evidence for a robust OARE effect.

6. Conclusion

We have analyzed the effect of free and reduced-fee online access to the environmental science literature via the OARE initiative on scientific productivity in Bolivia, Ecuador, Kenya, Nigeria and Peru. We provide empirical support for a positive marginal OARE effect of 29.6%. Given an average publication output by institution, discipline and quarter of 0.140 (see Table 1), this result suggests that, due to the OARE program, 2.07 extra articles were produced for an average institution publishing a total of 7.0 articles over the whole observation period. Looking at a 95% confidence interval this suggests that the OARE program adds between 1.3 and 2.8 extra articles for an average institution over the whole period of 50 quarters. The marginal OARE

[34] We thank an anonymous referee for this comment.

[35] From Appendix 1 we know that the mean number of publications for the single local author sample is 0.112. Hence, an average institution publishes 5.6 single-local authored articles over the observation period. Based on this, we obtain 1.315/5.6 = 0.2348 for column (7) of Table 9 and 1.362/5.6 = 0.2432 for column (1) of Table 9, respectively.

[36] Note that among the deleted observations are those institution/quarter/discipline pairs with $w > 80$ publications in the OARE field. Recall from Table 1 that the mean of w is 0.14. Hence, the exclusion of extreme outliers in terms of publication output in the OARE field explains that the OARE effect is lower than in the case where outliers are not excluded.

Table 7
Effect of OARE on Publication Input (References).

Model:	(1) Base	(2) + Article info	(3) +Rank	(4) +Population	(5) +Distance_1	(6) +Distance_2	(7) +Inst.-Disc. FE
Dependent variable:	*#References*	*#References*	*#References*	*#References*	*#References*	*#References*	*#References*
OARE treated (DDD)	2.017***	1.716***	1.711***	1.711***	1.711***	1.711***	1.722***
	(0.335)	(0.295)	(0.295)	(0.295)	(0.295)	(0.295)	(0.295)
# pages		1.774***	1.773***	1.774***	1.773***	1.773***	1.754***
		(0.0775)	(0.0775)	(0.0776)	(0.0776)	(0.0776)	(0.0791)
# co-authors USA		0.672***	0.672***	0.672***	0.672***	0.672***	0.660***
		(0.231)	(0.232)	(0.232)	(0.232)	(0.232)	(0.233)
# co-authors EUR		0.378*	0.378*	0.378*	0.378*	0.378*	0.372*
		(0.195)	(0.195)	(0.195)	(0.195)	(0.195)	(0.194)
Rank2: 5,000 < rank ≤ 10,000			−1.249	−1.296	−1.269	−1.416	
			(1.672)	(1.682)	(1.683)	(1.672)	
Rank3: 10,000 < rank ≤ 15,000			0.379	0.303	0.366	0.306	
			(2.085)	(2.079)	(2.084)	(2.087)	
Rank4: 15,000 < rank ≤ 25,000			0.380	0.436	0.370	0.216	
			(1.523)	(1.525)	(1.516)	(1.498)	
Rank5: rank > 25,000			−3.539***	−3.427***	−3.443***	−3.540***	
			(1.138)	(1.153)	(1.159)	(1.157)	
Pop1: 100 < pop ≤ 500, in 1,000				0.121	−0.133	−0.407	
				(0.720)	(0.750)	(0.767)	
Pop2: 500 < pop ≤ 1,000, in 1,000				0.266	−0.0109	−0.137	
				(0.683)	(0.721)	(0.706)	
Pop3: 1,000 < pop ≤ 5,000, in 1,000				−0.610	−0.435	−0.0620	
				(0.545)	(0.550)	(0.589)	
Pop4: pop > 5,000, in 1,000				−0.406	0.299	0.718	
				(0.732)	(0.885)	(0.870)	
Distance from largest domestic city, in 100 km					0.121		
					(0.0896)		
Distance from closest domestic city with > 1 mill. inhab., in 100 km						0.250**	
						(0.117)	
Constant	27.63***	11.63***	14.67***	14.68***	14.08***	14.41***	8.933***
	(0.462)	(0.741)	(1.430)	(1.456)	(1.511)	(1.457)	(0.503)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES	YES
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES
Observations	249,000	249,000	249,000	249,000	249,000	249,000	249,000
R-squared, overall	0.2884	0.6325	0.6339	0.6342	0.6344	0.6346	0.6283
Number of Inst_Discipline	4,980	4,980	4,980	4,980	4,980	4,980	4,980
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490

We use a balanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on the mean number of references (i.e., by institution, quarter and discipline) of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD estimation methods. We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is *rank* ≤ 5000. Reference population is *pop* ≤ 100. Random institution-discipline effects are included in columns (1) to (6). Robust standard errors clustered at the institutional level. **p* < 0.1, ***p* < 0.05, ****p* < 0.01.

effect is also positive and statistically significant when we run the regressions separately for Group A (free access) and Group B (reduced-fee access) countries, revealing at least 2.45 extra articles for institutions in group A countries and 1.21 extra articles for institutions in group B countries, respectively.[37] In addition, a robustness check analyzing a balanced panel with single local authors yields qualitatively similar results, i.e., we provide evidence for a statistically significant OARE effect of +23.5% with confidence interval (15.1%, 31.9%) using the most conservative specification. Overall, our results provide empirical support for the hypothesis that free online access to journals promotes research in developing countries. Moreover, we analyze the impact of OARE on the ability of researchers in the developing world to access academic knowledge. In line with Davis (2011)'s results we find a strong impact of the OARE initiative on both the overall number of references in articles to which an average institution contributed, i.e., a +8.4% increase of the number of references due to OARE, as well as the number of OARE references, i.e., a +14.5% increase in the number of OARE references due to OARE. However, our result of a significant OARE effect on publication output differs from Davis (2011) who finds no TEEAL effect on publication output. Arguably, the larger number of journals covered under the OARE initiative as compared to TEEAL (11,500 vs. 200 journals), OARE's lower subscription cost (ranging from free online access to a maximum of $ 1,000 per year vs. $ 5,000 for TEEAL), and the different means of access (online access for OARE and offline (CD-ROM) for TEEAL) provide an explanation why we find a substantial OARE effect on publication output while Davis (2011) finds no TEEAL effect on publication output.

Nevertheless, we find that there is potential for improvement on two grounds. First, we provide evidence that productive institutions benefit more from OARE and that the least productive institutions barely benefit from it. This result suggests that OARE increases the productivity difference between the most and least productive institutions. Under these conditions, the least productive institutions are ceteris paribus less likely to

[37] We estimate these numbers as follows. The average number of articles is 0.175 for group A and 0.0768 for group B countries. The OARE program increases publication output by 28% in group A and by 32% in group B countries. This adds up to 2.45 (1.21) extra articles for an average institution in group A (B) countries.

Table 8
Effect of OARE on Publication Input (OARE References).

Model: Dependent variable:	(1) Base *#OARE_References*	(2) + Article info *#OARE_References*	(3) +Rank *#OARE_References*	(4) +Population *#OARE_References*	(5) +Distance_1 *#OARE_References*	(6) +Distance_2 *#OARE_References*	(7) +Inst.-Disc. FE *#OARE_References*
OARE treated (DDD)	0.736*** (0.162)	0.622*** (0.154)	0.621*** (0.154)	0.621*** (0.154)	0.621*** (0.154)	0.621*** (0.154)	0.622*** (0.154)
# pages		0.425*** (0.0344)	0.425*** (0.0344)	0.425*** (0.0344)	0.425*** (0.0344)	0.425*** (0.0344)	0.423*** (0.0353)
# co-authors USA		0.297*** (0.110)	0.297*** (0.110)	0.298*** (0.110)	0.298*** (0.110)	0.298*** (0.110)	0.295*** (0.111)
# co-authors EUR		0.240** (0.116)	0.240** (0.116)	0.240** (0.116)	0.240** (0.116)	0.240** (0.116)	0.238** (0.117)
Rank2: 5,000 < rank ≤ 10,000			1.216 (0.865)	0.771 (0.870)	0.798 (0.876)	0.680 (0.871)	
Rank3: 10,000 < rank ≤ 15,000			0.600 (0.706)	0.0687 (0.693)	0.130 (0.700)	0.0710 (0.699)	
Rank4: 15,000 < rank ≤ 25,000			0.793 (0.767)	0.522 (0.749)	0.457 (0.738)	0.357 (0.728)	
Rank5: rank > 25,000			−0.228 (0.565)	−0.262 (0.566)	−0.278 (0.572)	−0.347 (0.572)	
Pop1: 100 < pop ≤ 500, in 1,000				−0.763** (0.369)	−1.011** (0.394)	−1.159*** (0.405)	
Pop2: 500 < pop ≤ 1,000, in 1,000				−1.083*** (0.372)	−1.354*** (0.389)	−1.386*** (0.380)	
Pop3: 1,000 < pop ≤ 5,000, in 1,000				−1.090*** (0.298)	−0.919*** (0.297)	−0.678** (0.312)	
Pop4: pop > 5,000, in 1,000				−2.271*** (0.366)	−1.581*** (0.436)	−1.427*** (0.419)	
Distance from largest domestic city, in 100 km					0.119** (0.0510)		
Distance from closest domestic city with > 1 mill. inhab., in 100 km						0.188*** (0.0682)	
Constant	4.873*** (0.166)	0.892*** (0.333)	0.995 (0.654)	1.880*** (0.670)	1.284* (0.710)	1.672** (0.676)	1.291*** (0.250)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES	YES
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES
Observations	249,000	249,000	249,000	249,000	249,000	249,000	249,000
R-squared, overall	0.1342	0.2907	0.2918	0.2995	0.3008	0.3017	0.2746
Number of Inst_Discipline	4,980	4,980	4,980	4,980	4,980	4,980	4,980
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490

We use a balanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on the mean number of OARE references (i.e., by institution, quarter and discipline) of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD estimation methods. We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is $rank \leq 5000$. Reference population is $pop \leq 100$. Random institution-discipline effects are included in columns (1) to (6). Robust standard errors clustered at the institutional level. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

catch up. Second, we find that not more than 5% of all eligible institutions joined OARE after a period of more than 5 years. This finding reveals the unused potential of the OARE initiative. Based on our results, policies aimed at increasing the awareness of free online access initiatives in developing countries should therefore be encouraged.

As a broader policy implication, our study suggests that an open access mandate or policy may promote scientific output – not only by research institutions in developing countries. Extending on the link between academic research and economic growth (see the literature discussed in Section 2) our findings may hence point to direct economic effects as a higher research output level stemming from OARE membership may result in new environmental innovations. A natural follow up is to explore the question of whether OARE has increased the number of patent applications using free or reduced-fee access throughout the OARE program. In addition, it would be interesting to investigate in more detail how (open) online access has changed the way scientists do research and collaborate internationally.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

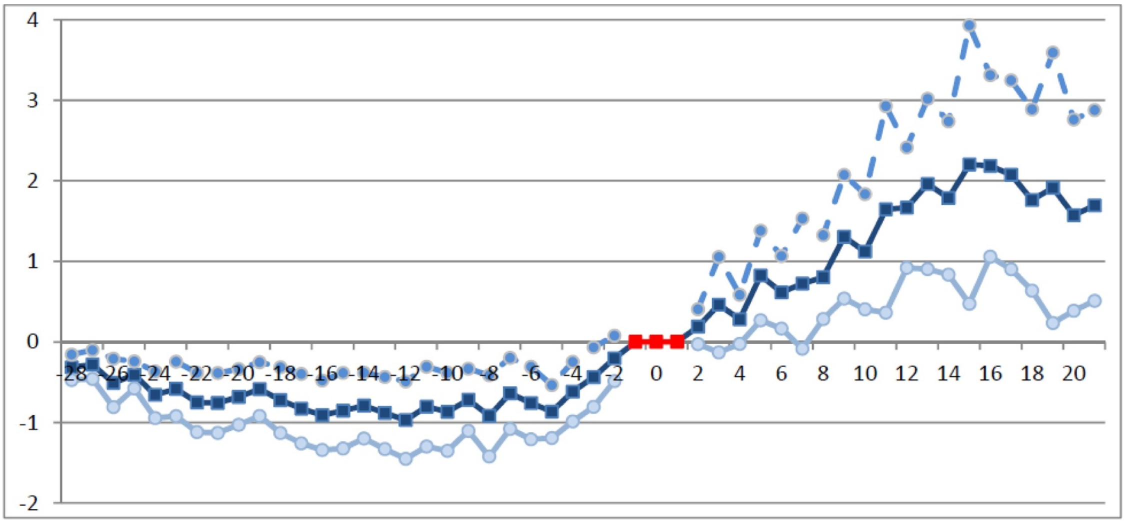


Fig. 3. PRE- AND POST-TREATMENT EFFECTS ON PUBLICATION OUTPUT *Notes:* Figure plots quarter-by-quarter pre-treatment and post-treatment effects on publication output, computed from OLS triple difference regressions with dummy variables for each quarter preceding and following OARE treatment (along with quarter fixed effects, institution-discipline fixed effects and article characteristics). The regressions reported in column (7) of Table 2 serve as the basis for the regressions from which create the figure. In contrast to the regressions reported in column (7) of Table 2, however, we replace the OARE treated dummy with the dummy variables for each quarter preceding and following treatment. The data represent each of the estimated marginal pre- and post-treatment quarter effects on publication output. All marginal effects are computed relative to the treatment quarter plus/minus one quarter. Outside lines bound the 95% confidence interval based on robust standard errors clustered at the institution level. Fig. 3 does not suggest that publication output follows a clear upward trend in the 28 quarters before the OARE treatment.

Table 9
OARE Effect (SINGLE Local Authors ONLY).

VARIABLES	(1) OLS Base	(2) + Article info	(3) +Rank	(4) +Population	(5) +Dist._1	(6) +Dist._2	(7) +Inst.-Disc. FE
OARE treated (DDD)	1.362***	1.347***	1.339***	1.339***	1.339***	1.339***	1.315***
	(0.249)	(0.245)	(0.244)	(0.244)	(0.244)	(0.244)	(0.240)
# pages		0.00679**	0.00598*	0.00591*	0.00594*	0.00594*	0.00995**
		(0.00317)	(0.00312)	(0.00313)	(0.00313)	(0.00313)	(0.00441)
# co-authors USA		−0.00484	−0.00501	−0.00504	−0.00507	−0.00511	−0.0123
		(0.0123)	(0.0122)	(0.0122)	(0.0122)	(0.0122)	(0.0163)
# co-authors EUR		0.0562	0.0558	0.0557	0.0558	0.0558	0.0681
		(0.0356)	(0.0356)	(0.0356)	(0.0356)	(0.0356)	(0.0414)
Rank2: 5,000 < rank ≤ 10,000			−0.638**	−0.641**	−0.642**	−0.637**	
			(0.259)	(0.259)	(0.260)	(0.260)	
Rank3: 10,000 < rank ≤ 15,000			−0.641***	−0.629***	−0.637***	−0.630***	
			(0.243)	(0.242)	(0.241)	(0.242)	
Rank4: 15,000 < rank ≤ 25,000			−0.0311	−0.0331	−0.0291	−0.0263	
			(0.373)	(0.369)	(0.370)	(0.370)	
Rank5: rank > 25,000			−0.742***	−0.754***	−0.752***	−0.751***	
			(0.219)	(0.221)	(0.221)	(0.221)	
Pop1: 100 < pop ≤ 500, in 1,000				−0.0483	−0.0259	−0.0333	
				(0.0396)	(0.0413)	(0.0425)	
Pop2: 500 < pop ≤ 1,000, in 1,000				0.0313	0.0576	0.0430	
				(0.0496)	(0.0543)	(0.0525)	
Pop3: 1,000 < pop ≤ 5,000, in 1,000				0.0319	0.0144	0.0139	
				(0.0466)	(0.0439)	(0.0468)	
Pop4: pop > 5,000, in 1,000				−0.0124	−0.0800*	−0.0486	
				(0.0300)	(0.0458)	(0.0334)	
Distance from largest domestic city, in 100 km					−0.0115**		
					(0.00575)		
Distance from closest domestic city with > 1 mill. inhab., in 100 km						−0.00773*	
						(0.00460)	
Constant	0.259***	0.166***	0.840***	0.842***	0.899***	0.851***	0.0453
	(0.0486)	(0.0538)	(0.213)	(0.215)	(0.216)	(0.214)	(0.0336)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES	YES

(*continued on next page*)

Table 9 (*continued*)

VARIABLES	(1) OLS Base	(2) + Article info	(3) +Rank	(4) +Population	(5) +Dist._1	(6) +Dist._2	(7) +Inst.-Disc. FE
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES
Observations	203,100	203,100	203,100	203,100	203,100	203,100	203,100
R-squared, overall	0.0917	0.0736	0.1044	0.1051	0.1060	0.1054	0.0571
Number of Inst_Discipline	4,062	4,062	4,062	4,062	4,062	4,062	4,062
Number of Inst	2,031	2,031	2,031	2,031	2,031	2,031	2,031

We use a balanced panel and take into account journal articles by single local authors only. Results on the impact of OARE membership (*treated*) on publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD estimation. We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is $rank \leq 5000$. Reference population is $pop \leq 100$. Random institution-discipline effects are included in columns (1) to (6). Robust standard errors clustered at the institutional level reported in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 10
Effect of OARE on Publication Output (excluding Outliers).

Model: Dependent variable:	(1) Base *w*	(2) + Article info *w*	(3) +Rank *w*	(4) +Population *w*	(5) +Dist._1 *w*	(6) +Dist._2 *w*	(7) +Inst.-Disc. FE *w*
OARE treated (DDD)	1.525***	1.529***	1.521***	1.521***	1.521***	1.521***	1.481***
	(0.0895)	(0.0892)	(0.0883)	(0.0883)	(0.0883)	(0.0883)	(0.0853)
# pages		0.00211**	0.00165	0.00160	0.00161	0.00161	0.00263**
		(0.00102)	(0.00103)	(0.00104)	(0.00104)	(0.00104)	(0.00117)
# co-authors USA		0.00565	0.00535	0.00535	0.00537	0.00534	0.00318
		(0.00490)	(0.00466)	(0.00465)	(0.00466)	(0.00466)	(0.00463)
# co-authors EUR		0.0163***	0.0160***	0.0159***	0.0160***	0.0159***	0.0173***
		(0.00396)	(0.00373)	(0.00373)	(0.00373)	(0.00373)	(0.00429)
Rank2: 5,000 < rank ≤ 10,000			−0.769*	−0.767*	−0.771*	−0.763*	
			(0.394)	(0.394)	(0.395)	(0.395)	
Rank3: 10,000 < rank ≤ 15,000			−0.894**	−0.876**	−0.884**	−0.876**	
			(0.348)	(0.345)	(0.345)	(0.345)	
Rank4: 15,000 < rank ≤ 25,000			−0.0633	−0.0588	−0.0501	−0.0514	
			(0.513)	(0.509)	(0.511)	(0.510)	
Rank5: rank > 25,000			−1.029***	−1.042***	−1.040***	−1.038***	
			(0.334)	(0.336)	(0.337)	(0.337)	
Pop1: 100 < pop ≤ 500, in 1,000				−0.0527	−0.0197	−0.0350	
				(0.0466)	(0.0489)	(0.0511)	
Pop2: 500 < pop ≤ 1,000, in 1,000				0.0565	0.0925	0.0700	
				(0.0631)	(0.0691)	(0.0681)	
Pop3: 1,000 < pop ≤ 5,000, in 1,000				0.0577	0.0350	0.0393	
				(0.0598)	(0.0554)	(0.0602)	
Pop4: pop > 5,000, in 1,000				0.00428	−0.0874	−0.0334	
				(0.0341)	(0.0596)	(0.0384)	
Distance from largest domestic city, in 100 km					−0.0157**		
					(0.00794)		
Distance from closest dom. city with > 1 mill. inhab., in 100 km						−0.00838	
						(0.00593)	
Constant	0.308***	0.278***	1.214***	1.205***	1.284***	1.214***	0.00205
	(0.0577)	(0.0582)	(0.337)	(0.338)	(0.342)	(0.338)	(0.00929)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES	YES
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES
Observations	244,291	244,291	244,291	244,291	244,291	244,291	244,291
R-squared, overall	0.0662	0.0641	0.0738	0.0734	0.0734	0.0736	0.0869
Number of Inst_Discipline	4,980	4,980	4,980	4,980	4,980	4,980	4,980
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490

We use a balanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD. Following Williams (2016), we delete observations that have at least one of the following characteristics: (a) value of standardized residuals > 3, (b) leverage > $2k/n$ where k is the number of independent variables in the regression and n is the number of observations, (c) Cook's Distance measure value > $4/n$. In total, 4,709 observations are deleted. 1,355 observations are deleted under (a), 13 under (b) and 3,341 under (c). We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is $rank \leq 5000$. Reference population is $pop \leq 100$. Random institution-discipline effects are included in columns (1) to (6). Robust standard errors clustered at the institutional level. Note that serial correlation is not an issue in our balanced panel because the large number of periods with 0 publications breaks any time correlation for any given institution. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 11
Effect of OARE on Publication Output (Unbalanced Panel).

Model: Dependent variable:	(1) Base *w*	(2) + Article info *w*	(3) +Rank *w*	(4) +Pop. *w*	(5) +Dist._1 *w*	(6) +Dist._2 *w*	(7) +Inst.-Disc. FE *w*
OARE treated (DDD)	4.951***	4.916***	3.723***	3.663***	3.654***	3.661***	2.944***
	(1.152)	(1.151)	(0.890)	(0.880)	(0.875)	(0.880)	(0.729)
# pages		−0.00344	−0.00225	−0.00367	−0.00306	−0.00321	−0.00159
		(0.00719)	(0.00710)	(0.00651)	(0.00662)	(0.00672)	(0.00352)
# co-authors USA		0.0966**	0.0938***	0.0942***	0.101***	0.0942***	−0.00524
		(0.0457)	(0.0332)	(0.0295)	(0.0302)	(0.0294)	(0.0136)
# co-authors EUR		0.0356	0.0229	0.0160	0.0193	0.0177	0.0308
		(0.0479)	(0.0487)	(0.0486)	(0.0486)	(0.0483)	(0.0310)
Rank2: 5,000 < rank ≤ 10,000			−1.242	−1.309	−1.208	−1.259	
			(1.327)	(1.282)	(1.299)	(1.284)	
Rank3: 10,000 < rank ≤ 15,000			−2.143**	−1.891*	−2.011**	−1.891*	
			(0.964)	(0.980)	(0.927)	(0.971)	
Rank4: 15,000 < rank ≤ 25,000			0.531	0.588	0.561	0.605	
			(1.519)	(1.427)	(1.394)	(1.421)	
Rank5: rank > 25,000			−1.757**	−1.984***	−2.010***	−1.978***	
			(0.724)	(0.719)	(0.716)	(0.719)	
Pop1: 100 < pop ≤ 500, in 1,000				−1.238**	−0.890*	−1.144**	
				(0.527)	(0.514)	(0.576)	
Pop2: 500 < pop ≤ 1,000, in 1,000				0.218	0.469	0.275	
				(0.702)	(0.712)	(0.724)	
Pop3: 1,000 < pop ≤ 5,000, in 1,000				0.190	−0.0835	0.0596	
				(0.648)	(0.595)	(0.677)	
Pop4: pop > 5,000, in 1,000				−0.456	−1.333*	−0.717	
				(0.470)	(0.691)	(0.479)	
Distance from largest domestic city, in 100 km					−0.161**		
					(0.0771)		
Distance from closest domestic city with > 1 million inhab., in 100 km						−0.0590	
						(0.0792)	
Constant	2.256***	2.243***	3.586***	3.847***	4.622***	3.917***	1.297***
	(0.428)	(0.413)	(0.808)	(0.831)	(0.891)	(0.839)	(0.274)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES	YES
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES
Observations	16,131	16,131	16,131	16,131	16,131	16,131	16,131
R-squared, overall	0.1218	0.1251	0.1565	0.1657	0.1724	0.1662	0.0687
Number of Inst_Discipline	3,229	3,229	3,229	3,229	3,229	3,229	3,229
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490

We use an unbalanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD. We use the *xtreg* command in STATA. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference country is Nigeria. Reference quarter is 36. Reference rank is *rank ≤ 5000*. Reference population is *pop ≤ 100*. Random institution-discipline effects are included in columns (1) to (6). Robust standard errors clustered at the institutional level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix 1: Summary Statistics (Single Local Authors)

	mean	sd	min	max	N
Dependent variables					
# publications	0.112	0.848	0	55	203,100
# references	19.80	21.77	0	293	203,100
# OARE references	4.243	7.383	0	135	203,100
Countries					
Kenya	0.256	0.436	0	1	203,100
Nigeria	0.240	0.427	0	1	203,100
Bolivia	0.134	0.341	0	1	203,100
Ecuador	0.137	0.344	0	1	203,100
Peru	0.233	0.423	0	1	203,100
Main variable of interest					
OARE treated (DDD)	0.008	0.0860	0	1	203,100
Article characteristics					
# co-authors USA	0.465	1.365	0	37.25	203,100
# co-authors EUR	0.569	1.727	0	54.33	203,100
# pages	6.081	6.547	0	73	203,100
Institutional characteristics					
Rank1: rank ≤ 5,000	0.0231	0.150	0	1	203,100
Rank2: 5,000 < rank ≤ 10,000	0.0162	0.126	0	1	203,100
Rank3: 10,000 < rank ≤ 15,000	0.0167	0.128	0	1	203,100
Rank4: 15,000 < rank ≤ 25,000	0.0246	0.155	0	1	203,100

Rank5: rank > 25,000	0.919	0.272	0	1	203,100
Rank, in 1,000	6.709	5.044	0.749	21.24	203,100
City characteristics					
Distance from largest domestic city, in 100 km	3.204	3.617	0	20.64	203,100
Distance from closest dom. city with > 1 million inhab., in 100 km	1.989	3.263	0	20.64	203,100
Pop0: pop ≤ 100, in 1,000	0.200	0.400	0	1	203,100
Pop1: 100 < pop ≤ 500, in 1,000	0.106	0.307	0	1	203,100
Pop2: 500 < pop ≤ 1,000, in 1,000	0.161	0.368	0	1	203,100
Pop3: 1,000 < pop ≤ 5,000, in 1,000	0.349	0.477	0	1	203,100
Pop4: pop > 5,000, in 1,000	0.184	0.387	0	1	203,100

We use a balanced panel and take into account journal articles by single local authors only. Data is aggregated at the institution-discipline-quarter level that constitutes our unit of observation.

Appendix 2: Summary statistics by country Group

A. Summary statistics for Group A countries

	mean	sd	min	max	N
Dependent variable					
# publications	0.175	1.420	0	93.82	159,900
Countries					
Kenya	0.398	0.490	0	1	159,900
Nigeria	0.401	0.490	0	1	159,900
Bolivia	0.201	0.401	0	1	159,900
Ecuador	0	0	0	0	159,900
Peru	0	0	0	0	159,900
Main variable of interest					
OARE treated (DDD)	0.008	0.0864	0	1	159,900
Article characteristics					
# co-authors USA	0.380	1.080	0	36	159,900
# co-authors EUR	0.499	1.633	0	57	159,900
# pages	6.064	6.047	0	66	159,900
Institutional characteristics					
Rank1: rank ≤ 5,000	0.0119	0.108	0	1	159,900
Rank2: 5,000 < rank ≤ 10,000	0.00813	0.0898	0	1	159,900
Rank3: 10,000 < rank ≤ 15,000	0.0188	0.136	0	1	159,900
Rank4: 15,000 < rank ≤ 25,000	0.0369	0.189	0	1	159,900
Rank5: rank > 25,000	0.924	0.264	0	1	159,900
Rank, in 1,000	8.883	5.197	0.907	21.79	159,900
City characteristics					
Distance from largest domestic city, in 100 km	3.523	3.568	0	20.64	159,900
Distance from closest dom. city with > 1 million inhab., in 100 km	2.031	2.915	0	20.64	159,900
Pop0: pop ≤ 100, in 1,000	0.257	0.437	0	1	159,900
Pop1: 100 < pop ≤ 500, in 1,000	0.0851	0.279	0	1	159,900
Pop2: 500 < pop ≤ 1,000, in 1,000	0.226	0.418	0	1	159,900
Pop3: 1,000 < pop ≤ 5,000, in 1,000	0.381	0.486	0	1	159,900
Pop4: pop > 5,000, in 1,000	0.0507	0.219	0	1	159,900

Data is aggregated at the institution-discipline-quarter level. The institution-discipline-quarter triplets constitute the unit of observation. We take into account journal articles by both single and multiple local authors in the Group A countries under study. Registered research institutions receive free OARE membership in Group A countries (GNI per capita below $1,600).

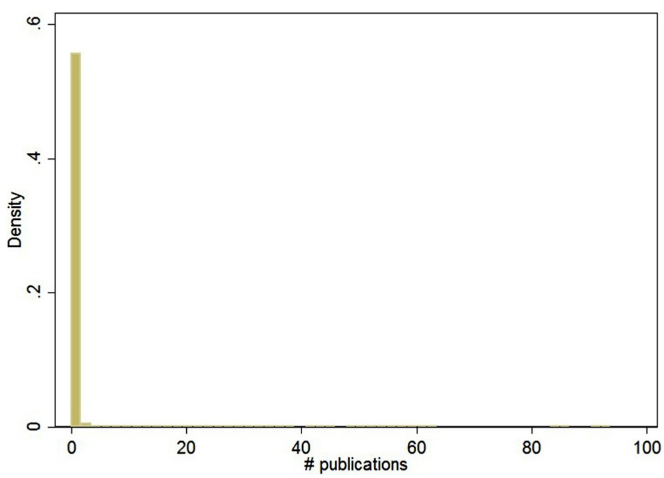
B. Summary statistics for Group B countries

	mean	sd	min	max	N
Dependent variable					
# publications	0.0768	0.598	0	29.53	89,100
Countries					
Kenya	0	0	0	0	89,100
Nigeria	0	0	0	0	89,100
Bolivia	0	0	0	0	89,100
Ecuador	0.364	0.481	0	1	89,100
Peru	0.636	0.481	0	1	89,100
Main variable of interest					
OARE treated (DDD)	0.006	0.0775	0	1	89,100
Article characteristics					
# co-authors USA	0.637	1.673	0	37.25	89,100
# co-authors EUR	0.621	1.681	0	39	89,100
# pages	6.457	6.738	0	120	89,100
Institutional characteristics					
Rank1: rank ≤ 5,000	0.0348	0.183	0	1	89,100
Rank2: 5,000 < rank ≤ 10,000	0.0303	0.171	0	1	89,100
Rank3: 10,000 < rank ≤ 15,000	0.0168	0.129	0	1	89,100

Rank4: 15,000 < rank ≤ 25,000	0.00449	0.0669	0	1	89,100
Rank5: rank > 25,000	0.914	0.281	0	1	89,100
Rank, in 1,000	4.137	1.714	0.749	21.39	89,100
City characteristics					
Distance from largest domestic city, in 100 km	2.771	3.643	0	15.73	89,100
Distance from closest dom. city with > 1 million inhab., in 100 km	1.920	3.693	0	15.73	89,100
Pop0: pop ≤ 100, in 1,000	0.153	0.360	0	1	89,100
Pop1: 100 < pop ≤ 500, in 1,000	0.145	0.352	0	1	89,100
Pop2: 500 < pop ≤ 1,000, in 1,000	0.0393	0.194	0	1	89,100
Pop3: 1,000 < pop ≤ 5,000, in 1,000	0.251	0.434	0	1	89,100
Pop4: pop > 5,000, in 1,000	0.412	0.492	0	1	89,100

Data is aggregated at the institution-discipline-quarter level. The institution-discipline-quarter triplets constitute the unit of observation. We take into account journal articles by both single and multiple local authors in the Group B countries under study. Registered research institutions receive reduced-fee OARE membership ($1,000 per year) in Group B countries (GNI per capita below $5,000).

Appendix 3: Histogram of the number of publications



Histogram of number of publications, *w*, at the institution-discipline-quarter level. *w* is the dependent variable in the regressions.

Appendix 4: Bayesian methodology

For robustness, we also estimate the OARE effect using Bayesian estimation techniques based on a data augmentation MCMC algorithm. There are two equations. The first equation determines self-selection in the OARE initiative using a latent variable framework. The second equation is identical to Eq. (1). We assume that the unobserved variables of both equations follow a bivariate normal distribution with correlation coefficient ρ. The MCMC algorithm simulates the latent variable of the first equation to generate the endogenous binary treatment effect. The Bayesian approach explicitly deals with the correlation between the unobserved variables of the two equations. If there are any unobserved variables that determine whether an institution self-selects into the OARE program, the Bayesian method accounts for its potential endogeneity on the estimation of the treatment effect. This comes at the cost of increasing computing time since the self-selection and productivity equations are estimated at the same time.

Eq. (A1) determines the outcome of the endogenous binary variable:

$$y_{1,i} = \begin{cases} 1, \; if \;\; w_{1,i} > 0 \\ 0, \; if \;\; w_{1,i} \le 0 \end{cases} \tag{A1}$$

where $w_{1,i} = x_{1,i}'\,\beta_1 + \varepsilon_{1,i}$, β_1 is of dimension k_1 and $x_{1,i}$ is a set of k_1 control variables.

Eq. (A2) explains the observed variable $w_{2,i}$ as a function of individual characteristics and the endogenous binary variable $z_{1,i} = 1$ if $w_{1,i} > 0$ and $z_{1,i} = 0$ if $w_{1,i} \le 0$,

$$w_{2,i} = z_{1,i}\delta_1 + z_{2,i}'\delta_2 + \varepsilon_{2,i} = x_{2,i}'\beta_2 + \varepsilon_{2,i} \tag{A2}$$

where δ_1 is the structural parameter associated with the binary endogenous variable z_1, $z_{2,i}$ is a set of k_2 explanatory variables not necessary identical to $x_{1,i}$ and δ_2 is a vector of parameters of dimension k_2, $x_{2,i} = (z_{1,i}, z_{2,i}')'$ and $\beta_2 = (\delta_1, \delta_2')'$. We assume that $\varepsilon_i = (\varepsilon_{1,i}', \varepsilon_{2,i}')'$ is normally distributed with mean $(0, 0)'$ and covariance Σ for $i = 1, \ldots, n$: $\sum = \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}$. Parameter ρ represents the correlation between the unobservable variables. Parameter σ^2 is the variance of $\varepsilon_{2,\,i}$. Since the probit Eq. (A1) is not identified, we chose to normalize the variance of the endogenous binary variable to 1. This is a standard restriction in probit models.

Let $\beta = (\beta_1', \beta_2')'$, $w_1 = (w_{1,1}, \ldots, w_{1,n})'$, $w_2 = (w_{2,1}, \ldots, w_{2,n2})'$ and define $w = (w_1', w_2')'$. We define ε_1, ε_2, and ε in a similar fashion.

The covariance of the unobservable variables is simply

$$\Omega = E\varepsilon\varepsilon' = \Sigma \otimes I_n$$

where I_n denotes the identity matrix of dimension $n \times n$. Thus Ω^{-1} is readily obtained. We similarly define

$$X = \begin{bmatrix} x_{11} & 0 \\ 0 & x_2 \end{bmatrix} 2n \times (k_1 + k_2)$$

The (partially) latent model can be written in matrix format:

$$w = X\beta + \varepsilon \quad \text{(A3)}$$

Hence conditional on w and Ω, the estimates of β are simply obtained by a generalized least-squares (GLS) regression of (A3).[38] Moreover, the matrices $X'\Omega^{-1}X$ and $X'\Omega^{-1}$ w required for the GLS estimates of the parameters of the model are easily computed. We use a uniform prior for β, ρ and a non-informative prior for σ: $p(\beta, \rho, \sigma) \propto 1/\sigma$.[39] The Metropolis-Gibbs sampling algorithm proceeds in 4 steps drawing from conditional distributions sequentially. The full procedure is described in Bounie et al. (2016). Parameters of the model are identified by the non-linearity of the probability of observing the endogenous equal to me, just as in the sample selection model. There is however no sample selection since we have observations for the two equations for all institutions in our sample. Moreover, the model is triangular and thus satisfies the principal assumption of Heckman (1978). Self-selection is accounted by the correlation between the unobservable self-selection equation and the productivity equation.

Appendix 5: OARE effect by country group

	(1)	(2)	(3)	(4)	(5)	(6)
Country Group:	Group A			Group B		
Model:	distance_1	distance_2	Inst.-Disc. FE	distance_1	distance_2	Inst.-Disc. FE
Dependent variable:	*w*	*w*	*w*	*w*	*w*	*w*
OARE treated (DDD)	2.472***	2.472***	2.446***	1.239***	1.239***	1.213***
	(0.542)	(0.542)	(0.536)	(0.409)	(0.409)	(0.402)
# pages	0.00628*	0.00628*	0.00873*	−0.000539	−0.000539	0.000764
	(0.00358)	(0.00358)	(0.00474)	(0.00471)	(0.00471)	(0.00614)
# co-authors USA	0.00842	0.00840	0.0102	0.00182	0.00181	0.000508
	(0.0151)	(0.0151)	(0.0172)	(0.0116)	(0.0116)	(0.0129)
# co-authors EUR	−0.00676	−0.00683	−0.00626	0.122**	0.122**	0.141**
	(0.00906)	(0.00905)	(0.0101)	(0.0531)	(0.0531)	(0.0567)
Rank2:	−0.679	−0.666		−0.844***	−0.845***	
5,000 < rank ≤ 10,000	(0.959)	(0.958)		(0.283)	(0.283)	
Rank3:	−1.214*	−1.195		−0.790***	−0.791***	
10,000 < rank ≤ 15,000	(0.726)	(0.729)		(0.262)	(0.262)	
Rank 4:	−0.298	−0.293		−0.892***	−0.892***	
15,000 < rank ≤ 25,000	(0.845)	(0.848)		(0.286)	(0.286)	
Rank5:	−1.359*	−1.346*		−0.815***	−0.815***	
rank > 25,000	(0.715)	(0.717)		(0.259)	(0.260)	
Pop1:	0.000493	−0.0373		−0.0643*	−0.0656*	
100 < pop ≤ 500, in 1,000	(0.0767)	(0.0789)		(0.0345)	(0.0348)	
Pop2:	0.150	0.0931		−0.0300	−0.0302	
500 < pop ≤ 1,000, in 1,000	(0.0944)	(0.0938)		(0.0341)	(0.0342)	
Pop3:	0.0189	0.0602		0.0462	0.0503	
1,000 < pop ≤ 5,000, in 1,000	(0.0746)	(0.0883)		(0.0461)	(0.0466)	
Pop4:	−0.265***	−0.128**		0.0378	0.0395	
pop > 5,000, in 1,000	(0.103)	(0.0653)		(0.0231)	(0.0242)	
Distance from largest domestic city, in 100 km	−0.0279**			0.00189		
	(0.0127)			(0.00232)		
Distance from closest dom. city with > 1 mill. inh., in 100 km		−0.00610			0.00220	
		(0.0127)			(0.00228)	
Constant	1.675**	1.524**	0.161***	0.793***	0.792***	−0.0646
	(0.708)	(0.704)	(0.0349)	(0.259)	(0.258)	(0.0431)
Quarter dummies	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	NO	YES	YES	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES	YES
Institution-discipline FE	NO	NO	YES	NO	NO	YES
Observations	159,900	159,900	159,900	89,100	89,100	89,100
R-squared, overall	0.1178	0.1162	0.0969	0.1333	0.1333	0.0584
Number of Inst_Discipline	3,198	3,198	3,198	1,782	1,782	1,782
Number of Inst	1,599	1,599	1,599	891	891	891

We use a balanced panel and take into account journal articles from both single and multiple local authors. Results on the impact of OARE membership (*OARE treated*) on publication output of research institutions by country group (Group A: Bolivia, Nigeria, Kenya; Group B: Ecuador, Peru) from OLS DDD estimation. We use Stata's *xtreg* command. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference quarter is 36. Reference rank is *rank* ≤ *5000*. Reference population is *pop* ≤ *100*. Random institution-discipline effects are included in columns (1), (2), (4), and (5). Robust standard errors clustered at the institutional level. Note that serial correlation is not an issue in our balanced panel because the large number of periods with 0 publications breaks any time correlation for any given institution. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

[38] Since each stage generally includes different sets of explanatory variables, we cannot estimate the seemingly unrelated regressions model with ordinary least-squares regression applied to each latent equation separately.

[39] The choice of the prior distribution does not matter much when there is a large number of observations. Moreover, using the uniform prior distribution provides a direct means of comparison with the maximum likelihood procedures.

Appendix 6: OARE effect by country

	(1)	(2)	(3)	(4)	(5)
Model:	+ Inst.-Disc. FE	+ Inst.-Disc. FE	+ Inst.-Disc. FE	+ Inst.-Disc. FE	+ Inst.-Disc. FE
Country:	Kenya	Nigeria	Bolivia	Ecuador	Peru
Dependent variable:	*w*	*w*	*w*	*w*	*w*
OARE treated (DDD)	2.017***	2.773***	0.811**	0.840*	1.364**
	(0.580)	(0.728)	(0.380)	(0.435)	(0.550)
# pages	0.00952	0.0122	0.00356	−0.0116	0.00877
	(0.00679)	(0.00969)	(0.00455)	(0.00751)	(0.00592)
# co-authors USA	0.0209	−0.0395	0.00804	−0.0291	−0.00821
	(0.0221)	(0.0498)	(0.0172)	(0.0498)	(0.0173)
# co-authors EUR	−0.00597	−0.0195	0.00511	0.276***	0.0780**
	(0.0195)	(0.0442)	(0.00922)	(0.0938)	(0.0394)
Constant	0.0472	0.0636	0.0345	−0.0703	−0.0374
	(0.0546)	(0.0679)	(0.0305)	(0.0479)	(0.0601)
Quarter dummies	YES	YES	YES	YES	YES
Country dummies	NO	NO	NO	NO	NO
Quarter-discipline dummies	YES	YES	YES	YES	YES
Institution-discipline FE	YES	YES	YES	YES	YES
Observations	63,700	64,100	32,100	32,400	56,700
R-squared, within	0.050	0.086	0.035	0.162	0.083
Number of Inst_Discipline	1,274	1,282	642	648	1,134
Number of Inst	637	641	321	324	567

We use a balanced panel and take into account journal articles from both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions by country from OLS DDD estimation. We use Stata's *xtreg* command. OLS regression coefficients reported. The institution-discipline-quarter triplets constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. Reference quarter is 36. Reference rank is $rank \leq 5000$. Reference population is $pop \leq 100$. Robust standard errors clustered at the institutional level. Note that serial correlation is not an issue in our balanced panel because the large number of periods with 0 publications breaks any time correlation for any given institution. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

References

Acs, Z.J., Audretsch, D.B., Feldman, M.P., 1994. R&D spillovers and recipient firm size. Rev. Econ. Stat. 76 (2), 336–340.
Annan, K., 2004. Science for all nations. Science 303, 925.
Audretsch, D.B., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. Am. Econ. Rev. 86 (3), 630–640.
Bergstrom, T., Bergstrom, C.T., 2004. The costs and benefits of library site licenses to academic journals. Proc. Natl. Acad. Sci. 101 (3), 897–902.
Bounie, D., Abel, F., Waelbroeck, P., 2016. Debit card and demand for cash. J. Bank. Financ 73, 55–66.
Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Res. Policy 23 (5), 487–521.
Davis, P.M., 2011. Do discounted journal access programs help researchers in Sub-Saharan Africa? A Bibliometric Analysis. Learned Publishing 24 (4), 287–298.
Dosi, G., 1988. Sources, procedures, and microeconomic effects of innovation. J. Econ. Lit. 26 (3), 1120–1171.
Eger, T., Scheufen, M., Meierrieks, D., 2015. The determinants of open access publishing: Survey evidence from Germany. Eur. J. Law Econ. 39 (3), 475–503.
European Commission, 2012. Towards better access to scientific information: Boosting the benefits of public investments in research. Brussels: European Commission.
Evans, J.A., Reimer, J., 2009a. Open access and global participation in science. Science 323, 1025.
Evans, J.A., Reimer, J., 2009b. Open access and global participation in science. Supporting Online Material. Science. pp. 323.
Frandsen, T.F., 2009. Attracted to open access journals: A bibliometric author analysis in the field of biology. J. Document. 65 (1), 58–82.
Furman, J.L., Stern, S., 2011. Climbing atop the shoulders of giants: The impact of institutions on cumulative research. Am. Econ. Rev. 101 (5), 1933–1963.
Gaulé, N., Maystre, N., 2011. Getting cited: Does open access help? Res. Policy 40 (10), 1332–1338.
Gaulé, P., 2009. Access to scientific literature in India. J. Am. Soc. Inform. Sci. Tech. 60 (12), 2548–2553.
Geroski, P.A., 2000. Models of technology diffusion. Res. Policy 29 (4–5), 603–625.
Griliches, Z., 1957. 'Hybrid corn': An exploration in the economics of technological change. Econometrica 25 (4), 501–522.
Griliches, Z., 1992. The search for R&D spillovers. Scand. J. Econ. 94, 29–47.
Hall, B., 2004. Innovation and diffusion, NBER, Cambridge, MA NBER Working Paper No. 10212.
Hall, B., Khan, B., 2003. Adoption of new technology, NBER. Cambridge, MA NBER Working Paper(9730) No.
Hanauske, M., Bernius, S., Dugall, B., 2007. Quantum game theory and open access publishing. Physica A. 382, 650–664.
Heckman, J.J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 46 (4), 931–959.
Jaffe, A.B., 1989. Real effects of academic research. Am. Econ. Rev. 79 (5), 957–970.
Jeon, D.S., Rochet, J.C., 2010. The pricing of academic journal: A two-sided market perspective. Am. Econ. J. Microecon. 2, 222–255.
Mansfield, E., 1961. Technical change and the rate of imitation. Econometrica 29 (4), 741–766.
Mansfield, E., 1963. Intrafirm rates of diffusion of an innovation. Rev. Econ. Stat. 45 (4), 348–359.
Merton, R.K., 1973. The sociology of science: theoretical and empirical investigations, Storer, N.W. (editor), University of Chicago Press: Chicago, IL.
McCabe, M.J., Snyder, C.M., 2005. Open access and academic journal quality. Am. Econ. Rev. Paper Proc. 95 (2), 453–458.
McCabe, M.J., Snyder, C.M., 2014. Identifying the effect of open access on citations using a panel of science journals. Econ. Inq. 52 (4), 1284–1300.
McCabe, M.J., Snyder, C.M., 2015. Does online availability increase citations? Theory and evidence from a panel of economics and business journals. Rev. Econ. Stat. 97 (1), 144–165.
McCabe, M.J., Snyder, C.M., 2018. Open access as a crude solution to a hold-up problem in the two-sided market for academic journals. J. Indust. Econ. 66 (2), 239–517.
Mueller-Langer, F., Scheufen, M., 2013. Academic publishing and open access. In: Towse, R., Handke, C. (Eds.), Handbook of the Digital Creative Economy. Edward Elgar, pp. 365–377.
Mueller-Langer, F., Watt, R., 2018. How many more cites is a $3,000 open access fee buying you? Empirical evidence from a natural experiment. Econ. Inq. 56 (2), 931–954.
Murray, F.E., Aghion, P., Dewatripont, M., Kolev, J., Stern, S., 2016. Of mice and academics: Examining the effect of openness on innovation. Am. Econ. J. Econ. Policy 8 (1), 212–252.
Ramello, G.B., 2010. Copyright and endogenous market structure: A glimpse from the journal-publishing market. Rev. Econ. Res. Copyright Issues 7 (1), 7–29.
Romer, P.M., 1986. Increasing returns and long-run growth. J. Polit. Econ. 94 (5), 1002–1037.
Romer, P.M., 1990. Endogenous technological change. J. Polit. Econ. 98 (5), 71–102.
Scheufen, M., 2015. Copyright versus open access: on the organisation and international political economy of access to scientific knowledge, Springer International Publishing.
Shavell, S., 2010. Should copyright for academic works be abolished? J. Legal Anal. 2 (1), 301–358.
Solow, R.M., 1956. A contribution to the theory of economic growth. Q. J. Econ. 70 (1), 65–94.
Sorenson, O., Fleming, L., 2004. Science and the diffusion of knowledge. Res. Policy 33 (10), 1615–1634.
Stephan, P.E., 1996. The economics of science. J. Econ. Lit. 34 (3), 1199–1235.
Suber, P., Arunachalam, S., 2005. Open access to science in the developing world, http://dash.harvard.edu/bitstream/handle/1/4725025/suber_worldinfocity.htm?sequence=1.
Suber, P., 2012. Open access. Cambridge: MIT Press.
Swan, T.W., 1956. Economic growth and capital accumulation. Econ. Rec. 32 (2), 334–361.
Williams, R., 2016. Outliers, University of Notre Dame, available at: https://www3.nd.edu/~rwilliam/.