Did Online Access to Journals Change the Economics Literature?

Mark J. McCabe
School of Information, University of Michigan, 105 S. State Street, Ann Arbor, MI 48104; email: mccabe@umich.edu

Christopher M. Snyder
Department of Economics, Dartmouth College, 301 Rockefeller Hall, Hanover, NH 03755; email: chris.snyder@dartmouth.edu

January 2011

Abstract: Does online access boost citations to an economics article? The answer has implications for a range of issues from the sustainability of open-access journals (which shift journal fees from subscribers to authors) to the nature of economists' citing behavior, to the impact and diffusion of new digital technologies. Using detailed panel data on citations to the universe of articles published in 100 of the top economics and business journals, we exploit exogenous variation in the timing of the online availability of print content to isolate the causal effect of online access from spurious selection effects plaguing previous studies. Controlling for quality with increasingly rich fixed-effects specifications reduces the measured effect of online access from the high levels (as much as 500%) found in previous studies to near zero. This aggregate zero effect masks substantial heterogeneity across platforms: JSTOR availability has a significant impact, boosting citations an average of 10%, whereas no impact is observed for Elsevier's Science Direct platform. The citation boost is uniform across the distribution of articles—from rarely cited articles in the "long tail" to "superstars". We explore further sources of heterogeneity in the citing author's geographic location and institutional affiliation, and the rank of the cited journal.

JEL Codes: L17 (open source products and networks), O33 (technological change: diffusion processes), A11 (role of economists)

Acknowledgements: The authors are grateful for helpful comments from Ajay Agrawal, Margo Bargheer, Irene Bertschek, William Bowen, Erik Brynjolfsson, Liz Cascio, Yan Chen, Paul Courant, Glenn Ellison, Joshua Gans, Patrick Gaule, Avi Goldfarb, Dietmar Harhoff, Justus Haucap, Thomas Hess, Tobias Kretschmer, Ethan Lewis, Jeffrey Mackie-Mason, Karen Markey, Thierry Penard, Martin Peitz, Roger Schonfeld, Tim Simcoe, Michael Smith, Doug Staiger, Scott Stern, Don Waters, Michael Ward, Harriet Zuckerman, Christine Zulehner; from seminar participants at Clemson University, JSTOR, LMU Munich, Mellon Foundation, University of Erlangen-Nuremberg, University of Goettingen, University of Michigan, University of Toronto, Yale University, ZEW Mannheim; and conference participants at the Workshop on Economic Perspectives on Scholarly Communication in a Digital Age (Ann Arbor), ZEW Workshop on the Economics of Information and Communication Technologies (Mannheim), International Industrial Organization Conference (Savannah), and Society for Scholarly Publishing Annual Meeting (San Francisco). We thank Mark Bard, Jamie Bergeson-Bradshaw, Yilan Hu, Ella Kim, Scot Parsley, and Kyle Thomason for their excellent research assistance as well as the many journal publishers and third-party platforms for assistance in constructing journal online histories. Special thanks to Roger Schonfeld for facilitating contacts with content providers and platforms, for his assistance in data collection, and for offering his insights on developments in scholarly communication. This research was generously supported by a grant from the Andrew W. Mellon Foundation.

1. Introduction

Could an economist quadruple his or her citation count just by publishing in an online journal rather than one available only in print? Our initial interest in this question was prompted by the huge effects of online access (full text of a volume's articles available via Internet subscription) and open access (full text of a volume's articles available via the Internet at no charge) found in previous empirical studies. For example, Lawrence (2001) studied a sample of nearly 1,500 conference proceedings in computer science that each published some articles online and the rest only in print. In the average proceedings, online articles received 336% more cites than print. Harnad and Brody (2004) studied the citation rates of published physics articles, some of which were also self-archived by the author on arXiv (a large, online repository offering free downloads of scientific manuscripts). Self-archived articles averaged 298% more cites than the others. Eysenbach (2006) studied the policy of the *Proceedings of the National Academy of Sciences* policy of providing open access to articles if the author pays a fee of $1,000. The citation rate for these articles was about 50% higher than their non-open-access counterparts.[1]

It would not be surprising if providing convenient online access to the full text of an article, reducing the fee for full-text access, or doing both would boost an article's citations.[2, 3] Enhanced access may expedite search, allowing citing authors to identify additional relevant articles, and may lower the cost of acquiring, reading, and ultimately citing the articles so identified. Yet the almost implausibly large results for the citation effect in these previous studies suggests they are biased upward. A likely source of this bias is that the effect of online or open access is confounded with article quality, which is unobservable to the econometrician and so is an omitted variable. For example, in Lawrence (2001), there is no mention that a random procedure was used to select articles for open-access publication. If instead of a random procedure, the best articles were published open access, the 336% effect on citations could just be picking up the difference in the citation rates of

[1] Other studies include Curti et al. (2001), a cross-sectional comparison of the cites to online versus print-only medical journals. They find an average citation boost of 54% over the period 1995-2000. Related to Eysenbach (2006), Walker (2004) studied an oceanography journal that allowed authors to buy open access for their articles, finding 280% more downloads for open-access articles than others. See Craig et al. (2007) for a survey of research on the citation boost from open access.

[2] The use of electronic means to access article information preceded the Internet. Digital bibliographic data became available for libraries in the 1970s, facilitating the searching of and access to academic articles (see Lancaster and Neway 1982). In the mid 1990s, popular Internet browsers such as Mosaic and Netscape Navigator allowed the literature searches previously conducted in libraries on a fee basis to be conveniently performed on a personal computer for free (Tenopir and Neufang 1995). Around the same time (1995 according to our data), academic journals began providing Internet access to some articles, allowing scholars instant access to the full text of these articles rather than having to visit the library or to wait for a print copy to arrive via a document-delivery service.

[3] In a related context, Agrawal and Goldfarb (2008) use the scholarly engineering literature to examine how the Internet influenced research collaborations among engineering faculty.

leading articles versus others. Similarly, the decision by authors to self-archive or to pay for open access may be plausibly correlated with article quality rather than random. Thus the large effects in Harnad and Brody (2004) and Eysenbach (2006) and the other papers cited in footnote 1 may be partly or even completely spurious.

In this paper we take on the econometric challenge of separately identifying the effect of online access from unobserved quality. However, we are not merely interested in providing a clever solution to an econometric puzzle. Understanding the market for academic journals is important to scholars because it is the one market in which they function as both producers and consumers.[4] Citations are the currency in this market, the prevailing indicator of the impact of scholars' research, advancing a scholar's prestige as well as salary.[5] If a small change in the convenience of access can cause a quadrupling of citations, then the typical citation may be of marginal value, used to pad the reference section of citing articles rather than providing an essential foundation for subsequent research. According to this view, citations would be at best a devalued currency, subject to manipulation through the choice of publication outlet. On the other hand, the finding of little or no citation boost would resuscitate the view of citations as a valuable currency and as a useful indicator of an article's contribution to knowledge.

The question of the citation boost provided by online and open access has real policy implications. Scholars and librarians have continued to debate the relative merits of the traditional versus the open-access model of journal pricing. The traditional model involves relatively low author fees but high reader fees (in the form of library subscriptions); the open-access model inverts this by allowing readers free access to articles over the Internet, making up for the revenue loss by increasing author fees. A recent theoretical literature (McCabe and Snyder 2005, 2007, 2010; Jeon and Rochet 2010) uses a two-sided-market model[6] to assess which model will come to dominate in equilibrium and which will generate the most social

[4] See Bergstrom (2001) and Dewatripont et al. (2006) for evaluations of the market for academic journals.

[5] We adopt the traditional approach in evaluating this "currency," i.e. the value of marginal citations is independent of their source. For a discussion of some recently developed alternative metrics see "New Measures of Scholarly Impact" published in *Inside Higher Ed* (Dec. 17, 2010), and available at www.insidehighered.com.

[6] In a so-called two-sided market, a platform intermediates transactions between the two sides, tailoring the price charged to each side to ensure sufficient participation on both sides. Participation on one side can benefit the other side, as the presence of many receivers in a telecommunication network exert a positive externality on a caller or the presence of many readers of an academic journal can exert a positive externality on an author. The two sides may not be able to pass fees or externalities through to each other through direct transfers, so the platform prices on the two sides has real economic consequences. See Armstrong (2006) and Rochet and Tirole (2006) for surveys of the theoretical literature on two-sided markets.

surplus. The answer to both questions hinges on the elasticities of demand on the author and reader sides. How much more an author would be willing to pay for better access by readers to his or her article depends on how this access translates into readership and citations. If online access quadruples citations, author demand is likely to be quite inelastic, enough to support the high author fees necessary for open access to be sustainable in long-run equilibrium and enough that this open-access equilibrium have desirable efficiency properties.[7] On the other hand, if the citation benefit is low, author demand may be so elastic that open access is unsustainable in equilibrium and/or socially inefficient.

The results also have implications for understanding the transformative value of new technology on scholarly communication. As will be seen, JSTOR will emerge from our analysis as the most important platform for online access to the economics literature in the citation period we study (1995-2005). Based on anecdotal evidence, many economists believe that JSTOR substantially enhanced research productivity just as EconLit in an earlier period and Google Scholar more recently. We provide the first systematic evidence of the impact of JSTOR on the economic literature. Facilitating scientific communication may have broader social welfare implications to the extent that better communication enhances research productivity, which in turn enhances overall economic productivity, as discussed in, e.g., Dosi (1998) and Freeman (1994).

We address the econometric challenge of separating the citation effect of online access from unobserved quality effects by assembling a large panel dataset, described in Section 2, on the citations indexed by Thomson ISI in the 1980-2005 period to all the articles published since 1956 in a sample of the top 100 economics and business journals. We merge in hand-collected data on the date that each journal volume was made available on the Internet, and if so, via which platform or platforms (i.e., the journal's own website, JSTOR, ProQuest, or several other major Internet platforms). The panel nature of the dataset allows us to control for unobserved quality effects by including fixed effects for journal volumes.[8] The considerable exogenous variation in the date of online access across journals allows us to account for secular trends in citations to various vintages of content in economics and business. Additional exogenous variation in the date of online access across volumes of the same journal allows us to account for the age profile of a volume's cites in a flexible way (allowing the age profile to differ not only across journals but across blocks of volumes for a given journal). It is vital to control for these secular trends and age profiles; otherwise they are easily confounded with the online indicators, which

[7] Author fees can be substantial: currently, the Public Library of Science (PLoS) charges an author fee of $2,900 for *PLoS Biology*, the highest-ranked journal in the ISI biology category.

[8] Although we have article-level data, because the date of online access does not vary across articles within a volume in our data, our analysis is conducted at the level of the journal volume. Hence the most refined set of fixed effects for content is are journal-volume fixed effects.

tend to "turn on" in later years and for certain ages of content (only after an embargo window, for example). As discussed in the literature review at the end of this introduction, this form of misspecification plagues several of the more recent articles that attempt to correct for the bias due to unobserved quality using panel data.

Because of the importance of the econometric specification for consistent estimates, in Section 3 we construct a model of the behavior of a representative citing author and use it to help inform the appropriate econometric specification, discussed in Section 4. Section 5 presents the results. The first set of results show that the same huge effects of online access found in the previous literature can be generated if fixed effects capturing the quality level of journal volumes are omitted. Once appropriate fixed effects are included, however, the aggregate result cannot be distinguished from zero. Thus much of the estimated effect of online or open access from the previous literature can be attributed to bias due to omitted quality. The second set of results is devoted to showing that absence of an estimated effect at the aggregate level masks substantial heterogeneity across platforms. While we find no effect for Elsevier's ScienceDirect platform among others, we find a positive and significant effect for JSTOR, providing a roughly 10% boost to citations on average.

We investigate other sources of heterogeneity in the online-access effect, for example whether the effect differs depending on the number of platforms offering online access, whether the effect differs across high- versus low-ranked journals, and whether the effect differs for citing authors in different ranked institutions or different countries. The regional analysis will allow us to address the policy question of whether facilitating access has a disproportionate effect for authors in developing countries where library resources may be limited.

Section 6 extends our investigation of heterogeneity to the article level by examining whether different articles within a journal volume benefit more or less from online access. Are the effects of online access concentrated among the most cited articles—the "superstars"—or the least cited ones—the "long tail"? Previous studies of online retailing suggest that the latter outcome is predominant: long-tail effects have been found in other markets including books (Brynjolfsson, Hu and Smith 2003), clothing (Brynjolfsson, Hu, and Simester 2007) and video sales (Elberse and Oberholzer-Gee 2008). To date, only one paper examines these issues in the context of scholarly communication. Evans (2008) reports that online access reduces the number of cited articles and increases the citation concentration of those articles that are cited, suggesting a superstar effect. As we explain later, there a number of reasons to doubt the robustness of his results.

The concluding section (Section 7) summarizes the results and draws out the implications of the results for the questions raised in this introduction. The appendix contains several tables providing further detail on our sample.

Several recent papers attempt to address the bias due to omitted article quality in estimating the effect on citations of online or open access. The closest to our approach are two articles in *Science*, Evans (2008) and Evans and Reimer (2009). These papers use the same basic approach as we do to controlling for quality by using panel data on citations to individual volumes and including volume fixed effects in their econometric model. Unfortunately their econometric model suffers from a different misspecification problem: the omission of time effects which should be included to account for secular trends in citations. In the absence of such time effects, recent secular increases in citations for certain journals might be picked up by an online or open-access indicator, which generally are turned on in later citation years, leading to an upward bias in the citation effect. We demonstrate the point concretely in Table 2, where we reproduce a similar estimate to the 26% for economics and business in Evans and Reimer (2009), but then show this estimate disappears in a subsequent regression when appropriate time effects are added.[9, 10]

Davis et al. (2008) conducted an experiment in which articles from 11 American Physiological Society journals were randomly selected to be openly accessible immediately upon publication, the rest receiving the usual treatment of restricted/fee online access for the first year and open access afterwards. Within-journal comparisons revealed no differences in citations or in the percentage of articles for the two types of access after one year (nor after three years, as shown in the follow-up study, Davis 2010). The randomized design eliminates any bias due to unobservable quality, so the finding of no effect suggests that the large results from the previous literature which did not control for quality are almost entirely spurious. Our finding of no aggregate effect is consistent with these experimental findings (though for the case of online access rather than open access). Our finding of heterogeneous effects, positive for JSTOR but not for other platforms, suggests a reason to doubt the generalizability of experimental findings for an isolated platform. JSTOR may provide a citation boost because of its desirable properties: it is well known among academics, it includes a large number of journals, and it includes all past articles

[9] Evans (2008) and Evans and Reimer (2009) have a number of other differences from our paper. They study a broader set of disciplines than economics and business and a larger set of journals within economics and business. This forces them to rely on a electronic database, Fulltext Sources Online, for information on online histories for journal in their sample, whereas we use hand-collected and cross-checked data. Our analysis of the Fulltext data suggested it is a useful tool to understand broad trends in online access, but that there are drawbacks to its use as a regression variable: it omits data for the earliest years of online access (1995-98), contains inaccuracies in online-access dates, and omits important access channels, e.g. JSTOR.

[10] A different approach using panel data was taken by, among others, Parker, Bauer, and Sullenger (2003) and De Groote, Shultz, and Doranski (2006). These articles examine citing behavior of authors at a single institution over time (Yale in the former paper, University of Illinois Chicago medical school in the latter paper), examining whether the change from print to online access for a journal at that institution caused the authors to increase cites to that journal. The evidence was mixed. The first paper found significantly positive effects at Yale. The second paper found no difference from journals that were print subscriptions throughout the study period at the University of Illinois medical school. A problem with the approach is that different journals may have different secular trends in citations, and these secular trends may be correlated with the access status of the journal. Addressing this problem would require including a set of time effects for journals, which cannot be estimated without data from multiple institutions.

for all listed journals. This is a more significant platform for readers than one which may not be well known outside of a subdiscipline, only makes a small number of journals available, and for these offers better access for only a scattered sample of articles for just one additional year. One might expect little citation boost from such a limited platform.

Contemporaneous work by Gaule and Maystre (2009) examines the effect of open access on citations to articles in the *Proceedings of the National Academy of Sciences* as did Eysenbach (2006), but attempts to control for the endogeneity involved in the author's paying $1,000 for open access by using instruments such as whether the article was published in the last fiscal quarter for the author's affiliated institution (under the presumption that research spending is less elastic then because of "use it or lose it" policies). Instrumenting in this way causes the open access effect to fall by 80% and become statistically insignificant. Gaule and Maystre's results provide yet more evidence that the large citation effects from earlier cross-sectional studies were largely spurious. As with Davis et al. (2008), the paper's finding of an insignificant access effect may be difficult to generalize because of its focus on a single online platform. Another difference with our paper is that we study the effect of online versus print access, whereas they study the effect of open versus fee access for a journal which is already online.[11]

2. Data

Our analysis is based on a selection of 100 journals in economics and business.[12] We focus on these disciplines because, as economists, we are interested in our own discipline's journal market and understand its institutional details better than other disciplines'. In addition, economics and business are arguably "representative" academic disciplines, between "soft" disciplines such as literature and history and "hard" disciplines such as chemistry and physics.[13] We restricted the sample to 100 journals because of the considerable expense and effort involved for each additional journal. Because each journal has

[11] Gargouri et al. (2010) try to identify the causal effect of self-archiving on citations (studied earlier by Harnad and Brody 2004) by examining institutions that require employees to post their publications in an open-access repository. Scholars operating under this mandate are exogenously more likely to self-archive than scholars at other institutions. Gargouri et al. (2010) conclude from their results that selection effects do not bias estimates of the effect of article access, contrasting our finding of huge selection effects. Unfortunately, they use the mandate variable as a regressor rather than as an excluded instrument, impairing the interpretation of their results.

[12] We followed the subject categorization of our data vendor, Thomson ISI. We included journals categorized as business along with economics because economists actively publish in many of these business journals (consider, for example, the *Journal of Finance* or *Management Science*). In addition, broadening the discipline allows us to enlarge the sample of journals with JSTOR coverage.

[13] We also collected a sample of 100 journals from each of history and biomedicine, using a similar selection procedure as with economics and business. Analysis of these other disciplines, not reported here for space considerations, suggests that economics and business is a suitable discipline on which to focus. Journal articles in history generate remarkably few subsequent citations in journals, possibly reflecting the fact that history is more of "book" than a "journal" discipline, leading to noisy and uninformative results. The results from biomedicine are similar in many respects to economics and business, providing some confidence in the latter's representativeness.

many volumes (35 on average in our sample) and because different volumes of the same journal can experience different patterns of online access, we will have many more "experiments" than the 100 journals would imply.

Appendix Table A1 lists the journals in our sample. The selection procedure was designed to achieve two goals. First, we wanted to focus on mainstream journals, among other reasons because they have generated enough cites to provide adequate variation in our left-hand-side variable. Second, we wanted to ensure that journals available on JSTOR were represented, based on our *a priori* belief that JSTOR provides a good source of exogenous variation in the timing of online access and in view of the availability of JSTOR subscription information by journal. The sample in fact includes all of the journals in economics and business that had at least some content posted on JSTOR by 2005 (30 in economics and 18 in business). The remaining journals were selected by first ranking them by the standardized ISI yearly impact factors averaged over the period 1985-2004 and then selecting the number of top-ranked journals in each subdiscipline so that the ratio of economics to business journals is the same among JSTOR as among non-JSTOR journals.[14] Overall, the sample includes 63 journals in economics and 37 in business.

Three different forms of data, described next in sequence, were merged together in the final dataset: (a) citations data, (b) online-availability histories, and (c) data on subscriptions to online channels. The citations data was acquired from Thomson ISI. For each of the 100 journals in our sample, ISI lists every article published since 1956. Each published article is linked to all cites from all of the over 8,000 ISI-indexed journals for each year from 1980 to 2005. The database includes detailed information on journal and article title, publication date, author name, affiliation, and location for both the citing article and the cited article.

To this basic citation data we merged hand-collected information on online availability of the full-text article. We first identified the major third-party aggregators which, in addition to the journal publisher's own website, may have been a channel of online access.[15] The major aggregators considered are JSTOR, EBSCO, ProQuest, Ingenta, Gale, OCLC, and DigiZeitschriften. Then we sought to determine the date on which each journal issue was made available online, if at all, through each channel. This was a painstaking process because information is only readily available regarding a journal issue's

[14] There was little conflict between the selection of journals based on rank versus the selection on JSTOR availability because JSTOR tends to include top-ranked journals. Only two JSTOR journals, the *Canadian Journal of Economics* and the *Journal of Risk and Insurance*, would not have been selected based on rank alone. Neither was ranked very far below the cutoff for inclusion in our sample, the former ranked 80 and the later 89 among economics journals.

[15] In addition to our own knowledge of the market, we used a number of sources to identify major third-party aggregators including electronic journal catalogs for a number of universities and consultations with market experts, one of whom worked for several of the major aggregators.

current online availability while our study requires historical information on the first date of availability and this at the issue (or at least the volume) level. With many journals and channels there was no systematic pattern to when the sequence of issues were placed online. To obtain this information, we contacted the publishers and aggregators and checked the information we received against independent resources including libraries' electronic journal catalogs and the "wayback machine" (www.archive.org), which provides regularly archived snapshots of large segments of the Web.[16]

The resulting dataset from these two sources includes observations for nearly 260,000 individual cited articles, indexed by i. The analysis is ultimately performed at a more aggregate level—the volume—comprising all of the articles a journal publishes in a given year. Aggregating in this way reduces the computational burden (the average volume contains 73 articles) without changing the results. The volume-level estimates are numerically identical to the article-level ones because none of the right-hand side variables vary at the article level within a volume. Let v index a volume, $j(v)$ index the journal title associated with the volume, and $p(v)$ index the year of the volume's publication. Our dataset has a panel structure because each volume receives cites each year over our sample period, from 1980 to 2005. Let t index the citation year. Note the distinction between the dataset's two time indexes: $p(v)$ indexes the year the *cited* article was published (from 1956 to 2005), and t indexes the year the *citing* article was published (from 1980 to 2005). Because each journal has many volumes, our sample of 100 journals yields over 3,500 volume observations; because each volume can have as many as 26 citation-year observations (one for each year 1980-2005), our panel yields over 60,000 volume-citation-year observations, the basic unit of analysis for our study.

Table 1 provides descriptive statistics for the dataset. All journals were founded by 1988, the earliest, the *Journal of Institutional and Theoretical Economics*, in 1844. Because some journals were founded after 1956, the average publication year for the over 3,500 volumes at (1985.7) is a few years later than the midpoint of the range of feasible publication years (1956-2005). Because some volumes were published after 1980, the average citation year for the over 60,000 volume-citation-year observations (1994.8) is also a few years later than the midpoint of the range of feasible citation years (1980-2005). The number of citations to a volume in a single year (CIT) is 35.7 on average, or about a half a cite per article. The standard deviation (59.4) is huge, as is the range, from 0 to a maximum of 771 (cites in 2004 to the 1982 volume of *Econometrica*). The heterogeneity in number of citations which was evident across volumes also is evident within a given volume. Only 21% of articles for the average volume are cited during the sample period. The standard deviation of this measure (0.21) is relatively

[16] Several studies (Evans 2008, Evans and Reimer 2009) have used the electronic database of dates of online availability provided by Fulltext Sources Online. The discrepancies and omissions compared to our hand-collected data led us to rely only on the latter.

large, and extreme values are observed (e.g., none of the articles in the 1958 volume of the *Review of Economics and Statistics* were cited in 1991; all of the articles in the 1997 *Quarterly Journal of Economics* were cited in 2003).

Figures 1 and 2 illustrate patterns in citations which are interesting in their own right but which will also be important to account for later in our estimation procedure. Figure 1 plots the profile of citations over the lifespan of the average journal volume. Technically, the figure plots the coefficients on a complete set of age indicators from a fixed-effects Poisson regression including fixed effects for journals and citation years.[17,18] There are 50 separate age indicators ranging from 0 (the year of publication) to 49 years (the oldest volumes in our sample were published in 1956 and cited in 2005, so 49 years old). Citations peak in the fifth year after publication, receiving 216% more than in the year of publication. After that, citations gradually fall each year, falling below the citations received in the year of publication beyond age 30. For the oldest volumes, the age profile asymptotes to about a 75% reduction in cites. In other words, a volume which received 40 cites in the year of publication would settle down to about 10 cites per year after age 45. Of course, this effect is an average across articles that are forgotten and ones that continue to be standards in the literature. The 95% confidence interval shows that the estimates start out very precise early on in the life cycle but become noisier with age.

Figure 2 plots secular trends in citations. The underlying regression is the same as that behind Figure 1, but here we are plotting the coefficients on the fixed citation-year effects rather than the age effects. Citations follow a steady upward trend over time, reaching a level by the end of the sample that is 120% higher than in the base year of 1980. This increase in citations is due to a number of factors including the increase in the number of citing journals indexed by ISI (reflecting an increase in the rate of indexing of existing journals and the entry of journals), an increase in articles per journal, and an increase in the number of references per citing article.

The last row of Table 1 provides information on the online-access indicator, OAC. For 26% of the observations, the full volume was available online through some channel for the full year. We will focus on full online access defined in this way throughout the analysis. The regressions will also include indicators for partial online access (only part of a volume's content available online during the year or all

[17] A more formal discussion of the estimation procedure, due to Wooldridge (1999), is postponed to Section 4.

[18] The impossibility of separately identifying age, cohort, and time effects, called the "identification problem" (Blalock 1966), appears in a variety of economic applications. See Heckman and Vytlacil (2001) for an application to education and McKenzie (2006) to international macroeconomics. The identification problem here is that age, volume, and citation-year effects cannot all be separately identified. The age profile is identified in the regressions behind Figure 1 by including journal rather than volume fixed effects. In essence, the identifying assumption is that volumes in the same journal have roughly similar citation levels after accounting for time effects. The citation-year profile in Figure 2 is identified using a similar strategy. As discussed in Section 4, the identification problem will be less of a concern in our later, primary regressions because the online-access variables of interest are identified after controlling for age, volume, and citation-year fixed effects even though the fixed effects cannot themselves be separately identified.

of its content available for only part of the year), but we will not focus on those results because partial access is a catch-all category combining observations with various degrees of online access. Figure 3 shows how the full-online-access indicator evolves over time in our sample. Full-text articles were only available online starting in 1995. There is a fairly consistent rise in the proportion of volumes that are online throughout the citation period except for the big jumps in 2002 and 2003. By the end of the citation period, online access was nearly universal in the sample, with 88% of volumes available online. The figure shows that there is considerable variation between volumes in their online availability for many years between 1995 and 2005 and suggests that there is considerable variation across time in a given volume's online status. This variation will help in identifying online-access effects from secular trends and age effects.

3. Model

We model the number of citations received by article i by first characterizing the behavior of a representative citing author with respect to the article and then aggregating over citing authors. Let $n = 1, \ldots, N_t$ index citing authors, where N_t is the total population who publish an article in year t. For simplicity, assume each citing author publishes one article (multiple articles per author or authors per article merely complicate the indexing). Let I_t be the population of economics articles that can be potentially cited in year t, so that the article index runs over the range $i = 1, \ldots, I_t$. A cite by author n to article i results from a three stage process: identifying the article and its value to one's project through search, acquiring a full-text copy of the article, and extracting the information that will make up the cited passage. To simplify we will collapse the last two stages into one by assuming that all acquired articles are cited; a more realistic approach with uncertainty about citation after acquisition would yield qualitatively similar results.

3.1. Search Stage

To characterize the search stage, let s_n be author n's expenditure of effort in searching through the economics literature. This search yields a probability $\pi_{ni} \in [0,1]$ that the author finds article i. Conditional on finding the article, the author learns its match quality for him or her, Q_{ni}. Match quality is an amalgam of a "vertical" dimension (a dimension of quality over which all authors agree that "more is better" such as rigor of the theoretical analysis, precision of the estimates, or novelty of the results) and a "horizontal" dimension (a dimension over which different citing authors have different opinions, such as subfield or topic).

The probability that author n finds article i has the following form:

$$\pi_{ni} = \pi(s_n, M_t, Q_{ni}, OAC_{it} l_{ni} c_{it}). \quad (1)$$

Assume π is increasing in its first argument, s_n, implying that an article is more likely to be found the harder the citing author searches the literature. The second argument, M_t, is a vector of overall "market" conditions, including the size of the pool of citable articles I_t (since π is the probability of finding a given "tree" in a "forest" of articles, it is plausibly declining in the size of the "forest"), the technology available for searching through these other articles, and so forth. Assume π is weakly increasing in its third argument, Q_{ni}, implying that the citing author can more easily find articles with good match quality. At one extreme, an author writing on a topic may be well acquainted with the handful of seminal articles on that topic even without having to search at all, in which case $\pi_{ni} = 1$ even if $s_n = 0$ for these. At the other extreme, it may be quite unlikely that an author will come across a given obscure article outside of the immediately relevant area even with extensive search. The last argument, which is the product of three factors, captures the contribution of online access to search. The first factor, OAC_{it}, is the indicator for online access to the channel under consideration (or a vector of indicators if all channels are being considered jointly); the second factor, l_{ni}, is an indicator for whether author n's library subscribes to the channel or channels providing online access; and the third, c_{it}, is a vector of characteristics of the channel or channels. The attributes of the online channel will affect the ease of search: browsing is easier the more journals a channel covers and the more complete set of volumes for each journal; browsing would have limited value via a channel covering one journal and a smattering of volumes for that journal.

3.2. Acquisition Stage

In the second stage, citing author n decides of which identified articles to acquire full-text copies, comparing the benefit of acquisition, Q_{ni}, to the cost, a_{ni}. While both benefit and cost are known to n at the time the decision is made in the second stage, in the first stage a_{ni} is uncertain, a continuous random variable on $[0, \infty)$ with distribution function $F(a_{ni}, OAC_{it} l_{ni} c_{it})$. The second argument of these functions allows the distribution of acquisition cost to depend on the same factors affecting online access as did the search process in the first stage. Assume in the second stage that better online availability (in the sense that the indicators OAC_{it} and l_{ni} are turned on or the characteristics c_{it} of the online channel are more favorable) will be assumed reduces the acquisition cost in the sense of first-order stochastic dominance. Author n acquires article i (and automatically cites it according to our simplifying assumptions) if $Q_{ni} > a_{ni}$ and not if the reverse inequality holds. Putting the probability of this condition together with the probability from the first stage that search identifies article i yields the following overall probability of citation from author n to article i:

$$\pi_{ni} F(Q_{ni}, OAC_{it} l_{ni} c_{it}). \tag{2}$$

By equation (1), the leading factor, π_{ni}, in (2) is a function of the endogenous search effort, s_n. Author n chooses s_n to maximize the benefit minus the cost of search:

$$\int_0^{I_t} \pi(s_n, M_t, Q_{ni}, OAC_{it} l_{ni} c_{it}) \left[\int_0^{Q_{ni}} (Q_{ni} - a) dF(a, OAC_{it} l_{ni} c_{it}) \right] di - s_n. \tag{3}$$

The outer integral in the first (benefit) term in (3) integrates over the population of citable articles at time t. Each article is taken to be infinitesimal in the sense that the characteristics of any one do not materially affect the author's search decision. The inner integral is the expected return of acquiring an identified journal which passes author n's benefit-cost test.

The optimal search effort $s_n^*(M_t)$ can be written as a function of market conditions alone because the characteristics (Q_{ni}, OAC_{it}, l_{ni}, c_{it}) of individual, infinitesimal articles integrate out. Only the distribution of these characteristics across articles (so the general online accessibility of articles, the general depth of author n's library's subscriptions, the attributes of online channels carrying the average article) matters; and this distribution can be considered a component of M_t. Substituting $s_n^*(M_t)$ into the expression for π_{ni} in equation (1) and then this value of π_{ni} into (2) gives a single author n's probability of citing article i. Summing over authors $n = 1, \ldots, N_t$ yields the following expression for total citations to article i at time t:

$$CIT_{it} = C\left(M_t, \{Q_{ni}\}_{n=1}^{N_t}, OAC_{it} \xi\left(\{l_{ni}\}_{n=1}^{N_t}, c_{it}\right)\right). \tag{4}$$

According to equation (4), the number of citations is a function of market factors M_t, the distribution of the quality of the match with article i in population of citing authors $\{Q_{ni}\}_{n=1}^{N_t}$, and the nature of online access. The nature of online access is given by the product of the online-access indicator OAC_{it} and a function ξ of the distribution across the population of citing authors of library subscriptions to the online channel or channels $\{l_{ni}\}_{n=1}^{N_t}$ and the characteristics of the online channel or channels c_{it}.

3.3. Comparative Statics

Equation (4) has straightforward comparative-statics properties. The effect of market forces M_t will depend on which component is being considered. An increase in the population of citing authors will tend to increase cites; an increase in the population of citable articles will tend to reduce cites to a given

article by diluting search effort; a change in professional norms toward including a larger literature review and bibliography in each article will tend to increase cites. An upward shift in the distribution of match qualities $\{Q_{ni}\}_{n=1}^{N_t}$ will generate more successful searches and induce more acquisition effort for identified articles, both effects tending to increase cites. Turning to the last argument of C, we expect online access to increase cites by making search and acquisition effort more productive for a given article. The online-access effect is mediated through several other layers—the citer's library and the channels to which this library subscribes—and the nature of these other layers will contribute to the magnitude of the online-access effect. The more libraries that subscribe to online channels and the better the attributes of the online channels, the more online access will boost cites. However, it is theoretically possible that online access could end up reducing cites if the publisher uses the occasion to raise prices enough to cause a large drop in library subscriptions.

Interactions between OAC_{it} and the first two arguments of C may lead to heterogeneous online-access effects. The effect of online access may gradually increase as citing authors grow familiar with the use of the Internet in their literature search. The online-access effect may vary with the quality of the article. Seminal articles may be so valuable that they are identified and acquired even if the procedure is fairly costly, so that online access would have little effect for them, corresponding to a negative cross partial between the second and third arguments of C. Conversely, low-quality articles may be of so little value that they garner few cites regardless of how efficient the search and acquisition process is, in which case online access may provide little citation boost for them, corresponding to a positive cross partial between the second and third arguments of C.

4. Empirical Methodology

Although the theoretical analysis in the previous section was conducted at the article level, the econometric analysis will use volume-level observations because, as mentioned, the estimates are numerically identical but involve less computation. Before diving further into the technical details, we will provide a broader overview of the implications of equation (4) for the empirical methodology.

4.1. Overview

Consistent estimation of the online-access effect requires controlling for the first two arguments of C. We control for market factors M_t by including a set of fixed effects for interactions between citation and publication years. This is an important control to include because otherwise the strong secular trends such as observed in Figure 2 might be confounded with online availability, which often occurs later in the

sample when secular trends are also highest.

We control for the second argument of C, the distribution of match qualities $\{Q_{ni}\}_{n=1}^{I_t}$, in two ways. Variation in match quality as a function of article age—the typical hump-shaped pattern shown in Figure 1—is captured by including a flexibly specified age profile. This is an important control to include in order, for example, to avoid confounding the natural peak in citations at age 5 with online access that might have started in that year. Time-invariant, "vertical" quality is picked up with volume fixed effects. The inclusion of volume fixed effects helps remove a source of bias that plagued many of the previous studies. These studies mainly made the cross-sectional comparison of whether articles available online received more cites than others. But higher quality articles may be more likely to be available online, in which case the online-access variable may be picking up quality differences between online and print-only articles. In terms of equation (4) the vector of match qualities across authors, $\{Q_{ni}\}_{n=1}^{N_t}$, may be correlated with OAC_{it}; the omission of variables measuring match quality may bias the coefficient on OAC_{it} upward. After controlling for age and volume effects, any residual dimensions of match quality become part of the error term. A crucial issue for the consistency of the results of interest is whether the online-access indicator is exogenous in the sense of being orthogonal to this error; we argue for this exogeneity below in Section 4.5.

4.2. Panel Count Data

To account for the count-data nature of citations in our panel-data setting, we estimate equation (4) using a fixed effects Poisson estimator with the following conditional mean:

$$E(CIT_{vt}|\alpha_v, x_{vt}) = \exp(\alpha_v + x_{vt}\beta), \tag{5}$$

where CIT_{vt} denotes citations to volume v in year t, α_v is a volume fixed effect, x_{vt} is a vector of regressors, and β is a vector of parameters to be estimated. As mentioned, including the volume fixed effects α_v controls for time-invariant aspects of volume quality that is a potential source of bias in earlier studies.

Wooldridge (1999) provides a Poisson quasi-maximum-likelihood (PQML) estimator which, as long as the conditional mean is specified correctly, produces consistent estimates of β under quite general conditions: (a) the conditional distribution of CIT_{vt} need not be Poisson, negative binomial, or any other specific distribution but can be any general positive distribution; (b) the function on the right-hand side of equation (5) need not be exponential but can be any increasing positive function.[19]

[19] We used Simcoe's (2007) implementation of the estimator in Stata, which computes robust standard errors clustered at the fixed-effect level, suggested by Wooldridge. The estimator is now available in updates to Stata 11.

4.3. Heterogeneity in Online Access

The most important regressor x_{vt} is the variable of interest, the online-access indicator OAC_{vt}, equaling 1 if volume v was available online in citation year t. We focus on the results for full online access, that is, availability of the entire volume's content for the entire year, but also include controls for partial online access. Our discussion of equation (4) above suggested that we may expect to find substantial heterogeneity in the online-access effect, varying with time, with the characteristics of the channels providing online access, and with the inherent quality of the volume.

We allow for different effects over time and for different vintages of content by estimating a matrix of ten different coefficients on online access, varying across blocks of publication years and citation years, each one a different shaded box in Figure 4.[20] While our initial regressions consider an aggregate indicator of online access, equaling 1 if the volume was available through any channel, subsequent regressions will allow for heterogeneous effects across different subsets of channels, for example, comparing access through a single channel versus multiple channels, or considering access through particularly important named channels such as JSTOR or Elsevier. The regressions for access through the specific channels JSTOR and Elsevier are run in two ways. First, they are run with the simple indicator variable as just described. Second, they are run with a continuous variable—the number of institutional subscribers to that channel—in place of the indicator. The number of subscribers will proxy for its higher-dimensional correlate appearing in equation (4): the distribution of library subscriptions across citing authors, $\{l_{ni}\}_{n=1}^{N_t}$,. The continuous specification allows us to distinguish between the impact of, for example, JSTOR for journals or time periods for which it only has a few subscribers to the impact in cases when it has more subscribers. We will take several approaches to determining whether the effect of online access is higher for popular or unpopular articles, devoting the whole of Section 6 to this question.

4.4. Other Controls

The regressors x_{vt} include several controls for the sorts of effects seen in Figures 1 and 2. We estimate journal *k*'s age profile with a quadratic specification:

$$\gamma_{1k} AGE_{vt} + \gamma_{2k} AGE_{vt}^2, \tag{6}$$

---

[20] A separate regression is run for each block of ten publication years, allowing the coefficients on all variables to vary across regressions. In each regression, two online-access coefficients are estimated, one for the early years in the online period (1995-99) and one for late years (2000-05).

where $AGE_{vt} = t - p(v)$ is the age of volume v of journal k in the year of citation, and γ_{1k} and γ_{2k} are coefficients which are allowed to vary not just across journals but across each of the five blocks of ten publication years shown in Figure 4. To account for secular trends in citations, the regressors also include a set of fixed effects for all of the interactions between individual citation years and publication years. This allows each publication year to have a different secular trend and allows the secular trend to have an arbitrary pattern.

Footnote 17 discussed the "identification problem," i.e., the impossibility of separately identifying age, cohort, and time effects. Here, volume plays the role of cohort and citation year plays the role of time, so translated into the present context, the identification problem regards the separate identification of age, volume, and citation-year effects. The formula for the age of a volume indicates why separate identification is impossible: age is the difference between citation year (already picked up by citation-year effects) and publication year (already picked up by volume fixed effects), so there is no variation left to identify age. Fortunately, the identification problem will not impair our ability to estimate the online-access effects of interest. The included age, volume, and citation-year fixed effects are not of direct interest themselves but are only included as controls to improve the estimation of the online-access variables. The identification problem will indeed have a symptom in that the fixed effects will not have independent interpretations as age, volume, or citation-year effects.[21] The results of interest on the online-access variables are unaffected by the identification problem because OAC_{vt} varies within all the sets of fixed effects.[22]

The variables of interest are not identified if we go as far as to include a different age profile for each volume. It would be impossible to tell if online access were having an effect or if the volume's cites happened decay more slowly than other journals' for intrinsic reasons. Identification is preserved with more aggregate age profiles; specifically, we specify a profile for each block of ten volumes rather than individual volumes. In essence, our identification assumption is that volumes of a journal that are published around the same time have similar age profiles. If, after netting out own-volume effects and secular trends, we see an increase in citations above this expected citation profile corresponding to when the online-access variable turns on, we attribute this effect to online access.

[21] The manifestation of this symptom is that more than one of some sets of fixed effects has to be dropped to avoid collinearity.

[22] The primary regressions in fact include a richer set of fixed effects than was used above to estimate Figures 1 and 2. The primary regressions will include volume rather than more aggregate journal fixed effects, interactions between publication years and citation years rather than just a set of age and a set of citation effects separately, and five different age profiles for each journal rather than one aggregate age profile.

4.5. Identification Challenges

Two challenges must be overcome for the regressor of interest (OAC_{vt} and related variables) to provide consistent estimates of the online-access effect. First, the online-access variable must be exogenous in the sense of being uncorrelated with the error term [the difference between the left- and right-hand sides of equation (5)]. As discussed previously in this section, one component of the error is the time-varying part of the distribution of match qualities across citing authors, $\{Q_{ni}\}_{n=1}^{I_t}$, that part which is not picked up by other controls (volume fixed effects, age profile, fixed effects for the interaction between publication and citation year, and the use of different blocks of data in the estimation). The online-access variable will be orthogonal to this error if journals did not choose when to place different volumes online based on their relative numbers of citations. The example of the *American Economic Review* (*AER*), shown in Figure 5, suggests that the online-access variable is plausibly exogenous. The journal was ultimately available online through two channels, JSTOR and the American Economic Association's website. In 1996, JSTOR placed a whole tranche of volumes online. After that, JSTOR put additional volumes online at the expiration of their "embargo" (the period during which recent content is only available from the publisher, presumably to maintain demand for journal subscriptions). In 2002, the American Economic Association began to make all recent content immediately available online through its own website. This pattern of large tranches and together with smaller streams is fairly typical and seems to be based more on technological convenience than the volumes' relative citations.

The second challenge is that the online-access variable must exhibit some independent variation from the other regressors. If volumes were placed online with a fixed lag, OAC_{vt} would be completely collinear with the volume's age. As Figure 5 shows, this is not typically the case. JSTOR began its coverage of the *AER* in 1996 by putting a large tranche (1956-89) of its backfiles online. More recent volumes were added in a stream, but the stream was not completely regular: JSTOR bunched some recent volumes together as part of its policy to shrink the "embargo" period from five to three years; the American Economic Association inaugurated its own website with three volumes. Paradoxically, this bunching of online availability provides a useful source of variation because simultaneous access affects different volumes at different points in their age profiles. For example, JSTOR's initial tranche of over 30 volumes of the *AER* is a shock to the 1956 volume in its fortieth year but the 1989 volume in its sixth. The less pronounced but still evident bunching for more recent content provides a similarly useful source of variation.

Online availability varies across journals as well as volumes within a journal. As shown in Figure 5, JSTOR added the big tranche of AER backfiles in 1996. While some other journals were added

to JSTOR at the same time, others were added at various points later (for example the *Economic Journal* in 1998 and the *Review of Economic Studies* in 1999).

5. Results

We begin by reporting aggregate results for the online-access effect. While the results are broken out by time blocks, different channels are combined as are all other sources of heterogeneity. We do this to make the broad point that the astonishingly high results found in many previous studies were due to their lack of adequate controls for quality. Any measurable effect of online access disappears, at least at the aggregate level, when rich enough controls are added. The section then goes on to restore a modest online-access effect by investigating possible sources of heterogeneity.

5.1. Alternative Specifications

Table 2 presents the results for our most aggregate measure of the online-access effect. All online channels are aggregated; our indicator is 1 if online access is provided through at least one channel. To demonstrate the importance of the various controls in our preferred specification, which is reported in the last column, the columns leading up to the last gradually enrich the included controls. There is some disaggregation allowed even for these "aggregate" results in that we report ten different online-access effects, one for each of the boxes in Figure 4, to allow for heterogeneity in the effect across blocks of publication and citation years. Technically, to allow for considerable flexibility in the coefficients on the included controls, we ran five different regressions, one for each ten-year block of publication years. Within each regression, the coefficient on online access is allowed to vary across early citation years (1995-99) and later ones (2000-05). We use shaded boxes as a device to indicate which results are coming from the same and which from separate regressions. The reported standard errors are robust to heteroskedasticity and clustered at the journal level. Only the results of interest (those for indicators for full online access) are reported; the hundreds of additional control variables are detailed in the notes for the table. Regression coefficients have been converted into a form interpretable as proportionate increases: a zero result corresponds to no measured effect from online access; a negative result corresponds to online access causing a reduction in cites; a positive result corresponds to online access causing an increase in cites. For example, a result of 0.2 corresponds to cites being 20% higher with online access than without.

An obvious pattern emerges scanning the table from left to right. Column 1 is run without journal or volume fixed effects to mimic the previous literature. Without these controls for quality we can reproduce the astonishingly high online-access effects found in many previous studies. For example, the

first coefficient of 5.195 has the interpretation that volumes published in 1956-65 received a more than a 500% fold boost in citations from online access in the years 1995-99 compared to having no online access. Similarly huge effects are seen for most of the other entries in the column. The median result for the column is 2.971, representing a nearly 300% boost in citations.

Column 2 adds journal fixed effects. Only two statistically significantly positive results remain, and their magnitudes have been reduced by more than an order of magnitude. Column 3 adds volume fixed effects, an even richer set of quality controls than journal fixed effects. The results are again reduced and only one statistically significantly positive result remains. The median effect is only 0.05, implying only a 5% citation increase from online access. Column 4 adds a quadratic age profile for each ten-year block of a journal's volumes to the specification in column 3. This further reduces the magnitude of the results toward zero from both directions. No statistically significant result remains; the median falls to a 1% effect of online access. The fairly tight standard errors suggest fairly precise zero effects from online access at the aggregate level.

Column 4 is our preferred specification, but we continue with two additional columns of results to provide a formal analysis of the misspecification in two important competing papers surveyed in the introduction: Evans (2008) and Evans and Reimer (2009). Column 5 is our attempt to reproduce the results from Evans and Reimer's results, which uses the more advanced specification of the two papers. While our underlying data source is different than Evans and Reimer's (see footnote 8), the controls on the right-hand side are the same, including lagged citations and volume fixed effects, which Evans and Reimer included to control for expected citations in the absence of a digitization effect. Importantly, the publication-year × citation-year fixed effects which we included in all of our specifications to control for secular trends are absent from column 5. Focus on the last row of results from column 5, because the citation period (2000-05) is most similar to theirs (1998-2005). We find a greater than 33% boost from online access, similar in magnitude to their finding of about 26% boost in citations from open access. Thus, in spite of the difference in underlying data, we are able to reproduce their result fairly closely.

Column 6 repeats the specification from 5 but adds publication-year × citation-year fixed effects to control for secular trends.[23] The digitization effect for the 2000-05 period disappears. This suggests that Evans and Reimer's (2009) results are spurious, generated by omitted-variable bias. Omitting time effects forces the open-access variable to pick up age effects (which are positive immediately to the right of the peak in Figure 1) and secular trends in citations (which are positive over the whole range in Figure 2). Evans and Reimer (2009) do not provide results for earlier periods to which ours can be compared;

[23] To be consistent with our preferred specification in column 4, we also add a quadratic age profile for each block of 10 volumes of a journal, but this inclusion causes less of a change in the results from column 5 than the publication-year × citation-year fixed effects.

we find statistically significant and substantial negative digitization effects for these earlier periods in column 5.

5.2. Number of Channels

Although the measured impact of online access may be negligible in our sample in the aggregate, the possibility remains that certain channels or combinations of them may have a more significant citation impact. We first investigate whether the number of online channels matters. The content from most publishers is available via multiple channels at least by the end of the sample, usually via the publisher's own website and one or more aggregators. Adding an online channel to an existing set can expand the number of citing authors with online access to the extent that some libraries gain online access through the new channel. Competition between duplicate channels may lower prices and increase subscriptions, although the publisher may exert some control over subscription prices depending on the contracting process.[24] The main exception to the rule is Elsevier, which allowed online access only through its own ScienceDirect website.

Figure 6 shows the proportion of observations in our sample available online through a single and multiple channels by the year in which the volume was published. The figure shows that both sole and duplicate access are represented in our sample, although sole access dominates for older content and duplicate access for more recent content. Columns 2 and 3 of Table 3 report parameter estimates for full online access via a single channel separately from full online access through multiple channels.[25] The aggregate results from column 5 of Table 2 are reproduced in column 1 for comparison purposes.

The results for sole access in column 2 are similar to the aggregate results in column 1, both small and statistically indistinguishable from zero. The online access effect is larger with multiple channels in column 3. For the earliest three decades of data (covering the period 1956-65, 1966-75, and 1976-85), we find positive and statistically significant effects from multiple-channel access, ranging from 9.5% to 21.8%. While consistent with the notion that multiple channels provide a greater citation effect, this evidence is not completely dispositive because other factors are correlated with the sole/duplicate access distinction. As mentioned, all of Elsevier's content is only available through a single channel, so the

[24] The publisher can be considered to be the upstream firm and aggregators to be downstream firms in a vertical industry structure serving the citing author as the consumer. Fauli-Oller and Sandonis (2010) show that an increase in the number of downstream firms may increase or decrease quantity and welfare in the final-good market, depending on the parameters, in a model in which an upstream monopolist offers of nonlinear wholesale tariffs. See Inderst (2008) for a related model.

[25] Besides the variables capturing partial-online access (access to only part of a volume's content or the volume's content for only part of a year), which are included in all reported regressions, the regressions in columns 2 and 3 and subsequent tables include indicators for hybrid online access: full access through some channels and partial access through others through hybrid channels. See the notes to the relevant tables for details regarding the exact specification.

difference between columns 2 and 3 may be picking up differences between Elsevier's and other online channels. This motivates further disaggregating the results by individual channel, described next.

5.3. Individual Channels

As the model [in particular equation (4)] indicated, the effect of online access may depend on the nature of the channel providing the access. All online channels may not be "created equal": the number of subscribers to the channel will affect its citation impact, as will the breadth of its offerings (a platform with more journals and/or more volumes per journal will be more valuable to the searcher), years of operation (citing authors gaining familiarity with the channel over time) and website design.

Our analysis will be restricted to the two individual channels for which we have the most data: JSTOR and Elsevier. Figure 7 shows the proportion of observations in our sample available online through these channels. JSTOR provided online access to a roughly constant fraction of the observations for each publication year, with a fall off for the most recent publication years as the embargo window was hit. Elsevier gains in importance in more recent publication years, providing online access for fully one quarter of the sample for publication years 1999 and 2000. About half of the sample journals were carried by JSTOR at some point and a third were published by Elsevier, so there is a substantial amount of data for each.

Table 4 reports results from online access through these two individual channels. For JSTOR, we specify a pair of indicator variables to capture the impact of online access via JSTOR alone (column 1) and the marginal effect of adding JSTOR access to one or more other channels (column 2). Since Elsevier's online content is available only through its own channel (ScienceDirect), one indicator is included to measure the impact of access to this channel (column 3). Also included but not reported are indicators for online access through other channels. The results for individual channels are extracted from the same regression (as before, one regression for each block of publication and citation years), so are collected together in the same shaded box. Crosses indicate parameters that cannot be estimated because no observations fit the cell.[26] The point estimates in column 1 for JSTOR as the sole channel are positive during the 2000-05 citation period for each of the five publication-year blocks. In three cases, 1956-65, 1966-75, and 1996-2005, the parameters are statistically significant, with implied citation increases ranging between 5.8% and 11.0%. The results for sole JSTOR access for the earlier citation period (1995-99) are generally smaller or even negative, and not statistically significant. The results in column 2 for the marginal effect of JSTOR access in the presence of other channels are imprecisely

[26] For example, the embargo window prevented JSTOR from posting content published in 1996-2005 early enough to be cited during the period 1995-99; Elsevier's pre-1995 volumes were not posted online before 2002, precluding any online effects during 1995-99.

estimated and not statistically significant. None of the Elsevier parameters are statistically significant; in fact, three of the four parameters associated with access during the 2000-05 period are negative.

The odd signs and general insignificance especially of the JSTOR results in the early citation period may be symptomatic of a problem with the indicator-variable specification in this case. Consider the graph of JSTOR subscription trends in Figure 8. In the 1995-97 period, although JSTOR had content posted online, it had hardly any institutional subscribers.[27] Subscriptions jumped in 1998, reaching 350 by 1999. Aggregating these two subperiods into a single "early citation period" may produce unreliable estimates. We thus move to a specification that uses the number of institutional subscribers as a continuous measure of the extent of online access to the channel. For JSTOR, we have data on the number of institutions subscribing to different "packages" of its online journals. For Elsevier, we have subscriptions for journal "backfiles," volumes of a journal published before 1995, for which an institution could pay a one-time fee for a perpetual access license. We do not have subscription data for Elsevier's more current content, so we restrict the reported Elsevier results just to its backfiles.[28] The trend in Elsevier backfile subscriptions is graphed along with JSTOR subscriptions in Figure 8.

Table 5 reports the results for the subscription specification, which is ultimately our preferred one. For ease of interpretation, the parameters have been converted into elasticities. For example, the first entry of 0.170 can be interpreted as saying that a doubling of JSTOR subscriptions with online access to the volume would result in a 17.0% increase in citations for that category (sole access through JSTOR to 1956-65 content). Again, because of these channel's different access policies, a pair of estimates are provided for JSTOR (separating cases in which JSTOR is the sole channel and in which JSTOR duplicates the access via some other channel); a single estimate is provided for Elsevier because it is the sole online channel for its backfiles. Because the subscription variable already picks up the main trends in citation effects, we save degrees of freedom by combining all the citation periods together rather than reporting an early and late citation-period effect.

The JSTOR estimates are now all positive, and most (eight of ten) are statistically significant. When JSTOR is the sole access channel for content (column 1), a doubling of subscriptions results in an increase in citations of between 3.7% and 17.0%, depending on the publication-year block. The magnitude and statistical significance of the results falls fairly consistently as one moves down column 1 from the oldest publication-year block (1956-65) to the most recent. When JSTOR provides access along with some other channel (column 2), this set of impacts is similar, ranging between 1.7% and 18.6%. The fact that access to other channel does not seem to impair the marginal contribution of JSTOR access

[27] For a comprehensive history of the creation and early evolution of JSTOR see Schonfeld (2003).

[28] We include but do not report indicators for online access to Elsevier's content apart from its backfiles, just as we do for other online channels besides JSTOR and Elsevier.

suggest that other channels are not good substitutes for JSTOR. In contrast to the JSTOR results, the Elsevier elasticities are small, sometimes negative, and statistically insignificant. Thus, online access through Elsevier's own website (Science Direct) appears to provide no citation boost; JSTOR appears to have a uniquely strong effect on citations, a effect which is generally strongest for the oldest content.

5.4. Other Sources of Heterogeneity

We next break the results into yet more fine categories to identify other sources of heterogeneity. The main theme of the analysis in this section will be to see whether online access has disproportionate effects for the fringe than the core of the discipline. Does online access open up opportunities for scholars in developing countries to read articles that they may not have in print in their libraries? Is the new mode of access more beneficial for scholars in institutions that traditionally do less citing than others? Do lower-tier journals receive a bigger citation boost than higher-tier ones, higher-tier journals perhaps being prestigious enough to generate wide print circulation and adequate search and acquisition effort regardless of the mode of access.

We try three different breakdowns, by location of the first citing author, by the ranking of the first citing author's institution, and finally by journal ranking. Table 6 provides some descriptive statistics for the different breakdowns. The reported results will again be for the two most important individual channels, JSTOR and Elsevier. We continue to use the preferred specification using institutional subscriptions to measure the extent of online access.

Consider first the breakdown of results by citing author's country of origin. There are a number of alternatives for defining the location of an article with several authors; for simplicity we use the location of the first citing author. Table 6 shows that articles with U.S.-based first authors are responsible for a majority (60%) of citations, reflecting the influence of that region in the fields of economics and business. The English-speaking West is responsible for 18% of cites, non-English speaking West for 16%, and the rest of the world for the remaining 6%.

Table 7 presents the regression results for the breakdown by region. Each column of boxes is directly comparable to Table 5; the specification is identical except for two differences. First, the left-hand side variable is the number of citations from the region rather than in aggregate. Second, the subscriptions variable on the right-hand side is the number of subscriptions of institutions located in the given region rather than in aggregate. Not surprisingly given the proportion of cites coming from the United States, the U.S. results are quite similar to the aggregate ones. JSTOR continues to have a generally positive and statistically significant online-access effect of roughly the same size as seen in Table 5. Elsevier continues to show no significantly positive online access effect; indeed, a negative elasticity (-0.104) for Elsevier content published in 1976-85 is significant at the 10% level. The pattern of

results is also similar for the other regions except for the non-English speaking West. In that region, JSTOR has no statistically significant positive effect. The only statistically significant results are the negative ones for the most recent content (1999-2005). The result does not seem to be due to the relative lack of subscriptions in this region: the measure of online access controls for the number of regional subscriptions in this specification; moreover, the number of subscriptions in this region is roughly equal to the number in the English-speaking West and also to the number in the rest of the world. Rather, the lack of a positive JSTOR effect seems instead to be due to greater reliance on national journals not represented in JSTOR by scholars in the non-English speaking West. Lubrano et al. (2003) found that the majority of 1991-2000 economics publications in the four largest non-English-speaking European countries appeared in national journals: 66% in Germany, 67% in Spain, 81% in Italy, and 85% in France. Further evidence is provided by Drèze and Estevan (2007), who found that 40 of the 57 journals (85%) appearing on the 2004 National Center for Scientific Research (CNRS) list of top journals ranked by peer opinion in France did not appear on the list of 68 top journals ranked by Lubrano et al. (2003) according to an objective citations measure.

The regional breakdown does not support the conclusion that developing countries benefit disproportionately from online access. The rest of the world does show a benefit from online access but not different in kind from the United States and the English-speaking West.

The next breakdown looks at citations by rank of the citing institutions. The idea is that the institutions responsible the most citing may already have a good infrastructure for their scholars in terms of rich library holdings, research assistance, and administrative support that would allow scholars there to search and acquire relevant articles with either print or online access, whereas scholars in institutions with less support may differentially benefit from online access. To avoid the problem of ranking institutions across different regions of the world, we focus on U.S. institutions only, ranking them by the proportion of cites in our data accounted for by first citing authors at that institution. Appendix Table A2 lists the top 100 citing institutions. Table 6 reveals the expected skewness in the distribution of citations, showing that the top 100 institutions are responsible for 73% of citations in our data.

Table 8 breaks the results down along this same dividing line, separately estimating the effect of online access on citations by the top 100 citing institutions from the effect on citations from the rest of U.S. institutions. For comparison, the aggregate results for all U.S. institutions from Table 7 is repeated in columns (1) and (2). Because of the generally insignificant results found for Elsevier so far, to save space in this and all subsequent tables we report only JSTOR results. While there are small differences in magnitude or significance, the general pattern looks quite similar between top-100 and other citing institutions. The effect of JSTOR appears to be fairly uniform across the two types of institution.

Last, we break the results down by the rank of the journal, where the same ISI impact factor used to rank journals for our data-selection procedure is used here to group the 100 journals into the top and bottom half, again stratified by subfield (economics vs. business). The lower entries in Table 6 again show the expected skewness in citations, with the top half of journals receiving 81% of cites in our sample.

Table 9 presents the JSTOR elasticities broken out this way in columns (3)-(6). Columns (1) and (2) repeat the aggregate results across all ranks of journals from Table 5 for comparison. While we see some differences in magnitude and significance between the top half and bottom half of journals, one does not appear to be systematically higher than the other. This may be due to the fact that our sample is already restricted to high-impact journals. In any event, we find that JSTOR provides a citation boost even for the very best journals, ranging as high as an elasticity of 18%.

6. Long-Tail Effects

The results in Table 9 examined whether online-access effects are *journal*-specific. In this section, we refine the analysis considerably, investigating whether these effects are *article*-specific. The idea that obscure or niche products might disproportionately benefit from internet search and acquisition was dubbed the "long-tail effect" in Anderson's (2004) famous *Wired* magazine article. In the market for academic journals, the long-tail effect might arise if obscure articles become easier to locate and acquire using the internet. Seminal articles might experience little effect because they would be well known and important enough to be acquired regardless of the access technology. Such long-tail effects have been found in other markets including books (Brynjolfsson, Hu and Smith 2003), clothing (Brynjolfsson, Hu, and Simester 2007) and video sales (Elberse and Oberholzer-Gee 2008). In terms of the model, a long-tail effect would show up as a negative interaction between the last two arguments of equation (4).

In theory, the effect could go the other way, with online access disproportionately boosting citations of the highest-cited articles, sometimes called a "superstar" effect. A superstar effect might arise if online access aids citing authors in identifying and acquiring articles outside of their subfields, but only the seminal articles outside of one's subfield are worth citing.

To date, only one paper examines these issues in the context of scholarly communication. Evans (2008) reports that online access reduces the number of cited articles and increases the citation concentration of those articles that are cited, suggesting a superstar effect.

Contrasting Evans' findings, we will show that, in the case of JSTOR, the effect of online access is fairly uniform across the distribution of articles, benefiting superstar and more obscure articles alike. In other words, the typical power law relationship between ranked articles and citation counts is shifted up

but its shape is not changed. In direct contrast to Evans (2008), we find that online access increases the fraction of articles receiving any cites.[29]

Our approach consists of two complementary estimation strategies. First, we bin the articles into different quintiles based on number of citations received at a certain age and then estimate the online-access effect in later years using a separate regressions for each quintile. Second, we focus further analysis on the least-cited articles by running regressions involving the proportion of articles in a volume receiving any cites. The first approach is described in Section 6.1 and the second in Section 6.2.

6.1. Quintile Analysis

The traditional approach to quintile analysis minimizes a sum of asymmetrically weighted absolute residuals to yield estimates of specific quintiles (see Koenker 2005). While this method has recently been extended to the case of count data (Machado and Silva 2005), no such estimator has been developed for panel count data. Our alternative consists of applying the Wooldridge (1999) PQML estimator to separate quintiles of articles ranked by the number of citations. In order to avoid bias due to selection of the sample based on residuals, we use a pre-period of citation years to form the quintile samples but then run the regressions using citation years separated from the pre-period by some gap in time. More specifically, we rank articles by citations in a pre-period window of length t_{vw} years, formed by taking the earliest t_{vw} years of citation data available for that article. The top 20% are placed in the highest quintile group, the next 20% in the next quintile, and so forth. We re-aggregate back to the volume level by collecting all the articles within a volume that fall into the same quintile. Notice that the regressions are ultimately run at the volume level, as we have done throughout the analysis, although article-level information was used to form quintiles. The end result is five samples of volume-level data, one subsample for each quintile. We estimate equation (5) separately for each of the five quintile subsamples after discarding observations in the pre-period window along with observations in an additional gap period of t_{vg} citation years. Thus the regressions are run only using citation years $t_v > t_{vw} + t_{vg}$ for volume v.

Our use of different citation periods for quintile selection and model estimation avoids a bias that would be present with a more naive approach that for each citation year assigns a volume's articles to quintiles based on their observed citation performance for that same citation year. If the regression errors are serially uncorrelated, omitting observations from the per-period window of t_{vw} citation years will produce consistent estimates. Since we include volume fixed effects in each of the separate quintile

[29] The discrepancy may have its source in methodological problems in Evans (2008) discussed in the introduction.

regressions, our method will also produce consistent estimates if there is a unit root in the error term for each volume-quintile. The only difficulty that arises for the method is for the intermediate case in which the error term follows an AR(1) process with autocorrelation coefficient $\rho \in (0,1)$. In this case, omitting the gap period of t_{vg} citation years between quintile selection and estimation will attenuate bias due to selection on the autocorrelated disturbance. While we report results with a two-year window used for quintile selection ($t_{vw} = 2$) and with no additional gap before estimation ($t_{vg} = 0$), as a specification check we also estimated the regressions using different combinations of selection windows and gaps (t_{vw} ranging from 1 to 3 and t_{vg} ranging from 0 to 4). The results were similar across these alternatives, suggesting that a bias due to an intermediate level of serial correlation in the errors is not a concern.

The results are reported in Table 10. The specification is identical to that in Table 5, the only differences being the ones just mentioned, that the aggregated results in Table 5 are disaggregated by quintile here and that fewer citation years are used here to allow for different quintile selection and estimation periods. Again, the reported results are converted into subscription elasticities, and just the JSTOR elasticities are reported for space considerations.[30]

The results for the highest-citation (80-100) quintile in columns 9 and 10 are very similar to the corresponding aggregate results in Table 5 in both magnitude and significance. Thus, the aggregate results appear to be driven by the most-cited articles. The citation boost from a doubling of JSTOR subscriptions when JSTOR is the sole channel for online access ranges from 2.8% to 13.6% and is significant in three out of the five publication-year blocks. Similar results are seen for the marginal effect of JSTOR when it is added to access through other channels. The implication is that the most popular articles receive a citation benefit from JSTOR access.

Looking at the results for lower quintiles in columns 1-8, we also see positive and statistically significant effects of JSTOR access for less popular articles. For some blocks of publication and citation years, the results are higher and more significant for lower than the highest quintile and for others the reverse is true. An examination of the standard errors across each row indicates that the estimates become increasingly noisy as one moves from the highest to the lowest quintile. For the lowest (0-20) quintile, very few results are statistically significant. Still, there is at least some evidence of positive JSTOR effects in all quintiles and nothing to suggest that the proportional effects are greater for the 80-100 quintile. Our interpretation is that positive JSTOR effects can be observed throughout the distribution of articles, from less popular ones in the long tail to the superstars.

[30] Although not reported for space considerations, online access to Elsevier backfiles produces no statistically significant citation effects even when broken down at the quintile level, reinforcing the conclusions from the aggregate analysis in Table 5.

6.2. Fraction of Articles Cited

Given the noise in the estimates for the lowest (0-20) quintile, we take another, complementary approach to studying the effect of online access on the least cited articles, determining if online access affects the proportion of a volume's articles that are cited. Articles that receive no cites in a print world are the true long tail. To quantify the presence of such articles, we go back to the disaggregated, article-level data to construct a new variable, $FCIT_{vt}$, measuring the fraction of articles in volume v receiving at least one cite in year t. Descriptive statistics for this variable are provided in Table 1. Figure 9 shows that the age profile for this variable is similar to that for the mean number of cites shown in Figure 1. The percentage of articles cited peaks at age 4 with about 30% more cited articles than in the baseline age 0. After that point, the percentage of cited articles falls with age, dipping below that for the baseline at about age 25.

To deal with a dependent variable having a fractional-response form in a panel-data setting with relatively large cross-sectional and small time-series dimensions, we employ the pooled fractional probit (PFP) estimator proposed by Papke and Wooldridge (2008). The PFP estimator assumes a conditional mean of the following general form

$$E(FCIT_{vt}|\alpha_v, x_{vt}) = \Phi\left(\alpha_v + x_{vt}\beta + \bar{x}_{j(v)t}\xi\right) \quad (7)$$

where α_v is a volume fixed effect, here assumed to have a normal distribution conditional on the regressors x_{vt}, Φ is the standard normal cumulative distribution function, $\bar{x}_{j(v)t}$ is the mean value of regressors x_{vt} across volumes for the same journal, and β and ξ are parameter vectors. The estimator can be implemented in Stata by regressing $FCIT_{vt}$ on a constant, regressors x_{vt}, and regressor means $\bar{x}_{j(v)t}$, using a generalized linear model with a binomial "family" and Bernoulli "link function". Papke and Wooldridge (2008) emphasize the need to cluster the errors at the fixed-effect level, the volume level in our setting. We take a more conservative approach and cluster at the journal level; this also allows us to be consistent with the clustering strategy used previously with the PQML estimator. The reported results Table 11 are average partial effects, obtained by differentiating (7) with respect to the variable of interest, online-channel subscriptions, and then summing the result across observations in the sample. Average partial effects are converted into elasticities for comparison to results from previous methods.

Table 11 again reports the results just for the important online channels with subscription data: JSTOR and Elsevier backfiles. Although involving a different left-hand-side variable than in Table 5—fraction cited articles rather than total number of citations—the results are remarkably similar in both size and statistical significance. As column 1 shows, a doubling of JSTOR subscriptions increases the fraction of cited articles by 17.6 percentage points for the earliest content (1956-65 publication years). This effect gradually becomes smaller with more recent content, but is positive for all but the last block of

publication years and statistically significant for the first three blocks. Similar effects are seen in column 2 for JSTOR when it is operating alongside another online channel. Again the effects appear to diminish as the content becomes more recent. Consistent with previous findings, online access to Elsevier backfiles has no measurable effect on the fraction of cited articles.

Overall, the results from Table 11 indicate a significant long-tail effect of JSTOR access. JSTOR access leads to significantly fewer uncited articles. The effect is strongest for the earliest content and gradually disappears for the most recent. These results support the conclusions from the quintile analysis from the previous subsection that the effects of JSTOR access increase citations throughout the distribution of articles, for both popular and obscure ones. By contrast, there is no significant effect of line access to Elsevier backfiles at any point in the ranking of articles by citations.

7. Conclusions

Our results for the effect of online access on citations to economics articles can be read as a play in two acts. The first act is destructive. By including fixed effects for journal volumes as controls for unobservable quality of the articles in the volume, the estimate of the online-access effect was reduced from the huge levels found in the previous literature, over 500% in some cases, down to a precisely estimated value of zero. We conclude that the huge estimates found previously are largely spurious, due to these earlier studies' use of cross-sectional data which prevented them from controlling for unobservable quality. We went on to show that the few recent studies (e.g., Evans and Reimer 2009) which attempt to use panel data to get around the bias due to unobservable quality in the earlier literature generally introduce their own specification problem in that they generally lack adequate controls for journal volume age and secular trends in citations. Significant aggregate citation effects disappear when an age profile and time effects are included. We conclude that careful specification of the econometric model is as crucial as careful dataset construction in identifying the effect of journal access on citations.

The second act is constructive. We show that the zero effect of online access in the aggregate masks substantial heterogeneity across platforms. While some platforms including Elsevier's ScienceDirect exhibit no online effects, JSTOR shows significantly positive effects, averaging around a 10% subscription elasticity (meaning that a doubling of JSTOR subscriptions causes a 10% increase in citations). JSTOR has a number of attractive features that may have contributed to its relative importance as a platform: it contains a cross-section of many important journals, it offers access to the entire backfile history up to the embargo window, and it was an early entrant in the market. Indeed, JSTOR offered online access to backfiles five years before ScienceDirect, a long time for users to learn to use the

platform and share their experience with colleagues.[31] JSTOR effects tended to be especially large for the earliest content in our sample, that is, articles published between 1956 and 1975. This is consistent with our model of article search and acquisition: under a range of conditions, the benefits from online access should be greatest for the content that was heretofore more difficult to access in print. Print access was indeed likely to be more difficult for older content because archival content is often stored in hard-to-access satellite facilities, and EconLit, the major tool for searching the economics literature before Google, did not include information about content published before 1969. We also found that the marginal impact of JSTOR was not diminished if duplicate access was provided by other platforms such as Ebsco or ProQuest. In most cases these alternative platforms placed backfiles online after JSTOR, and often in a piecemeal fashion, likely reducing the relative value of these platforms to citing authors. Overall, while economically meaningful and statistically significant, the JSTOR effect is still modest compared to the huge effects found in the previous literature which did not control for article quality.

We disaggregated the results in other dimensions, focusing mainly on JSTOR because this is the case most likely to give nontrivial results. The JSTOR results were surprisingly uniform across these other breakdowns. Based on our finding that older content received the greatest proportional citation boost from online access, we expected that the effects would be larger for categories that might be viewed as suffering a disadvantage which online access might help overcome. For example, articles in lower-tier journals might be more costly and less valuable to access, so an increase in the convenience of access might have a particularly big effect for them. Likewise, authors from lower-ranked institutions or from countries outside of core for economics publishing might show a disproportionate increase in citations from online access. Instead we found that the citation benefit was fairly uniform across cited journals by rank, and across citing authors by rank of citing institutions or region. The one anomaly was the effect of JSTOR on citing authors in non-English-speaking Europe. Whereas JSTOR had a significant positive effect on citing authors in most other regions including the U.S., JSTOR had no effect on the citations from authors in non-English-speaking Europe. One explanation is that authors in this region relied more on national journals rather than the English-language journals available on JSTOR, consistent with the findings of Lubrano et al. (2003) and echoing Drèze and Estevan's (2007) call for economists in Germany, France, Italy, and Spain to increase their publishing in mainline English-language journals. We find no evidence that citing authors in developing countries would receive a disproportionate benefit from more convenient access than, say, those in the U.S.

We further disaggregated the results by binning the articles into quintiles based on citation rank in a pre-period. We found positive online effects throughout the quintiles. We also found that online access

[31] For further discussion of the relative merits of JSTOR, see Harley, et. al. (2010). Note that Google Scholar did not appear on the scene until the very end of our sample period, in late 2004.

decreases the percentage of articles within a volume that do not receive any cites. Taken together, these results suggest that "superstar" articles as well as articles residing in the "long tail" benefit from online access. Thus, the typical power-law relationship between ranked articles and citation counts is shifted up but its shape is not changed. This result contrasts with studies of long-tail effects in online retail markets, such as books and clothing, where niche products benefit disproportionately from use of internet search capabilities.[32] The difference between the two domains may stem from the different search objectives: whereas retail customers typically search for the single best product match, citing authors search for a bundle of references. Lower-cost access may increase cites to more obscure articles in the author's area of specialization as well as superstar articles outside the author's narrow subdiscipline, simultaneously broadening and deepening the use of the scientific literature.

Tying the results back to the broader policy issues considered in the introduction, the lack of online access effects at the aggregate level and the modest effects at the channel level resuscitate the view of citations as a valuable currency and useful indicator of an article's contribution to knowledge. At the same time the modest size of these effects, and the current lack of evidence that free online access performs better, implies that the citation benefits of open-access publishing have been exaggerated by its proponents. Even if publishing in an open-access journal were generally associated with a 10% boost in citations, it is not clear that authors in economics and business would be willing to pay several thousand dollars for this benefit, at least in lieu of subsidies. Author demand may not be sufficiently inelastic with respect to submission fees for two-sided-market models of the journal market (e.g., McCabe and Snyder 2005, 2007, 2010; Jeon and Rochet 2010) to provide a clear-cut case for the equilibrium dominance of open access or for its social efficiency.

The analysis confirms the anecdotal impression that JSTOR was the most important innovation of its time in providing access to the economics and business literature. Whether the social value in terms of scholarly productivity—and the spillover benefits for economic productivity more broadly—exceed the cost of constructing and maintaining the channel are beyond the scope of this paper, but at least a plausible case can be made. The greater impact of JSTOR relative to other channels does support the utility of some of JSTOR's attributes, including its stability and its coverage of a large number of journals and a complete set of backfiles for most of these. While they may not revolutionize the scholarly literature, next-

[32] Hervas-Drane (2010) provides a model in which long-tail and superstar effects operate simultaneously in retail markets.

generation technologies such as Google Scholar advance some of these same desirable features and thus promise to continue making measurable contributions to scholarly productivity.

References

Agrawal, Ajay and Avi Goldfarb. (2008) "Restructuring Research: Communication Costs and the Democratization of University Innovation," *American Economic Review* 98: 1578-1590.

Anderson, Chris. (2004) "The Long Tail," *Wired* Issue 12:10, October.

Armstrong, Mark. (2006) "Competition in Two-Sided Markets," *Rand Journal of Economics* 37: 668-691.

Bergstrom, Theodore. (2001) "Free Labor for Costly Journals?" *Journal of Economic Perspectives* 15: 454-474

Blalock, Hubert M. (1966) "The Identification Problem and Theory Building: The Case of Status Inconsistency," *American Sociological Review* 31: 52-61.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Duncan Simester. (2007) "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," MIT Sloan School working paper.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith. (2003) "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49: 1580-1596.

Craig, Iain D., Andrew M. Plume, Marie E. McVeigh, James Pringle, and Mayur Amin. (2007) "Do Open Access Articles Have Greater Citation Impact? A Critical Review of the Literature," *Journal of Informetrics* 1: 239-248.

Curti, Moreno, Vanna Pistotti, Gabriella Gabutti, and Catherine Klersy. (2001) "Impact Factor and Electronic Versions of Biomedical Scientific Journals," *Haematologica* 86:1015-1020.

Davis, Philip M. Bruce V. Lewenstein, Daniel H. Simon, James G. Booth, and Mathew J. L. Connolly. (2008) "Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial," *British Medical Journal* 337: 568-573.

Davis, Philip. (2010) "Does Open Access Lead to Increased Readership and Citations? A Randomized Controlled Trial of Articles Published in APS Journals," *The Physiologist* 53: 197-201.

De Groote, Sandra L., Mary Shultz, and Marceline Doranski. (2005) "Online Journals' Impact on the Citation Patterns of Medical Faculty," *Journal of the Medical Library Association* 93: 223-228.

Dewatripont, Mathias, *et al*. (2006) *Study on the Economic and Technical Evolution of the Scientific Publication markets in Europe.* Brussels: European Commission Directorate General for Research.

Dosi, Giovanni. (1988) **"**Sources, Procedures, and Microeconomic Effects of Innovation," *Journal of Economic Literature*, 26: 1120-1171.

Drèze, Jacques H. and Fernanda Estevan. (2007) "Research and Higher Education in Economics: Can We Deliver the Lisbon Objectives?" *Journal of the European Economic Association* 5: 271-304.

Elberse, Anita and Felix Oberholzer-Gee. (2008) “Superstars and Underdogs: An Examination of The Long Tail Phenomenon in Video Sales,” Harvard Business School working paper no. 07-015.

Evans, James. (2008) “Electronic Publication and the Narrowing of Science and Scholarship,” *Science* 321: 395-399.

Evans, James and Jacob Reimer (2009) “Open Access and Global Participation in Science,” *Science* 323: 1025.

Eysenbach, Gunther. (2006) “Citation Advantage of Open Access Articles,” *PLoS Biology* 4: 692-698.

Fauli-Oller, Ramon and Joel Sandonis. (2010) “On the Profitability and Welfare Effects of Downstream Mergers,” University of Alicante working paper.

Freeman, Chris. (1994) “The Economics of Technical Change,” *Cambridge Journal of Economics* 18: 463-514.

Gargouri, Yassine, Chawki Hajjem, Vincent Larivière, Yves Gingras, Les Carr, Tim Broday, and Steven Harnad. (2010) “Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research,” *PLoS One* 5 (e13636) 1-12.

Gaule, Patrick and Nicholas Maystre. (2009) “Getting Cited: Does Open Access Help?” University of Geneva working paper, SSRN abstract no. 1427763.

Harley, Diane, Sophia Krzys Acord, Sarah Earl-Novell, Shannon Lawrence, and C. Judson King. (2010) *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. UC Berkeley: Center for Studies in Higher Education. Retrieved from http://escholarship.org/uc/cshe_fsc.

Harnad, Steven and Tim Brody. (2004) “Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals,” *D-Lib Magazine*, 10 no. 6.

Heckman, James and Edward Vytlacil. (2001) “Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return to Schooling,” *Review of Economics and Statistics* 83: 1-12.

Hervas-Drane, Andres. (2009) “Word of Mouth and Taste Matching: A Theory of the Long Tail” NET Institute working paper no. 07-41.

Inderst, Roman. (2008) “Wholesale Price Determination under the Threat of Demand-Side Substitution,” University of Frankfurt working paper.

Inside Higher Education , "New Measure of Scholarly Impact," Dec. 17, 2010.

Jeon, Doh-Shin and Jean-Charles Rochet. (2010) “The Pricing of Academic Journals: A Two-Sided Market Perspective,” *American Economic Journal: Microeconomics* 2: 222-255.

Koenker, Roger. (2005) *Quantile Regression.* Cambridge: Cambridge University Press.

Lancaster, F. W. and Julie M. Neway. (1982) "The Future of Indexing and Abstracting Services," *Journal of the American Society for Information Science* 33: 83-89.

Lawrence, Steve. (2001) "Free Online Availability Substantially Increases a Paper's Impact," *Nature* 411: 521.

Lubrano, Michel, Luc Bauwens, Alan Kirman, and Camelia Protopopescu. (2003) "Ranking Economics Departments in Europe: A Statistical Approach," *Journal of the European Economic Association* 1: 1367-1401.

Machado, José A. F. and J. M. C. Santos Silva. (2005) "Quantiles for Counts," *Journal of the American Statistical Association* 100: 1226-1237.

McCabe, Mark J and Christopher M. Snyder. (2005) "Open Access and Academic Journal Quality," *American Economic Review Papers and Proceedings* 95: 453-458.

McCabe, Mark J and Christopher M. Snyder. (2007) "Academic Journal Prices in a Digital Age: A Two-Sided Market Model," *The B.E. Journal of Economic Analysis & Policy* 7: Issue 1 (Contributions), Article 2.

McCabe, Mark J and Christopher M. Snyder. (2010) "The Economics of Open Access Journals," Dartmouth College working paper.

McKenzie, David J. (2006) "Disentangling Age, Cohort, and Time Effects in the Additive Model," *Oxford Bulletin of Economics and Statistics* 68: 473-495.

Papke, Leslie E. and Jeffrey M. Wooldridge. (2008) "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of Econometrics* 145: 121-133.

Parker, Kimberly, Kathleen Bauer, and Paula Sullenger. (2003) "E-Journals and Citation Patterns: Is It All Worth It?" *Serials Librarian* 44: 209-213.

Rochet, Jean-Charles and Jean Tirole. (2006) "Two-Sided Markets: A Progress Report," *Rand Journal of Economics* 37: 645-667.

Schonfeld, Roger, *JSTOR: A History*, Princeton University Press, 2003.

Simcoe, Tim. (2008) "XTPQML: Stata Module to Estimate Fixed-Effects Poisson (Quasi-ML) Regression with Robust Standard Errors," Statistical Software Components, Boston College Department of Economics, http://econpapers.repec.org/RePEc:boc:bocode:s456821.

Tenopir, Carol and Ralf Neufang. (1995) "Electronic Reference Options: Tracking the Changes," *Online* 16: 67-73.

Walker, Thomas. (2004) "Open Access by the Article: An Idea Whose Time Has Come?" *Nature Web Focus* Article 13, April 15.

Wooldridge, Jeffrey M. (1999) "Distribution-Free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics* 90: 77-97.

Table 1: Descriptive Statistics

	Level of Statistics	Obs.	Mean	Std. Dev.	Min.	Max.
Year journal founded	$j(v)$	100	1956.4	28.0	1844	1988
Publication year $p(v)$	v	3,558	1985.7	13.1	1956	2005
Citation year t	vt	60,453	1994.8	7.1	1980	2005
Cites to volume in year *CIT*	vt	60,453	35.7	59.4	0	771
Fraction of volume's articles cited *FCIT*	vt	60,453	0.21	0.21	0	1
Online indicator *OAC*	vt	60,453	0.26	0.44	0	1

Notes: Dataset comprised of journal volumes (indexed by v) observed each year (indexed by t) during the citing period. The journal that publishes volume v is denoted $j(v)$.

Table 2: Alternative Specifications for Aggregate Results

Publication Years	Citing Years	(1)	(2)	(3)	(4)	(5)	(6)
1956-65	1995-99	5.195*** (2.177)	-0.003 (0.071)	-0.007 (0.055)	0.004 (0.033)	-0.061*** (0.013)	0.010 (0.026)
	2000-05	5.589*** (2.409)	0.172 (0.182)	0.123 (0.137)	0.073 (0.051)	-0.007 (0.022)	0.056 (0.044)
		n = 8,944	n = 8,840	n = 8,372	n = 8,372	n = 8,025	n = 8,025
1966-75	1995-99	5.014*** (2.106)	0.033 (0.060)	0.019 (0.059)	-0.010 (0.033)	-0.118*** (0.025)	0.003 (0.027)
	2000-05	3.436*** (1.526)	0.067 (0.094)	0.054 (0.090)	0.025 (0.024)	-0.157*** (0.021)	0.032 (0.029)
		n = 12,506	n = 12,506	n = 12,168	n = 12,168	n = 11,700	n = 11,700
1976-85	1995-99	4.085*** (1.773)	0.041 (0.083)	0.051 (0.086)	0.003 (0.017)	-0.155*** (0.016)	0.031** (0.015)
	2000-05	2.505*** (1.034)	0.051 (0.083)	0.052 (0.084)	0.001 (0.018)	-0.190*** (0.020)	0.025 (0.018)
		n = 18,441	n = 18,441	n = 18,394	n = 18,394	n = 17,618	n = 17,618
1986-95	1995-99	0.855*** (0.324)	0.026 (0.044)	-0.027 (0.021)	-0.010 (0.020)	-0.013 (0.017)	-0.004 (0.017)
	2000-05	1.709*** (0.459)	0.144*** (0.045)	0.109*** (0.032)	-0.008 (0.018)	-0.039** (0.017)	0.007 (0.015)
		n = 15,062	n = 14,907	n = 14,789	n = 14,789	n = 13,815	n = 13,815
1996-2005	1995-99	0.281 (0.225)	-0.073 (0.061)	-0.143*** (0.050)	0.025 (0.048)	0.019 (0.079)	-0.008 (0.045)
	2000-05	1.944*** (0.478)	0.182** (0.077)	0.004 (0.049)	0.026 (0.037)	0.334*** (0.085)	0.009 (0.032)
		n = 5,500	n = 5,390	n = 5,292	n = 5,292	n = 4,312	n = 4,312
Fixed Effect for Source		None	Journal	Volume	Volume	Volume	Volume
Interacted Time Effects		Yes	Yes	Yes	Yes	No	Yes
Quadratic Age Profile		No	No	No	Yes	No	Yes
Lagged Citations		No	No	No	No	Yes	Yes

Notes: Results from Wooldridge's (1999) PQML procedure. Dependent variable is cites to a volume in a citing year. Each box reports results of interest from a separate regression for each block of ten publication years. Shown are results for coefficients on the interaction between a full-online-access variable and two citation-year blocks. Results converted into marginal effects given by $\exp(\beta) - 1$, where β is the Poisson regression coefficient and $\exp(\beta)$ is the incidence rate ratio. Regressions include online-access variables analogous to those reported in the table, but reflecting partial access (access only to part of a volume's content or only for part of the year). Bottom of table lists other included variables; "interacted time effects" refers to the inclusion of a full suite of publication-year x citation-year fixed effects. Robust standard errors clustered at the journal level reported in parentheses. Number of observations given at bottom of each box; some observations may be dropped when moving to a richer specification if cites are constant within a fixed-effect group. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 3: Aggregate Results Broken out by Number of Duplicate Channels

Publication Years	Citing Years	Any Online Access (1)	Single Channel (2)	Multiple Channels (3)
1956-65	1995-99	0.004 (0.033)	0.006 (0.032)	†
	2000-05	0.073 (0.051)	0.073 (0.047)	0.218*** (0.081)
1966-75	1995-99	-0.010 (0.033)	-0.003 (0.026)	†
	2000-05	0.025 (0.024)	0.034 (0.028)	0.201*** (0.079)
1976-85	1995-99	0.003 (0.017)	0.008 (0.017)	†
	2000-05	0.001 (0.018)	0.005 (0.019)	0.095* (0.053)
1986-95	1995-99	-0.010 (0.020)	-0.009 (0.019)	-0.019 (0.020)
	2000-05	-0.008 (0.018)	-0.009 (0.019)	-0.022 (0.021)
1996-2005	1995-99	0.025 (0.048)	0.029 (0.047)	0.022 (0.067)
	2000-05	0.026 (0.037)	0.029 (0.039)	0.002 (0.042)

Notes: For comparison, column 5 from Table 2 repeated here as column 1. The remaining notes apply to columns 2 and 3. Results from Wooldridge's (1999) PQML procedure. Dependent variable is cites to a volume in a citing year. Each box reports results of interest from a separate regression for each block of ten publication years. Shown are results for coefficients on the interactions between a full-online-access variable and separate indicators for sole and multiple access. Results converted into marginal effects given by $\exp(\beta) - 1$, where β is the Poisson regression coefficient and $\exp(\beta)$ is the incidence rate ratio. Regressions include the following variables not reported in the table: journal-volume fixed effects, publication-year × citing-year fixed effects, and a quadratic age profile for each journal. Also included is an indicator for partial online access only (access only to part of a volume's content or only for part of the year) and an indicator for hybrid online access (full access through exactly one channel and partial online access through at least one other channel). Number of observations given in column (5) of Table 2. Robust standard errors clustered at the journal level reported in parentheses. †No observations fit this category. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 4: Marginal Effects for Selected Channels

Publication Years	Citing Years	JSTOR: Sole Channel (1)	JSTOR: With Other Channels (2)	Elsevier (Current Content and Backfiles) (3)
1956-65	1995-99	0.009 (0.034)	†	†
	2000-05	0.091* (0.050)	-0.025 (0.057)	†
1966-75	1995-99	-0.001 (0.025)	†	†
	2000-05	0.058* (0.036)	0.092 (0.067)	-0.048 (0.052)
1976-85	1995-99	0.016 (0.024)	†	†
	2000-05	0.032 (0.035)	0.030 (0.034)	-0.087 (0.060)
1986-95	1995-99	-0.017 (0.013)	-0.015 (0.019)	-0.087 (0.052)
	2000-05	0.004 (0.021)	0.018 (0.023)	-0.026 (0.045)
1996-2005	1995-99	†	†	0.073 (0.070)
	2000-05	0.110*** (0.039)	0.009 (0.013)	0.101 (0.080)

Notes: Results from Wooldridge's (1999) PQML procedure. Dependent variable is cites to a volume in a citing year. Shown are results for indicators of full online access through the JSTOR and Elsevier; JSTOR results further interacted with indicators for sole access and full access through one other channel. Each box reports results of interest from a separate regression for each block of ten publication years. Results converted into marginal effects given by $\exp(\beta) - 1$, where β is the Poisson regression coefficient and $\exp(\beta)$ is the incidence rate ratio. Regressions include the following variables not reported in the table: journal-volume fixed effects, publication-year × citing-year fixed effects, and a quadratic age profile for each journal. Also included are the following indicators (which apply to online access through channels other than JSTOR or Elsevier and which are aggregated across these other channels): partial access only (access only to part of a volume's content or only for part of the year), hybrid access (full access through exactly one channel and partial access through at least one other channel), full access through exactly one channel, and full access through two or more channels. Also included are partial access indicators for JSTOR and Elsevier. All unreported online-access indicators are interacted with the two citing-year blocks. Number of observations approximately that in column 5 of Table 2. Observations dropped if among ten or fewer for which any online-access indicator is positive. Robust standard errors clustered at the journal level reported in parentheses. †No observations fit this category. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 5: Subscription Elasticities for Selected Channels

Publication Years	JSTOR: Sole Channel (1)	JSTOR: With Other Channels (2)	Elsevier Backfiles (3)
1956-65	0.170** (0.072)	0.138 (0.089)	†
1966-75	0.104** (0.047)	0.186** (0.078)	0.023 (0.065)
1976-85	0.062* (0.036)	0.068* (0.036)	-0.049 (0.054)
1986-95	0.037 (0.027)	0.062** (0.025)	-0.012 (0.047)
1996-2005	0.067*** (0.022)	0.017 (0.011)	†

Notes: Results from Wooldridge's (1999) PQML procedure. Dependent variable is cites to a volume in a citing year. Each box reports the results of interest from a separate regression for each block of ten publication years. Shown are results on the interaction between subscriptions to the channel of interest (JSTOR or Elsevier backfiles), an indicator for full online access through the channel, and indicators for either access through that channel and one other; JSTOR results further interacted with indicators for sole access and full access through one other channel. Results converted into elasticities in two steps. First, a marginal effect is computed as $\exp(\beta) - 1$, where β is the Poisson regression coefficient and $\exp(\beta)$ is the incidence rate ratio. Second, the marginal effect is converted into an elasticity by multiplying by the subsample mean number of subscribers. Regressions include the following variables not reported in the table: a complete set of journal-volume fixed effects, a complete set of publication-year × citing-year fixed effects, and a quadratic age profile for each journal. Also included are the following indicators for online access through Elsevier and separate indicators for online access through channels other than JSTOR or Elsevier: partial access only (access only to part of a volume's content or only for part of the year), hybrid access (full access through exactly one channel and partial access through at least one other channel), full access through exactly one channel, and full access through two or more channels; these indicators are aggregated across individual channels and all interacted with two citing-year blocks. Also included are indicators for full online access through JSTOR, interacted with indicators for partial access through some other channel, interacted with JSTOR subscribers. Also included are indicators for partial access through JSTOR, each interacted with two citing-year blocks. Observations dropped if among ten or fewer for which any online-access indicator is positive. Number of observations approximately that in column 5 of Table 2. Robust standard errors clustered at the journal level reported in parentheses. †No observations fit this category. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 6: Breakdown of Cites into Various Categories

A. Breakdown by Location of First Citing Author (Percentage of Cites from All Regions)

USA		60%
English-Speaking West		18%
Non-English Speaking West		16%
Rest of World		6%
	Total:	100%

B. Breakdown by Ranked U.S. Institutions (Percentage of U.S. Cites)

Top 100 U.S. Citing Institutions		73%
Other U.S. Institutions		27%
	Total:	100%

C. Breakdown by Journal Ranking (Percentage of Cites from All Regions)

Top Half of Journals		81%
Bottom Half of Journals		19%
	Total:	100%

Notes: Percentages may not sum down the column to 100% due to rounding. In part A of the table, "West" refers to the United Nations Regional Group "Western Europe and Others". Country classified as English-speaking if English is one of its official languages. Specifically, "English-Speaking West" includes Australia, Canada, Ireland, Israel, New Zealand, and United Kingdom; and "Non-English Speaking West" includes Austria, Belgium, Denmark, France, Germany, Greece, Italy, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Spain, Sweden, Switzerland, and Turkey. In part B of the table, ranking based on institutions doing the most citing in each discipline as measured by cites from articles with first citing author located in the U.S. See Appendix Table A2 for institution list. In part C of the table, ranking based on average of standardized ISI impact factors from 1985-2005.

Table 7: Subscription Elasticities Broken out by Region

	USA			English-Speaking West			Non-English Speaking West			Rest of World		
	JSTOR			JSTOR			JSTOR			JSTOR		
Publication Years	Sole Access (1)	With Other Channels (2)	Elsevier Backfiles (3)	Sole Access (4)	With Other Channels (5)	Elsevier Backfiles (6)	Sole Access (7)	With Other Channels (8)	Elsevier Backfiles (9)	Sole Access (10)	With Other Channels (11)	Elsevier Backfiles (12)
1956-65	0.183** (0.086)	0.093 (0.078)	†	0.137 (0.101)	0.223* (0.146)	†	0.080 (0.098)	-0.182 (0.131)	†	0.286*** (0.141)	0.509*** (0.180)	†
1966-75	0.122** (0.055)	0.175*** (0.060)	0.054 (0.067)	0.080 (0.069)	0.277** (0.140)	-0.030 (0.100)	0.052 (0.065)	0.087 (0.130)	-0.080 (0.117)	0.145* (0.093)	0.254** (0.122)	0.468* (0.342)
1976-85	0.056 (0.043)	0.067* (0.041)	-0.104* (0.051)	0.042 (0.037)	0.092** (0.073)	0.053 (0.086)	0.042 (0.047)	0.039 (0.052)	-0.029 (0.090)	0.080* (0.051)	0.048 (0.074)	0.100 (0.152)
1986-95	0.039 (0.031)	0.089*** (0.025)	-0.008 (0.054)	0.063* (0.035)	0.075* (0.043)	-0.017 (0.061)	-0.014 (0.038)	0.019 (0.041)	-0.052 (0.060)	0.044 (0.034)	0.093** (0.047)	0.033 (0.087)
1996-2005	0.081** (0.036)	0.015 (0.012)	†	0.108* (0.060)	0.060*** (0.024)	†	-0.249** (0.091)	-0.075*** (0.024)	†	0.112 (0.118)	0.032 (0.042)	†

Notes: Dashed boxes indicate regressions which, to obtain convergence, omit the quadratic term in the journal-age profile and retain only the linear term. Additional specification notes from Table 5 apply here.

Table 8: JSTOR Subscription Elasticities Broken out by Rank of U.S. Citing Institutions

	All U.S. Cites		Top 100 U.S. Institutions		Other U.S. Institutions	
Publication Years	Sole Access (1)	With Other Channels (2)	Sole Access (3)	With Other Channels (4)	Sole Access (5)	With Other Channels (6)
1956-65	0.183**	0.093	0.186***	0.056	0.199*	0.057
	(0.086)	(0.078)	(0.064)	(0.067)	(0.112)	(0.100)
1966-75	0.121**	0.173***	0.129***	0.211***	0.092*	0.092
	(0.055)	(0.060)	(0.054)	(0.088)	(0.056)	(0.061)
1976-85	0.058	0.074*	0.090**	0.069	-0.043	0.077
	(0.043)	(0.042)	(0.047)	(0.051)	(0.037)	(0.048)
1986-95	0.047	0.096***	0.032	0.077***	0.011	0.061*
	(0.030)	(0.024)	(0.026)	(0.020)	(0.042)	(0.031)
1996-2005	0.081**	0.015	0.053	0.005	0.015	-0.028
	(0.036)	(0.011)	(0.043)	(0.011)	(0.056)	(0.031)

Notes: Dependent variable is number of cites from articles whose first citing author's first listed institution is in the indicated group. U.S. institutions ranked according to which did the most citing (see Appendix Table A2 for list). Dashed boxes indicate regressions which, to obtain convergence, omit the quadratic term in the journal-age profile and retain only the linear term. Additional specification notes from Table 5 apply here. The Elsevier backfile variables reported in Table 5 are also included here but not reported for space considerations.

Table 9: JSTOR Subscription Elasticities Broken out by Journal Rank

Publication Years	All Journals		Top-Ranked Half		Bottom-Ranked Half	
	Sole Access	With Other Channels	Sole Access	With Other Channels	Sole Access	With Other Channels
	(1)	(2)	(3)	(4)	(5)	(6)
1956-65	0.170**	0.138	0.180**	0.106	0.184***	0.316***
	(0.072)	(0.089)	(0.091)	(0.105)	(0.067)	(0.129)
	n = 8,372		*n* = 4,576		*n* = 3,795	
1966-75	0.104**	0.186**	0.091*	0.172*	0.182**	0.134
	(0.047)	(0.078)	(0.048)	(0.098)	(0.076)	(0.135)
	n = 12,168		*n* = 6,604		*n* = 5,556	
1976-85	0.062*	0.068*	0.076*	0.082*	-0.223***	-0.144**
	(0.036)	(0.036)	(0.040)	(0.045)	(0.036)	(0.069)
	n = 18,394		*n* = 9,030		*n* = 9,349	
1986-95	0.037	0.062**	0.045	0.054*	-0.073*	0.066*
	(0.027)	(0.025)	(0.030)	(0.028)	(0.042)	(0.035)
	n = 14,789		*n* = 7,428		*n* = 7,345	
1996-2005	0.067***	0.017	0.031	0.013	0.077***	0.009
	(0.022)	(0.011)	(0.037)	(0.011)	(0.025)	(0.049)
	n = 5,292		*n* = 2,700		*n* = 2,592	

Notes: For comparison, results from Table 7 repeated here in columns 1 and 2. Number of observations given at bottom of each box. Regressions for subsets of journals involve a subset of observations; a few additional observations dropped if the number of cites becomes constant within a fixed-effect group. Additional specification notes from Table 5 apply here. The Elsevier backfile variables reported in Table 5 are also included here but not reported for space considerations.

Table 10: JSTOR Subscription Elasticities by Quintile

	0-20 Quintile		20-40 Quintile		40-60 Quintile		60-80 Quintile		80-100 Quintile	
Publication Years	Sole Access (1)	With Other Channels (2)	Sole Access (3)	With Other Channels (4)	Sole Access (5)	With Other Channels (6)	Sole Access (7)	With Other Channels (8)	Sole Access (9)	With Other Channels (10)
1956-65	‡		‡		‡		0.323* (0.171)	0.322 (0.320)	0.136* (0.078)	0.105 (0.105)
1966-75	0.171 (0.114)	0.222** (0.113)	0.011 (0.087)	0.290* (0.173)	0.309*** (0.089)	0.497** (0.239)	0.164** (0.079)	0.200* (0.110)	0.101* (0.060)	0.186*** (0.069)
1976-85	-0.023 (0.080)	-0.034 (0.115)	0.047 (0.046)	0.147** (0.075)	0.049 (0.054)	0.064 (0.107)	0.106** (0.048)	0.107 (0.082)	0.056 (0.040)	0.070 (0.052)
1986-95	-0.037 (0.052)	0.041 (0.035)	0.036 (0.041)	0.061 (0.040)	0.057 (0.043)	0.026 (0.038)	0.025 (0.035)	0.072** (0.032)	0.028 (0.029)	0.063** (0.032)
1996-2005	0.004 (0.089)	-0.002 (0.033)	0.067 (0.076)	0.017 (0.032)	0.015 (0.081)	0.101*** (0.034)	-0.158*** (0.054)	0.003 (0.022)	0.081** (0.041)	0.015 (0.021)

Notes: Quintiles formed by ranking articles within volume by citations in earliest two citing years available (years used for ranking omitted from regressions). Dashed boxes indicate regressions which, to obtain convergence, omit the quadratic term in the journal-age profile and retain only the linear term. ‡Regressions had too few observations with positive citations to produce a non-singular variance-covariance matrix even omitting the linear term in the journal-age profile. Additional specification notes from Table 5 apply here. The Elsevier backfile variables reported in Table 5 are also included here but not reported for space considerations.

Table 11: Elasticities of Proportion of Cited Articles with Respect to Subscriptions to Selected Channels

Publication Years	JSTOR: Sole Channel (1)	JSTOR: With Other Channels (2)	Elsevier Backfiles (3)
1956-65	0.176** (0.069)	0.200** (0.081)	†
1966-75	0.111*** (0.039)	0.401*** (0.056)	0.071 (0.085)
1976-85	0.050* (0.027)	0.078** (0.036)	-0.020 (0.045)
1986-95	0.030 (0.029)	0.089*** (0.032)	0.065 (0.042)
1996-2005	-0.034 (0.033)	-0.024 (0.026)	†

Notes: Results from pooled fractional probit estimator developed by Papke and Wooldridge (2008) for panel fractional-response data. Dependent variable is proportion of articles in a volume cited in a given year. Each box reports the results of interest from a separate regression for each block of ten publication years. Shown are results for an indicator for full online access through the selected channels (JSTOR and Elsevier backfiles) interacted with subscribers to those channels, separately interacted with indicators for sole access and full access through some other channel. Coefficients first converted into marginal effects following Papke and Wooldridge's equation (3.10), except we compute the effect at subsample covariate means rather than computing average partial effects. Marginal effects then converted into elasticities by scaling by the ratio of subsample means of the independent variable to that of the dependent variable. Subsample mean of subscribers (rather than the interaction of subscribers with online access) used as scale factor for independent variable. See Table 5 for list of additional variables included but not reported, notes about number of observations, specification of standard errors, and definition of symbols.

Appendix Table A1: Journals in Dataset

Economics	Economics (con't)	Business
Amer. Econ. Rev. (J)	*J. Institutional & Theoretical Econ.*	*Acad. Manag. J.* (J)
Amer. J. Agricultural Econ. (J)	*J. Int. Econ.* (E)	*Academy Manag. Rev.* (J)
Brookings Papers Econ. Activity (J)	*J. Int. Money & Fin.* (E)	*Accounting Org. & Soc.* (E)
Cambridge J. Econ.	*J. Labor Econ.* (J)	*Accounting Rev.* (J)
Canadian J. Econ. (J)	*J. Law & Econ.* (J)	*Admin. Science Q.* (J)
Econometric Theory	*J. Law Econ. & Org.* (J)	*Bus. Hist.*
Econometrica (J)	*J. Mathematical Econ.* (E)	*Bus. Hist. Rev.* (J)
Econ. Dev. & Cultural Change (J)	*J. Monetary Econ.* (E)	*California Manag. Rev.*
Econ. Geography (J)	*J. Money Credit & Banking* (J)	*Fin. Manag.*
Econ. Hist. Rev. (J)	*J. Political Econ.* (J)	*Harvard Bus. Rev.*
Econ. Inquiry	*J. Public Econ.* (E)	*IEEE Trans. Engineering Manag.*
Econ. J. (J)	*J. Retailing* (E)	*J. Accounting & Econ.* (E)
Economica (J)	*J. Risk & Insurance* (J)	*J. Accounting Res.* (J)
Econ. & Philosophy	*J. Risk & Uncertainty*	*J. Advertising*
Econ. & Society	*J. Urban Econ.* (E)	*J. Advertising Res.*
European Econ. Rev. (E)	*Kyklos*	*J. Banking Fin.* (E)
Explorations Econ. Hist. (E)	*Land Econ.* (J)	*J. Bus. & Econ. Statistics* (J)
Int. Econ. Rev. (J)	*National Tax J.*	*J. Business* (J)
Int. J. Forecasting (E)	*Oxford Bull. Econ. & Statistics*	*J. Bus. Venturing* (E)
Int. J. Industrial Org. (E)	*Oxford Econ. Papers* (J)	*J. Consumer Res.* (J)
J. Agricultural Econ.	*Q. J. Econ.* (J)	*J. Fin.* (J)
J. Applied Econometrics (J)	*Rand J. Econ.* (J)	*J. Fin. & Quantitative Analysis* (J)
J. Comparative Econ. (E)	*Regional Science & Urban Econ.* (E)	*J. Fin. Econ.* (E)
J. Development Econ. (E)	*Rev. Econ. Stud.* (J)	*J. Futures Markets*
J. Econometrics (E)	*Rev. Econ. & Statistics* (J)	*J. Int. Bus. Stud.* (J)
J. Econ. Behavior & Org. (E)	*Scandinavian J. Econ.* (J)	*J. Manag.* (E)
J. Econ. Dynamics & Control (E)	*World Bank Econ. Rev.*	*J. Manag. Stud.*
J. Econ. Hist. (J)	*World Development* (E)	*J. Marketing* (J)
J. Econ. Literature (J)		*J. Marketing Res.* (J)
J. Econ. Perspectives (J)		*J. Product Innovation Manag.* (E)
J. Econ. Psychology (E)		*Manag. Science* (J)
J. Econ. Theory (E)		*Marketing Science* (J)
J. Environ. Econ. & Manag. (E)		*Org. Dynamics* (E)
J. Health Econ. (E)		*R&D Manag.*
J. Human Resources (J)		*Rev. Fin. Stud.* (J)
J. Industrial Econ. (J)		*Strategic Manag. J.* (J)

Notes: (J) indicates a journal available on JSTOR at some point during sample and (E) one published by Elsevier at some point in sample. Classification into economics versus business subdisciplines according to ISI primary subject.

Appendix Table A2: U.S. Institutions That Do the Most Citing

Rank	Institution	Rank	Institution	Rank	Institution
1.	Harvard	34.	Purdue	67.	Georgetown
2.	Penn.	35.	Rochester	68.	Nebraska
3.	Texas	36.	Georgia	69.	Oklahoma
4.	Wisconsin	37.	IMF	70.	Washington State
5.	Illinois	38.	UC Davis	71.	Tennessee
6.	Chicago	39.	Arizona State	72.	Miami
7.	Stanford	40.	Colorado	73.	George Mason
8.	Michigan	41.	Boston Univ.	74.	Federal Reserve
9.	UC Berkeley	42.	Arizona	75.	Syracuse
10.	NYU	43.	Missouri	76.	SUNY Buffalo
11.	Northwestern	44.	Virginia	77.	George Washington
12.	Columbia	45.	U Conn	78.	Temple
13.	MIT	46.	South Carolina	79.	Georgia Tech
14.	UCLA	47.	Iowa	80.	Brown
15.	Ohio State	48.	Pittsburgh	81.	So. Illinois
16.	Indiana	49.	Vanderbilt	82.	UC Santa Barbara
17.	Minnesota	50.	LSU	83.	Oregon
18.	Cornell	51.	UC Irvine	84.	Case Western
19.	Maryland	52.	UC San Diego	85.	Auburn
20.	Penn State	53.	Georgia State	86.	Iowa State
21.	North Carolina	54.	VA Tech	87.	Utah
22.	NBER	55.	Emory	88.	Tulane
23.	Michigan State	56.	Florida State	89.	Wayne State
24.	USC	57.	NC State	90.	Kansas
25.	Duke	58.	Washington Univ.	91.	Baruch
26.	Texas A&M	59.	Boston College	92.	Delaware
27.	Yale	60.	So. Methodist	93.	Brigham Young
28.	World Bank	61.	Houston	94.	Johns Hopkins
29.	Florida	62.	Dartmouth	95.	Clemson
30.	Rutgers	63.	Kentucky	96.	SUNY Albany
31.	Univ. Washington	64.	U Mass	97.	Wyoming
32.	Carnegie Mellon	65.	Alabama	98.	Rice
33.	Princeton	66.	Notre Dame	99.	Oklahoma State

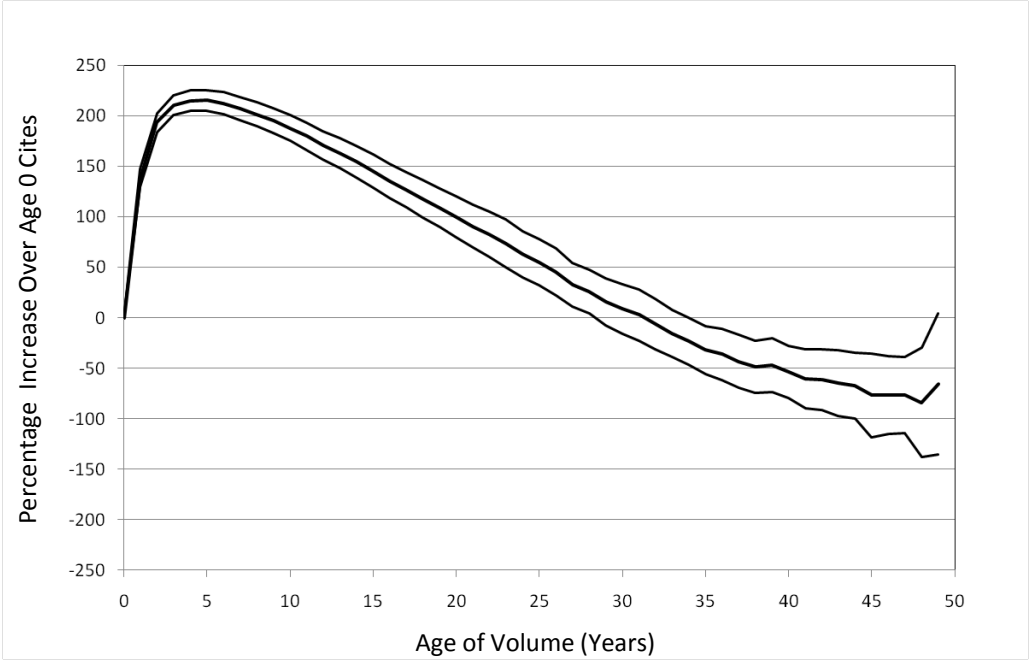

Figure 1: Citation Age Profile. Middle curve plots a set of fixed age effects from Wooldridge's (1999) Poisson quasi-maximum-likelihood procedure. Regression also includes a citation-year and journal fixed effects. Outside lines bound 95% confidence interval based on robust standard errors clustered by journal.

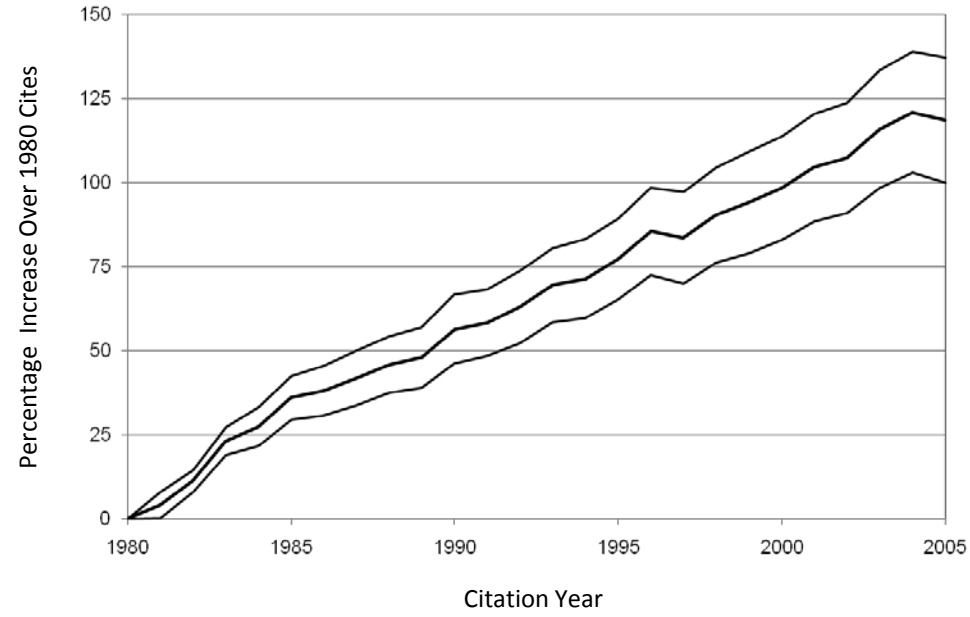

Figure 2: Secular Trend in Citations. Middle curve plots a set of fixed citation-year effects from Wooldridge's (1999) Poisson quasi-maximum-likelihood procedure. Regression also includes a set of age and journal fixed effects. Outside lines bound 95% confidence interval based on robust standard errors clustered by journal.

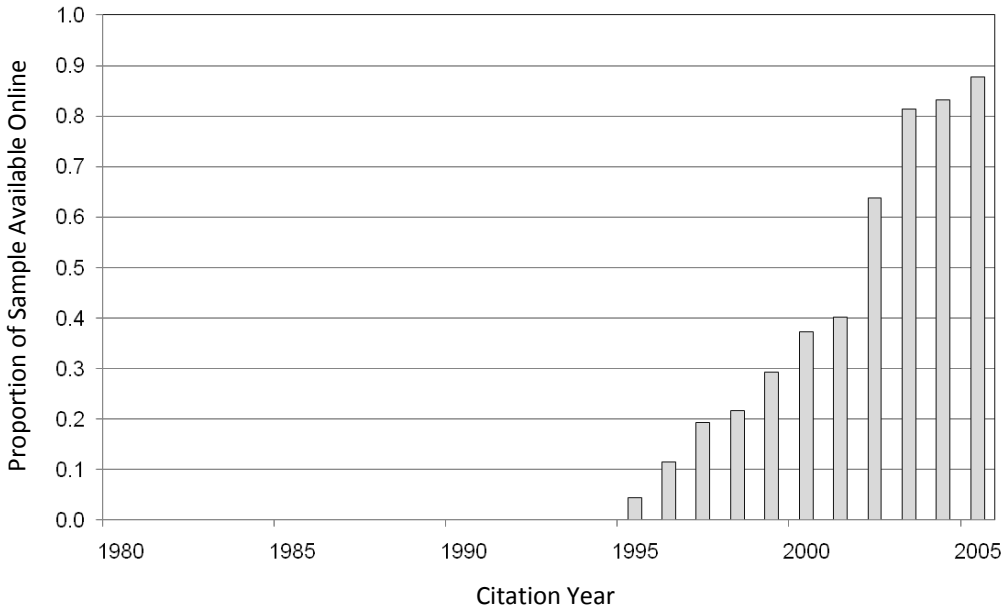

Figure 3: Trends in Online Availability. Mean (across volumes in sample available to be cited in given year) of indicator for online access.

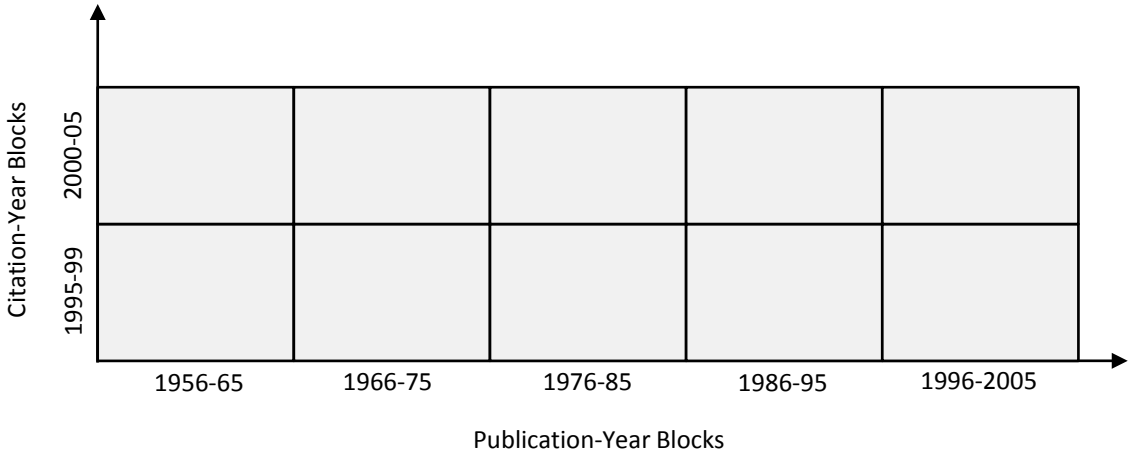

Figure 4: Matrix of Estimated Online Access Effects

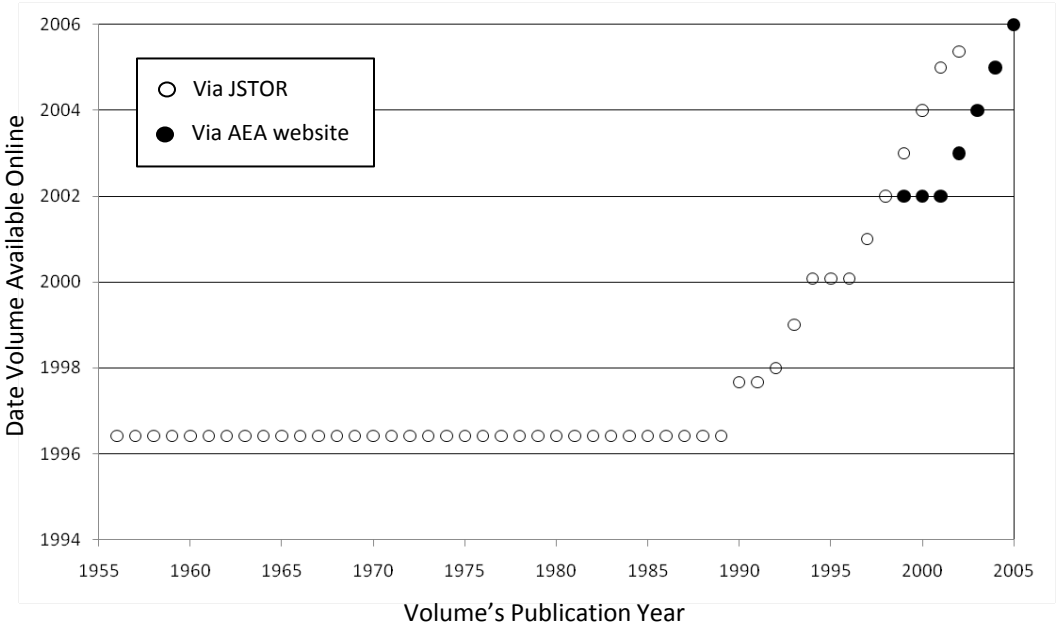

Figure 5: Online Availability in *American Economic Review* Example

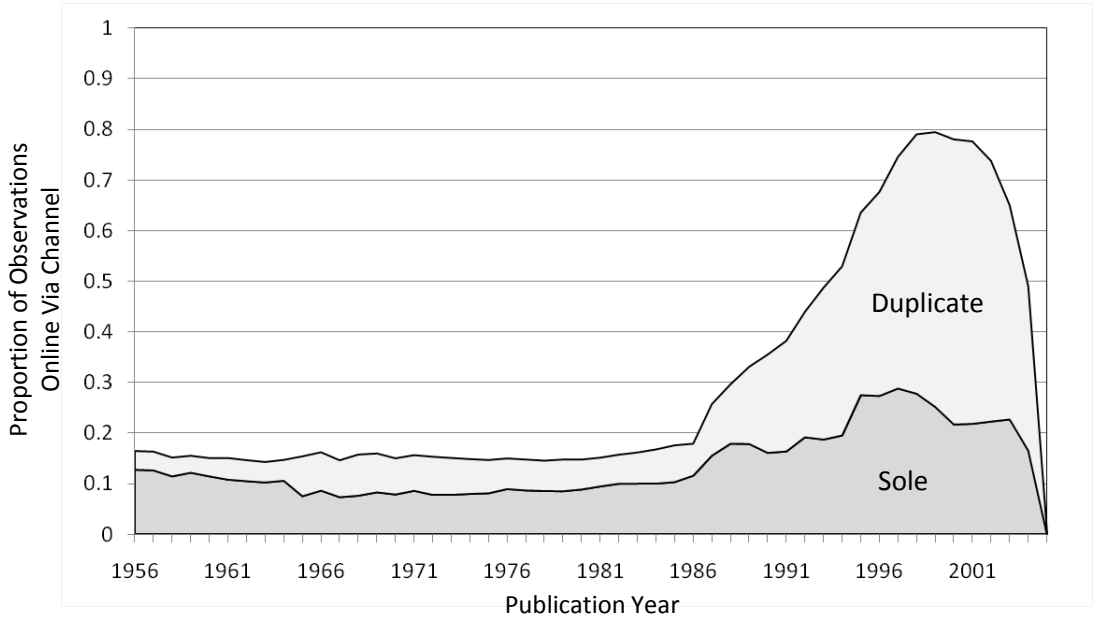

Figure 6: Sole Versus Duplicate Online Access Channel for Various Publication Years. Mean value of indicator of online access of indicated sort for each publication year taken across journals and across citation years.

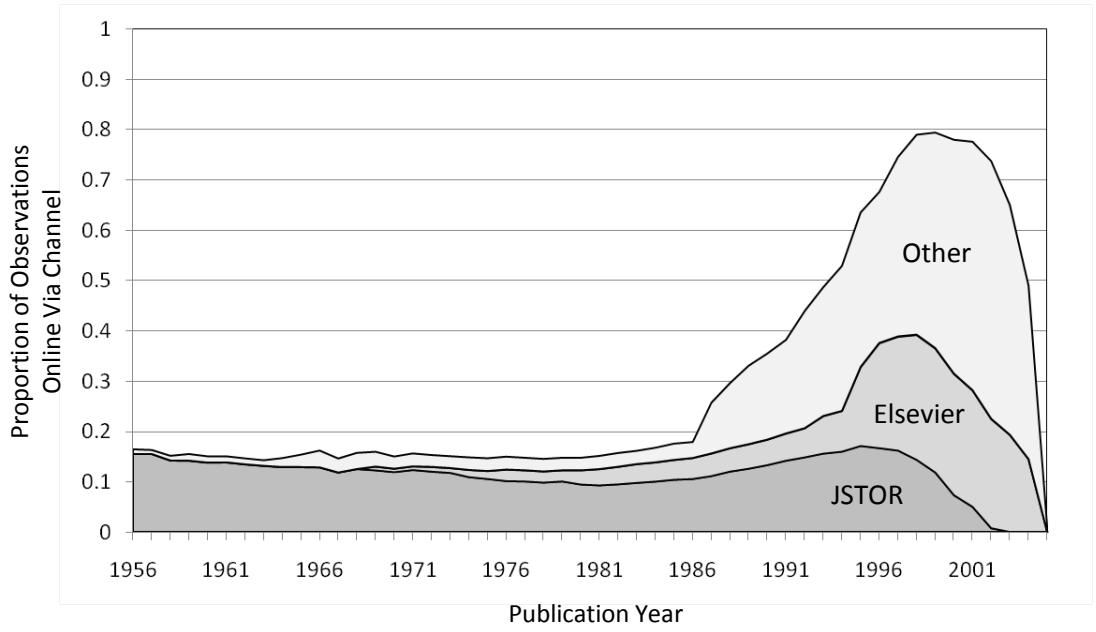

Figure 7: Online Access by Channel for Various Publication Years. Mean value of indicator of online access via JSTOR or Elsevier for each publication year (including both sole access through these channels and duplicate access through some other channel as well). Means taken across journals and citation years. "Other" is a residual category equal to the mean of an indicator for online access but not through JSTOR and Elsevier.

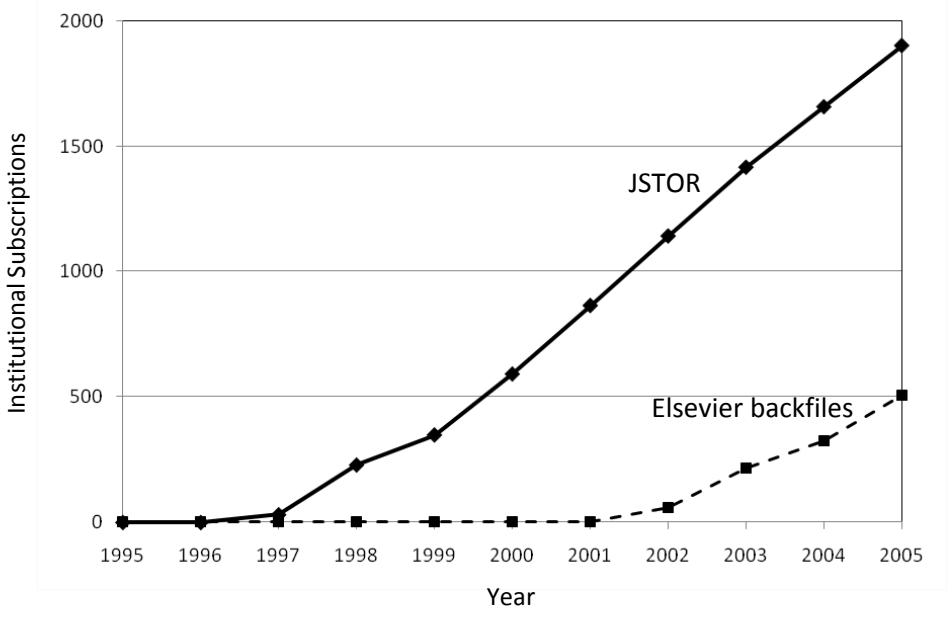

Figure 8: Subscription Trends. Maximum number of institutions subscribing to an online package provided by the indicated channels containing at least one of the journals in our sample. "Elsevier backfiles" refers to institutional purchases of archive of pre-1995 content, which focus on for Elsevier in our subscription analysis.

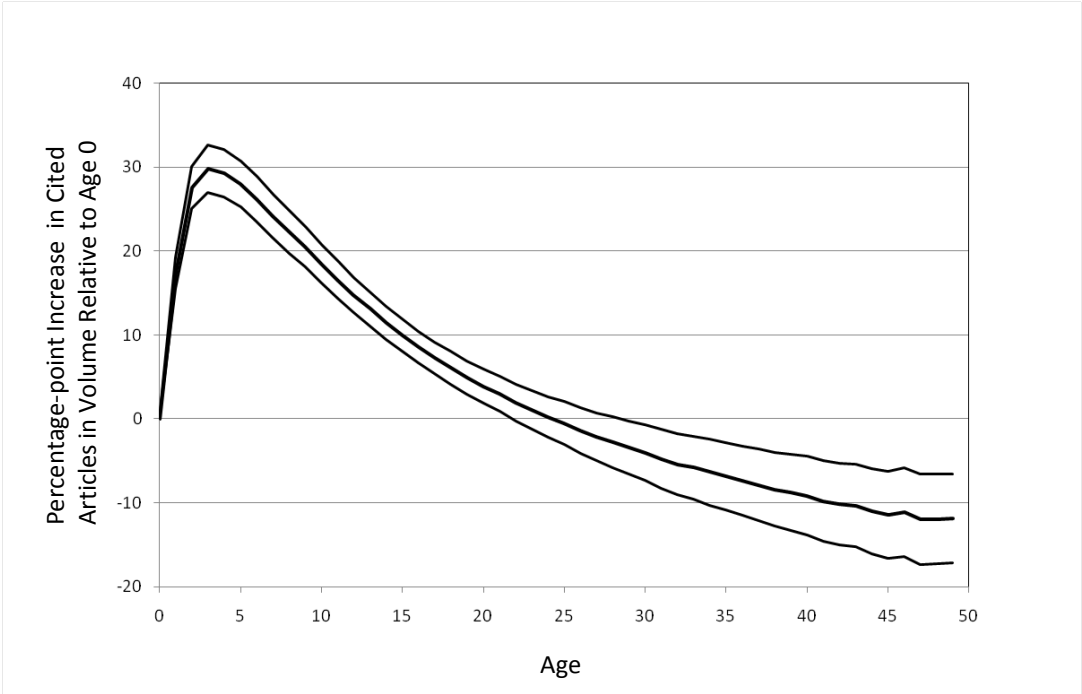

Figure 9: Age Profile for Long-Tail Effect. Middle curve plots of a set of fixed age effects from linear, panel-data regression. Regression also includes a set of citation-year and journal fixed effects. Outside lines bound the 95% confidence interval based on robust standard errors clustered by journal.