The Cost of Knowledge: Academic Journal Pricing and Research Dissemination

Yonghong An Michael A. Williams Mo Xiao*

December 22, 2024

Abstract

The academic community is deeply concerned about restricted access to published research. To assess the effects of various access barriers, we compile a comprehensive database covering articles, journals, and publishers across three primary academic fields: economics, physics, and electronic engineering. Exploiting variations in publishers' product portfolios that do not directly affect an article's outcome for identification, we find that a 1% rise in journal price leads to a 0.77% decline in citations for an economics article within five years of publication. Elevated prices and substantial publisher power consistently hinder citation and collaborative research metrics, with these effects being more severe for lower-ranked institutions and in developing countries. In addition, we highlight subscription disparity using library-publisher contracts and utilize the timing of journals exiting publisher paywalls to demonstrate an immediate increase in article citations.

Keywords: Academic Publishing; Knowledge Dissemination, Market Power

JEL Classifications: L1, L13, L31

*Yonghong An: Department of Economics, Texas A&M University; email: yonghongan@tamu.edu. Michael A. Williams: Berkeley Research Group; email: michael.williams@thinkbrg.com. Mo Xiao: Department of Economics, Eller College of Management, the University of Arizona, Tucson, AZ 85721; email: mxiao@arizona.edu. We thank Florian Ederer, Mark McCabe and Christopher Snyder and many seminar participants for comments and suggestions. We have especially benefited from discussions with Eric Hartnett from the Department of Electronic Resources at Texas A&M University. Yongzhi Xu provided excellent research assistance. All errors are our own.

1 Introduction

The academic publishing industry serves as the nexus between knowledge dissemination and knowledge creation. Efficient functioning of this industry concerns not only the intellectual community but also the general public. Recently, however, there has been a surge of criticism about exorbitant journal prices and increased access restrictions, which has often escalated to battles between the research community and major publishers. In 2012, a renowned mathematician, Timothy Gowers, called for a boycott of Elsevier, an academic publishing powerhouse. The boycott developed into the "Cost of Knowledge" movement, which has garnered close to 20,000 researchers' signatures.[1] In 2019, the University of California system had a dispute with Elsevier about bundled access and universal free access to articles written by the University's researchers, which led to a well-publicized breakdown of their negotiation process.

Compared with other industries, the academic publishing industry has distinctive features that could lead to internal conflicts. First, this industry has a unique production function. Researchers, who are the ultimate consumers of published knowledge, also provide the main inputs to the knowledge production process. In addition to creating knowledge, researchers provide unpaid services, such as refereeing and editing, to get others' research published.[2] Second, the general public is effectively paying for knowledge twice: the first time when research is conducted using public funding; the second time when publishers charge prices for access. Third, as technology progresses, the duties traditionally performed by publishers (e.g., proofreading, typesetting, and copy editing) are either fully assumed by researchers who submit their manuscripts, or rendered unnecessary due to the advent of digital distribution (e.g., printing). As a result, major publishers have been able to cut their variable costs significantly and maintain a high margin in their pricing models.[3]

The internal conflicts embedded in the production and consumption of academic journals lead to "a captive work force and a captive audience."[4] This captivity is exacerbated by the highly concentrated market structure in the academic publishing industry. The inelastic demand (from academic libraries, which make the largest chunk of publishing revenue) and

[1] The movement cited three reasons for the boycott: 1) high prices for subscriptions to individual journals; 2) bundling subscriptions to journals of different value and importance; and 3) the publisher's support for laws and regulations that restrict the free exchange of information.

[2] Ellison (2002) provides a detailed account of the painfully drawn-out production and review process in the economics profession.

[3] Larivière, Haustein, and Mongeon (2015) document that Elsevier's operating profits rose from 665 million US dollars in 1991 to more than two billion US dollars in 2013, and its profit margin rose from 17% to about 24%. Overall, the profit margins of leading publishers were on a comparable level with the most profitable drug, bank and auto companies during this period.

[4] *Boston Globe* "Why scientists are boycotting a publisher," by Gareth Cook, Feb 12, 2012.

the high concentration in supply are often listed as the main culprits behind high prices of journal subscriptions.[5] The research community is deeply concerned about the consequences and repercussions of such high prices because the inability to pay for subscriptions restricts access to research to privileged individuals and institutions. This may create an imbalance in the infrastructure required by research and collaboration opportunities, widening the gap in knowledge dissemination and knowledge creation across institutions, communities and countries. There has been anecdotal evidence about the access disparity and its negative implications, but systematic evidence has yet to emerge.

To fill this gap, we constructed a comprehensive database, dated 2009 to 2018, of articles' citation outcomes, journal prices and attributes, and publishers' product portfolio in three academic fields: economics, electronic engineering, and physics. We focus on two measures of access barriers — journal retail prices and publishers' market shares of published articles, both of which contribute to the limited distribution of published research. For each field, we conduct a separate analysis to quantify the effects of increasing prices and publishers' market power on the volume and distribution of article citations and research collaboration across institutions and countries. We exploit variations in our data across journals and publishers and over time for the identification of causal effects. In particular, to address the endogeneity problem of journal price setting (based on time-varying unobserved journal attributes), we adopt an instrumental variable approach in which the average attributes of rival publishers' journals, interacted with the focal publisher's nonprofit status, serve as instruments for journal prices. The idea is that publishers set prices based on how their rivals set prices, but competition pressure affects for-profit and nonprofit publishers differently; at the same time, these instrumental variables should not directly affect the citation outcomes of an article published in the focal journal.

We present results from economics as our benchmark. In economics, academic publishing has been dominated by a few large firms, led by for-profit global publishing powerhouses Elsevier and Wiley. Deflated retail prices of journals roughly doubled in the ten years that we study. We show that in five years of publication, an article's citations and the number of citing authors of these citations were substantially lower with higher journal prices and stronger publishers' market power. Under our preferred specification, a 1% increase in journal prices leads to a 0.77% decrease in an article's citation count and a 1.08% decrease in its citing author count. This negative effect is much larger for citations from lower-ranked institutions and from developing countries. Furthermore, both measures of access barriers hinder researcher collaboration, captured by the number of co-authors in a citation and

[5] For example, McCabe (2002) shows that biomedical journal prices increased substantially due to frequent mergers among publishers of biomedical journals in the 1990s.

collaboration across institutions/countries. Such a hindrance effect is roughly the same within the same institution and across different ones (for example, collaboration between a researcher from Texas A&M and one from the University of Arizona), but much more pronounced across institution tiers (for example, collaboration between a Harvard researcher and a University of Arizona one).[6]

A natural question is that whether these detrimental results are, indeed, due to restricted access. Are researchers really affected by the retail prices often printed on the journal covers? How much do universities actually pay for journal subscriptions? Are university subscriptions sensitive to price changes? How does publishers' market dominance affect journal access? To answer these questions, we conduct three additional analyses.

First, we make use of a sample of subscription contracts between public universities and Elsevier in the early part of our data periodto study how access at the university level is determined. We discover substantial heterogeneity across these contracts, especially across higher- and lower-ranked universities. Lower-ranked universities subscribe to many fewer journals and pay much more dispersed prices for the journal bundle they subscribe to. We also find a strong, positive correlation between the contracted prices and retail prices of the journals in these contracts. Based on these facts and previous studies on the negotiation process between universities and publishers, we conclude that universities negotiate highly-customized contracts with publishers, with the retail prices serving as the basis for the prices of basic bundles and add-on individual journals. In this negotiation process, each party's bargaining power affects the prices set and journals included. Large publishers such as Elsevier and Wiley tend to have the upper hand, negotiating more favorable terms for themselves. Through the negotiation process, journals' retail prices and publisher market power act together to determine research distribution and access (or lack thereof).

Second, we investigate the extent to which a journal article experiences a citation boost immediately after the journal moves out of the publisher's paywall. If there is little boost, the paywall should take no blame for restricting access. In economics, JSTOR is a popular access point for many researchers, offering a subscription-based access model for institutions and much lower prices than publishers' charge. Economics journals typically move into JSTOR with a delay of between zero and ten years (named "JSTOR moving wall"), varying significantly across journals. We exploit the timing of the moving wall in a difference-in-difference framework, estimating the gain in citations after an article becomes JSTOR-accessible. The JSTOR treatment is binary and staggered: the treated group contains journals accessible at JSTOR after their moving wall (of various lengths) ends, and the control group contains

[6]As a comparison, this hindrance effect is about the same for collaborations within a country, across countries and across country development statuses.

journals that are not yet or are never available at JSTOR. We allow heterogeneous, dynamic treatment effects in two-way fixed effects regressions based on De Chaisemartin and d'Haultfoeuille (2020) and De Chaisemartin and d'Haultfoeuille (2024). We find an immediate, sizable boost in the citation outcomes of articles with the JSTOR treatment. The boost grows stronger over the post-publication observation period, and is particularly noticeable for citations coming from lower-ranked institutions, which tend to experience more access limitations. As the predetermined length of the moving wall does not depend on the time-varying unobserved heterogeneity at the article (or journal) level, we believe that this finding supports our claim that accessibility is the key mechanism behind our results. This finding, coupled with the fact that Elsevier, the largest publisher in economics, does not allow its journals to enter JSTOR, also suggests that publisher market power is the bottleneck of expanding accessibility of academic research.[7]

Lastly, we expand our analyses to physics and electronic engineering for a cross-field comparison to shed light on both diagnoses and remedies. In these two fields, as in economics, subscription-based journals serve as a main media for knowledge dissemination, and journal publication is a major criterion for a researcher's professional development. In the three academic fields we study, however, the research input, researchers' outside option to access journals,[8] funding channels, collaboration format and publication cycles differ substantially from one another. Moreover, the role of nonprofit publishers can be vastly different across fields; for example, the *Institute of Electrical and Electronics Engineers (IEEE)*, a nonprofit publisher, dominates the publications in electronic engineering. We consider expanding our investigation to different fields necessary to generalize our results and to understand the extent of access barriers under different production functions of knowledge.

Indeed, the barriers we study indeed function differently in physics and electronic engineering, but the big picture on access barriers remains the same. In physics, high price is the main hindrance factor, while publisher market power has a statistically insignificant effect; in electronic engineering, the results are just the opposite: prices play a negligible role, while publisher market power has a huge negative effect. These results reflect the fundamental differences in how markets are organized in these three fields: in economics, for-profit publishers dominate, with Elsevier and Wiley leading the pack; in physics, prices are much higher, and the market structure is dispersed among a handful of for-profit companies; in electronic engineering, the market is again very concentrated (perhaps even more so than in

[7]The only exception is that Elsevier Masson SAS, a publisher created by the 2005 Masson and Elsevier merger, has a single journal *Sociologie du Travail* at JSTOR.

[8]For example, in physics, pre-print dissemination via the *arXiv* is the norm, but in economics, access to the NBER or IZA working paper series requires institution membership. In electronic engineering, conference proceedings could have a higher impact than journals to distribute the newest knowledge and discoveries.

economics), but it is dominated by one large nonprofit publisher (IEEE).

Diagnosing the problems of the academic journal market is the pre-step to finding potential remedies. A further look at our results leads to discoveries on two frequently-debated remedies: the role of non-for-profit publishers and that of open access. We find little evidence to support elevating the status of nonprofit publishers as a solution. Across all three fields, the nonprofit status of publishers seems to have a negligible role in reducing the negative effects of higher prices and stronger publisher power across fields. Results from electronic engineering also suggest that simply replacing market powers with nonprofit organizations is not the solution to overcome barriers. Open access, however, seems to be very effective at boosting citations in economics: an article with open access has a 28% gain in the five-year citation count and a 34% gain in the five-year citing authors count. In physics and electronic engineering, the open-access effects can be considered quite sizable as well, going from 10% to 20% depending on outcomes and specifications. We show that a 30% decrease in prices in economics (a 5% decrease in physics), roughly corresponding to \$300 to \$400 in both fields, achieves roughly the effect of open access in the field, but this price reduction remedy needs to be field-specific and location-specific.

Researchers across multiple disciplines complain about high prices and market power in the academic publishing industry(Bergstrom (2001); Edlin and Rubinfeld (2004)), but there is only scant evidence on credible causal links between access barriers and knowledge dissemination. This paper is the first to directly estimate the effects of academic journal prices and publishers' market power on knowledge dissemination. We have constructed a comprehensive database across multiple academic fields on journal articles, citations, citing authors' networks, journal prices, journal attributes and publisher attributes. Our identification strategy makes use of the variation in publishers' product portfolios that does not directly affect an article's impact — a widely recognized identification strategy in economics. We further provide direct evidence to demonstrate the effects of enhanced access on article citations, exploiting the timing of economics journals moving out of publishers' paywall. In summary, our work brings new data and a new identification strategy to provide much-needed evidence about the effects of access barriers on knowledge dissemination and knowledge creation.

Our research contributes to two branches of closely-related literature, the first attempts to diagnose the problems of academic publishing and to find remedies (e.g. the push for open access) and the second one concerns the process of knowledge diffusion and creation.

Increasing prices and high ownership concentration have been noted in the literature(Albee and Dingley (2001)), despite the data being a bit dated. Our own data shows that the average prices (in 2020 dollars) of academic journals in economics, physics and electronic engineering roughly doubled from 2009 to 2018; these three fields have been consistently dominated by a few power house publishers. Furthermore, there has been a stark difference between for-profit and nonprofit pricing. Bergstrom et al. (2014) collected the contracts between publishers and public universities and documented that commercial publishers charge much higher prices per citation for academic journals than nonprofit societies charge.[9]

Researchers have cited both demand- and supply-side reasons for these facts and trends. On the demand side, libraries cannot easily switch away from over-priced journals because researchers (the primary consumers) face no budget constraints and are therefore unlikely to respond to prices (Nevo, Rubinfeld, and McCabe (2005), McCabe, Nevo, and Rubinfeld (2006)).[10] On the supply side, McCabe (2002) shows that mergers of commercial publishers led to higher journal prices in a broadly defined portfolio market; Edlin and Rubinfeld (2004) and Björk (2021) attribute the price increases to publishers' "big deal" bundling practices.[11] This literature focuses on understanding the pricing tactics of publishers, which is a prerequisite to potential remedies. Our work has the same theme but a different focus; that is, we study the outcomes instead of the determinants of high prices and market power. In this sense, our work represents the other side of the coin in the literature on the effects of open access (of a journal or an article), which acts as an access facilitator.[12]

Going beyond this specific marketplace, this research contributes to the broader literature of knowledge diffusion and creation. We follow in the footsteps of Iaria, Schwarz, and Waldinger (2018), who find that a disruption in global knowledge exchange triggered by World War I led to a marked decrease in scientific productivity, highlighting the importance of access to published research for scientific and technological advancement. Conversely, reducing the cost of obtaining existing research significantly fosters innovation and knowledge creation. Murray et al. (2016) demonstrate that a late-1990s agreement by the National Institute of Health, which lowered access costs for certain genetically engineered mice, allowed new researchers to enter the field and encouraged exploration of novel research directions. Berkes and Nencka (2021) document a 7 to 11% increase in patenting within two decades

[9]For example, Elsevier's prices per citation for large PhD-granting institutions are almost three times those charged by nonprofit publishers, while other for-profit publishers charge even higher rates.

[10]The authors did find inelastic demand but also found journal prices were lower than most static pricing models would predict. They suspected the dynamics of journal pricing at play.

[11]Many publishers offer libraries large bundles of journals across academic fields and across print and electronic versions at a discount. This is usually a long-term arrangement that locks in libraries, leaving them little room to purchase journals from competitors, especially new entrants.

[12]Armstrong (2015) provides an in-depth analysis of the rationale and trade-offs of open access policies, and most of this literature tries to quantify the effects of open access on the number of citations of a journal article using different data and identification strategies. McCabe and Snyder (2005) discuss the relationship between open access and journal quality. McCabe and Snyder (2014) employ a panel data to estimate that open access increases citations by 8%. McCabe and Snyder (2015) show that JSTOR's availability increases citations to journals substantially. More recently, McCabe and Snyder (2021) demonstrate the heterogeneous effects of open access on the least- and most-cited articles.

following the establishment of Carnegie-funded public libraries, which made knowledge more accessible to the public. By providing credible empirical results on the effects of access barriers in a modern-day setting, our research fills a void in the current discussion of the malfunctions of the academic publishing.

We will proceed by describing the institutional details of academic publishing that are relevant to our research, and the multiple data sets we collected to compile the panel we use for analysis. We then present the econometric framework and identification strategy. Among the three fields that we explore, we present the results from economics first and in more detail. Lastly, we present results from physics and electronic engineering for comparison and discussions. We provide an online appendix to document data-processing procedures and show details and robustness analyses.

2 Institutional Background and Data

2.1 A Brief History and a Recent Rift

Academic journal publishing is a sub-field of publishing that distributes academic research articles and theses. The first academic journal, *Journal des Sçavans*, published in January 1665, marked the birth of the academic journal publishing industry. The industry changed the process of scholarly communication from personal correspondence, society meetings and books to systematic distribution, recording and archiving. Up until the end of World War II, academic publishing was supported almost entirely by nonprofit academic societies. After 1945, for-profit publishers began to acquire journals from academic societies. The for-profit publishers' share of academic publishing increased dramatically due to the rapid growth of academic research and the availability of generous funding for university libraries.

Over multiple decades, the number of researchers and the market size of scholarly publishing have grown steadily. The Web of Science recorded 24,974 journals in October 2021, and the market value of scholarly publishing was $26.5 billion in 2020. The market has recently been highly concentrated. Based on Journal Citation Reports (JCR) 2014-18 data, Kim and Park (2020) find that 56.3% of journals and 57.3% of articles were published by five for-profit publishers: Elsevier, Springer, Wiley, Taylor & Francis, and Sage. Waves of consolidation in the publishing industry in the last few decades contributed to this high concentration(McCabe (2002)).

Academia has complained about the publishing industry for some time. The prices of for-profit academic journals have increased rapidly from the 1980s to the 2000s and have shown no sign of stopping (Albee and Dingley (2001); Kim and Park (2020)). Frequently mentioned complaints also include the bundles of journals that are forced upon universities and restrictions on authors' rights to share their articles(Edlin and Rubinfeld (2004); Björk (2021)). In general, the public resents the access restrictions set by the very profitable publishing business, claiming that knowledge has been held hostage by one or by several publishing houses. In the last decade, this resentment has escalated to disputes, lawsuits, and frequent breakdowns of contract negotiations. In 2015, Elsevier filed a law suit against Sci-Hub, a website that provides free access to published articles, infringing on publishers' copyright and bypassing publishers' paywalls. In 2018, about 3000 people, mostly academic computer scientists, signed a petition promising not to submit, review, or edit articles for *Nature Machine Intelligence*, a new journal from the publisher Springer Nature, which was set to begin publication in January 2019. Recently, several university libraries, including those of the University of California, Harvard, and the University of Konstanz, stopped negotiations or temporarily canceled subscriptions with major for-profit publishers.

The core of this conflict is the public good nature of knowledge. As Armstrong (2015) points out, a dollar of subscriber payment is worth more than a dollar of publisher profit; furthermore, much of the knowledge has been created with public funding. The open access movement has been brewing for the last two decades, culminating in the Biden administration's announcement in 2022 that, by the end of 2025, federal agencies will be mandated to make papers that describe taxpayer-funded work freely available to the public as soon as the final peer-reviewed manuscript is published. Bergstrom (2001) discusses three other approaches to deal with the conflicts: expanding nonprofit journals; supporting new, competing electronic journals; and punishing overpriced journals. All these solutions, however, are often still blueprints (and sometimes just the wishful thinking of the academia), the implementation of which requires a concerted effort from parties with very different interests and incentives.

2.2 Choice of Academic Fields to Study

We establish two criteria to select academic fields for our study. First, subscription-based journal articles are major outputs to evaluate the quality of a scholar's research in the field. The arts and humanities, for this reason, is not a good candidate because journal articles are not the most important research output in this area.[13] Second, there have been significant variations in journal prices and market concentration over the time span we study.

Based on these criteria, economics, two sub-fields of physics (atomic, molecular, & chem-

[13]Crossick (2007) summarizes data from the 2001 Research Assessment Exercise in the UK to show that journal articles comprise only 37% of scholars' output in the arts and humanities, while books comprise 50%. In contrast, journal articles account for 96% and 78% in science and engineering, respectively.

ical and condensed matter, hereinafter physics),[14] and electronic engineering (EE) stand out, each representing a significant research field in social sciences, sciences, and technology. In addition, existing studies on open access suggesting that scholars in these fields are more likely than those in other fields to have limited access to the literature(Tennant et al. (2016)).

Among the three fields, we use economics as our benchmark because the discipline of social sciences has the highest level of publisher concentration[15] and because we have relatively more knowledge and experience as researchers or practitioners in this field. We then use physics and EE as comparison groups because research input, researchers' outside option to access journals, funding channels, collaboration format and publication cycles are substantially different in these two fields. In particular, these two fields each depart from economics in one prominent way: in physics, journal prices, on average, exceed $5,000 for an annual retail subscription, while the equivalent is about $1,000 in economics; in EE prices are roughly in the same ballpark as economics, but EE features a dominant nonprofit publisher — IEEE — while economics' leading publishers are Elsevier and Wiley, both for-profit. Investigation across these distinctly different fields allows us to evaluate the generalizability of our results and to show how dysfunctions in the publishing industry interact with each field's idiosyncrasies.

2.3 Data

We assemble data on journal articles, journals, publishers, and library-publisher contracts spanning eleven years (roughly from 2009 to 2018) from multiple sources. We use the main data sets (three data sets in economics, physics and EE, respectively) to quantify the impacts of journal prices and publishers' market power on knowledge dissemination and research collaboration. We use complementary data sets to enhance our understanding of the institutions, industry practices, and mechanisms behind our results.

For each of the three academic fields, we compile a list of journals that were ever shown in the Social Sciences Citation Index (SSCI) or Science Citation Index Expanded (SCIE) from 2009 to 2019. For each journal on this list, we collect the following data:

- *Journal articles, citations, and citing authors.* We collect all the articles published in the journal between January 2009 and December 2018. For each article, we record

[14] Physics is a very broad field, with 469 journals in total. To make data collection feasible, upon the advice of a University of Arizona physicist, we chose two sub fields in which academic journals play a more important role in knowledge dissemination.

[15] Larivière et al. (2015) report that, in 2013, 70% of published papers in the social sciences were from the top five publishers.

whether the article is open-access[16] and the article's citations dated from its publication to December 2019. For each citation of a journal article, we record each co-author's name, affiliation, and country of affiliation. We collect these data by web scraping from the Incites-Clarivate (owned by the Web of Science). The citation count, citing author count and citing authors' collaboration relationships are our main outcome variables — doing this, we follow the tradition of using bibliometric data to measure knowledge generation and diffusion(Griliches (1990); Moser et al. (2018)).[17]

- *Journal retail prices for annual subscription.* Whenever possible, we recorded pricing information manually from publishers and various online sources. If no results can be found this way, we directly contact publishers for such information. After cleaning, we have retail prices (in 2020 dollars) for print-only, electronic-only, and print-and-electronic-bundled access for the majority of journals in all three fields.[18] The first two subsections of Appendix A describe the publishers we work on and the construction process of the price variables.

- *Journal attributes.* For each of the journals collected in Incites, we web-scrape the journal's publisher, the percentage of citable items, the ranking, and various measures of the journal's impact (e.g., the journal's Impact Factor). The last subsection of Appendix A describes various journal attributes we use in our study.

- *University Ranking.* We obtain the ranking of global top universities in 2020 from Quacquarelli Symonds (QS), a British company specializing in the analysis of global higher education market. QS World University Rankings is regarded as one of the three most influential university rankings in the world. We use this information when we assess the differential impact of the academic journal market on education institutions in different quality tiers.

We merge the above data sets by journal and university names. In the final product, each entry is an article, published in a certain year by a journal that is owned by a publisher. For each article, we observe its aggregate citation and citing author outcomes within a certain

[16] An article is open-access when the journal is an open-access journal or when the authors of the article purchased open access from the subscription-based journal.

[17] Collection and cleaning of this data set have been extremely time-consuming due to the size of the data (9.6 million citation items for economics; other fields are of similar size by estimation), the frequently changed layout of Incites-Clarivate that interrupts the web crawler, and the inconsistent spelling of institutions and countries, which requires manual cleaning.

[18] We have prices data for 309 out of 364 non-open-access journals in economics(the total number of journals is 423), 90 out of 121 non open access journals in physics (the total number of journals is 133), and 192 out of 303 non-open-access journals in EE(the total number of journals is 387). Note that the number of journals varies from year to year. In EE, for example, there were 266 journals in 2019.

time frame (specifically, within two, five and eight years since publication), the number of citing authors in each citation, and each citing author's institution/country affiliation. In addition, we measure the extent of co-authorship across institutions and countries in the fashion of Azoulay et al. (2019). At the article's year of publication, we then merge in journal prices and journal attributes, as well as publisher attributes (nonprofit, number of journals owned, number of articles published, and market shares based on journals or articles).

Beyond the main data, we collect one additional piece of information for the field of economics — the time at which the economics journal is deposited at JSTOR. JSTOR is a nonprofit digital library that provides delayed bundled access to academic journals at a low price point for subscribing institutions. The length of delay (called "Moving Wall" by JSTOR) can vary from zero to ten years and is negotiated by JSTOR and journal publishers. We use the Moving Wall length to construct a difference-in-differences identification strategy to evaluate the effect of increased access at lower prices on citation outcomes.

Lastly, we obtain a sample of subscription contracts between Elsevier and public universities. These contracts were requested by Professor Ted Bergstrom at University of California, Santa Barbara under the Freedom of Information Act (Bergstrom et al. (2014)). We have 70 contracts spanning 2001 to 2014, with the majority from 2007 to 2009.[19] These contracts often represent several universities' collective bargaining with Elsevier. For example, the Arizona Board of Regents negotiated with Elsevier, representing Arizona State University, the University of Arizona and Northern Arizona University. These contracts record the list of journals to which a university subscribed from Elsevier and the negotiated price for each journal. These contracts allow us to study the relationship between journals' retail prices and their negotiated ones and to analyze different subscription access by universities of different tiers.

3 Summary Statistics: Facts and Trends

In this section, we present stylized facts about articles, journals and publishers, which will help us to pin down empirical specifications of our study. The main text reports statistics on economics, while Appendix B reports on physics and EE. In addition, we analyze the university-publisher contracts to understand negotiated pricing and subscribed access.

[19]This collection originally had 360 contracts, but only Elsevier contracts recorded detailed information on negotiated prices.

3.1 Articles and Citation Outcomes

We report summary statistics for journal articles and their citation outcomes in the field of economics in Table 1. The top panel reports the number of aggregate citations and citing authors within two, five and eight years since the article's publication. The number of observations decreases along these three time frames because we do not have a sufficiently long observation window for articles published in later years. The middle and bottom panels single out the five year time window, for which we have six cohorts of articles (articles born in 2009, 2010, ..., and 2014), and we consider five years a milestone time frame to measure an article's impact following norms in academic publishing.

Table 1 shows a few notable patterns. First, not surprisingly, citations built up gradually, with large variations across articles. Second, the mean of the number of citations and citing authors was much higher than the median, suggesting that the distributions of citations and citing authors were highly skewed to the right. Third, there were substantial variations across different tiers of institutes and countries. Lastly, the 2014 cohort displays significant difference from the 2009 cohort: a mere five years later, the 2014 cohort receives more citations, most of the gain driven by lower-ranked institutions and developing countries; citations are more likely from multi-author (≥ 3) collaborations in various forms (within the same institutions/country, across institutions/country, and across institution/country tiers). Although only briefly reported in Table Appendix B.1, the citation outcomes in physics and EE are fairly similar to those in economics. The strong cohort effect and huge variations across articles necessitate the inclusion of year- and journal- fixed effects in our empirical specification.

3.2 Journals and Publishers

Table 2 reports summary statistics on economics journals (Panel A) and publishers (Panel B). The data come with the caveat that we do not have prices for all 3,525 journal-year combinations: instead, our price data cover all major publishers (five for-profit and seven nonprofit), which represent 67% of all journal-year combinations and 80% of all articles published during the time span. The most striking feature is the huge variation across journals. Among the three price formats, the print- and electronic-access bundle costs the most, followed by print-only, and then by electronic-only. The bundle prices range from a mere \$100 to almost \$14,000, with an average of slightly above \$1,000. The journal attributes also display substantial variations in the number of articles published and the quality measures.

The publisher market in economics can be best described as an oligopoly with a fringe

Table 1: Articles and Citations in Economics

Variable	# obs.	Mean	Median	SD	2009 Mean	2014 Mean
Open Access	240,208	0.266	0	0.442	0.212	0.294
Citations within a time frame after article publication						
# Citations in 2 years	213,183	2.946	1	5.817	2.684	2.831
# Citations in 5 years	135,425	9.115	3	18.229	8.673	9.331
# Citations in 8 years	61,927	16.127	6	35.787	15.481	n.a.
# Citing authors in 2 years	213,183	7.929	2	19.479	6.118	7.731
# Citing authors in 5 years	135,425	24.277	8	55.907	21.114	26.508
# Citing authors in 8 years	61,927	42.245	12	103.122	39.062	n.a.
Among citing authors within 5 years of publication						
# from rank $1^{st} - 100^{th}$	135,425	4.257	0	11.771	4.037	4.316
# from rank $101^{st} - 500^{th}$	135,425	5.978	1	14.227	5.223	6.365
# from rank $501^{st} - above$	135,425	2.618	0	6.532	2.242	3.021
# from unranked	135,425	11.428	3	28.783	9.611	12.806
# from developed countries	135,425	18.232	5	44.617	16.745	18.628
# from developing countries	135,425	5.875	0	17.574	4.116	7.784
Among citations within 5 years of publication						
# w. single author	135,425	2.171	1	4.831	2.482	1.886
# w. two authors	135,425	2.950	1	6.096	2.980	2.867
# w. three authors	135,425	2.244	1	5.073	1.945	2.417
# w. $\geq$ four authors	135,425	1.750	0	5.210	1.266	2.161
# w. same inst. colab.	135,425	2.283	1	5.040	2.024	2.396
# w. across inst. colab.	135,425	1.414	0	3.193	1.279	1.532
# w. across inst. tiers colab.	135,425	3.247	1	7.256	2.888	3.517
# w. same country. colab.	135,425	4.529	1	9.633	4.127	4.761
# w. across country colab.	135,425	1.469	0	3.525	1.322	1.538
# w. across country tiers colab.	135,425	0.946	0	2.528	0.743	1.146

Notes: Each observation is a journal article published in an economics journal between 2009 to 2018. There are 240,208 articles in ten cohorts, and for each cohort, we have an observation window of different length. Articles published in later years do not have citation outcomes in the five-year or eight-year observation window. For the last two columns, the number of journal articles in 2009 is 19,996, and the number in 2014 is 25,269.

Table 2: Journals and Publishers in Economics (2009 to 2018)

Variable	# obs.	Mean	Median	SD	Min	Max
Panel A: Journal Attributes						
Prices (in 2020 $)						
Bundle: print + electronic	2,366	1,026.911	741.633	977.895	102.788	13,873.36
Print only	1,620	949.099	634.382	841.908	85.380	5,203.36
Electronic only	2,352	920.076	628.001	919.732	75.828	12,139.56
# Article published	3,525	68.144	40	213.768	1	4,952
% Citable Items	3,174	98.678	100	5.167	0	100
Impact Factor	3,178	1.224	0.920	1.087	0	11.775
Immediacy Index	2,839	0.300	0.188	0.413	0	5.833
Eigenvector Score	3,187	0.005	0.002	0.010	0	0.137
Cited Half Life	2,823	8.403	8.800	2.919	0.800	30.200
Web of Science Rank	3,187	163.551	161	96.245	1	363
Panel B: Publisher Attributes						
nonprofit (= 1 if yes)	620	0.773	1	0.420	0	1
# Journals owned	620	5.685	1	14.251	1	74
# Article published	620	387.432	45	1,471.574	2	12,895
Market share in journals	620	0.018	.003	0.045	0.003	0.233
Market share in articles	620	0.016	.002	0.060	0.0001	0.478

Notes: Panel A: 3,525 journal-year combinations (363 journals × 10 years, with some journals entering in the middle of our time span and missing information in prices or journal attributes). Panel B: 620 publisher-year combinations (64 publishers × 10 years, with five publishers entering in the middle of our time span). The last subsection of Appendix A defines journal attributes reported in Panel A.

of small firms, with 77% of publishers being nonprofit organizations. There were mega publishing houses such as Elsevier and Wiley, which owned 50 to 75 journals and took a big chunk of market shares in published articles;there were also hundreds of tiny publishers that owned a single journal and accounted for almost nothing in market shares.

Figure 1 shows a steady to steep increase in prices (in $2000) from 2008 to 2018 for four tiers of economics journals, labeled "Top 5," " Field +," "Lower" and "Lowest" (tiers are based on Combes and Linnemer (2010)). For "Lower" and "Lowest" journals, the annual subscription prices roughly doubled. Figure 2 shows that Elsevier and Wiley were market leaders, followed closely by the next five. The right panel of this figure is marked by a dramatic increase in Elsevier's market share at the expense of Wiley's, despite only moderate fluctuations in the number of journals published by Elsevier and Wiley.[20]

Appendix B reports journal and publisher attributes for physics and EE. Both fields were much larger than economics in terms of the number of articles published, with 389,421 journal articles in physics and 495,403 in EE (compared to 240,208 in economics). Both fields, however, had fewer publishers and journals: physics had 133 journals by 26 publishers; EE had 387 journals by 41 publishers. Two facts represent notable departures from economics. In physics, journals were much higher-priced, averaging around $5,000 for annual subscriptions. In EE, there was a single dominant player in the market — IEEE — a nonprofit organization that published over 50% of articles in the field. Elsevier was only a distant second.

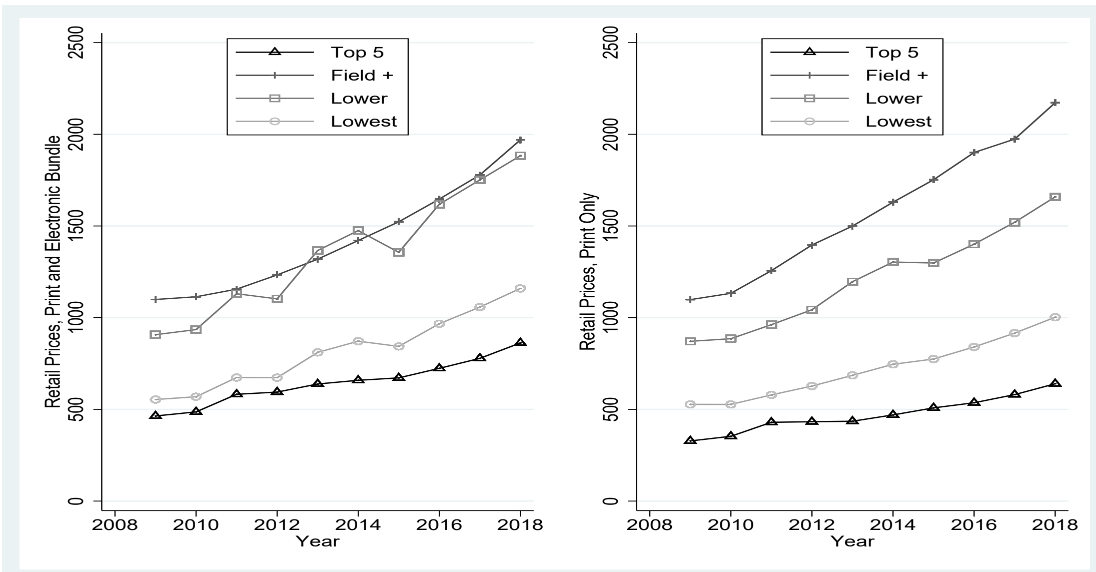
These stylized facts and trends in prices and publishers' market shares not only illustrate the rapidly-increasing barriers to journal access over time, but also give us variations necessary for the identification of the impact parameters we are interested in.

3.3 Elsevier-University Library Contracts

We construct two measures of access barriers: journals' retail prices for annual subscriptions and publishers' market power as measured by their share of articles. Do these two measures capture the access problem in this market? In particular, does anyone at all subscribe to journals at their retail prices? And why does the concentration of the publisher matter? After all, the majority of the research community accesses academic journals through university libraries, which typically negotiate multi-year contracts with publishers (Bergstrom et al. (2014)). If such contracts contained similar-sized bundles of journals of similar quality, then researchers affiliated with subscribing universities would have equitable access to academic journals. If negotiated prices had no relations with retail prices, then retail prices would

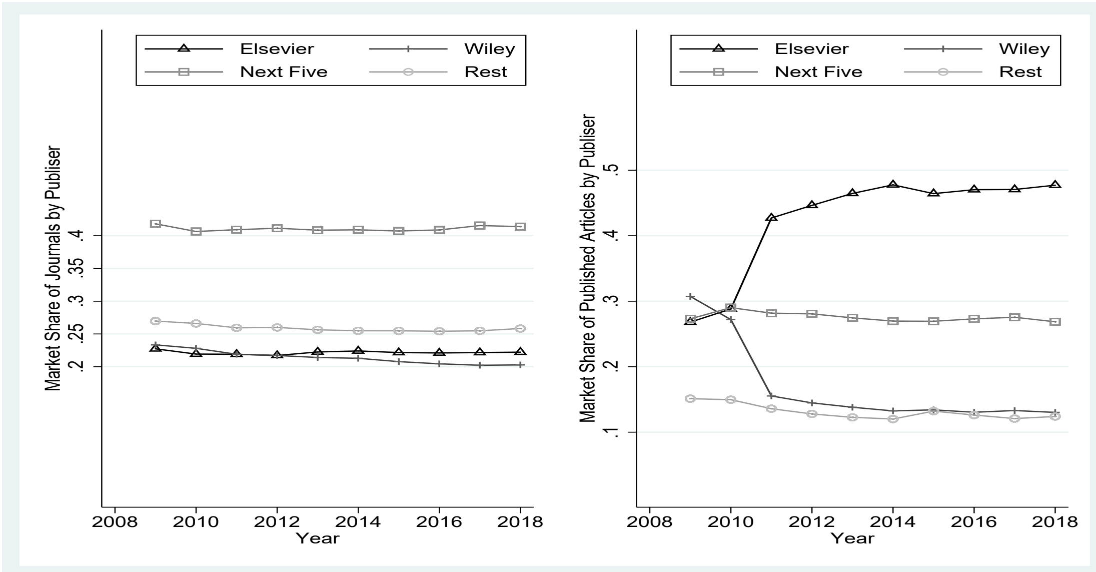
[20] A journal named *Value in Health* switched from Wiley to Elsevier in 2011. This journal published 2,567 to 4,952 articles annually during our data span. This is why the market share of Elsevier increased dramatically, while the share of Wiley dropped by roughly the same amount in 2011.

Figure 1: Economics Journal Prices Over Time



Notes: From left to right, we show trends in the print/ electronic bundle and print only journal prices for annual subscription. The trends in electronic only prices were very similar to the left panel of the graph. We classify economic journals into four categories according to the ranking developed in Combes and Linnemer (2010) — the CL-index — which classifies journals into four categories AAA, AA, A, and B and lower. Category AAA (our "Top 5") includes the top five journals, and AA (our "Field +") includes 15 next-tier journals, four AEJ journals, *Quantitative Economics*, and *Theoretical Economics*. Category A (our "Lower") contains 82 journals, but seven of them are not on our list (*Journal of Business*, *Journal of the American Statistical Association*, *American Political Science Review*, *Water Sources Research*, *Journal of Financial Intermediation*, *Industrial and Labor Relations Reviews*, and *Journal of International Money and Finance*). All other journals are classified into our "Lowest."

Figure 2: Evolution of Publishers: Economics



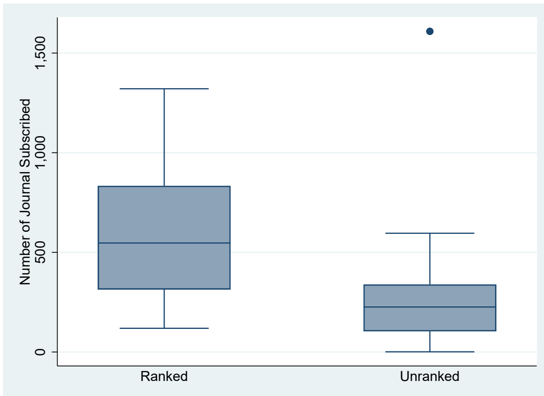
Notes: Elsevier and Wiley are the largest two publishers based the article market share, followed by the "Next Five:" Taylor & Francis, Springer, Oxford, Cambridge and the American Economic Association.

have no bearings on journal access and knowledge dissemination.

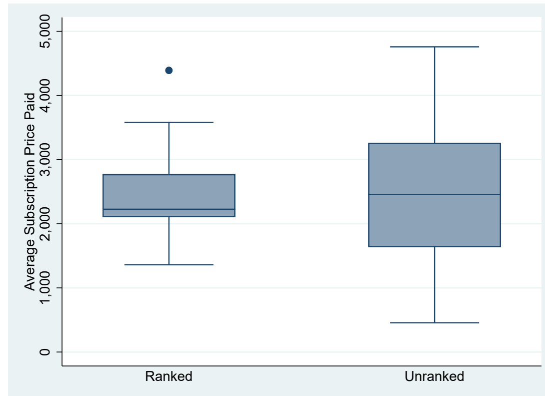
We look into the contracts between Elsevier and major public universities to investigate whether there is a basis for such claims. Figure 3 shows otherwise. We partition the universities in the contract data into two sets: those ranked by QS in 2020 (47 in total) and those not ranked (23 in total). For each contract, we tally the number of journals contained and the average price across subscribed journals. The left panel of Figure 3 shows that unranked universities subscribed to fewer than half of the journals (with the exception of California State University) that their ranked counterparts subscribe to. The right panel shows that although unranked universities paid only slightly more, on average, than ranked ones, the prices they paid had a much larger dispersion. This latter fact could be attributed to multiple causes: the unranked universities might have chosen journals of different quality levels; they might have paid a price premium for small-sized bundles; or they might have paid very different prices for similar-sized and similar-quality journals. Combining these facts, we find that researchers at unranked universities did have unequal footing when accessing academic journals — their libraries had substantially fewer journals and, perhaps, very different journals at different price points.

Looking into the individual journals in each contract, we find retail prices and contracted prices to be highly correlated. Table 3 reports the results of OLS regressions of contracted prices against retail prices, with different sets of fixed effects added in different columns.

Figure 3: What's in the Contract: Ranked vs. Unranked Universities



(a) Number of Journal Subscribed



(b) The Average Price of Journals Subscribed

Notes: Panel (a) is based on 70 Elsevier-University Library contracts. Outlier: California State University Libraries, though unranked in QS in 2020, subscribed to more than 1,500 journals. Panel (b) is based on 48 such contracts as not all contracts report prices for journals subscribed. Outlier: The University of Iowa paid more than $4,000 per journal, doubling the average journal price paid by the other ranked universities.

Table 3: Contracted Prices and Retail Prices

Dependent Variable	Journal Prices in Elsevier Contracts			
	(1)	(2)	(3)	(4)
Retail Price	0.906***	0.892***	0.891***	0.891***
for Annual Subscription	(0.023)	(0.017)	(0.017)	(0.008)
Library-year Fixed Effects	Yes	Yes	n.a	n.a
Field Fixed Effects	No	Yes	n.a	n.a
Library-field-year Fixed Effects	No	No	Yes	Yes
Standard errors clustering	Library	Library	Library	Field
Adjusted R^2	0.967	0.977	0.977	0.977
Observations	7,941	6,149	5,750	5,750

Notes: Each observation is a journal-library-year combination in an Elsevier contract. Standard errors are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Taking our trusted specifications in columns (3) and (4), which include library-field-year fixed effects,[21] we observe that contracted prices are highly correlated with retail prices: a \$1 increase in retail prices corresponded to a \$0.891 increase in contracted prices.

Beyond the scope of Elsevier contracts, we examine 88 contracts between university libraries and Wiley, another major publisher in numerous academic disciplines. The Wiley contracts in our collection lack specific names and contracted prices for individual journals, precluding an analysis comparable to Table 3. Nevertheless, Wiley contracts frequently detail specific terms that allow a university library to tailor the journal bundle on top of a baseline collection. For instance, in the 2014 agreement with Auburn University, Wiley clarified that "Member Institutions may add new journals to their Core Collection at the full rate institutional online-only subscription price." This indicates that the listing prices are utilized to determine the total price of a customized package. Across universities, these bundles demonstrate a significant range in the number of journals subscribed to and the prices paid per title. Within our sample of Wiley contracts, the average number of subscribed journals is 249, with a standard deviation of 283. In 2014, the University of Texas at Brownsville subscribed to 18 Wiley journals, the University of Kansas to 274, the University of Tennessee to 539, and Yale to 1,429. The prices paid per journal similarly exhibited considerable, right-skewed variation, with a median cost of \$1,400 per title, an average of \$6,100, and a standard deviation of \$8,700.

Based on the stylized facts derived from the contracts with Elsevier and Wiley, we contend that the journals' retail prices and the publishers' market dominance can be indicative of the extent of access barriers in the academic journal sector. These two factors influence both the actual subscription costs and the selection of journals by university libraries. Each library negotiates a highly customized contract with publishers, rather than selecting from a standard bundle of various sizes predetermined by the publishers. Typically, the negotiation involves selecting a base bundle and then adding individual journals. The retail prices of journals determine the price of the base bundle and the negotiated prices for additional journals. During these negotiations, the concentration of ownership among publishers tilts the battlefield toward large publishers. Large publishers have more negotiating power against university libraries in setting prices and configuring bundles. They can advocate for "big bundles" for their own journals, which strain library budgets, limiting their ability to purchase journals from other publishers (Edlin and Rubinfeld (2004)). Essentially, our perspective on the impact of retail pricing and publisher power on knowledge dissemination aligns with the argument presented by Reimers and Waldfogel (2022). This argument ex-

[21] Column (3) allows the error terms to cluster at the library level and column (4) clusters at the academic field level. The choice of different clustering affects the standard errors of estimated coefficients.

amines how public libraries' holdings of both physical and digital books depend on book prices of different formats, and how these holdings influence the circulation of books. In the following analyses, we use journals' retail prices and the publishers' market share as proxies for access barriers to published research.

4 An Empirical Framework to Measure the Effects of Journal Prices and Publishers' Market Power

4.1 Econometric Model

In the above discussion, we rationalize our choices of journal retail prices and publisher market power as measures of access restrictions in the academic journal market. Higher journal prices directly reduce journal subscriptions due to libraries' budget constraints; publishers' greater market power elevates price levels in the negotiation process, weakens competition from new entrants, and exacerbates the already-constrained choices of university libraries.

Combining these measures and the measures of knowledge dissemination and academic collaboration, we adopt the following empirical framework to estimate the impact of access barriers in the academic publishing industry:

$$\begin{aligned} \log(Y_{ajst}+1) = \beta_0 + \beta_1 \log(Price_{jt}) + \beta_2 PublisherShare_{st} + \beta_3 OpenAccess_{ajt} \\ + X_{jt}\beta_x + \lambda_j + \eta_t + YearTrends_{journaltier} + \epsilon_{ajst}. \end{aligned} \tag{1}$$

In Equation (1), we use subscript a for article, j for journal, s for publisher, and t for the year the article was published. Y_{ajst} is an outcome of article a published by journal j (owned by publisher s) in year t. In the baseline model, we use the number of citations, citing authors, and the nature of collaboration among citing authors within a certain time frame to measure different outcomes of knowledge dissemination. $\log(Price_{jt})$ is the natural logarithm of the price for printed- and electronic- bundled access. $PublisherShare_{st}$, measured from 0 to 1, is the publisher's market share of all published articles in the field in year t. $OpenAccess_{ajt}$ is an indicator function that equals 1 if the article has open-access status. X_{jt} is a vector of time-varying journal attributes that measures the journal's academic quality, including % Citable Items, Impact Factor, Immediacy Index, Eigenvector Score, Cited Half Life and Web of Science ranking.[22] We also include journal-level fixed effects λ_j and year-level fixed

[22]We omitted a few typical journal impact measures such as five-year Impact Factor and Article Influence Score because they are highly correlated with the journal attributes we include in the regressional analysis.

effects η_t, as well as linear year trends by journal ranking tiers.[23] Lastly, ϵ_{ajst} are *i.i.d* shocks to the outcome variable. Parameters of interest include β_1, β_2 and β_3, measuring the effects of prices, publisher market power, and open access to articles respectively.

4.2 Identification

In Equation (1), we use journal- and time- fixed effects to control for time-invariant journal attributes and time-varying trends in journals and research fields. We think this panel data identification strategy alleviates our concern of endogenous publisher power and articles' open-access status. Time-varying unobserved heterogeneity at the journal level is unlikely to correlate with publisher power given that many publishers own multiple journals of different attributes. Furthermore, the author team of an article is unlikely to respond to time-varying shocks at the journal level to make decisions on open access. Still, there could exist time-varying unobserved heterogeneity that affects the causal inference on the main dependent variables — journal prices. We can think of two mechanisms driving a correlation between prices and such time-varying unobserved heterogeneity. A publisher could set a journal's price based on the journal's constantly-evolving academic reputation, which is not fully captured by all the observed measures or by the fixed effects.[24] Both citation outcomes and journal prices respond to journal reputation, confounding the interpretation of estimates of Equation (1). At the same time, a journal editor selects papers to maximize the expected quality of the journal, as in Card and DellaVigna (2020). The editor observes many attributes of the journals (time-invariant and time-varying, including prices) and decides which papers to accept based on the paper's content, the referees' report and the information they observe. Both the publisher's price-setting process and the editor's paper-selection process could create a correlation between journal prices and the unobserved quality of the journal or the article in question.

To alleviate this endogeneity concern, we adopt an instrumental variable approach for the price variable, exploiting the fact that multi-product publishers engage in price competition with rivals based on the extent of product differentiation in the product space. In particular, we instrument journal prices with the one-year lag[25] of the average attributes of competing journals by rival publishers, interacted with the nonprofit status of the focal journal.

[23] We classify economic journals into four tiers according to the ranking developed in Combes and Linnemer (2010). Otherwise, we use journal ranking for categorization: for physics, tier 1 covers rank 1 to 5, tier 2 rank 6 to 20, tier 3 rank 21 to 40, and tier 4 rank 41 and above; for electronic engineering, tier 1 covers rank 1 to 10, tier 2 rank 11 to 50, tier 3 rank 51 to 100, and tier 4 rank 101 and above.

[24] For example, *Econometrica* and the *Quarterly Journal of Economics* publish articles in different styles and have different reputations, a fact that is well known in research communities, but this fact is not captured by the measures we include in X_{jt}, and this unobserved heterogeneity can change over time.

[25] Presumably, publishers adjust prices based on lagged information on competing journals.

This identification strategy is motivated by how publishers set prices in this particular industry. A notable feature of the academic publishing industry is the prevalence of nonprofit publishers, which typically set prices much lower than their for-profit counterparts. Despite that their responses to competition may differ based on their respective objective functions, both nonprofit and for-profit publishers must consider a journal's location among competing products in the product space. This identification strategy is, in principle, the same strategy proposed by Berry, Levinsohn, and Pakes (1995) (commonly known as the "BLP" instrument in the industrial organization literature). They propose that a firm sets a product's pricing based on the other products with which the focal product competes.[26] However, the focal product's unobserved heterogeneity is not correlated with the observed attributes of competing products.

Note that the typical "BLP" instruments include the average attributes of other journals in the same publishers' product portfolio. We deliberately drop this set of potential instruments because after 2000's a common practice in this industry is that a publisher negotiates a "big bundle" with university libraries. A journal may be included in the bundle because a library wants to subscribe to the flagship journal by a publisher (for example, *Nature* by Springer Nature) or to the higher-quality journals by this publisher. The "big bundle" practice gives rise to the possibility that product attributes of journals by the same publisher affect the access to the focal journal directly, which violates the exclusion restriction that is necessary for the validity of the instruments.[27] This concern is exacerbated when the library faces a budget constraint and has to decide on a cutoff quality threshold when negotiating with a publisher, as illustrated by Nevo, Rubinfeld, and McCabe (2005) and McCabe, Nevo, and Rubinfeld (2006).

The validity of our "adapted BLP" instrumental variables depends on two additional assumptions. First, an individual journal's time-varying unobserved attributes are uncorrelated with competing journals' attributes. This can happen if a publisher does not have control over the time-varying unobserved heterogeneity of a journal. Second, the editor of an individual journal accepts papers based on the focal journal's attributes, but the selection process does not depend on the attributes of other journals (that is, the editor acts like a local monopoly). We believe that these assumptions are reasonable in the academic publishing industry, providing a solid basis for our instrumental variable strategy.

In Appendix Table C.1, we report, field by field, estimates of regressions of $\log(Price_{jt})$ on

[26] Dubois et al. (2007) illustrates how journal characteristics, including impact factors, enter academic publishers' pricing decisions in France using this framework.

[27] This logic breaks down with a rival's bundle, because the attributes of a rival's bundled journals should affect the price of the focal publisher's bundle instead of the accessibility of an individual journal by the focal publisher.

publisher market shares in journal articles, own journal attributes, one-year lag of average attributes of same-field journals by rival publishers and their interactions with the focal publisher's nonprofit status). All regressions include journal- and year- fixed effects, as well as journal-tier year trends.[28] The first 12 variables are the instruments we constructed based on the above reasoning. The results show that many instruments enter the determination of journal prices, although the factors that play more prominent roles vary across fields.

5 Evidence from Economics

In this section, we report results from the field of economics. Economics is a major branch of the social sciences in which journal publication is the most important measure of academic output. Researchers in this field are well aware of the impediments and limitations associated with accessing published research.

5.1 Overall Impact on Research Dissemination

We start from the overall impact, measured by the natural logarithm of the number of citations and that of citing authors within five years of an article's publication year plus one. The number of citations is widely recognized as a measure of an article's impact; viewed from the receiving end, it is also a measure of knowledge dissemination as it is a function of how widely the focal article has been accessed by the research community. The number of citing authors is an alternative measure of this same dissemination process, focusing on the number of researchers exposed to the focal article.

Table 4 shows how different specifications affect our understanding of the effects of rising prices and publisher market power in this industry. Going from column 1 to 4, we add in year fixed effects, journal fixed effects and linear year trends by journal tiers, which could correlate with journal prices. When we get to column (3), the estimated coefficient for $\log(Prices)$ becomes negative but statistically insignificant. Adding linear year trends by journal tiers in column (4) does not change estimates much. The use of instrument variables, as discussed in Section 4.2, changes this estimate drastically, as reported by column (5).[29] The estimated coefficients for *Publisher Share* and *Article Open Access*, however, do not change much from column (3) to column (5), suggesting that these two variables are less susceptible to confounding factors in the unobservables of Equation (1).

Combining columns (5) and (6), we have two main findings: First, price increases nega-

[28] Note that Appendix Table C.2, instead of Table C.1, reports the first-stage results for Equation (1).

[29] Appendix Table C.2 reports the first-stage estimates and F-test statistics for this column.

Table 4: Knowledge Dissemination Affected by Journal Access

Dependent Variable	Follow-up Research within Five Years of Article Publication					
	log(# Citations)					log(# Citing Authors)
	OLS	OLS	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
log(Journal Prices)	0.298***	0.333***	-0.107	-0.110	-0.768**	-1.084**
	(0.094)	(0.107)	(0.090)	(0.091)	(0.299)	(0.440)
Publisher Share	-0.214	-0.102	-0.217**	-0.200*	-0.263**	-0.314**
	(0.358)	(0.315)	(0.108)	(0.112)	(0.121)	(0.154)
Article Open Access	0.474***	-0.451***	0.281***	0.281***	0.281***	0.339***
	(0.156)	(0.144)	(0.043)	(0.043)	(0.043)	(0.052)
Journal Attributes† :	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes	Yes	Yes	Yes
Journal Fixed Effects	No	No	Yes	Yes	Yes	Yes
Journal Tier Year Trends	No	No	No	Yes	Yes	Yes
Adj. R^2	0.232	0.235	0.475	0.475	n.a	n.a
# observations	92,772					

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. † : Time-varying journal attributes included in some specifications: % citable items, impact factor, immediacy index, eigenfactor score, cited half life, and journal rank. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

tively affect research dissemination with about unitary elasticity. If journal prices increase by 1%, the number of total citations within five years would decrease by 0.77%, and the number of total citing authors within five years by 1.08%. Second, the rise in publishers' market power negatively affects research dissemination. Column (5) of Table 4 shows that a 1% increase in a publisher's article market share leads to a 0.26% decrease in citations and a 0.31% decrease in citing authors.

As citations often take time to accumulate, we follow the citation outcomes over time in our observation window. Appendix Table D.1 presents estimation results for two, five, and eight years. These estimates reveal a clear and intuitive pattern. The citations and citing authors within two years of article publication drop by 0.44% and 0.71%, respectively, if journal prices increase by 1%. The drop amounts to 0.77% and 1.08% within five years of article publication. Eight years after articles are published, citation and citing authors drop only 0.30% and 0.40%, respectively. The change in estimated effects suggests that the full effects of access restriction are cumulative and, perhaps, maximized around five years after article publication. The effects of publishers' market power follow a similar pattern over time. The declining impact over an eight-year period suggests that researchers gain access to content beyond the publisher's paywall as time passes.

Finally, Table 4 illustrates that an journal article gaining open access significantly improves its citation metrics. Specifically, when an article is made open access, it can experience a 28.1% increase in citations and a 33.9% rise in the number of citing authors.[30] As the majority of previous studies study journal-level open access, our findings cannot be directly compared with theirs. Authors may choose open access for higher-quality articles, resulting in larger impacts, while a journal's overall open access status might expand its subscription and readership beyond what any single article can achieve through open access. Our results on economics articles are derived from across articles variations within the same journal as we incorporate journal fixed effects and year-specific trends within journal tiers to mitigate the issue of unobserved heterogeneity. Using the same approach (presented in Section 6.2), we estimate an article-level open access effect of approximately 10% in physics and 16% in electronic engineering, which aligns more closely with the findings of McCabe and Snyder (2014) and McCabe and Snyder (2021) based on science journal data.

5.2 Distributional Impact on Research Dissemination

As we show in Section 3.3, universities and institutions negotiate distinct packages for journal subscriptions, resulting in varying degrees of access restrictions for their affiliated researchers. These variances may have minimal significance when comparing the University of Arizona to Harvard, but for a non-R1 university,[31] access to journals could experience a drastic decline. When transitioning from developed to developing countries, a similar, if not more pronounced, decline can also be anticipated.

The estimates in Table 5 report the distributional impact on research — we take Equation (1) to each of the four tiers of universities and institutions and each of the two tiers of countries.[32] Across columns, we can see that the results are highly consistent with the results reported by Table 4. Columns (4) and (6) show, however, that price elasticities (with respect to the number of citing authors) are greater than unitary elasticities for citing authors from unranked institutions and developing countries. If journal prices were to increase by 1%, the number of citing authors from unranked universities and universities in developing countries would decrease by 1.00% and 1.22%, respectively. This result, especially the result about unranked institutions, is unlikely driven by the log-log specification in Equation (1) — if the number of authors from unranked institutions were smaller than that from other tiers, then we might have recovered elasticities from the steeper part of a log curve. However,

[30] Table 5 indicates that open access leads to 15.4%-29.1% more citing authors across institutions at varying ranks, with increases of 16.7% in developing countries and 35.5% in developed countries.

[31] In the U.S., there are 146 institutions that are classified as "R1: Doctoral Universities – Very high research activity" in the Carnegie Classification of Institutions of Higher Education as of the 2021 update.

[32] We lump transitional economies and developing countries into one category.

Table 5: Disparate Effects of Journal Access: IV results

Dependent Variable	log(# Citing Authors) within Five Years of Article Publication					
	From Academic Institutions Ranked from				From Countries	
	$1^{st} - 100^{th}$	$101^{st} - 500^{th}$	$501^{st}+$	Unranked	Developed	Developing
	(1)	(2)	(3)	(4)	(5)	(6)
log(Journal Prices)	-0.737***	-0.876***	-0.768***	-1.000**	-0.901***	-1.216**
	(0.262)	(0.334)	(0.250)	(0.438)	(0.344)	(0.490)
Publisher Share	-0.232**	-0.290***	-0.173**	-0.240*	-0.366**	-0.148
	(0.106)	(0.109)	(0.084)	(0.137)	(0.147)	(0.127)
Article Open Access	0.266***	0.266***	0.154***	0.291***	0.355***	0.167***
	(0.040)	(0.038)	(0.026)	(0.046)	(0.052)	(0.031)
# Observations	92,772					

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications use instrumental variables, and include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

the number of authors from unranked institutions accounts for almost half of the authors citing the published economics articles that have five-year citation outcomes; if anything, the reported elaticities from the log-log specification may have underestimated how high prices have disproportionately hurt researchers affiliated with unranked institutions.[33]

Table 5 also indicates that authors affiliated with highly ranked institutions experience negative impacts due to access barriers to published research. If journal prices were to rise by 1%, there would be a 0.74% decrease in the number of citing authors from universities ranked among the top 100 worldwide. As libraries at these institutions typically have ample resources and few restrictions in journal subscriptions, this negative effect is likely due to diminished collaboration between researchers at top-tier universities and those at lower-ranked schools. The latter group, facing restricted access, might be unable to collaborate effectively with the former, consequently reducing the research output from the higher-ranked group as well. The next subsection provides evidence substantiating this argument.

5.3 Impact on Research Collaboration

Varying degrees of exposure to cutting-edge literature in their respective fields may affect researchers' capacity to collaborate with one another, but it is difficult to know in which direction this effect goes *a priori*. Several forces could counteract each other. First, re-

[33]Another observations is that companies or research institutes are typically counted as "unranked," — the role of academical journals should be more important because researchers in these places lack unofficial channels such as seminars, conferences, and working papers to access knowledge.

Table 6: Research Collaboration across Individuals Affected by Journal Access

Dependent Variable	log(# Citations) within Five Years of Article Publication			
	with # Co-authors			
	= 1	≥ 2	≥ 3	≥ 4
	(1)	(2)	(3)	(4)
log(Journal Prices)	-0.149	-0.896***	-1.025***	-1.003***
	(0.118)	(0.339)	(0.390)	(0.373)
Publisher Share	-0.158*	-0.242**	-0.195*	-0.194**
	(0.085)	(0.119)	(0.107)	(0.091)
Article Open Access	0.200***	0.264***	0.225***	0.151***
	(0.030)	(0.041)	(0.035)	(0.026)
# Observations	92,772			

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications use instrumental variables, and include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table 7: Collaboration Across Schools and Countries Affected by Journal Access

Dependent Variable	log(# Citations) within Five Years of Article Publication					
	Co-authored					
	same inst.	across inst.	across inst. tiers	same country	across countries	across country tiers
	(1)	(2)	(3)	(4)	(5)	(6)
log(Journal Prices)	-0.649***	-0.608***	-0.732***	-0.791***	-0.553***	-0.546**
	(0.216)	(0.218)	(0.284)	(0.303)	(0.163)	(0.228)
Publisher Share	-0.149*	-0.159**	-0.198**	-0.208**	-0.138*	-0.122**
	(0.081)	(0.073)	(0.094)	(0.104)	(0.076)	(0.059)
Article Open Access	0.178***	0.162***	0.224***	0.222***	0.193***	0.107***
	(0.028)	(0.026)	(0.034)	(0.036)	(0.028)	(0.018)
# Observations	92,772					

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications use instrumental variables, and include year- and journal fixed effects as well as journal-tier linear year trends. Institution tiers are the four tiers defined in Table 5. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

searchers within the same institution may all face knowledge access limitations, leading to reduced collaborative research within institutes. Second, the access limitation may lead to incentives to seek collaborations from outside the institutions because external collaborators bring in new knowledge. Third, disparities in knowledge can complicate effective communication between institutions or across countries, reducing incentives for collaboration. These forces may take effect simultaneously; as a result, we cannot predict whether access barriers hinder or incentivize collaboration.

By examining all the citing authors for each citation of a published article, we can create collaboration metrics within the network of citing authors. We investigate the impacts of journal prices, publishers' market power, and open access on various measures of collaboration based on the specifications of Equation (1). Table 6 focuses on the number of coauthored articles, and Table 7 goes one step further to study the nature of these collaborations.

Table 6 reports that high prices have a statistically insignificant impact on the number of single-authored citations. The adverse high price effect, however, is apparent for co-authored citations and intensified when more coauthors are involved. Publisher power and open access have similar effects across single-authored and co-authored articles. Table 7 first investigates the effects on collaboration across tiers of institutions. The negative effects are consistent across the board, with the most pronounced effects on collaboration across institution tiers. For a 1% increase in journal prices, the number of citations from collaborating authors at institutions from different tiers drops by 0.73%, while the numbers are 0.65% and 0.61% for collaboration from different and the same institutions, respectively. Furthermore, the effect of publisher market power and open access are also the largest across institution tiers. We also estimate the effects on collaboration within the same countries, between different countries, and among varying levels of country development. In each instance, journal prices and publishers' market power exert significant hindrance effects on research collaborations. However, the effect on collaboration between countries with different development levels is comparatively less pronounced. This may be attributed to the fact that access barriers more significantly impede collaborations between institutions within the same country, where most collaborative efforts occur.

5.4 Robustness

We conduct a battery of robustness checks of our main findings and present some of the results in Appendix Table E.1. As measuring the outcomes and the access barriers is key to our results, in this table, we focus on the robustness checks when we use alternative dependent variables, prices and publisher market power. In column (1), as the outcome variable we use a binary variable indicating whether the focal article has any citation. This variable measures knowledge dissemination at the extensive margin. The absence of statistically significant findings indicates that access barriers primarily affect the intensive margin, meaning that they diminish the distribution of articles that garner positive citations. In the case of articles that receive no citations, there is nothing to diminish. In a sense, this result confirms that our primary results are not driven by spurious correlations in our data. In column (2), we count only unique authors in the number of citing authors measure. In columns (3) and (4), we use alternative journal prices — i.e., electronic-only or print-only prices — to replace bundle prices. In columns (5) and (6), we redefine publisher market share based on the number of journals or the number of journals weighted by the Journal Impact Factor, while in the benchmark analysis, we base the publisher market share variable on the number of published articles. Results from column (2) to (6) suggest that our qualitative findings are robust to all these changes.[34]

Throughout our specifications, we employed a "log-like" transformation on citation outcomes, represented by $\log(Y + 1)$, which behaves like $\log(Y)$ for large values of Y and remains defined at zero. Chen and Roth (2024) point out that this transformation might introduce arbitrary unit-dependence in result magnitudes if treatment affects the extensive margin. Fortunately, as demonstrated in column 1 of Appendix Table E.1, the extensive margin effect in our findings is minimal. Nonetheless, to address this concern, we adopt the recommendations of Chen and Roth (2024) by clearly defining the intensive and extensive margins. Appendix Table E.2 outlines results from this analysis. From columns (1) to (4), we equate the effect of moving from 0 to 1 citation (or citing authors) to a 100% increase in the outcome; from columns (5) to (6), we interpret this shift as a 200% increase. In comparison, the shift from 0 to 1 equals a 69% increase in the $\log(Y + 1)$ transformation. Both definitions acknowledge diminishing returns in treatment effects and provide straightforward interpretation of coefficients as approximate percentage (log point) effects, where an increase from 0 to 1 citation (or citing authors) equates to 100 or 200 log points. This table demonstrates that our results remain robust under various log transformation schemes applied to key specifications in Table 4; in fact, these alternative transformations slightly amplify the effects of access barriers like high prices and elevated publisher share, and their removal via open access articles.

[34] Appendix Table E.1 reports results for only a few specifications, but we perform robustness checks using the above alternative variables for all the regression tables reported so far. Our results are very qualitatively robust — we do not report these results in this paper only due to the constraint on paper length.

6 Mechanism, Generalizability, and Remedies

6.1 JSTOR Moving Wall: Access is Key

We argue in this paper that high prices and publisher power act as barriers to researchers' access to published knowledge. We now provide direct evidence to show that accessibility is a key mechanism contributing to a published article's citation outcome. Specifically, to show that gained access increases citation outcomes significantly, we use the timing of an economics journal entering the JSTOR distribution channel, which provides a low-price access alternative to individuals and institutions.

"JSTOR" stands for "Journal Storage." Founded in 1995, JSTOR is a nonprofit digital archive and online platform that provides access to a vast collection of academic journals, books, images and primary sources across many academic disciplines. JSTOR offers a subscription-based access model for institutions with prices much lower than those charged by the publishers. The subscription fees are based on the subscribing institution's size, type, and country. For a large U.S. university or four-year college, the range is typically \$3,000 to \$6,000 annually for a multi-discipline collection. Thirteen thousand schools, universities, and institutions around the world subscribe to JSTOR.

JSTOR typically does not provide immediate access to the most recent issues of journals. Instead, there is a delay, or a "moving wall," between the publication date and the availability of that content in JSTOR. The JSTOR website defines a moving wall as the gap of content between the archival (or past) and current (most recent) issues of a journal. The moving wall delay is set by a journal's publisher and ranges from zero to ten years, although the majority of journals in the JSTOR archive have a moving wall of three to five years. The availability of a published article on JSTOR after the journal's moving wall ends indicates an expanded accessibility to only researchers subscribing to JSTOR. This expansion in access can be considered a "treatment"(Rubin (1974); Card and Krueger (1994)), which is not correlated with an individual article's unobserved attributes or quality level.

The JSTOR moving wall setup provides a natural setting for us to evaluate the effects of accessibility on citation and collaboration. We use a difference-in-differences (DID) approach to estimate the treatment effects of an article gaining JSTOR access. Journals that are not collected by JSTOR constitute the control group, and journals collected by JSTOR enter the treatment at different timing. This setting approximates a staggered rollout of a policy in different locales. To estimate the treatment effect, we specify Equation (2) for all economics articles published in a specific year.

$$\log(Y_{ajn} + 1) = \xi_j + \mu_n + \sum_{\ell=0}^{L} \gamma_{1\ell} 1\{\ell = n - JSTOR_j\} + \epsilon_{ajn}. \quad (2)$$

In Equation (2), we follow article a, published by journal j, through years after its publication (indexed by n). We measure outcome Y_{ajn} using the increase of the number of citations/citing authors from year $n-1$ to year n. For all articles born in a certain year, the treatment is moving to JSTOR in a staggered fashion (depending on the length of the moving wall). An article published in journal j receives the JSTOR treatment at year $W_j + 1$ after the publication year, with W_j being the length of JSTOR moving wall — i.e., $JSTOR_j = W_j + 1$ for journal j.[35]

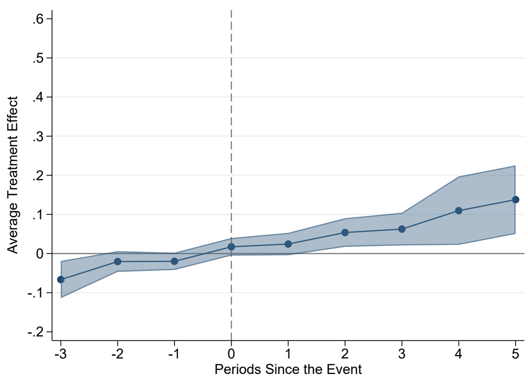
The treatment effect, captured by $\gamma_{1\ell}, \ell \geq 0$ can be heterogeneous for each treatment group. By design, it is different for every W_j. If W_j is 1, an article gains better exposure right after publication; if W_j is 5, an article has limited exposure in the first five years of publication. Moreover, journals with different moving walls may be inherently different and benefit from JSTOR access differently. Furthermore, the treatment effects can also vary over time, as past treatments affect the current outcomes — a paper gets exposure to the research community after the moving wall ends and gained citations invite more citations over time. We use methods developed by De Chaisemartin and d'Haultfoeuille (2020); De Chaisemartin and d'Haultfoeuille (2024) to estimate the heterogeneous, dynamic treatment effects.

This identification may fail, if the journal articles under JSTOR treatment experienced different trends from those who were in the control group. To assess this concern, we follow De Chaisemartin and d'Haultfoeuille (2024) to conduct placebo tests for $\ell = 1, 2, 3$ years before the treatment. We compare an article's from $JSTOR_j - (\ell + 1)$ to $JSTOR_j - \ell$ after its publication, where $JSTOR_j$ is the treatment time and ℓ is a pseudo timing that asks the question, "What if the treatment happens at ℓ periods before the actual treatment?" If the placebo tests returns a statistically significant result, it suggests that our results are due to different trends in the treatment and control groups, rather than the causal effect of the treatment. This placebo test basically tests the parallel trends assumption, which is essential to our identification strategy.

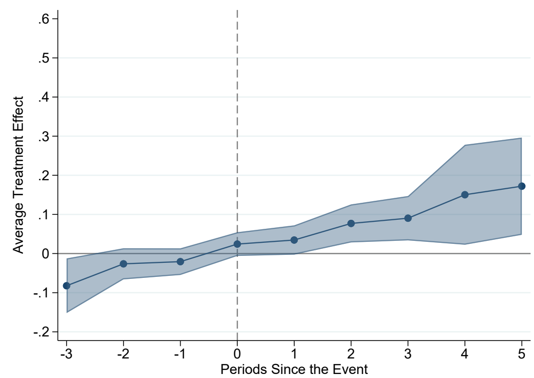
Figure 3 presents the average treatment effects (averaged across different treatment groups) of Equation 2. This figure reports both the dynamic treatment effects and the placebo test results. The distance to the first treatment changes is on the x-axis; the DID estimators are on the y-axis to the right of zero (including zero); and the placebo estimators are on the y-axis to the left of zero. As shown in Figure 3, after an article enters JSTOR circulation, it experiences a significant increase in citations and citing authors. This positive effect grows over time: for the 2009 cohort of articles, two years into the treatment, the

[35] JSTOR's website states: "The Wall resets, or moves forward, after a complete year, every year in early January. This is when another year of content is added into the JSTOR archive. The Moving Wall calculation does not include the current year. So in 2019 for a journal with a Moving Wall delay set to 5 years, archival content goes up to 2013. (5 years from the previously completed year, which would be 2018)."

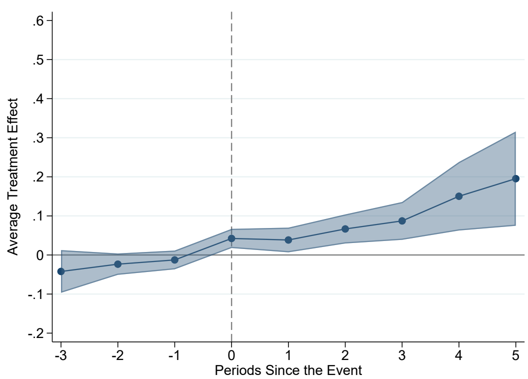
Figure 3: Dynamic Treatment Effects of JSTOR Circulation



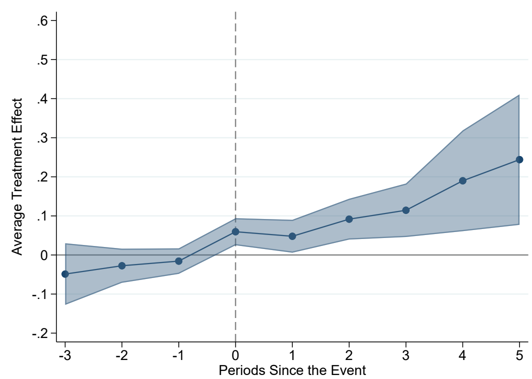
(a) log(Δ# Citations), 2009 Articles



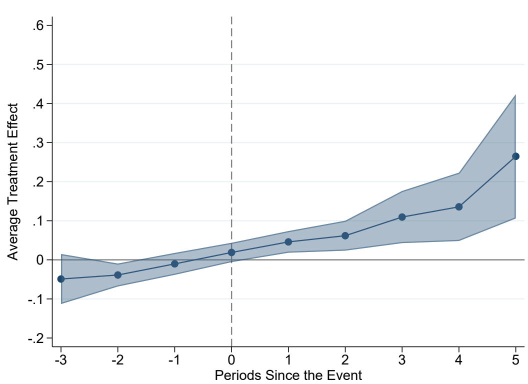
(b) log(Δ# Citing Authors), 2009 Articles



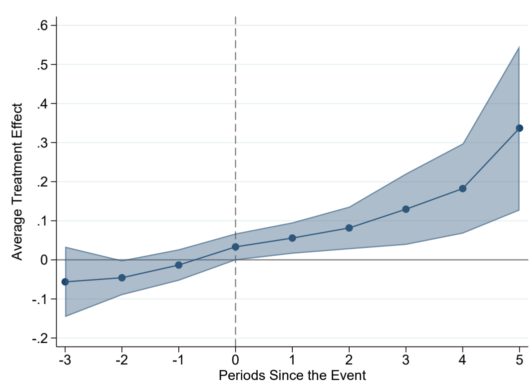
(c) log(Δ# Citations), 2010 Articles



(d) log(Δ# Citing Authors), 2010 Articles



(e) log(Δ# Citations), 2011 Articles



(f) log(Δ# Citing Authors), 2011 Articles

Notes: This figure presents the average of the heterogeneous, dynamic treatment effects of an article entering JSTOR circulation.

annual increase in citations that year were boosted by roughly 5%; five years into the treatment, 15%. Such patterns are similar for the number of citing authors, and consistent over three cohorts of journal articles. To the left of zero, the placebo estimators show confidence intervals containing zero; that is, we cannot reject the null hypothesis that the placebo test has zero effects. This is strong support for the "common trends" assumption of our DID identification strategy. Appendix Table E.3 presents results if we assume homogeneous treatment effects with no dynamics, using articles published in economics journals in 2009, 2010 and 2011. Columns (3) and (6) of this table report a 5.1% boost in the annual increase in citations and a 6.9% boost in the annual increase in citing authors, in the ballpark of the average of the dynamic effects of the treatment reported by Figure 3.[36]

McCabe and Snyder (2015) conduct a prior investigation on how access to JSTOR influences economics journal citations. They use panel data on citations to journal volumes, combined with manually gathered data on the number of institutions subscribing to various JSTOR package offerings. Their findings indicate that doubling JSTOR subscriptions, granting online access to a journal volume, could lead to an increase in citations by 6% to 17%. This effect varies based on the journal's cohort and various confounding variables. In contrast, we implement a different empirical approach that taps into alternative sources of variation and assesses the effect of JSTOR availability on the year-over-year citation growth of individual articles. While our results are not directly comparable to those of McCabe and Snyder (2015), they align within the same magnitude of their estimated range.

Appendix Table E.4 reveals that the boost in citations and citing authors due to JSTOR access becomes more pronounced as one descends the journal ranking hierarchy. This table presents findings under the assumption of homogeneous treatment effects absent dynamics, with similar results observed when considering dynamic heterogeneous treatment effects. For economics papers published in 2009, there is a 2.8% boost in the annual increase of citing authors from top-ranking institutions, with effects of 5.1%, 3.1%, and 7.2% noted for institutions within the subsequent three ranking tiers. This trend remains consistent for the cohorts of economics articles published in 2010 and 2011. The result corroborates our previous findings derived from university library contracts and pricing data, indicating that institutions most disadvantaged by access barriers would derive the greatest benefit from the reduction of such impediments.

[36] These estimated effects, however, cannot be directly compared with those from our journal pricing analysis. This is because the JSTOR analysis focuses on the annual increase in citations, whereas the pricing analysis results pertain to the total number of citations within a defined period.

6.2 Physics and EE: What Holds, What Breaks?

The production function for academic research exhibits significant variation across diverse academic fields. Key components influencing this function encompass research input, funding channels, collaboration formats, publication cycles, and researchers' alternative avenues for accessing journals. Importantly, the organization of the academic publishing market is influenced by traditions, norms, and idiosyncrasies. What serves as a barrier in one field may be inconsequential in another, and vice versa.

We expand our empirical analyses to the fields of physics and electronic engineering (EE) to explore the external validity of our baseline results. Table 8 reports results for physics, and Table 9 for EE. Table 8 shows that the negative effects of journal prices on citations and citing authors are much larger for physics than for economics. If physics journal prices increase by 1%, the number of total citations within five years would decrease by 1.92%, and the number of total citing authors within five years by 2.48% (in economics, the numbers are 0.77% and 1.08% respectively). However, publishers' market shares have no statistically significant impacts across columns. The effects of open access are not as prominent as in the field of economics, but the effects are still sizable. There are two key points to explain these discrepancies: first, the cost of journal subscriptions in physics is considerably higher, averaging $5,500 annually, compared to $1,000 in economics. Second, within the field of physics, the collective major publishers ranked within the "next five" posed strong competition to Elsevier. In contrast, within economics, Elsevier maintained a leading and more substantial dominance over its competitors.

EE is distinctive in its own way. The conference proceedings are the most important channel for knowledge dissemination, and *IEEE*, a nonprofit organization, holds a dominant position in both academic conferences and journals. Elsevier, the for-profit powerhouse in both economics and physics, is only a distant second. As Table 9 reports, prices have no statistically significant effect, but the negative effects of publisher market power is huge: a 1% increase in a publisher's article market share leads to a 4.67% decrease in citations and a 6% decrease in citing authors (in economics: 0.26% and 0.31%, respectively). In EE, the effects of article open access is similar to those in physics.

The findings indicate that elevated journal prices and publisher dominance manifest distinctively across fields. In physics, high prices had tangible consequences, whereas in electronic engineering (EE), it was primarily a matter of market power. Despite these variations, we can consistently identify high prices and publisher market dominance as common contributors to access barriers, recognizing that the specific culprit may vary across different academic domains.

Table 8: Physics — Prices have a Real Bite

Dependent Var.	Citation Outcomes within Five Years of Article Publication							
	log(# Citations)		log(# Citing Authors)		log(# Citing Authors)		log(# Citations w. Co-authors)	
					Unranked Inst.	Developing Countr.	≥ 2	across inst.
	OLS	IV	OLS	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Journal Prices)	-0.325**	-1.916***	-0.421*	-2.477***	-2.397***	-1.737**	-1.872***	-1.180***
	(0.157)	(0.504)	(0.218)	(0.720)	(0.762)	(0.716)	(0.487)	(0.364)
Publisher Share	0.089	-0.708	0.348	-0.682	0.030	-0.720	-0.734	-0.021
	(0.598)	(0.681)	(0.824)	(0.918)	(0.921)	(0.952)	(0.658)	(0.523)
Article Open Access	0.100**	0.099**	0.129**	0.127**	0.023	-0.165***	0.100**	0.029
	(0.050)	(0.049)	(0.061)	(0.060)	(0.062)	(0.056)	(0.049)	(0.038)
# Observations	106,595							

Notes: Each observation is a journal article published in a physics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table 9: Electronic Engineering — All about Market Power

Dependent Var.	Citation Outcomes within Five Years of Article Publication							
	log(# Citations)		log(# Citing Authors)		log(# Citing Authors)		log(# Citations w. Co-authors)	
					Unranked Inst.	Developing Countr.	≥ 2	across inst.
	OLS	IV	OLS	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Journal Prices)	0.174	-0.053	0.220	-0.026	-0.087	0.533	0.003	0.017
	(0.141)	(0.398)	(0.193)	(0.558)	(0.505)	(0.501)	(0.402)	(0.225)
Publisher Share	-4.450***	-4.671***	-5.777***	-6.015***	-5.286***	-5.024***	-4.730***	-1.998***
	(1.098)	(1.271)	(1.592)	(1.820)	(1.582)	(1.704)	(1.303)	(0.699)
Article Open Access	0.161***	0.161***	0.197***	0.196***	0.130***	-0.030	0.160***	0.088***
	(0.020)	(0.020)	(0.027)	(0.027)	(0.023)	(0.027)	(0.020)	(0.012)
# Observations	147,807							

Notes: Each observation is a journal article published in a physics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

6.3 What are Potential Remedies?

Pooling findings from the realms of economics, physics, and electronic engineering, we underscore the detrimental impact of access barriers on the dissemination of knowledge and collaborative research. This prompts the next inquiry: can this diagnosis guide us toward discovering effective remedies? In this subsection, we focus on two lessons learned from our investigation regarding potential remedies.

Nonprofit Publishers: Is it the Way Out?

Disseminating knowledge and advancing scientific discoveries are typically viewed as public goods. A common argument contends that the profit-driven incentives of for-profit publishers are fundamentally at odds with the public goods nature of academic publishing. In our data, nonprofit publishers indeed typically charge much lower prices than for-profit ones; Bergstrom et al. (2014) documents the same pattern. However, elevating the status of nonprofit publishers may not be a straightforward solution. Two pieces of evidence challenge this notion. First, insights from the field of EE reveal that a dominant nonprofit publisher (IEEE) does not mitigate publisher power as an access barrier; in fact, the adverse effects of publisher power are even more pronounced than in economics and physics, in which for-profit publishers hold market leadership. Second, our analysis finds no indication that the nonprofit status of a publisher alleviates the negative impacts of high prices. Appendix Table E.5 demonstrates that increasing prices of for-profit and nonprofit publishers led to qualitatively similar decreases in citations and citing authors. This pattern is consistent across all three fields under examination.

Two noteworthy pieces of positive evidence about non-profit publishers deserve attention. First, as seen in columns (3), (4), and (6) of Table E.5, a substantial market share for non-profit publishers correlates with increased citations from unranked institutions and developing countries, alongside more inter-institutional collaboration. Second, having nonprofit status is linked to a significantly greater positive impact of article open access on citations and citing authors. These findings contrast with the absence of similar effects on journal prices. Due to the mixed nature of our evidence and the lack of causal inference, we are hesitant to advocate for non-profit publishers.

Lower Prices to Achieve Open-access Effects

Consistently across all three fields, an article's open-access status significantly enhances citations and collaborative research. The importance of open access is widely recognized by academia, policy advocates, and policy makers. Recently, the Biden Administration took significant steps to promote open and equitable research. Among the initiatives proposed in 2022, a standout is the open-access mandate for federally funded research. By December

31, 2025, federal agencies will be required to make papers describing taxpayer-funded work freely accessible to the public as soon as the final peer-reviewed manuscript is published, and the underlying data must also be made available without delay.[37] In 2023, European Union member states are on the verge of adopting a call to make immediate open access the default, with no author fees.[38]

However, it is important to note that the open access mandate, while applauded by the research community, covers only a subset of academic research. Advocating open access for cases in which no public funding is involved can be challenging, as resistance is anticipated from publishers whose revenue models rely on paywalls. Complete open access may not be imminent, and a multi-channel approach that compels publishers to allocate a portion of their substantial profits to the public interest may serve as a more effective compromise. The question, then, becomes how much publishers need to reduce their prices to achieve the desired open-access effects.[39]

Our results contribute to evaluating these trade-offs. Using the findings in Table 4 and Table 8, we perform a back-of-the-envelope calculation. In economics, roughly, a 30% to 35% reduction in prices is necessary to achieve the effects of open access, while in physics, it is approximately 5%.[40] Interestingly, the dollar amount corresponding to this percentage decrease is about \$300 to \$400 for both economics and physics. Several observations can be made about these results. First, the impact of lowering prices varies across fields, making the effectiveness of price reduction field-specific. For example, reducing prices for electronic engineering journals may not yield significant benefits. Second, we demonstrate that accessibility has no effect after eight years of publication, suggesting that university libraries should negotiate more assertively for lower prices for older journals for archival purposes. Third, pricing remedies should be location-specific, as high prices disproportionately harm lower-tier institutions and developing countries. Lastly, persuading publishers to voluntarily reduce their prices may be challenging; actively curbing market power, such as blocking mergers, could be a more effective method to ultimately lower prices.

[37] *The White House.* (2022, August 25). Retrieved on December 20, 2024, from https://www.whitehouse.gov/ostp/news-updates/2022/08/25/ostp-issues-guidance-to-make-federally-funded-research-freely-available-without-delay/.

[38] *Council of the European Union.* (2023, May 23). Retrieved on December 20, 2024, from https://www.consilium.europa.eu/en/press/press-releases/2023/05/23/council-calls-for-transparent-equitable-and-open-access-to-scholarly-publications/.

[39] There are precedents of policymakers pressuring firms to lower prices; for instance, in 1995-96, Congressman Samuel Gjedenson of Connecticut and Senator Charles Schumer of New York successfully employed public policy strategies to compel firms to reduce prices of ready-to-eat breakfast cereals.

[40] Based on estimates reported in columns 5 and 6 of Table 4, $\widehat{\beta}_1 \cdot (\Delta p/p) = \widehat{\beta}_3$, then $\Delta p/p = -0.281/0.768 = -0.366$ for citations and $\Delta p/p = -0.339/1.084 = -0.313$ for citing authors. We perform the same calculation using Table 8's results for physics.

7 Concluding Remarks

This research sheds light on the dysfunctional nature of the academic publishing industry. Elevated prices and market concentration hinder access to published research, impeding researchers' capacity to build upon existing knowledge and collaborate effectively. Even though open-access archives, working paper collections, inter-library loans, and pirate platforms act as informal means to mitigate these access obstacles, the negative impacts remain considerable. Our conclusions are drawn from comprehensive data and rigorous identification methods across three distinct academic fields. Collectively, these findings provide a quantitative measure of the impact of issues within academic publishing, extending beyond a single industry and beyond the conventional outcomes typically examined by industrial organization economists.

These "non-market" consequences pertain to the sharing of knowledge and collaborative research. On an individual level, unhindered access to research is a prerequisite for a scholar's research productivity. Moreover, dismantling barriers to knowledge enhances the well-being of disadvantaged individuals and communities, fostering diversity and prosperity in knowledge creation and technological advancement. On an aggregate scale, knowledge dissemination is pivotal for generating new knowledge, with cumulative knowledge serving as a linchpin for technological advancement and economic growth. The effective dissemination of knowledge is influenced by societal institutions, structures, and policies (Furman and Stern (2011); Berkes and Nencka (2021)). Our work underscores one critical restriction and proposes potential solutions with the overarching aim of fostering creativity, growth, and development.

References

Albee, B. and B. Dingley (2001). Us periodicals prices-2001. *American Libraries 32*(5), 72–79.

Armstrong, M. (2015). Opening access to research. *The Economic Journal 125*(586), F1–F30.

Azoulay, P., C. Fons-Rosen, and J. S. Graff Zivin (2019). Does science advance one funeral at a time? *American Economic Review 109*(8), 2889–2920.

Bergstrom, T. C. (2001). Free labor for costly journals? *Journal of Economic Perspectives 15*(4), 183–198.

Bergstrom, T. C., P. N. Courant, R. P. McAfee, and M. A. Williams (2014). Evaluating big deal journal bundles. *Proceedings of the National Academy of Sciences 111*(26), 9425–9430.

Berkes, E. and P. Nencka (2021). Knowledge access: The effects of carnegie libraries on innovation. *Available at SSRN 3629299*.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Björk, B.-C. (2021). Why is access to the scholarly journal literature so expensive? *Portal: Libraries and the academy 21*(2), 177–192.

Card, D. and S. DellaVigna (2020). What do editors maximize? evidence from four economics journals. *Review of Economics and Statistics 102*(1), 195–217.

Card, D. and A. B. Krueger (1994). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review 84*(4), 772–793.

Chen, J. and J. Roth (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics 139*(2), 891–936.

Combes, P.-P. and L. Linnemer (2010). Inferring missing citations: A quantitative multi-criteria ranking of all journals in economics. *halshs-00520325*.

Crossick, G. (2007). Journals in the arts and humanities: their role in evaluation. *Serials 20*(3), 184–187.

De Chaisemartin, C. and X. d'Haultfoeuille (2024). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics, Forthcoming*.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

Dubois, P., A. Hernandez-Perez, and M. Ivaldi (2007). The market of academic journals: Evidence from data on french libraries. *Journal of the European Economic Association 5*(2-3), 390–399.

Edlin, A. S. and D. L. Rubinfeld (2004). Exclusion or efficient pricing-the big deal bundling of academic journals. *Antitrust LJ 72*, 119.

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

Furman, J. L. and S. Stern (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review 101*(5), 1933–63.

Griliches, Z. (1990). *Patent Statistics as Economic Indicators: A Survey, Part II.* National Bureau of Economic Research.

Iaria, A., C. Schwarz, and F. Waldinger (2018). Frontier knowledge and scientific production: Evidence from the collapse of international science. *The Quarterly Journal of Economics 133*(2), 927–991.

Kim, S.-J. and K. S. Park (2020). Influence of the top 10 journal publishers listed in journal citation reports based on six indicators. *Science Editing 7*(2), 142–148.

Larivière, V., S. Haustein, and P. Mongeon (2015). The oligopoly of academic publishers in the digital era. *PloS one 10*(6), e0127502.

McCabe, M. J. (2002). Journal pricing and mergers: a portfolio approach. *American Economic Review 92*(1), 259–269.

McCabe, M. J., A. Nevo, and D. L. Rubinfeld (2006). The pricing of academic journals.

McCabe, M. J. and C. M. Snyder (2005). Open access and academic journal quality. *American Economic Review Papers and Proceedings 95*(2), 453–458.

McCabe, M. J. and C. M. Snyder (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry 52*(4), 1284–1300.

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

McCabe, M. J. and C. M. Snyder (2021). Cite unseen: Theory and evidence on the effect of open access on cites to academic articles across the quality spectrum. *Managerial and Decision Economics 42*(8), 1960–1979.

Moser, P., J. Ohmstedt, and P. W. Rhode (2018). Patent citations—an analysis of quality differences and citing practices in hybrid corn. *Management Science 64*(4), 1926–1940.

Murray, F., P. Aghion, M. Dewatripont, J. Kolev, and S. Stern (2016). Of mice and academics: Examining the effect of openness on innovation. *American Economic Journal: Economic Policy 8*(1), 212–252.

Nevo, A., D. L. Rubinfeld, and M. McCabe (2005). Academic journal pricing and the demand of libraries. *American Economic Review 95*(2), 447–452.

Reimers, I. C. and J. Waldfogel (2022). The first sale doctrine and the digital challenge to public libraries. Technical report, National Bureau of Economic Research.

Rubin, D. (1974). Estimating causal effects to treatments in randomised and nonrandomised studies. *Journal of Educational Psychology 66*(5), 688–701.

Tennant, J. P., F. Waldner, D. C. Jacques, P. Masuzzo, L. B. Collister, and C. H. Hartgerink (2016). The academic, economic and societal impacts of open access: an evidence-based review. *F1000Research 5*.

Appendix — Not for Publication

This appendix contains details of our data and additional empirical analyses.

A Variable Definition and Construction

A.1 Sources of Price Data

We collected historical price data from the following publishers by either manually scraping data from various online sources or contacting the publishers via email or phone.

- Economics
  - nonprofit: American Economic Association (AEA), Cambridge University Press, Duke University Press, MIT Press, Oxford University Press, University of Wisconsin Press, and an anonymous publisher
  - For-profit: Elsevier, Sage, Springer, Taylor & Francis, and Wiley
- Physics (Atomic, Molecular & Chemical, and Condensed Matter)
  - nonprofit: American Chemistry Society (ACS), the American Institute of Physics Publishing (AIP), Institute of Electrical and Electronics Engineers (IEEE), the Institute of Physics Publishing (IOP)
  - For-profit: Elsevier, Sage, Springer, Taylor & Francis, Wiley, Nature,[41] and World Scientific
- Electronic Engineering
  - nonprofit: Institute of Electrical and Electronics Engineers (IEEE), Cambridge University Press, Institution of Engineering and Technology (IET), the Institute of Physics Publishing (IOP), Institute of Electronics, Information and Communication Engineers (IEICE)
  - For-profit: Elsevier, Sage, Springer, Taylor & Francis, Wiley, and World Scientific

A.2 Construction of Price Variables

We perform the following steps to ensure that prices from different publishers are consistently recorded:

[41] Nature was published by Nature Publishing Group (NGP), which was part of the Macmillan Science and Education Division. In 2015, Springer and the majority of Macmillan Science and Education merged to form Springer Nature.

- We deflated all prices to 2020 U.S. dollars using the Consumer Price Index for All Urban Consumers (CPI-U) from the U.S. Bureau of Labor Statistics. If a publisher lists prices in different currencies across regions, we use the prices for the U.S. regions for our specifications.[42]

- A publisher typically lists journal prices for print-only, electronic-only and print-and-electronic bundled access. We use the bundled prices for our baseline specifications and the other two prices for robustness checks.

- Publishers occasionally package a subset of their journals and set a price for each package instead of for the individual journals in the package. For example, Wiley sells *Papers In Regional Science* and *Regional Science Policy and Practice* in the *Regional Science* Package. The number of observations from economics journals sold in a package account for 3.9% of our total observations in this field. Lacking data on individual journal prices, we use the package price as the price for the individual journals in the package.

A.3 Description of Journal Attributes

In this appendix, we provide definitions for the journal attributes used in our analysis. We obtain all journal attributes from the Incites platform.

- **Citable items.** Citable items are those identified in the Web of Science as articles, reviews or proceedings papers, which are considered substantive articles that contribute to the body of scholarship in a particular research field and are most likely to be cited by other articles. Other forms of journal content, such as editorial materials, letters, and meetings abstracts, are not considered citable items. The number of citable items is the denominator for calculating various journal impact factors.

- **% Articles in Citable Items.** The % of Articles in Citable Items is the percentage of journal articles published in a year that count toward the total citable items. This measure emphasizes the amount of a journal's original research.

- **Journal Impact Factor.** The Journal Impact Factor is defined as all citations to the journal in the current Thomson Reuters' Journal Citation Reports (JCR) year to items published in the previous two years, divided by the total number of citable items published in the journal in the previous two years. It measures the average number of times journal articles published in the past two years have been cited in the JCR

[42] AEA and Wiley list region-specific retail prices; Elsevier list prices for the U.S and the rest of the world for a small subset of their journals. The U.S. prices are highly correlated with the regional prices. The U.S. prices, however, are typically lower than the price averaged across regions. For example, in economics and EE, respectively, 61% and 80% of U.S. prices are lower than the region-average prices, conditional on regional prices being listed.

year. The Five-year Journal Impact Factor is defined similarly, using journal articles published in the past five years instead of two years.

- **Immediacy Index.** The Immediacy Index is calculated by dividing the number of citations to articles published in a given year by the number of citable items published in that year. It measures how quickly journal articles are cited. The Immediacy Index can provide a useful perspective for comparing journals specializing in cutting-edge research.

- **Eigenfactor.** The Eigenfactor is based on the number of times journal articles published in the past five years have been cited in the JCR year. The Eigenfactor Score is essentially a ratio of the number of citations to the total number of citable items. Different from the Five-year Journal Impact Factor, the Eigenfactor Score excludes journal self-citations and weights citations from different journals. Journal weights are based on a stochastic measure of the amount of time researchers spend reading the journal. Eigenfactor scores are scaled so that the sum of the Eigenfactor scores of all journals listed in Thomson Reuters' Journal Citation Reports (JCR) is 100.

- **Cited Half-Life.** The Cited Half-Life is the median age of the citations received by a journal during the JCR year, with a citation's age defined as JCR year minus the publication year of the cited item. By definition, half of a journal's earned citations are to items published more recently than the Cited Half-Life, and the other half are to items published before the Cited Half-Life. For example, a 2015 Cited Half-Life value of 7.0 for Journal X means that half of the Journal X papers that were cited in 2015 were published in the last seven years. JCR caps Cited Half-Life at ten years — any journal scoring over ten years will display as > 10 in the product. A low Cited Half-Life suggests citation activity that peaks and drops off quickly; a high Cited Half-Life suggests citation activity that peaks and drops off more slowly.

- **Journal Rank.** The Journal Rank is based on the Five-year Journal Impact Factor. The lower the rank number is, the higher is the journal ranked.

B Summary Statistics for Physics and Electronic Engineering

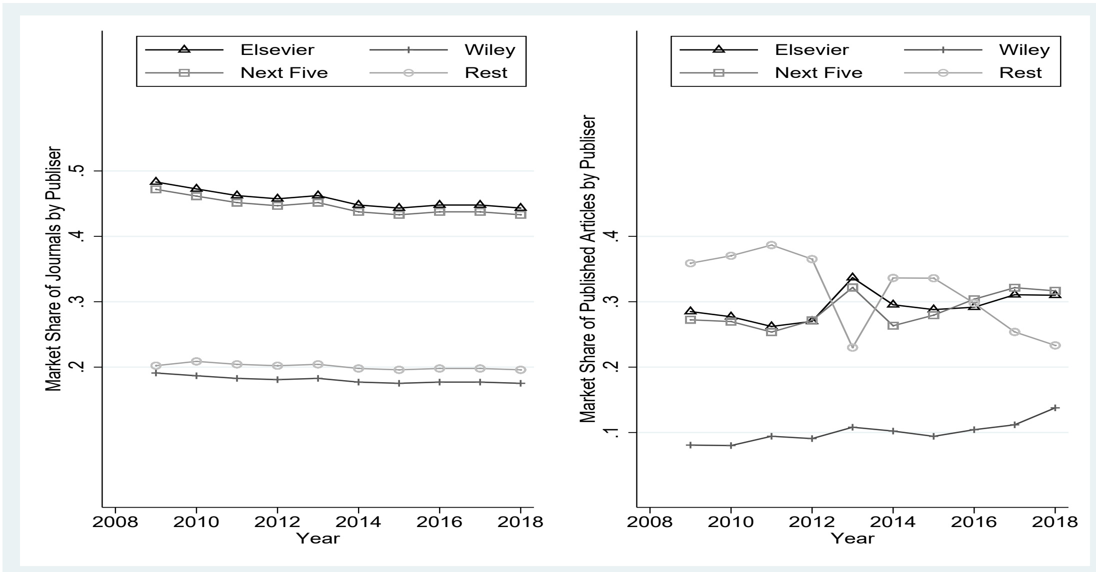
In this section, we present summary statistics for journals in physics and EE.

Table B.1: Articles, Journals and Publishers: Physics and EE

	Physics			Electronic Engineering		
	# obs.	Mean	SD	# obs.	Mean	SD
Panel A: Article Attributes						
Open access	389,421	0.145	0.352	495,403	0.157	0.364
# citations in 5 years	225,415	17.214	45.778	254,627	12.315	30.295
# citing authors in 5 years	225,415	85.047	255.662	254,627	48.216	121.424
Panel B: Journal Attributes						
Prices (in 2020 $)						
Bundle: print + electronic	571	5,482.583	4,358.165	1,222	1,869.516	1,789.644
Print only	403	5,416.068	4,062.633	1,115	1,366.546	1,096.427
Electronic only	619	5,257.631	4,067.102	1,399	1,557.736	1,540.765
# Article published	942	413.398	639.197	2,531	195.734	228.386
% Citable Items	900	87.405	30.703	2,396	98.678	6.881
Impact Factor	903	3.675	5.526	2,399	1.780	1.634
Immediacy Index	900	0.831	1.279	2,397	0.334	0.549
Eigenvector Score	908	0.036	0.083	2,404	0.009	0.013
Cited Half Life	903	7.721	4.187	2,277	6.308	2.502
Web of Science Rank	908	27.510	18.354	2,404	124.031	72.137
Panel C: Publisher Attributes						
nonprofit (= 1 if yes)	199	0.548	0.499	394	0.594	0.492
# Journals owned	199	6.075	9.894	404	6.545	18.265
# Article published	199	1,956.889	2,765.594	404	1,226.245	4,155.782
Market share in journals	199	0.065	0.105	404	0.026	0.072
Market share in articles	199	0.050	0.071	404	0.025	0.081

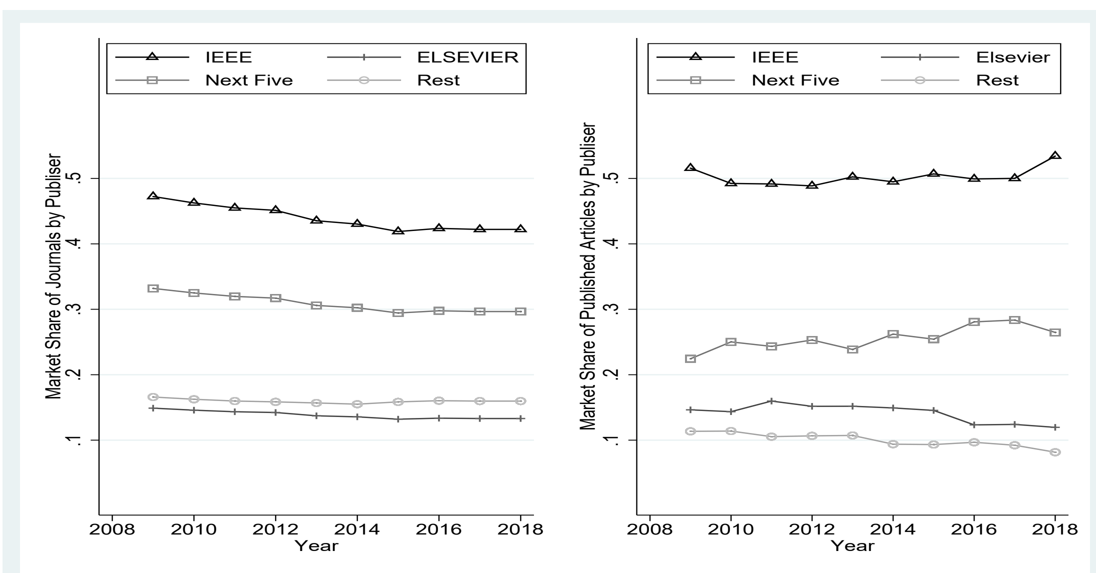
Notes: Each observation is a journal article published in a Physics or EE journal between 2009 and 2018.

Figure 1: Evolution of Publishers: Physics



Notes: Elsevier and Wiley are the two largest publishers based on article market share, followed by the "Next Five:" Springer, Taylor & Francis, American Chemical Society, Institute of Physics, and World Scientific Publishing Company.

Figure 2: Evolution of Publishers: Electronic Engineering



Notes: IEEE (Institute of Electrical and Electronics Engineers) and Elsevier are the two largest publishers based on article market share, followed by the "Next Five:" Springer, Wiley, Institution of Engineering Technology, Taylor & Francis and the Institute of Electronics, Information and Communication Engineers.

C Journal Pricing

In this section, we present two tables on journal pricing. Table C.1 reports, field by field, estimates of regressions of $\log(Price_{jt})$ on publisher market shares in journal articles, own journal attributes, one-year lag of average attributes of same-field journals by rival publishers and their interactions with the focal publisher's nonprofit status). All regressions include journal- and year- fixed effects, as well as journal-tier year trends. Table C.2 reports the first-stage regression results of the IV approach — the specifications are roughly the same as those for Table C.1, but the regression is at the article level. In both tables, the first 12 variables are the instruments we constructed based on our identification strategy.

Table C.1: Journal Pricing: Economics, Physics and EE

Dependent Variable	log(Journal Prices)		
	Economics (1)	Physics (2)	EE (3)
Attribute Average of Rival Publishers' Journals, Lag One Year			
% Citable Items	-0.024	0.062***	0.216***
	(0.026)	(0.021)	(0.036)
Immediacy Index	0.323	-0.460**	-0.998**
	(0.466)	(0.214)	(0.410)
Impact Factor	-0.014	-0.236**	-0.500***
	(0.291)	(0.100)	(0.172)
Eigenvector Score	151.880***	-19.422*	96.928
	(50.283)	(10.176)	(78.850)
Cited Half Life	-0.071	0.027	-0.324
	(0.061)	(0.052)	(0.230)
Web of Science Rank	0.017***	-0.033	(0.252)
	(0.003)	(0.025)	(0.009)
Above Variables, Interacted with Nonprofit (NP) Status of the Focal Publisher			
% Citable Items $\times NP$	0.005	-0.003	-0.151
	(0.003)	(0.041)	(0.120)
Immediacy Index $\times NP$	0.184	-0.156	1.189
	(0.509)	(0.263)	(1.002)
Impact $\times NP$	-0.196	0.023	-0.741*
	(0.143)	(0.133)	(0.378)
Eigenvector Score $\times NP$	-54.717	3.041	99.274
	(34.562)	(7.555)	(95.113)
Cited Half Life $\times NP$	0.008	0.014	-0.196
	(0.033)	(0.069)	(0.235)
Web of Science Rank $\times NP$	-3.45e-06	-0.071	-0.003
	(0.001)	(0.082)	(0.004)
Journal Attributes of the Focal Journal			
Publisher Article Share	-0.244***	0.398	0.512
	(0.057)	(0.508)	(0.509)
% Citable Items	0.0004	-0.0001	0.0002
	(0.0003)	(0.001)	(0.002)
Immediacy Index	0.003	-0.006	0.031**
	(0.006)	(0.012)	(0.013)
Impact Factor	0.007	-0.0004	0.001
	(0.008)	(0.005)	(0.009)
Eigenvector Score	-2.823	3.314***	2.845**
	(1.961)	(1.097)	(1.144)
Cited Half Life	-0.001	-0.005	-0.012*
	(0.002)	(0.005)	(0.006)
Web of Science Rank	-0.001**	-0.002	-0.0001
	(0.0002)	(0.002)	(0.0002)
# Observations	2,105	548	1,159
R-squared	0.992	0.985	0.992

Notes: Each observation is a journal-year combination. All specifications include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are clustered at the journal level and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table C.2: First-stage Regression Results: Economics, Physics and EE

Dependent Variable	log(Journal Prices)		
	Economics (1)	Physics (2)	EE (3)
Excluded Instrumental Variables, All Attribute Variables Lag One Year			
Rival Publisher % Citable Items	0.003	-0.013	0.222*
	(0.020)	(0.117)	(0.018)
Rival Publisher Immediacy Index	-0.186	0.192	0.180
	(0.942)	(0.251)	(3.868)
Rival Publisher Impact Factor	0.538	0.031	-0.726
	(0.425)	(0.120)	(0.512)
Rival Publisher Eigenvector Score	-92.870***	-25.945***	291.994**
	(35.668)	(7.187)	(128.354)
Rival Publisher Cited Half Life	-0.293**	-0.233**	-0.814***
	(0.116)	(0.096)	(0.302)
Rival Publisher Web of Science Rank	0.004	-0.019	-0.005***
	(0.003)	(0.020)	(0.005)
Rival Publisher % Citable Items $\times NP$	0.014	0.026***	-0.033
	(0.013)	(0.006)	(0.062)
Rival Publisher Immediacy Index $\times NP$	-0.568	-0.273	-1.014
	(0.546)	(0.350)	(3.599)
Rival Publisher Impact Factor $\times NP$	0.573	0.110**	1.180**
	(0.808)	(0.048)	(0.559)
Rival Publisher Eigenvector Score $\times NP$	-55.988	-28.196***	197.152
	(78.464)	(7.934)	(437.424)
Rival Publisher Cited Half Life $\times NP$	-0.109	-0.135	0.311
	(0.193)	(0.130)	(0.353)
Rival Publisher Web of Science Rank $\times NP$	-0.002	-0.016	0.003***
	(0.003)	(0.020)	(0.014)
Included Instrumental Variables			
Publisher Article Share	0.091	-0.616	-0.205
	(0.121)	(0.518)	(0.568)
Article Open Access	0.001**	-0.001	-0.001
	(0.00046)	(0.001)	(0.001)
% Citable Items	0.0002	-0.0003	-0.004*
	(0.0007)	(0.003)	(0.002)
Immediacy Index	0.002	0.064***	-0.003
	(0.012)	(0.021)	(0.028)
Impact Factor	-0.014	0.022***	0.002
	(0.009)	(0.006)	(0.008)
Eigenvector Score	-1.146*	1.307***	3.070***
	(1.912)	(0.467)	(0.615)
Cited Half Life	-0.001	-0.012	-0.0009
	(0.004)	(0.010)	(0.005)
Web of Science Rank	-0.0004*	0.002	-0.0003
	(0.0002)	(0.001)	(0.0004)
# Observations	92,772	106,595	147,807
R-squared	0.997	0.998	0.995
F-test of Excluded Instruments	13.56	16.90	12.74

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. All specifications include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table D.1: Citation Outcomes over Time

Dependent Variable	log(# Citations)			log(# Citing Authors)		
	2 years (1)	5 years (2)	8 years (3)	2 years (4)	5 years (5)	8 years (6)
log(Journal Prices)	-0.446*	-0.768**	-0.300**	-0.707*	-1.084**	-0.398**
	(0.271)	(0.299)	(0.139)	(0.422)	(0.440)	(0.183)
Publisher Share	-0.221**	-0.263**	-0.229	-0.279*	-0.314**	-0.273
	(0.110)	(0.121)	(0.147)	(0.160)	(0.154)	(0.178)
Article Open Access	0.195***	0.281***	0.298***	0.268***	0.339***	0.338***
	(0.018)	(0.043)	(0.070)	(0.020)	(0.052)	(0.087)
Observations	154,100	92,772	39,265	154,100	92,772	39,265

Notes: In columns (1) and (4), each observation is a journal article published in an economics journal between 2009 and 2017; in columns (2) and (5), each observation is a journal article published in an economics journal between 2009 and 2014; in columns (3) and (6), each observation is a journal article published in an economics journal between 2009 and 2011. All specifications are the same as in columns (4) and (5) in Table 4 — all specifications use instrumental variables, and include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

D Citation Outcomes over Time

Table D.1 presents IV results of citations and citing authors within two, five, and eight years after article publication for the field of economics. The specifications are the same as in columns (5) and (6) in Table 4, only with different time frames for the citation outcomes.

E Robustness Results

We provide five supplementary tables that are analyzed within the paper. The first two tables report the outcomes of robustness checks conducted using alternative measurements of key variables and different log transformation methods. The next two report the static difference-in-differences (DID) findings for the JSTOR analysis. The final table reports the impact of nonprofit status on citation metrics.

Table E.1: Results Robustness to Alt. Measurement of Key Variables)

Dependent Variable	Citation Outcomes within Five Years of Article Publication					
	Citations ≥ 1	log(Citing Authors) No Duplicate	log(# Citations)		log(# Citations)	
	(1)	(2)	(3)	(4)	(5)	(6)
log(Journal Price), Bundle	-0.029	-1.139**			-1.186***	-0.764**
	(0.059)	(0.447)			(0.450)	(0.355)
log(Journal Price), Electronic Only			-1.096***			
			(0.240)			
log(Journal Price), Print Only				-0.681***		
				(0.184)		
Publisher Share, Article	-0.037	-0.321**	-0.150	-0.258		
	(0.035)	(0.154)	(0.164)	(0.160)		
Publisher Share, Journal					-2.524**	
					(1.138)	
Publisher Share, Journal Weighted						-0.119*
						(0.071)
Article Open Access	0.042***	0.327***	0.353***	0.387***	0.282***	0.281***
	(0.009)	(0.050)	(0.054)	(0.074)	(0.043)	(0.043)
# Observations	92,772	92,772	94,620	75,257	92,772	92,772

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications use instrumental variables, and include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table E.2: Results Robustness to log Transformation

Dependent Variable	within Five Years of Article Publication					
	# Citations		# Citing Authors		# Citations	# Citing Authors
	OLS (1)	IV (2)	OLS (3)	IV (4)	IV (5)	IV (6)
log (Journal Prices)	-0.123	-0.920**	-0.186	-1.199**	-0.950**	-1.229**
	(0.115)	(0.382)	(0.141)	(0.507)	(0.430)	(0.556)
Publisher Share	-0.259*	-0.336**	-0.265	-0.363*	-0.373**	-0.400*
	(0.143)	(0.154)	(0.171)	(0.185)	(0.184)	(0.214)
Article Open Access	0.342***	0.343***	0.392***	0.394***	0.385***	0.435***
	(0.053)	(0.053)	(0.062)	(0.062)	(0.061)	(0.070)
Adjusted R^2	0.491	n.a.	0.479	n.a.	n.a.	n.a.
log Transformation	Outcome 0 to 1 equals 100% increase				Outcome 0 to 1 equals 200% increase	
# Observations	92,772					

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table E.3: Citation Gain after JSTOR Circulation

Dependent Variable	log(Δ # Citations)			log(Δ# Citing Authors)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Articles Published in 2009, # observations: 159,968						
JSTOR Treatment	0.338***	0.345***	0.051***	0.417**	0.413**	0.069***
	(0.126)	(0.133)	(0.016)	(0.172)	(0.183)	(0.023)
Adjusted R^2	0.019	0.020	0.338	0.015	0.015	0.336
Panel B: Articles Published in 2010, # observations: 162,344						
JSTOR Treatment	0.340***	0.341***	0.059***	0.422***	0.412**	0.075***
	(0.114)	(0.120)	(0.016)	(0.158)	(0.169)	(0.023)
Adjusted R^2	0.018	0.019	0.321	0.014	0.015	0.324
Panel C: Articles Published in 2011, # observations: 151,466						
JSTOR Treatment	0.378***	0.381***	0.052***	0.469**	0.463**	0.068***
	(0.132)	(0.139)	(0.016)	(0.185)	(0.196)	(0.024)
Adjusted R^2	0.020	0.020	0.337	0.014	0.015	0.340
Year Fixed Effects	No	Yes	Yes	No	Yes	Yes
Journal Fixed Effects	No	No	Yes	No	No	Yes

Notes: Each panel reports results for journal articles published in an economics journal in a different year, and each observation in a regression is a journal article published in the n^{th} year after publication. For column (1) to column (3), the dependent variable is the natural logarithm of the increase in the number of citations of the published articles from year $n-1$ to year n plus one; For column (4) to column (6), the dependent variable is the natural logarithm of the increase in the number of citing authors of the published articles from year $n-1$ to year n plus one. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table E.4: Citation Gain after JSTOR Circulation, Disparate Effects

Dependent Variable	log(Δ# Citing Authors) From Academic Institutions Ranked from			
	$1^{st}-100^{th}$	$101^{st}-500^{th}$	501^{st}+	Unranked
	(1)	(2)	(3)	(4)
Panel A: Articles Published in 2009, # observations: 159,968				
JSTOR Treatment	0.028**	0.051***	0.031*	0.072***
	(0.012)	(0.016)	(0.017)	(0.025)
Adjusted R^2	0.249	0.245	0.162	0.269
Panel B: Articles Published in 2010, # observations: 162,344				
JSTOR Treatment	0.038***	0.065***	0.033**	0.059**
	(0.012)	(0.017)	(0.016)	(0.024)
Adjusted R^2	0.240	0.236	0.160	0.265
Panel C: Articles Published in 2011, # observations: 151,466				
JSTOR Treatment	0.037***	0.055***	0.040**	0.063***
	(0.012)	(0.017)	(0.017)	(0.024)
Adjusted R^2	0.255	0.245	0.160	0.270

Notes: Each panel reports results for journal articles published in an economics journal in a different year, and each observation in a regression is a journal article published in the n^{th} year after publication. For column (1) to column (4), the dependent variable is the natural logarithm of the increase in the number of citing authors from differently ranked institutions of the published articles from year $n-1$ to year n plus one. All specifications include year and journal fixed effects. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Table E.5: Do nonprofit Publishers Alleviate the Journal Access Problem?

Dependent Variable	Citation Outcomes within Five Years of Article Publication					
	log(# Citations)	log(# Cit. Authors)	log(# Citing Authors) Unranked Inst.	log(# Citing Authors) Developing Countries	log(# Citations) w. Co-authors ≥ 2	log(# Citations) w. Co-authors across inst.
	(1)	(2)	(3)	(4)	(5)	(6)
log(Journal Prices)	-1.448***	-2.097***	-2.023***	-2.309***	-1.707***	-1.132***
	(0.334)	(0.426)	(0.377)	(0.437)	(0.337)	(0.228)
log(Journal Prices) × nonprofit	0.028	0.044	0.033	0.036	0.033	0.010
	(0.041)	(0.056)	(0.051)	(0.051)	(0.041)	(0.025)
Publisher Share	-0.254*	-0.299	-0.227	-0.135	-0.231	-0.154*
	(0.143)	(0.185)	(0.169)	(0.157)	(0.144)	(0.089)
Publisher Share × nonprofit	6.179	9.173	10.884*	11.627*	7.606	6.291*
	(5.400)	(7.284)	(6.069)	(6.335)	(5.380)	(3.226)
Article Open Access	0.250***	0.304***	0.257***	0.139***	0.235***	0.138***
	(0.042)	(0.052)	(0.046)	(0.028)	(0.040)	(0.024)
Article Open Access × nonprofit	0.231**	0.252*	0.246**	0.209**	0.212**	0.171**
	(0.112)	(0.130)	(0.115)	(0.085)	(0.105)	(0.074)
# Observations	92,772					

Notes: Each observation is a journal article published in an economics journal between 2009 and 2014. Time-varying journal attributes included in all specifications are the same as in Table 4. All specifications use instrumental variables, and include year- and journal fixed effects as well as journal-tier linear year trends. Standard errors are adjusted for clustering within journals and are provided in parentheses. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.