2-20-2014

Identifying the Effect of Open Access on Citations Using a Panel of Science Journals

Mark J. McCabe
SKEMA Business School

Christopher M. Snyder
Dartmouth College, Christopher.M.Snyder@dartmouth.edu

Identifying the Effect of Open Access on Citations Using a Panel of Science Journals

Mark J. McCabe

University of Michigan School of Information
University of Göttingen
SKEMA Business School

Christopher M. Snyder

Dartmouth College
National Bureau of Economic Research

October 2013

Abstract: An open-access journal allows free online access to its articles, obtaining revenue from fees charged to submitting authors or from institutional support. Using panel data on science journals, we are able to circumvent problems plaguing previous studies of the impact of open access on citations. In contrast to the huge effects found in these previous studies, we find a more modest effect: moving from paid to open access increases cites by 8% on average in our sample. The benefit is concentrated among top-ranked journals. In fact, open access causes a statistically significant reduction in cites to the bottom-ranked journals in our sample, leading us to conjecture that open access may intensify competition among articles for readers' attention, generating losers as well as winners.

Acknowledgments: The authors are grateful for helpful comments from Brett Danaher, Robert Johnson, Elizabeth Kirk, Andreas Moxnes, Frank Müller-Langer, Nina Pavcnik, Jane Quigley, Giovanni Ramello, Joel Waldfogel, and participants at the 2012 Academia and Publishing Conference (Torino), the 2012/13 Internet and Society Seminar Series (Göttingen), the 2013 International Industrial Organization Conference (Boston), and the 2013 Max Planck Workshop for Junior Researchers on the Law & Economics of Intellectual Property and Competition Law (Munich) for helpful comments. The paper benefitted from the suggestions of editor Ted Bergstrom and the referees. The authors thank Mark Bard, Jamie Bergeson-Bradshaw, Yilan Hu, Ella Kim, Scot Parsley, and Kyle Thomason for assistance in constructing the dataset. Construction of the dataset was supported by a grant from the Andrew W. Mellon Foundation. Research for this paper was supported by a grant from the Alfred P. Sloan Foundation. The authors are grateful for the generous funding from these sources.

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

Contact Information: Mark McCabe, School of Information, University of Michigan, 105 S. State Street, Ann Arbor, MI 48104, email mccabe@umich.edu. Christopher Snyder, Department of Economics, Dartmouth College, 301 Rockefeller Hall, Hanover, NH 03755, email chris.snyder@dartmouth.edu.

1. Introduction

Academic journals facilitate communication of research between scholars in their dual role as authors and readers. The traditional business model is for journals to earn most of their revenue from the reader side through library subscription fees. Library subscription fees have been quite high, especially for commercial publishers (Bergstrom 2001, Bergstrom and Bergstrom 2004, Dewatripont *et al.* 2006). That subscription fees remain high despite the advent of the Internet, which effectively reduces the cost of distributing the journal to readers close to zero, has led to dissatisfaction with the traditional business model and to the proposal of an alternative: the open-access model. An open-access journal allows free online access to its articles, obtaining revenue from institutional support or fees charged to submitting authors.

An active policy debate surrounds open-access journals. The European Union recently announced that recipients of the expected $100 billion in grants over the next decade would be required to publish their research results in open-access journals, following similar requirements issued by the United Kingdom, the U.S. National Institutes of Health, and other funding agencies (*The Economist* 2012). Whether such requirements, along with other policies such as subsidies to cover the operating costs of open-access journals and fees charged to submitting authors, improve the functioning of the market for academic journals and improve scholarship more generally is a controversial policy question.

The empirical literature measuring the impact of open access on citations is a mixture of optimistic claims and contradictory evidence. Early studies, relying on cross-sectional data, report citation benefits as large as several hundred percent.[1] The extraordinary size of the

[1] Lawrence (2001) studied a sample of articles in the proceedings of a computer-science conference, some of which were available only in print, some openly accessible online. The open-access articles received 336% more cites. 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. Walker (2004) studied an oceanography journal that allowed authors to buy open access for their articles, finding 280% more downloads for open-access articles.

estimated effects in these studies prompts suspicion that they are biased upward. A possible source of this bias is that the effect of open access is confounded with article quality, which is unobservable to the econometrician and so is an omitted variable. Recent papers have employed a variety of methods to circumvent this specification problem including panel-data methods (Evans and Reimer 2009), instrumental-variables methods (Gaule and Maystre 2011), and field experiments (Davis *et al.* 2008). However, each of these papers exhibits some drawbacks as well.[2]

In this paper, we investigate the causal impact of the move from paid to open access on citations by applying a carefully designed econometric specification to rich data for a panel of science journals. Our dataset, described in Section 2, includes all the citations indexed by Thomson ISI between 1996 and 2005 to all the articles published during that period in a sample of the top 100 titles in ecology, botany, and multidisciplinary science and biology. We add hand-collected information on the dates each volume of each journal was made available online, when if ever it was made freely available, and on which platforms. Our econometric specification is outlined in Section 3. The panel nature of the dataset allows us to control for unobserved quality using fixed effects. The variation in the date of open access across journals allows us to account for secular trends in citations affecting various vintages of content. Additional exogenous variation in the date of open access across volumes of the same journal

---

Eysenbach (2006) studied the effect of open access on citations to *Proceedings of the National Academy of Sciences* (*PNAS*) articles. See Craig et al. (2007) for a survey of research on the citation boost from open access.

[2] Evans and Reimer's (2009) specification does not adequately control for citation age profiles. We discuss this and other specification issues in detail in Section 4, replicating their method in our data to measure the direction and magnitude of the bias. Gaule and Maystre (2011) instrument for authors' endogenous decision to pay a $1,000 fee to have their *PNAS* articles openly accessible using the timing of budget cycles. However, *PNAS* may not be the best test case given that most citing scholars have institutional access to this top-flight journal and that the fee only moves the date of open access up by six months, after which there is open access to all *PNAS* articles. Davis et al. (2008) conduct an experiment in which articles from American Physiological Society journals were randomly selected to be openly accessible immediately upon publication, the rest receiving the usual fee access for the first year. The randomized design solves the problem of separating the open-access effect from unobservable quality. However, offering better access to a scattered sample of articles does not replicate the effect of providing open access to structured content on a broad platform. See Davis (2011) for a field experiment with more journals followed for longer periods of time.

allows us to account for the age profile of a volume's cites in a flexible way. It is vital to control for these secular trends and age profiles; otherwise they are easily confounded with the open-access indicator, which tends to "turn on" in later years and for certain ages of content (only after an embargo window, for example). 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.

Our first set of results, presented in Section 4, highlights the importance of the carefully designed specification. We show that the same huge effects of open access found in the early cross-sectional literature can be generated if fixed effects capturing the quality level of journal volumes are omitted. Including increasingly rich fixed effects substantially reduces the estimated open-access effect. In our preferred specification, we estimate that moving an online journal from paid to open access increases cites by around 8% for the average volume. To obtain this estimate requires us to specify a journal-specific quadratic age profile for citations. Without these age profiles, the estimate of the open-access effect is biased downward because open access tends to come during the declining portion of the age profile, leading us to find no effect.

Obtaining reliable estimates of causal effect of open access on citations is crucial for policy. As shown by the theoretical literature based on two-sided market models (Jeon and Rochet 2010, McCabe and Snyder 2005, 2007; McCabe, Snyder, and Fagin 2013), whether the open-access model comes to dominate in equilibrium against the traditional model, and whether open access is socially more efficient hinges on the elasticities of demand on the author and reader sides. But the elasticity of author demand—i.e., how much more an author would pay for readers to have better access to his article—depends on how this access translates into readership and citations. If open access quadruples citations as the early empirical literature suggested, 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. 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. Yet more important, understanding best practice for facilitating scholarly communication can have broad implications for economic growth.[3]

Our last set of results moves beyond the estimation of a single open-access effect to an exploration of possible heterogeneity in the open-access effect. Perhaps the most provocative result is that the benefit of open access appears concentrated in the higher-tier journals in our sample, a "superstar" effect. Lower-tier journals suffer a statistically significant drop in cites from open access.

That open access could actually cause a reduction in cites is surprising. We conjecture that the effect of open access depends on the channel via which it is delivered. If open access is provided by placing the article on a broad platform such as PubMed Central, which allows efficient cross referencing toward and away from the article, this may intensify the competition for citing authors' attention. Some lower quality articles may be harmed by this competition much as low productivity firms are harmed by the opening of international trade in the models of Melitz (2003) and Chaney (2008). Further evidence supporting this conjecture is provided by breaking down the effect of open access by whether it was provided via PubMed Central or just via the narrower platform of the journal's own website.

Besides the previously cited literature, the paper is most closely related to some previous research of our own, McCabe and Snyder (2013). That paper uses similar panel-data techniques to estimate the citation effect of moving from print to online access for a sample of economics

[3] 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 (see Freeman 1994 and Dosi 1998). Development of a microeconomic foundation for the relationship between scientific publication and innovation is in its nascent stage. Empirical work by Murray and Stern (2007) finds that patenting ideas first published in scientific articles reduces cites to these articles. Additional work by Fehder, Murray and Stern (2012) suggests that the reduction in cites associated with patenting is concentrated early in the life of a journal; over time, as a journal's reputation for publishing high quality scholarship increases, this negative citation impact disappears. In other words, intellectual property rights may have limited influence on knowledge shared through established two-sided journal platforms. Theoretical work by Gans, Murray and Stern (2011) considers the strategic tradeoffs involved in disclosing new knowledge via publications, patents, or both.

journals. Because none of those journals offered open access during the period studied, that sample would be unsuitable for the study of open access which is our central focus here. The present paper does provide some ancillary results that overlap with the previous paper's findings. In our preferred specification, we find a fairly precisely estimated zero for the effect of moving from print to online access. This finding echoes the finding in McCabe and Snyder (2013) of no aggregate effect of online access. In more disaggregated analysis, McCabe and Snyder (2013) do find evidence of an online-access effect through select channels, chiefly JSTOR.

Our finding of heterogeneity in the open-access effect across journals of different ranks contributes to the literature measuring the effect of the Internet on the distribution of transactions across popular and obscure products. In the journals market, McCabe and Snyder (2013) show the increase in citations from being added to JSTOR is fairly uniform across article qualities. Studies of a broader range of retail markets find that online retailing boosts sales more for products in the long tail, in markets ranging from clothing (Brynjolfsson, Hu, and Simester 2007) to video sales (Elberse and Oberholzer-Gee 2008). In the conclusion, we reconcile our finding of a "superstar" effect in the present paper with some of the contrasting results in the literature.

2. Data

Our analysis is based on a sample of 100 journals in ecology, botany, and multidisciplinary science and biology. Appendix Table A1 provides a list of the journals, which we selected as follows. We included all the journals primarily categorized as ecology by Thomson ISI in their set of indexed journals. This accounts for 60% of the titles. Of the remaining 40%, 60% were taken from botany, the most closely related subfield to ecology, and 40% from multidisciplinary science and biology, presuming that some ecology and botany research is published in such general-interest journals. We select the top journals from each category, ranked based on the standardized ISI yearly impact factors averaged over the period 1985-2004. We focus on

ecology among the subfields of hard science because it involves a manageable number of journals and because it experienced substantial growth of open access. We restricted the sample to 100 journals because of the considerable expense and effort involved for each additional journal.

The dataset merges citations data together with historical information on online availability. The citations data was acquired from Thomson ISI. For each of the 100 journals in our sample, ISI lists every article published since 1996. Each published article is linked to all cites from all of the over 8,000 ISI-indexed journals for each year from 1996 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 whether the full-text article was available online or open access. To determine online availability, we sought the date on which each journal issue was placed online either on the journal's own website or one of the major digital aggregators (JSTOR, EBSCO, ProQuest, Ingenta, Gale, and OCLC). This was a painstaking process because information is only readily available regarding current online availability, while our study requires the first date of online availability for each volume. To obtain this information, we contacted the publishers and aggregators, cross-checking their reports using libraries' electronic journal catalogs and the Internet Archive (www.archive.org), which provides regularly archived snapshots of large segments of the Web. We collected information on open access for each volume in a similar way, contacting publishers and cross checking with the Internet Archive.

The resulting dataset from these two sources includes observations for over 200,000 individual cited articles. 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 over 200 articles—without changing the results—the volume-level estimates are numerically identical to the article-level

ones because none of our right-hand side variables will 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 1996–2005. Let t index the citation year. Note the distinction between the dataset's two time indexes: $p(v)$ indexes the year the *cited volume* was published, while t indexes the year the *citing article* was published. Because most journals published ten volumes over the 1996-2005 period, our sample of 100 journals yields almost 1,000 volume observations; because the average volume is cited over a five-year span in our sample, our panel yields over 5,000 volume-citation-year observations, the basic unit of analysis for our study. These volumes received 4.8 million cites from ISI-indexed articles over our sample period.

Table 1 provides descriptive statistics for the dataset. All journals were founded by 1991. One, *Philosophical Transactions*, was the first journal devoted exclusively to science ever published, in 1665. The average volume in our sample receives almost 900 cites in a year, about four cites per article. Yearly cites to a volume has a huge standard deviation (3,928.1) as well as range, from a low of 0 (the 1996 volume of *Natural History* received no cites in 2004) to a high of 32,589 (received by the 2002 volume of the *Proceedings of the National Academy of Sciences* in 2004).

Figures 1 and 2 illustrate patterns in citations which, while interesting in their own right, will 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. Citations peak in the second year after publication, receiving 30% more than the baseline year. After that, citations gradually fall each year, falling below the baseline after six years. The pattern is quite different from what McCabe and Snyder (2013) find for economics journals. Citations to economics journals peak later, not until the fifth year after publication, and decay more slowly, taking 15 years to return to the level

in the first year after publication. Figure 2 plots secular trends in citations. Citations have a significant upward trend, rising by about 20% from 1996 to 2005. An increase in indexed journals, articles per journal, and cites per article all contribute to the trend.

Returning to Table 1, the last row provides information on indicators for online and open access. For more than half of the observations, the full volume was available online through some channel for the full year. A much smaller fraction of observations, 6%, were freely available online for the full volume for the full year. We will focus on *full* online and open access throughout the analysis. The regressions will also include indicators for *partial* online and open access—only part of a volume's content available in the indicated way during the year or all 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 varying degrees of access.

Figure 3 shows the growth in online and open access in the sample. Full-text articles started to be posted online in 1995. Online access grew quickly, becoming ubiquitous by the end of our sample, with over 80% of volumes available online in whole or part online in 2005. Open access grew more slowly. By 2005, 10% of the volumes were available via open access.

3. Methodology

To account for the count-data nature of citations in our panel-data setting, we use a fixed effects Poisson estimator with the following conditional mean:

$$E(Cites_{vt}|Age_{vt}, Access_{vt}, p(v), j(v)) = \exp\left(\alpha_v + \beta_{p(v)t} + \gamma^1_{j(v)} Age_{vt} + \gamma^2_{j(v)} Age^2_{vt} + \delta Access_{vt}\right), \quad (1)$$

where, recall, v indexes the volume (our unit of observation), $p(v)$ is the volume's publication year, and $j(v)$ is the journal in which the volume appears. $Cites_{vt}$ denotes the number of cites

received by journal volume v in year t, $Age_{vt} = t - p(v)$ the volume's age, and $Access_{vt}$ a vector of variables capturing the nature of access to the volume (whether print or online, paid or free, full or partial, etc.). The remaining variables are parameters to be estimated: α_v is a journal fixed effect, $\beta_{p(v)t}$ is a time effect possibly varying for each publication year × citation year combination, $\gamma^1_{j(v)}$ and $\gamma^2_{j(v)}$ are coefficients on a quadratic age profile separately estimated for each journal, and δ a vector of parameters capturing access effects.[4] Wooldridge (1999) provides a Poisson quasi-maximum-likelihood (PQML) estimator for equation (1), which, as long as the conditional mean is specified correctly, produces consistent estimates of the parameters for any positive conditional distribution of $Cites_{vt}$ (Poisson, negative binomial, or other). PQML is thus robust to overdispersion (higher variance than mean) or an excess of zeros relative to a Poisson distribution.[5]

Including volume fixed effects α_v in equation (1) helps remove the bias that plagued previous cross-sectional studies of the open-access effect. If higher quality articles are more likely to be published open access, the open-access coefficient in previous studies may just be picking up quality differences between open- and gated-access articles. The quadratic age profile controls for the hump-shaped pattern of cites shown in Figure 1. The flexible specification allows for an individual profile for each journal. It is important to control for the age profile to avoid, for example, attributing the natural decline in citations after age 2 with open access that might have started then. The time effects $\beta_{p(v)t}$ control for secular trends in citations such as observed in Figure 2. Without such controls, the secular growth in cites could confound estimates of the effect of online and open access, both of which tend to occur later in the sample. Estimating independent effects for each publication year × citation year combination allows the

[4] Allowing higher order polynomials up to a quartic to control for each journal's citation age profile did not appreciably change the results of interest. We report the results of the more parsimonious, quadratic, age profile.

[5] We use Simcoe's (2008) implementation of this estimator in Stata.

secular growth in citations to vary by vintage of content and for the pattern of growth to have an arbitrary shape.

The regressors of interest are contained in $Access_{vt}$ in equation (1), including an online-access indicator, equaling 1 if volume v was available online in citation year t and an open-access indicator, equaling 1 if the volume was available via open access in the citation year. As mentioned, we focus on the results for full online or open access, i.e., access in the specified way to the entire volume's content for the entire year, but also include controls for partial access.

The section concludes with a discussion of some possible threats to identifying the access effects of interest and how our methodology addresses these threats. The impossibility of separately identifying age, cohort, and time effects, called the "identification problem" (Blalock 1966), familiar from many contexts in applied microeconomics, arises here in that age, volume, and citation-year fixed effects cannot all be separately identified. Fortunately, the problem will not impair our ability to estimate the coefficients of interest. The specified volume, age profile, and publication × citation-year interaction variables are not of direct interest themselves but are only included as controls to improve the estimation of the online- and open-access variables. Estimation of these access variables is not impaired by the identification problem because the access variables vary within these controls.[6]

The access variables 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, for example, open access was having an effect or if the volume's cites happened to decay more slowly than others' for intrinsic reasons. Identification is preserved by specifying that volumes of the same journal share the same age profile. In essence, our identification assumption is that volumes of a journal that are published

[6] The age profile in Figure 1 and the citation-year pattern in Figure 2 cannot be identified if volume fixed effects are included. We identify them by including journal rather than volume fixed effects, essentially assuming that journals maintain a consistent quality level over the sample period. We do not make this assumption in our preferred specification [column (4) of Table 2 as well as all regressions reported in Table 3] because we use the finer—volume—fixed effects there.

during our relatively short sample period 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 an online-access or open-access variable turns on, we attribute this effect to the indicated change in access.

Two further threats must be overcome for our access indicators to provide consistent estimates. First, the access variables must be exogenous, i.e., orthogonal to the residual after including all the other controls on the right-hand side of equation (1). To make the discussion concrete, focus on the open-access indicator. Intuitively, this indicator will be exogenous if it is not subject to reverse causation which would arise if a publisher granted open access to a volume based on information about deviations from an expected citation profile, arising for example if the publisher pushed up the date of open access for a volume if it saw cites to the volume growing unexpectedly quickly. With volume fixed effects, the open-access indicator remains orthogonal to the error even if only publishers of highly-cited journals decided to offer open access. The average level of cites received by the content over time is swept out by the volume fixed effects. The open-access coefficient is effectively estimated from the difference in cites before and after the volume is available via open access (at the same time controlling for secular trends with fixed time effects and controlling the expected age profile). The open-access indicator also remains orthogonal to the error if authors make submission decisions based on the journal's access policy. For example, suppose the most-cited authors value open access more than do others and thus tend to submit to journals that are, or are expected to be, open access. Sweeping out the mean citations over time *for each volume* with volume fixed effects will control for this quality effect. For this sort of submission behavior to lead to an endogeneity problem, the time-series profile—not merely the level—of cites to highly cited authors would have to differ from the profile for other authors, and this difference would have to be correlated with the timing of open access.

The example of *Plant Physiology*, shown in Figure 4, helps allay such concerns about the endogeneity of open access in our specification. In 2001, the journal allowed open access to a whole tranche of volumes through 1999. After that, the journal maintained a policy of making articles available open access after a two-year "embargo" behind a pay wall. This pattern of maintaining a fixed embargo period combined with episodes in which a tranche of back issues is made openly accessible is fairly typical and seems to be based more on technological convenience than on innovations in the time series of a volume's cites. The example does not appear consistent with the possibility that the authors decided to submit to *Plant Physiology* based on when they expected their citations to peak relative to when the volume was granted open access. It is doubtful that authors published in the 1996-99 volumes understood that the whole tranche would be granted open access together in 2001 when they made their submission decisions. It is even less plausible that each successive year, submitting authors anticipated shifts in their citation peaks coinciding precisely with a reduction in the delay before open access was granted: a delay of five years for 1996 articles, four years for 1997 articles, three years for 1998 articles, and so forth. *Plant Physiology* is just one example; the picture for most other journals would be similar.

The second threat to identification is that the access variables must exhibit some independent variation from the other regressors. If all volumes of a journal were made openly accessible after the same embargo period, the open-access indicator would be completely collinear with the volume's age. As Figure 4 shows, this is not typically the case. Paradoxically, the tranche of 1996-99 volumes that *Plant Physiology* made openly available in 2001 helps identify the effect of open access on cites because simultaneously turning on the open-access indicator hits different volumes at different points in their age profiles. As mentioned, the 1996 volume is first openly accessible in its fifth year after publication, the 1997 in its fourth year, the 1998 in its third, and so forth. The 1996 volume provides information on what the citation age

profile should look like through the fourth year in the absence of open access. If the 1998 volume deviates from this pattern in 2001, say experiencing a jump relative to expectations, this jump can be attributed to the effect of the start of open access in that year. For this identification strategy to be valid, one must be able to purge secular time effects using data from other journal volumes of around the same vintage having a different pattern of open access. Our data satisfy this requirement. First, most journals in our sample are never openly accessible. For those that are, the timing of open access follows idiosyncratic patterns. In the case of the *Proceedings of the National Academy of Sciences* (*PNAS*), also shown in Figure 4, the 1996 and 1997 volumes were already open access by 2001. *PNAS* granted open access to slightly different tranche of volumes in 2001 than *Plant Physiology*.

Note that *PNAS* did not maintain a perfectly regular embargo period after 2001. While open access to the 2001 was allowed after one year; full open access to the 2002 volume was not allowed for two years. Our methodology exploits both irregularities in the embargo period and tranches of volumes being made openly available at the same time to identify the open-access effect.

4. Results

Discussion of the results is organized around two tables. Table 2, discussed in Section 4.1, demonstrates the importance of saturating the specification with a rich set of controls as does our preferred specification, showing that less rich models can produce unreliable results. Table 3, discussed in Section 4.2, interacts the variable of interest, the open-access indicator, with a suite of additional variables to uncover sources of heterogeneity in the open-access effect.

4.1 Alternative Specifications

Table 2 presents the coefficients of interest from specifications of a count-data model along the lines of equation (1), experimenting with alternative sets of fixed- and time-effect controls. The reported variables are simple indicators for full online and open access, also including analogous indicators for partial access, as well as the controls listed at the bottom of the table. To demonstrate the importance of the controls in the shaded column (4), containing the preferred specification, the columns leading up to it gradually enrich the included controls. The reported standard errors are robust to heteroskedasticity and clustered at the volume level. Regression coefficients have been converted into marginal effects interpretable as proportionate increases: a zero marginal effect, say for open access, corresponds to open access having no measured effect; a negative marginal effect corresponds to open access causing a reduction in cites; a positive marginal effect corresponds to open access causing an increase in cites. For example, a marginal effect of 0.2 corresponds to cites being 20% higher with open access than without.

Scanning the first row of the table, corresponding to the online-access effect, from left to right reveals a clear pattern. Column (1) is run without journal or volume fixed effects to mimic the early literature. Without these controls for quality we can reproduce the extraordinarily high online-access effects found in these studies. The first marginal effect, 6.436, has the interpretation that the average volume receives a 643.6% boost in citations from online compared to print access. Column (2) adds journal fixed effects, reducing the marginal boost from online access several orders of magnitude to 22.9%, still statistically significantly different from 0 at the 1% level. Column (3) adds volume fixed effects, an even richer set of quality controls than journal fixed effects, picking up changes in a journal's quality over time. The results are further reduced, to 14.7%. Column (4) adds a journal-specific quadratic age profile to the specification in column (3). This further reduces the marginal effect of online access to around 0. The

standard error falls as controls are added moving from column (1) to (4), resulting in a fairly precisely estimated 0 for the marginal effect of online access in column (4).

Next, consider the second row, corresponding to the open-access effect. (Since the online indicator equals 1 when online access is provide through both paid and open channels, the open-access indicator measures the additional citation boost from open access above and beyond online access.) Here again, the specification with few controls in column (1) leads to enormous estimates, a marginal effect of 662.4% from open access. Adding journal and then volume fixed effects in columns (2) and (3) causes the marginal effect to fall, indeed becoming a large negative number, -8.7%, in column (3), although it is imprecisely estimated. Adding the quadratic age profile in column (4) reverses the sign and increases the precision, leading to our preferred estimate of the marginal effect of open access of 8.1%, significantly different from 0 at the 1% level.

Evidently, controlling for the age profile is vital for consistent and precise estimates of the open-access effect. Figure 1 suggests why. Cites fall with increasing rapidity after age 3. If this fall is not controlled for, it will be attributed to the open-access indicator, which turns on for later citation years for most of our sample. The average volume observation is two years old in our sample, an additional 1.5 years older if online, and nearly a year older yet if available open access. Open access thus tends to be observed during the period of declining cites for the typical volume, explaining why adding a quadratic age profile leads to an increase in the measured effect. By contrast, online access typically turns on earlier, near the citation peak, explaining why controlling for the age profile reduces the estimated effect of online access, at the same time it increases the estimated effect of open access.

Although column (4) reports our preferred specification, we continue with two additional columns of results to provide a further understanding of specification issues. Column (5) examines the value of including the full set of publication year × citation year interaction terms.

Removing these terms considerably increases the estimate of the marginal effect of online access (to 121.1%) and decreases the estimate of the marginal effect of open access (to 1%). Column (6) reproduces the methodology from Evans and Reimer (2009), who attempt to control for time and age effects, not as we do, but by including lagged citations. This specification results in a significantly negative marginal effect of open access. This result can be explained by the inadequacy of lagged citations as a control for the omitted age profile. The hump-shaped age profile seen in Figure 1 means that cites sometimes rise and sometimes fall from one year to the next; on average the change in cites from year to year may be close to 0. However, as we discussed, the open-access variable turns on late in the sample when cites are falling with age. If this fall is not picked up by controls for the age profile, it will show up as a negative open-access coefficient, as we see in column (6).

4.2. Expanded Analysis of Preferred Specification

Table 3 reports on further details of the analysis using our preferred specification. For reference, column (1) reports the same regression as in column (4) of Table 2.

Taking advantage of the space to present more detail on the regressions, Table 3 reports coefficients on partial access, which were also included in the regressions in Table 2 but not reported there for brevity. The result for partial online access in column (1) is similarly small and insignificant as its full online access analogue. The result for partial open access (4.4%) is about half the size of its full open access analogue (8.1%) and is statistically significant only at the 10% level. Although we did not collect information that would allow us to measure exactly how much access was afforded by the average year of partial access, the estimates are consistent with partial access affording about half the access (either in terms of amount of content, time the content was available, or some combination) of full access. These findings for partial access hold across all the columns in the table.

The remaining columns in Table 3 look for heterogeneity in the open-access effect by providing separate estimates of the marginal effect of full open access for different conditions. Column (2) allows the marginal effect of open access to differ depending on whether or not there is also online access to the journal through some paid channel. Given that the regression includes a control for online access, the interacted open access indicators need to be interpreted carefully. The interaction of full open access with no paid access measures what could be labeled a "conversion" effect: the marginal effect of converting an existing paid online channel into an open-access channel. The interaction of full open access with some paid access measures what could be labeled an "addition" effect: the marginal effect of adding an open-access channel to an existing paid online channel. The difference between conversion and addition is that the latter case readers can access the content through more channels; so, in theory, the addition effect should be weakly larger than the conversion effect. In practice, the finding of essentially no online-access effect in the column (1) regression suggests there may not be much difference between the conversion and addition effects because an additional paid online channel may not be expected to provide a measureable additional citation boost. The results bear this out. Equality of the marginal effects for the interactions with no paid access and some paid access, 8.2% and 7.4% respectively, cannot be rejected. Both are similar to the 8.1% effect observed when the two cases were estimated together.

Column (3) allows the marginal effect of full open access to differ between the 50 top-ranked journals in our sample and the remaining 50. The journals in our sample were ranked relative to each other using the same ISI impact factor used in the procedure to select our sample described in Section 2. Appendix Table A1 provides the ranks. The marginal effect for the top-50 journals, 8.6%, is similar to the basic result we obtained before dividing journals by rank. The marginal effect for the bottom-50 journals is quite different, significantly negative, with open access leading to a 18.5% reduction in cites for these journals. Column (3) thus provides

evidence of a "superstar" effect of open access, i.e., open access benefits higher-quality journals more than lower-quality.[7]

That any category of journal would *suffer* from being made openly accessible is initially surprising and begs explanation. One possible explanation hinges on the fact that open access does not just reflect a reduction in the price of accessing an article through the existing channel but in some cases represents a fundamental change in the technology used to access the article. Open access allows access to articles directly from an Internet search (for example using Google) rather than having to go through the journal's own website. Conversion from closed to open access effectively changes the platform that readers use to access content from the narrow one of the journal's own website to a platform as broad as the Internet itself. Exposure on this broad platform improves readers' ability to find the article but also facilitates substitution away from it toward articles on other open access platforms competing for the reader's attention. In a similar way that exposing domestic firms to international trade may create winners of productive firms because of the opening of export markets and losers of unproductive firms because of competition from imports (see, e.g., Melitz 2003 and Chaney 2008), exposure on a broad platform may increase cites to high-quality articles and reduce cites to low-quality articles. This would explain the gains from open access for the top-50 journals and the losses for the bottom-50 journals found in column (3) of Table 3.

If this explanation is correct, substitution effects are likely to be important the more and broader are the platforms through which access is offered. In our sample, open access was either provided solely through the journal's website or additionally through PubMed Central, a large open-access archive. Although articles posted on PubMed Central are visible to Google and

[7] To gauge the robustness of the journal-rank results, as an alternative to the step-function specification in column (3), we allowed the effect to be a linear function of rank by including an indicator for full open access (giving the intercept) and the interaction of this indicator with the continuous rank variable (giving the slope). Consistent with column (3), the line ranges from 0.093 for the rank-1 journal to -0.227 for the rank-100 journal, the intercept and slope both significant at better than the 1% level.

other external searches, PubMed Central has enhanced capabilities for search within the platform; indeed, its internal search capabilities were considerably better than Google's during our sample period (Young 2004). Thus any substitution effects toward high-quality and away from low-quality articles were likely magnified when open access was available through PubMed Central compared to when it was solely available through the journal's own website.

To explore this idea, in column (4) we estimated separate marginal effects for open access via PubMed Central (in addition to the journal's website) and open access solely through the journal's own website. While access solely through a journal website continues to have a significantly positive effect, additional access through a potentially broader platform (PubMed Central) is significantly smaller and indeed is not significantly different from 0. Evidently PubMed Central encourages substitution toward other articles on the platform reducing the otherwise significantly positive effect of open access.

The last column of Table 3 explores possible heterogeneity in the open-access effect across subfields in our sample: ecology, botany, and multidisciplinary science and biology. The results show surprising differences. The open-access effect is significantly positive for multidisciplinary science and biology (8.4%) and botany (7.2%) but significantly negative for ecology (-10.6%). This difference is surprising given that the subfields are fairly closely aligned, so one may not have expected their readers to respond differently to open access. Further analysis provides an alternative explanation. The open-access journals in ecology are generally among the bottom-ranked whereas those for the other subfields are among the top-ranked. This can be seen more formally by looking at the correlation between the open-access indicator and journal rank across subfields. The correlation is 0.11 for ecology, -0.28 for botany, and -0.16 for multidisciplinary science and biology, all significant at better than the 1% level. Thus the heterogeneity across subfields in column (5) may just be reflecting heterogeneity in the open-

access effect across different journal ranks seen in column (3) rather than some other inherent differences across the subfields.

To analyze whether the fundamental source of heterogeneity is across journal ranks or subfields, we conducted a formal comparison of the model in column (3) against that in column (5) based on their Akaike information criterion (AIC) values. As shown in the last row of Table 3, $AIC_{(3)} = 43{,}712$ and $AIC_{(5)} = 43{,}790$, where the subscript refers to the model that that value comes from. The statistic called the relative likelihood, given by the formula

$$RL = \exp\left(\frac{AIC_{(3)} - AIC_{(5)}}{2}\right),$$

is interpreted as the probability that using the model (5) to represent the true model results in less loss of information than model (3) (see Bernham and Anderson 2002). In our case, $RL = 1.7 \times 10^{-17}$, implying there is essentially no chance model (5) involves no relative information loss. A related test using the Bayesian information criterion (BIC) would produce an even more extreme result because BIC penalizes models with more variables more than AIC.[8,9] Model (3) manages to fit better than model (5) while using one fewer parameter.

Overall, whether measured by the log-likelihood or *AIC* reported at the bottom of Table 3, the model allowing heterogeneity in the open-access effect across journal ranks in column (3)

[8] The Davidson and MacKinnon (1981) *J* test adds predicted values from model (3) as a regressor in model (5), tests the significance of this added regressor, then repeats the process reversing the roles of the two models. Unfortunately, as is often the case, this test could not be used because of colinearity between the additional regressor and included controls, causing some variables to be dropped.

[9] Another way to compare the models formally follows Vuong (1989). In the case of so-called overlapping models (models which share a number of covariates but include a subset that cannot be nested within each other), the relevant test statistic is the likelihood ratio $LR = -2\left[LL_{(5)} - LL_{(3)}\right]$, where LL is the sum of the log likelihoods across observations in the model referred to in the subscript. A positive value indicates that model (3) fits better and a negative value that (5) fits better. From information at the bottom of Table 3, one can compute $LR = 75.3$. While the test statistic itself is straightforward to compute, its distribution is extremely complicated: Vuong (1989) shows it is the weighted sum of χ^2 random variables, where the weights are the eigenvalues of a complex matrix of moments. There are only rare situations where the distribution has been computed, ours not among them. We bootstrapped the Vuong statistic 500 times using a procedure that allowed for random draws from volume clusters. Fewer than 3% of the replications generated the negative values of LR that would indicate model (5) fits better than (3). The rest of the replications were consistent with model (3) having a better fit, allowing us to conclude that model (3) fits better than (5) at better than the 5% level of statistical significance according to this test.

produces the best fit of any of the models in Table 3. The next best fitting is in column (4), allowing for heterogeneity across different open-access platforms (the narrow platform of the journal's own website and the broad platform of PubMed Central). In sum, the main sources of heterogeneity we have identified involve journal rank and the nature of the open-access platform.

5. Conclusions

Our first set of results in Table 2 provides a dramatic illustration that an appropriate econometric specification is required to identify the causal effect of open access on citations using panel data. When we omit fixed effects for journal volumes as controls for unobservable quality of the articles in the volume, we can replicate the extraordinary effects found in the previous literature, in our case finding an over 600% citation boost caused by open access. When volume fixed effects are included along with a rich set of time effects and controls for the volume's age, the estimate of the causal effect of open access falls to 8%. This positive effect is statistically significant at the 5% level, so we conclude that open access does boost cites; but the effect is much more modest than many previous estimates. Our analysis suggests 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.

Table 2 also showed 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. A lagged-citations variable does not appear to be an adequate control on its own because when it is substituted for these richer controls, the results of interest change dramatically. 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.

In analysis allowing the open-access effect to differ across different categories, the biggest source of heterogeneity was journal rank. Open access caused a significant increase in cites to the top-50 journals and a significant decrease in cites to the bottom-50 journals in our sample. One explanation for this surprising negative effect for lower-tier journals is that open access changed not just the price of accessing a volume but also the platform on which the volume is available. Placing the volume on a broad platform allows more efficient cross referencing both toward and away from it. The broader platform intensifies competition for readers' attention, possibly benefitting high-quality articles and harming low-quality articles in the same way that the opening of international trade benefits productive domestic firms and harms unproductive ones in Melitz (2003) and Chaney (2008). As suggestive evidence for this hypothesis, we found that open access only provided a significant increase for those volumes made openly accessible via the narrow channel of their own websites rather than the broader PubMed Central platform. In future work, we hope to construct a formal theory of competition for readers' attention in which we can derive comparative-statics results relating the change in citations from open access to the breadth of the open-access platform. Empirically verifying that more nuanced comparative-statics results are also supported by the data may increase the confidence that our explanation of the provocative negative result estimated here for lower quality journals.

Tying the results back to the broader policy issues considered in the introduction, the modest open-access benefit we estimate should lead to a reconsideration of the benefits of and future prospects for the open-access model. If open access were to boost citations by more than 600% as found in some of our specifications mimicking the previous cross-sectional literature, then any reasonable estimate of author demand with respect to submission fees would be so inelastic that, when plugged into two-sided-market models of the journal market (e.g., Jeon and Rochet 2010; McCabe and Snyder 2005, 2007; McCabe, Snyder, and Fagin 2013), would

generate a clear-cut case for the equilibrium dominance of open access and for its social efficiency. Our positive and significant result is not inconsistent with these possibilities, but the modest size means the case is less clear-cut.

Our finding that top-50 journals benefitted more than bottom-50 journals can be viewed as supporting a "superstar" rather than a "long-tail" benefit from enhanced journal access. By contrast, McCabe and Snyder (2013) found that being added to JSTOR resulted in a fairly uniform increase in citations across quintiles of article quality. By further contrast, long-tail benefits from the growth of Internet retailing have been found in recent studies of markets outside of journals including Brynjolfsson, Hu, and Simester's (2007) study of clothing sales and Elberse and Oberholzer-Gee's (2008) study of video sales. These contrasting findings can be reconciled if the presence of superstar or long-tail effects turns out to depend on fine market details. For example, some but not all platforms may have design features that facilitate substitution away from products in the long tail as much as toward them, possibly reducing demand for the long tail when these platforms are opened. Alternatively, reducing consumer search frictions may have different effects depending on the nature of product differentiation, possibly increasing demand for products that happen to be in the long-tail because of horizontal differentiation (i.e., unique items) but decreasing demand for products that happen to be in the long-tail because of vertical differentiation (i.e., lower-quality items). The future work discussed above may help sort out these possibilities.

References

Bergstrom, Theodore. (2001) "Free Labor for Costly Journals?" *Journal of Economic Perspectives* 15: 183–198.

Bergstrom, Theodore and Carl T. Bergstrom. (2004) "The Costs and Benefits of Library Site Licenses to Academic Journals," *Proceedings of the National Academy of Sciences* 101: 897–902.

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.

Chaney, Thomas. (2008) "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review* 98: 1707–1721.

Craig, Iain D., *et al*. (2007) "Do Open Access Articles Have Greater Citation Impact? A Critical Review of the Literature," *Journal of Informetrics* 1: 239–248.

Davidson, Russell and James G. MacKinnon. (1981) "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica* 49: 781–793.

Davis, Philip M., *et al*. (2008) "Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial," *British Medical Journal* 337: 568–573.

Davis, Philip M. (2011) "Open Access, Readership, Citations: A Randomized Controlled Trial of Scientific Journal Publishing," *FASEB Journal* 25: 2129–2134.

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: 1120–1271.

The Economist. (2012) "Brought to Book: Academic Journals Face a Radical Shake-Up," July 21. Accessed on July 31, 2012 from http://www.economist.com/node/21559317.

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 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.

Fehder, Daniel C., Fiona E. Murray, and Scott Stern. (2012) "Intellectual Property Rights and the Evolution of Scientific Journals as Knowledge Platforms," Working Paper.

Freeman, Chris. (1994) "The Economics of Technical Change," *Cambridge Journal of Economics* 18: 463–514.

Gans, Joshua S., Fiona E Murray and Scott Stern. (2011) "Contracting Over the Disclosure of Scientific Knowledge: Intellectual Property and Academic Publication," SSRN working paper abstract number 1559871.

Gaule, Patrick and Nicholas Maystre. (2011) "Getting Cited: Does Open Access Help?" *Research Policy* 40: 1332–1338.

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.

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.

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

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," *B.E. Journal of Economic Analysis & Policy* 7: Issue 1 (Contributions), Article 2.

McCabe, Mark J and Christopher M Snyder. (2013) "Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals," forthcoming, *Review of Economics and Statistics*.

McCabe, Mark J, Christopher M. Snyder, and Anna Fagin. (2013) "Open Access versus Traditional Journal Pricing: Using a Simple 'Platform Market' Model to Understand Which Will Win (and Which Should)," *Journal of Academic Librarianship* 39: 11–19.

Melitz, Marc. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica* 71: 1695-1725.

Murray, Fiona E., and Scott Stern. (2007) "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-commons Hypothesis," *Journal of Economic Behavior & Organization* 63: 648–687.

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, econpapers.repec.org/RePEc:boc:bocode: s456821.

Vuong, Quang H. (1989) "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica* 57: 307–333.

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.

Young, Jeffrey R. (2004) "Google Tests Search Engine for Colleges' Scholarly Materials," *Chronicle of Higher Education,* April 23, p. A36.

Table 1: Descriptive Statistics

	Level of Statistics	Obs.	Mean	Std. Dev.	Min.	Max.
Year journal founded	$j(v)$	100	1936.2	51.1	1665	1991
Publication year $p(v)$	v	967	2000.5	2.9	1996	2005
Citation year t	vt	5,361	2002.0	2.4	1996	2005
Cites to volume in year	vt	5,361	894.0	3,928.1	0	32,589
Online-availability indicators						
Full online availability	vt	5,361	0.55	0.50	0	1
Partial online availability	vt	5,361	0.16	0.36	0	1
Open-access indicators						
Full open access	vt	5,361	0.06	0.29	0	1
Partial open access	vt	5,361	0.02	0.15	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: Marginal Access Effects in Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Full online access	6.436***	0.229***	0.147***	0.006	1.211***	0.067
	(3.903)	(0.082)	(0.067)	(0.029)	(0.372)	(0.046)
Full open access	6.624***	0.038	-0.087	0.081***	0.007	-0.160***
	(5.803)	(0.082)	(0.100)	(0.027)	(0.089)	(0.062)
Fixed effect for source	No	Journal	Volume	Volume	Volume	Volume
Publication-year x citation-year time effects	Yes	Yes	Yes	Yes	No	No
Journal-specific quadratic age profile	No	No	No	Yes	Yes	No
Lagged citations	No	No	No	No	No	Yes

Notes: Results from Wooldridge's (1999) PQML procedure. Dependent variable is cites to a volume in a citing year. 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- and open-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. In all columns but (5), robust standard errors clustered at the journal level reported in parentheses. In column (5), robust standard errors are clustered at the volume level because the variance matrix associated with clustering at the journal level was not invertible. Re\gressions run on sample of 5,361 observations; some observations may be dropped when moving to a richer specification if cites are constant within a fixed-effect group. Column for preferred specification shaded. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 3: Detailed Analysis of Access Results

	(1)	(2)	(3)	(4)	(5)
Partial online access	-0.005 (0.021)	-0.006 (0.021)	-0.004 (0.021)	-0.008 (0.021)	-0.006 (0.021)
Full online access	0.006 (0.029)	0.006 (0.029)	0.010 (0.029)	0.001 (0.029)	0.006 (0.029)
Partial open access	0.044* (0.024)	0.042* (0.023)	0.046* (0.024)	0.039* (0.023)	0.045* (0.024)
Full open access	0.081*** (0.027)				
A. Interacted with no paid channels		0.082*** (0.027)			
B. Interacted with some paid channels		0.074*** (0.021)			
A. Interacted with top-50 journal			0.085*** (0.027)		
B. Interacted with bottom-50 journal			-0.185*** (0.059)		
A. Interacted with availability on PubMed and journal website				0.046 (0.033)	
B. Interacted with availability only on journal website				0.072*** (0.085)	
A. Interacted with multidisciplinary science and biology					0.084*** (0.027)
B. Interacted with botany					0.072** (0.034)
C. Interacted with ecology					-0.106* (0.060)
χ^2 test statistic for A = B or A = B = C		1.7	17.1***	7.3**	10.0***
Log-likelihood (*LL*)	-21,661	-21,654	-21,612	-21,639	-21,650
Akaike information criterion (*AIC*)	43,808	43,795	43,712	43,766	43,790

Notes: Column (1) is the same regression as in column (4) of the previous table, also displaying results for partial access omitted from previous table for brevity. All regressions in this table use Wooldridge's (1999) PQML procedure. Dependent variable is cites to a volume in a citing year. 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 fixed effects for individual journal volumes, publication year x citation year effects, and a quadratic age profile for each journal. Robust standard errors clustered at the journal level reported in parentheses. Regressions run on sample of 5,361 observations; some observations may be dropped when 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.

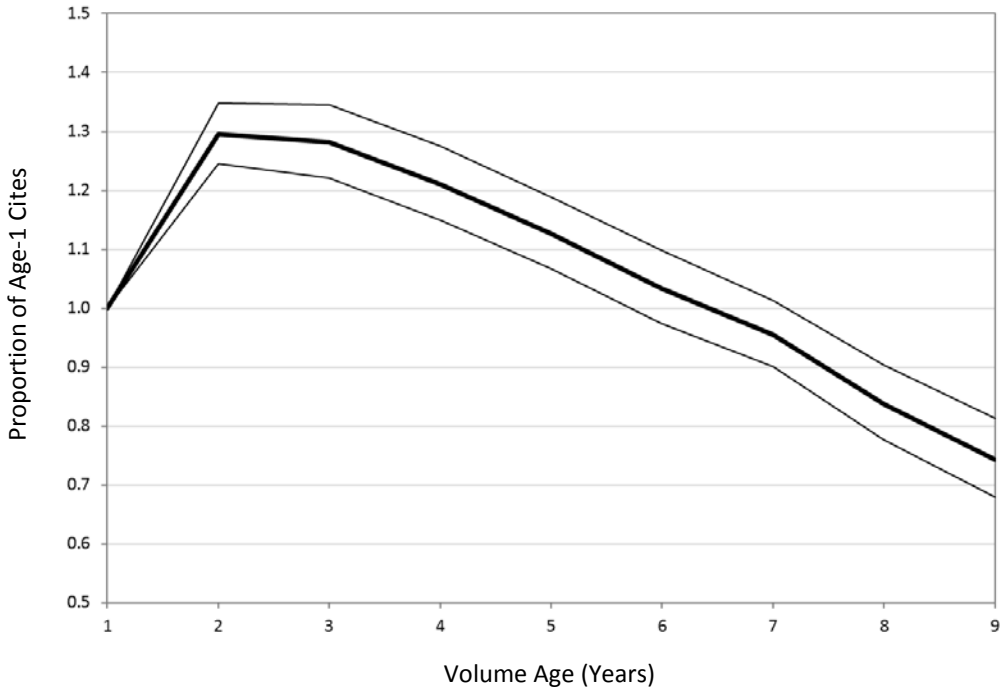

Figure 1: Age Profile of Sample Citations. Bold curve is plot of a set of fixed age effects from Poisson quasi-maximum-likelihood procedure suggested by Wooldridge (1990) for panel count data, implemented by Simcoe (2007). Coefficients converted into incidence rate ratios (IRRs) before graphing. Regression also includes a set of citation-year fixed effects and a set of journal fixed effects. Lighter outside curves bound the 95% confidence interval based on robust standard errors on the IRRs clustered by journal.

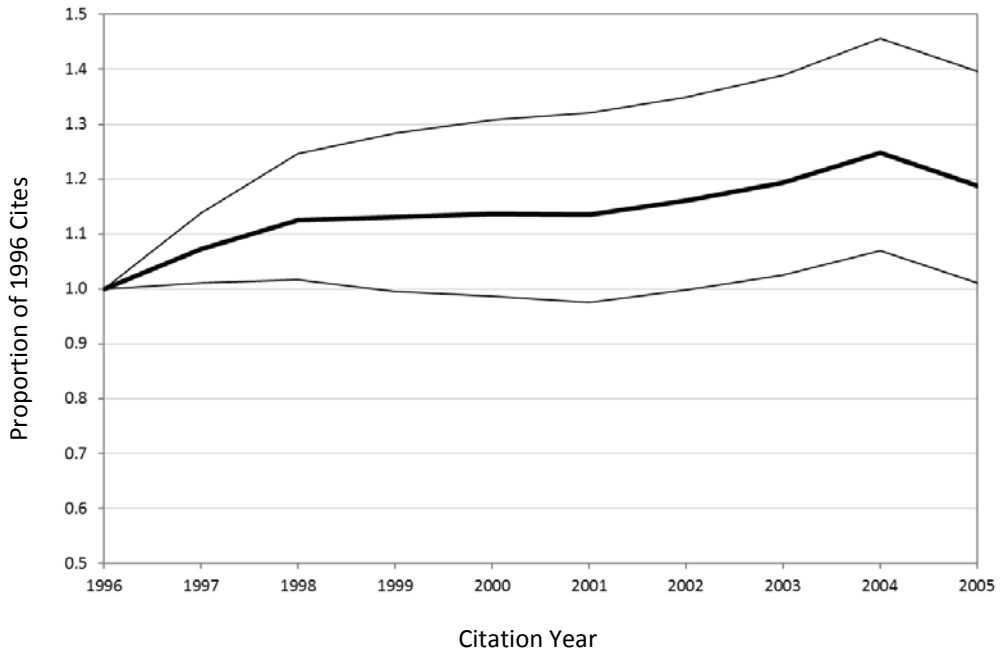

Figure 2: Secular Trends in Sample Citations. Bold curve is plot of a set of fixed citation-year effects from Poisson quasi-maximum-likelihood procedure suggested by Wooldridge (1990) for panel count data, implemented by Simcoe (2007). Coefficients converted into incidence rate ratios (IRRs) before graphing. Regression also includes a set of volume-age fixed effects and a set of journal fixed effects. Lighter outside curves bound the 95% confidence interval based on robust standard errors on the IRRs clustered by journal.

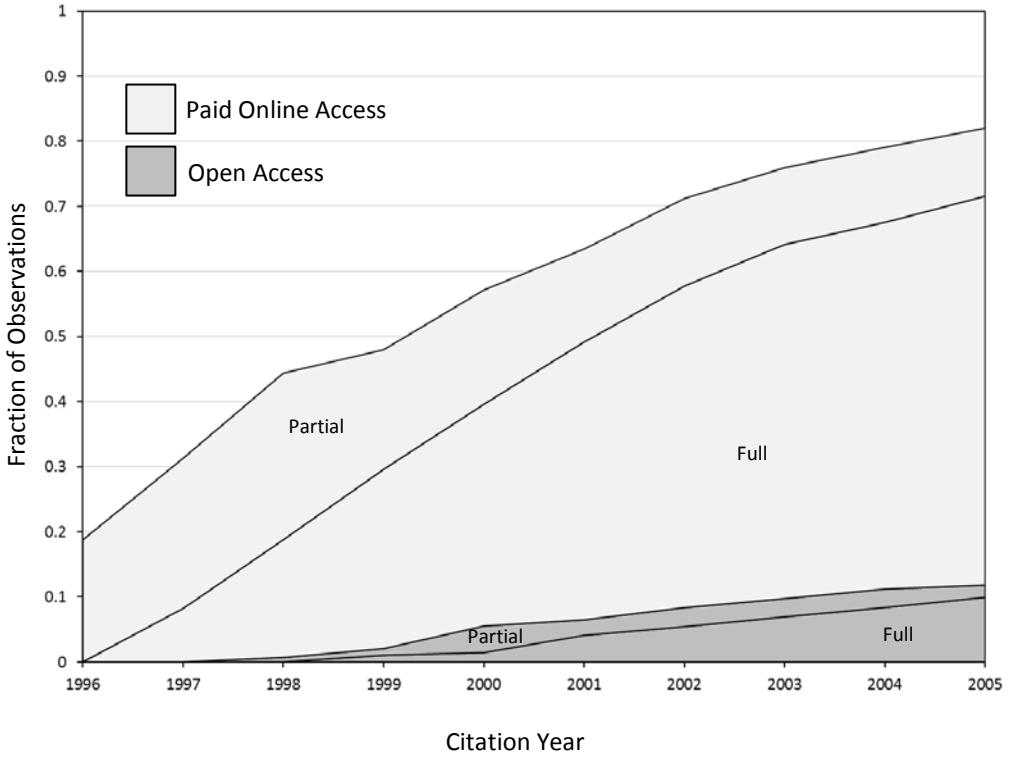

Figure 3: Growth of Online and Open Access in Sample. Total shaded region is fraction of volume observations in that citation year having some online availability, whether open or paid. "Full" denotes availability of all articles for entire year via indicated channel. "Partial" denotes some access via indicated channel but not all articles and/or not for entire year.

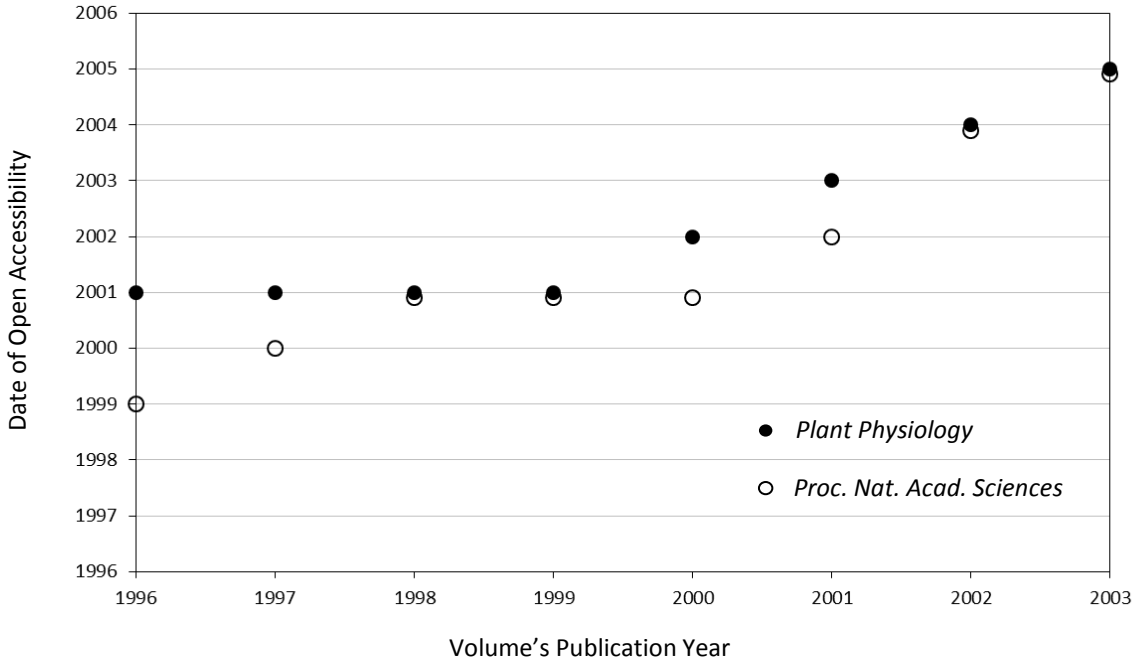

Figure 4: Pattern of Open Accessibility for Example Journals

Appendix Table A1: Journals in Sample

Ecology				Botany		Multidisciplinary Science and Biology	
Rank	**Journal**	**Rank**	**Journal**	**Rank**	**Journal**	**Rank**	**Journal**
5	*Annual Rev. Ecol. & Systematics*	63	*Sarsia*	4	*Plant Cell*	1	*Proc. National Acad. Sciences*
6	*Advances Ecol. Res.*	64	*Environ. Bio. Fishes*	9	*Annual Rev. Phytopathology*	2	*Nature*
7	*Ecol. Monographs*	65	*New Zealand J. Ecol.*	10	*Plant Physiology*	3	*Science*
8	*Trends In Ecol. & Evolution*	66	*Ecol. Modelling*	12	*Plant Molecular Bio.*	16	*Proc.: Bio. Sciences*
11	*Amer. Naturalist*	68	*Acta Oecologica*	15	*Planta*	17	*Philosophical Trans.: Bio. Sciences*
13	*Evolution*	69	*J. Tropical Ecol.*	18	*Molecular Plant-Microbe Interactions*	49	*Amer. Scientist*
14	*Ecol.*	70	*Agricultural Ecosystems & Environ.*	19	*Plant Cell & Environ.*	55	*Annals New York Acad. Sciences*
20	*J. Animal Ecol.*	71	*Pedobiologia*	21	*Botanical Rev.*	56	*Naturwissenschaften*
22	*Behavioral Ecol. & Sociobiology*	72	*Biochemical Systematics & Ecol.*	24	*Photosynthesis Res.*	58	*Comptes Rendus Acad. Sciences Serie III*
23	*J. Ecol.*	74	*J. Soil & Water Conservation*	28	*Theoretical & Applied Genetics*	60	*Proc. Japan Acad. Series B*
25	*Marine Ecol.*	76	*Amer. Midland Naturalist*	29	*New Phytologist*	67	*Trans. Royal Soc. South Africa*
26	*Paleobiology*	77	*Rangeland Ecol. & Manag.*	32	*Plant & Cell Physiology*	73	*J. Royal Soc. New Zealand*
27	*Ecol. Applications*	78	*J. Arid Environ.*	33	*Protoplasma*	82	*South African J. Science*
30	*Oecologia*	79	*J. Natural Hist.*	34	*J. Experimental Botany*	90	*Current Science*
31	*Oikos*	80	*Wildlife Soc. Bull.*	35	*Physiologia Plantarum*	92	*Interciencia*
37	*Microbial Ecol.*	81	*Proc. Acad. Natural Sciences Phila.*	36	*J. Phycology*	94	*Archives Sciences*
38	*J. Applied Ecol.*	83	*Population Ecol.*	39	*Amer. J. Botany*	96	*Ohio J. Science*
42	*J. North Amer. Benthological Soc.*	84	*J. Freshwater Ecol.*	40	*Phytopathology*		
43	*Functional Ecol.*	85	*African J. Ecol.*	41	*Annals Missouri Botanical Garden*		
44	*Theoretical Population Bio.*	86	*Rev. Ecol.-La Terre Et La Vie*	48	*Physiological & Molec. Plant Pathology*		
45	*J. Evolutionary Bio.*	87	*South African J. Wildlife Res.*	51	*Systematic Botany*		
46	*J. Experimental Marine Bio. & Ecol.*	88	*Revista Chilena Hist. Natural*	62	*Int. J. Plant Sciences*		
47	*Conservation Bio.*	89	*Northwest Science*	75	*Functional Plant Bio.*		
50	*J. Chemical Ecol.*	91	*Canadian Field-Naturalist*	100	*J. Torrey Botanical Soc.*		
52	*Evolutionary Ecol.*	93	*Western North Amer. Naturalist*				
53	*J. Biogeography*	95	*Bull. Amer. Museum Natural Hist.*				
54	*Polar Bio.*	97	*Biocycle*				
57	*J. Wildlife Manag.*	98	*Natural Hist.*				
59	*Bio. Conservation*	99	*Russian J. Ecol.*				
61	*Biotropica*						

Notes: Classification into ecology versus botony versus general science according to ISI primary subject. Journals ranked 1-100 within our sample using ISI impact factor averaged over 1984-2004.