UNC CHARLOTTE ECONOMICS WORKING PAPER SERIES

SITED, SIGHTED, AND CITED: THE EFFECT OF JSTOR IN ECONOMIC RESEARCH

Michael R. Ward
Craig A. Depken, II

Working Paper No. 2016-018

THE UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
BELK COLLEGE OF BUSINESS
DEPARTMENT OF ECONOMICS
9201 University City Blvd
Charlotte, NC 28223-0001
October 2016

ABSTRACT

By increasing the ability to discover, access, and use academic journal articles, the Internet has become the dominant mode by which scholars stay abreast of the scholarly literature. This new technology may have impacted the referencing pattern as well as the research productivity of scholars. These hypotheses are tested in the area of economics using a natural experiment of access to the JSTOR article archiving service. We find evidence that access to journals through JSTOR leads economists to refer more to JSTOR journals at the expense of non-JSTOR journals, that is, JSTOR access induces substitution away from journals not available in the JSTOR archive. Furthermore, greater JSTOR access increases the amount of economic research at an institutional level.

Sited, Sighted, and Cited:

The Effect of JSTOR in Economic Research

Michael R. Ward
University of Texas at Arlington
and
Zentrum für Europäische Wirtschaftsforschung

Craig A. Depken, II
University of North Carolina at Charlotte

ABSTRACT

By increasing the ability to discover, access, and use academic journal articles, the Internet has become the dominant mode by which scholars stay abreast of the scholarly literature. This new technology may have impacted the referencing pattern as well as the research productivity of scholars. These hypotheses are tested in the area of economics using a natural experiment of access to the JSTOR article archiving service. We find evidence that access to journals through JSTOR leads economists to refer more to JSTOR journals at the expense of non-JSTOR journals, that is, JSTOR access induces substitution away from journals not available in the JSTOR archive. Furthermore, greater JSTOR access increases the amount of economic research at an institutional level.

***JEL* Codes**: O30, D20, L96

Keywords: Internet, Scholarly Communication, Academic Research

We wish to thank Emek Basker, Jeff Desimone, Tobias Klein, Keld Laursen, Babur De Los Santos, Mahmut Yasar, and workshop participants at DRUID, Telecom ParisTech, UT Arlington, UNC Charlotte, and ZEW for helpful comments. This research was conducted in part while Michael Ward was a visitor at Zentrum für Europäische Wirtschaftsforschung GmbH (ZEW) and it benefited greatly from discussions with scholars there. All remaining errors are our own.

Sited, Sighted, and Cited:
The Effect of JSTOR in Economic Research

I. Introduction

Academic scholars who earned their degrees within the past decade are often at a loss to describe how they would stay current with the work in their field without using the Internet, as their elders once did. Many fail to conceive how scholarship could have commenced in those dark ages before the light shined down fiber-optic wires. Just as the Internet has transformed book retailing (Brynjolfsson and Smith, 2000), music retailing (Zentner, 2008), concerts (Krueger, 2005), and the insurance industry (Brown and Goolsbee, 2002), it appears the Internet has had a major effect on the "research industry." The Internet has had profound and lasting effects on the way academics disseminate the knowledge they create, how they discover knowledge created by other researchers, and how they communicate with each other. Have Internet-enabled tools measurably affected how research is conducted or the productivity of the average researcher? To address these questions, we exploit a natural experiment in which scholars obtained access to a major online scholarly tool at different times and with different levels of functionality.

Specifically, we examine the impact of one particular Internet tool, the JSTOR journal archive, on one particular discipline, economics. JSTOR is the first large scale Internet-based journal article storage, search, and retrieval service.[1] Scholars at research institutions that subscribe to JSTOR can easily find and read at their desktops archived articles published in hundreds of journals over the past century or more. We exploit the fact that the time of first subscription for institutions and the number of journals available to scholars at those institutions from JSTOR

[1] JSTOR is not the only service of this kind, but it is the oldest and, perhaps, the most widely known.

varied across institutions when the service began in 1997 and JSTOR was a near monopoly in this type of service. We find that once a journal's previously published articles become available to economists at an institution, these economists refer to these journals more often and refer to excluded, i.e., non-JSTOR, journals less often. Thus, JSTOR appears to have lowered the relative cost of the former causing a substitution away from the later. Moreover, we find evidence that JSTOR led to an increase in the research productivity of these economists as measured by the rate at which they publish but that this increase was experienced only at lower ranked institutions.

This study does not attempt to gauge the social welfare implications of the impact of JSTOR in the economics discipline. Yet, the value to society of increased research productivity across all areas of knowledge creation might be immense. Granted, our application focuses on the production of economic research which rarely leads to a demonstrated link to commercialization via new products or processes.[2] As the Internet has been embraced by almost all academic disciples, if similar mechanisms have been at work in engineering, biology, physics or medicine, they could be helping to increase the pace at which academic research output in these fields generates ideas that are commercially exploitable. Moreover, this mechanism is likely quickening the pace of academic research output as continuous development of newer Internet applications allow for ever richer scholarly communication and collaboration. If so, the pace of new inventions emanating from this research may be accelerating.

II. The Internet, the Academy and JSTOR

The academy was instrumental in the development, use, and popularization of the Internet. Many of the pioneering applications of the early Internet were developed on university campuses

[2] There are a few notable exceptions including the Beta from the Capital Asset Pricing Model, the Black-Scholes option pricing model, the prisoner's dilemma, and, perhaps, the game Monopoly.

such as the Archie search engine at McGill University in 1990, the Gopher document linking system at the University of Minnesota in 1991, and the Mosaic browser at the University of Illinois in 1993. Non-technologists in academia were early adopters of these and other Internet tools. The effect of the Internet on scholarly communication is evident in its facilitation of collaboration between distant scholars, improved arrangements for conferences and seminars, the development of course websites and online courses, the creation of searchable working paper archives, as well as published journal article retrieval. This early adoption by universities was significant enough that students attending universities during this period became conduits through which others would gain exposure to the Internet (Goldfarb, 2006).

There have been investigations of the effect of the Internet on university research as well. There is evidence that the Internet has lowered many geographical and international barriers that hampered economics and finance researchers outside of elite universities (Kim, Morse and Zingales, 2009). Bitnet adoption (an early version of the Internet) at universities appears to have led to changes in electrical engineering research productivity, especially at lower tier schools (Agrawal and Goldfarb, 2008). Another factor is that the Internet created alternative venues to peer-reviewed journal articles for the dissemination of research for the top researchers (Ellison, 2011).

However, many Information and Communication Technology (ICT) applications provide entertainment value the consumption of which might crowd out productive time (Belo, 2014). Hamermesh and Oster (2002) provide evidence suggesting that ICT provides researchers with "toys" as well as "tools" and may merely serve to add to the consumptive value of being an academic without enhancing research productivity. Examining the effect of JSTOR, exhibiting

little entertainment value, allows us to test for productivity effects distinct from entertainment effects.

JSTOR, as an Internet application, is hypothesized to have enhanced research productivity at universities who have access. JSTOR was initially conceived in 1993 as a digital solution to the then-growing problem of space constraints at many research libraries. As binding space constraints were overmatched by an ever-increasing knowledge base available in various media, there was a strong demand for a way to reduce library possession of printed, bound, shelf-riding, and dust-gathering journals without sacrificing access to the knowledge encapsulated in them.

As a panacea to the binding space constraints, JSTOR appears to have failed. Many research libraries have reduced their possession of physical copies of many of JSTOR archived journals but have added additional non-electronic holdings in other areas. However, JSTOR's success as a research resource facilitating scholars' access to scholarly literature has exceeded the original expectations of the founders of JSTOR. Although JSTOR began in 1997 with only ten archived journals and a dozen "test bed" institutions as subscribers (Schonfield, 2002), as of January, 2013, the archive contained over 50 million pages from nearly 1,700 academic journals with 800 participating publishers and more than 10,000 participating subscribers from 160 countries. Usage has steadily grown to the point that users downloaded more than 74 million articles in 2010. If traffic to the web site is any indication, it appears evident that increasing numbers of publishers, subscribing institutions, and scholars have benefited from the development of the JSTOR archive.

The first journals and institutions included in JSTOR tended to be more research oriented. JSTOR management consciously decided to first archive the journals that were most widely read and had the largest number of older volumes so to maximize the physical amount of shelf space

released. Similarly, subscribership diffused from the leading research institutions to progressively weaker research institutions. Many of the leading US institutions were charter members at the time of JSTOR's launch and there were almost 200 US subscribers by the end of 1997. Some non-US institutions obtained access during 1997 but non-US subscribership only accelerated in 1999-2000. Among the non-US subscribers too, the leading institutions tended to be earlier adopters. By the middle of the first decade of the 21st century, many publishers were beginning to provide Internet access to the archive of their published journal articles.[3] To avoid misattribution, our analysis focuses on the period before this became common.

These patterns of journal incorporation and institutional access to JSTOR, from the most research intensive journals and institutions to those less so, have implications for our estimation strategy. First, it is important to account for journal quality when measuring JSTOR's effect on the likelihood of referencing a journal. This will typically be done with journal fixed effects. Second, the distribution of JSTOR to institutions was not random. Thus, it is possible that JSTOR effects will be biased since early adopters of JSTOR were more research intensive. Again, we will generally include institution fixed effects or intertemporal changes so that our estimates reflect only the increased referencing and publishing due to JSTOR for a given institution.

III. A Simple Model of Research Production

Despite the tremendous growth in Internet traffic, it is not immediately clear whether JSTOR or other online "tools" actually enhance research output, either in quality or quantity. We adopt a simple model of the academic research production process using standard neo-classical theory. The model provides a framework in which to develop testable hypotheses regarding the

[3] Some publishers do this by redirecting from their site to the JSTOR archive.

impact of JSTOR on the quantity and quality of economic research. Consider that researchers choose among multiple inputs to a research project, including co-authors, colleagues, graduate students, statistical software, library resources, their stock of human capital, human capital they may acquire for the project, and combine them in a rather complicated manner to produce research findings, usually presented in the form of a peer-reviewed journal article.

Part of the process of producing the final output is to address how previous authors have dealt with the problem and how the current project relates to the existing literature. To accomplish this, authors usually refer to recent and not-so-recent papers published elsewhere. We view JSTOR as lowering the costs of accessing JSTOR archived journals relative to journals not included in the JSTOR archive, thereby potentially altering the mix of references used by a researcher in her pursuit of new knowledge. As such, standard isocost/isoquant analysis can be used to determine the expected effects on research inputs and output.

Consider an academic research production function $q = f(x_1, x_2, \ldots, x_N)$, where q represents the amount of research produced by a researcher, the x's represent the various inputs used to produce research, and f() represents a production function with standard properties. Research output has both quantity and quality dimensions and fully specifying the production function is difficult as it may involve collaboration effects from colleagues and students as well as scale or scope economies. These considerations are beyond the scope of this analysis. For our purposes, we assume that library resources, and the literature review in general, are separable from the other inputs used in research production.

We assume researchers face shadow prices of inputs, w, and are rewarded according to some shadow price of output, p. Note that the prices and costs need not be those incurred by the institution. For example, the researcher usually faces a zero pecuniary cost to using JSTOR. The

relevant costs for our analysis are the time and effort required to locate and use the relevant prior literature. Similarly, the reward, p, to the researcher need not be the same as to his or her institution. It likely includes advancement toward promotion and merit raises, but could include, for example, income from grants, travel opportunities, and possible future consulting fees. We assume that incentive problems are sufficiently addressed so p is positive and that researchers maximize a shadow profit function:

$$\Pi = pf(x_1, x_2, \ldots x_N) - \sum_{i=1}^{N} w_i x_i.$$

That is, researchers face an optimization problem analogous to the optimization problem facing any neo-classical firm. Quite generally, researchers equate the marginal rate of transformation, MP_i/MP_j, with the ratio of factor input prices, w_i/w_j, for $i \neq j, i = 1...N, j = 1...N.$. Let x_1 and x_2 be the processes of searching for, reading, and incorporating an article from journals 1 and 2 into one's research. These searches have marginal benefits of MP_1 and MP_2 and costs of w_1 and w_2. Access to journal 1 through JSTOR, but not journal 2, is assumed to reduce w_1, but not change w_2. As a consequence, we expect the researcher to make more use of articles in journal 1. This will involve a substitution effect away from articles found in journal 2 (see Figure 1). Since the costs of production for any level of research will have declined, we expect a scale effect as researchers produce more and/or better research. This could be the case either if a more thorough understanding of past research improves the quality of current research or if time saved through easier access to past research allows a researcher to work on more projects. The net effect on substitute inputs is ambiguous but the direct effect on JSTOR accessible journals is unambiguously toward greater usage. The goal of our analysis is to determine if we can detect 1) an increase in usage of JSTOR accessible journals (the direct effect), 2) a decrease in usage of journals not

accessible from JSTOR (an indirect effect), and 3) an increase in research output and/or quality (a scale effect).

IV. Journal and JSTOR Data

The data for the analysis come from JSTOR's own records of journals archived and institutions' access arrangements and from ISI's Social Science Citation Index (SSCI) database for the economics discipline from 1985 through 2006. After matching these data sources by institution, journal and year, we limited the sample to the years 1991-2006, the top 100 journals, and the top 500 institutions worldwide.[4] This set of journals insures a balanced panel of journals continuously in operation over the sample period. Since this is a broad definition of research producing institution, some analyses are performed for just the top 100 institutions. Our complete sample includes over 57,000 articles generating almost 500,000 backward citations.

Information about research institutions' access to JSTOR economics and business collections was made available by JSTOR.[5] Institutions could subscribe to any of seven different collections that include economics related journals archived by JSTOR (Arts & Sciences I, II, III, IV, and Complement, and Business I and II). Each collection included a set of specified journal titles that increased in number over time and were not necessarily mutually exclusive. Scholars at these institutions had access to a covered journal's complete archive except for a few years prior to the present as dictated by the journal's 'moving wall.' Most journals had opted to embargo the most current issues, usually three years' worth, from JSTOR to avoid cannibalizing journal

[4] Our sample period ends in 2006 because after that time several other online resources became available and more popular. For instance, Google Scholar was launched in 2004 but did not gain much of its functionality until after 2006. CiteSeer was launched in 1998 but focused primarily on a reduced universe of documents posted on publicly available websites. Since 2006, More recently, RePEC (Research Papers in Economics), SSRN (Social Science Research Network), and Google Scholar have become popular sites for searching archived economics publications and working papers.

[5] We thank Andrew McLetchie at JSTOR for his assistance.

subscriptions and sales. We obtained information about the dates that different institutions subscribed to the different collections as well as the date that different journals were included into each collection. While JSTOR was the clear first-mover for this service, by 2006 many journals were also beginning to offer access to their archived past issues (McCabe and Snyder, 2015). Those journals generally regarded as the most important were usually archived by JSTOR first with less highly cited journals being added to collections over time. From this information we can generate a three way electronic access dummy variable by institution, journal, and year.

Most of the journal titles eventually archived by JSTOR are also among the more than 160 economics journal titles indexed by ISI. The sample of articles we use includes the top 100 journal titles as determined by the ratio of forward to backward citations over the entire sample. These include all of the most important general economics journals, e.g. *The American Economic Review* and *The Journal of Political Economy,* and top field journals, e.g., *The Journal of Money, Credit and Banking,* and *The Rand Journal of Economics*. Out of these titles, 31 were archived in JSTOR by the end of the sample. Table 1 lists the included journals and when they were first available through JSTOR.

The JSTOR sample includes 3,602 institutions that subscribed to any of the collections that eventually included the different economics-related journal titles. These subscribers included most research universities worldwide but also included lesser-known colleges, government agencies, non-governmental agencies, private consultancies, and many high schools. Since our focus is on the 'production' of journal articles, most of those entities that are primarily consumers of journal articles are not included in this analyses. Ultimately, the sample includes top research institutions and some institutions not particularly known for their research output. For our broad view of what constitutes a research institution, our sample included the top 500 institutions in the world in terms

of publication output over the sample period.[6] Even though JSTOR and ISI were begun in the US and have primarily an English language focus, about one-third of these institutions are outside of the US. Figure 2 indicates that research institutions had access to more journals over time and more so for higher ranked institutions. At the same time, Figure 3 indicates that as late as 2006, there was still considerable variation in JSTOR availability within a tier of research institution.

Information about each research institution's scholarly output comes from ISI's "Web of Knowledge" service that contains the Social Science Citation Index (SSCI). For all issues of all of the included journals, we have access to general bibliographic and citation information. We include only 'articles' and 'notes' as distinct from 'letters,' 'front matter,' or any other designation. This represents nearly 60,000 articles over the sample period of which over two thirds were authored by scholars at the top 500 institutions and these articles collectively made close to half a million citations. For the purpose of this study, variables of interest for an article include the journal title, date of publication, the authors' institutional affiliations, and, for each of the article's references, the journal referred to and the year of the referred-to publication.[7] Also available for each article is a variable indicating the number of citations it has received.[8]

We distinguish "elite" institutions from other researching institutions based on within-sample calculations of publishing production. We suspect that access to journals through JSTOR will have a larger impact on researchers at institutions less well-known for their research. These institutions tend to have less extensive library facilities, tend not to have as many researcher colleagues, and host fewer research presentations. In a sense, JSTOR may serve to reduce the

[6] This measure is described more fully below. Essentially, each author of an article with N authors is attributed with $1/N$ authorship. Moreover, articles are weighted be the ratio of incoming to outgoing cites to the journal. In our sample, we rarely observe zero publications for an institution and year.
[7] In fact, due to data limitation issues, we include only the first 200 citations made by an article. Fewer than 10 articles, usually survey articles, include more than 200 citations.
[8] The citations are as of August, 2007 which introduces data truncation.

comparative advantage of elite research institutions. For some of our analyses, we divide the sample of 500 top institutions into the top 100 and the 101st through 500th institutions.

V. The Effect of JSTOR on Referencing Patterns

We first document that JSTOR affected the referencing patterns of scholars writing in peer reviewed economics journals. We construct a balanced panel measuring the number of references made by scholars at institution *i* to journal *j* in year *t*, $Referenes_{ijt}$. Because of the "moving wall" in which the most recent articles, usually those published in the previous three years, were not available through JSTOR, our count measure omits these references. Our sample includes the top 500 institutions, the top 100 journals, and the years 1991 through 2006, for a total of 800,000 observations. Likewise, we construct a dummy variable indicating whether researchers at institution *i* had JSTOR access to journal *j* at time *t*, $JSTOR_{ijt}$. We expect that, all else equal, JSTOR access will increase referencing to a journal. In our sample, no institutions had access to any journals at the beginning of the sample and some did not at the end. Moreover, some institutions gained access to different sets of journals and did so in different years. Thus, we have three-dimensional variation in our variables of interest.

As a first pass, we simply graph the number of references to journals for years before and after an institution obtains access to the journal.[9] Figure 4 depicts the number of references aggregated across journals and institutions for elite and other institutions. It appears that there are more references to a journal once it is available through JSTOR. However, the effect also appears to be larger the longer a journal has been available through JSTOR. This could simply reflect a secular increase in referencing over time or that scholars more fully adopted JSTOR as they gained

[9]We exclude references to material published within the previous three years to account for JSTOR's "moving wall."

more experience with it. This could explain why this post-adoption trend is stronger with the more elite institutions that tended to be earlier adopters.

In addition to a direct effect, we also hypothesize an indirect effect in which a researcher will be less likely to reference any particular journal when she has access to an increasing number JSTOR archived journals. Except for references to seminal papers, support for a proposition a researcher wishes to make may be found in multiple previously published papers. In this sense, references "compete" with each other across journals. References to any one paper may fall when the "cost" to locating and referencing a substitute paper falls. To test this hypothesis, we construct a variable equal to the number of other journals that researchers at institution *i* have access to in year *t*, $NumJSTOR_{ijt}$. If references compete with each other, then the number of references to journal *j* should fall as $NumJSTOR_{ijt}$ rises.

Our specification relates the count of references to our measures of JSTOR access using the negative binomial estimator.[10] The count of references differs across journals, as some are generally more highly cited, across institutions, as some exhibit more research productivity than others, and over time, as there is a gradual secular increase in the number of references per article. We are not interested in modeling the characteristics of journals or institutions that are the cause of this variation in references. Instead, we adopt three-way fixed effects that proxy for the specific bundle of characteristics of a journal, institution, and year. Implicitly, we are assuming that these bundles of characteristics do not change over the 16 years of the sample period, or do not do so in a way correlated with JSTOR access. Our estimating equation becomes:

$$f\left(References_{ijt}\right) = exp\left(\beta_0 JSTOR_{ijt} + \beta_1 NumJSTOR_{ijt} + \delta_i + \delta_j + \delta_t\right), \qquad (1)$$

[10] Since over ninety percent of all institution, journal, and year cells have zero references, we also estimated a logit model where the dichotomous dependent variable took a value one if there was any reference. These are similar to the negative binomial specification and are available from the authors upon request.

where we expect $\beta_0 > 0$ and $\beta_1 < 0$ and $\beta_0 > |\beta_1|$. Note that this specification simply tests for differences in levels and does not try to model the learning that appears to be present in Figure 4.

Table 2 reports the summary statistics for the sample used to estimate equation (1). The three columns are for the sample that includes all top 500 institutions, and this group divided between the top 100 and the rest. To address the "moving wall," we only examine references to articles that have been published three years prior to the referring article. For the complete sample, on average a journal receives about one-fourth of a reference from researchers at an institution per year. However, due to their higher level of journal article production, top 100 institutions are about 10 times more likely to reference a journal than the next 400. Across all institutions and years, including years before JSTOR existed, researchers had access to a journal about 3% of the time. By 2006, the fraction of the journals to which these researchers had JSTOR access had risen to 9%.

Table 3 reports the results of the estimation of equation (1) using these data. As in table 2, the first column is for the entire sample of 500 top institutions while the next two are for the top 100 institutions and the rest. The first column's specification includes fixed effects for each of 16 years, 100 journals, and 500 institutions. While the individual estimated coefficients are not reported, each set of dummies is jointly significantly different from zero in all three columns. These sets of dummies indicate a secular increase in the number of references over time, indicate that journals generally perceived to be higher quality are referenced more often, and indicate that authors from institutions general perceived to be of higher quality are reference more.

The variables of interest are those relating to JSTOR access. For all three samples, JSTOR access to a journal significantly increased references to that journal and an increase in the number of other journals to which an institution has access significantly reduced references to the journal.

Moreover, the magnitude of the first effect is substantially larger than the latter effect.[11] As hypothesized, obtaining JSTOR access to a journal increased references to that journal by almost 20% and decreased references to another journal by about three-quarters of a percent on average. In addition, it appears that the referencing patterns at top 100 institutions were affected to a larger extent than at non-top 100 institutions. Finally, the direct effect of JSTOR access to this journal and indirect effect from access to alternative journals roughly offset each other when calculated at the sample mean. This is consistent with large distributional effects but no appreciable aggregate effect on referencing (McCabe & Snyder, 2015).

VI. The Effect of JSTOR on Research Productivity

The above results are consistent with previous findings. Researchers appeared to use JSTOR to keep current with the relevant prior literature. Indeed, it would be surprising if researchers had not made use of Internet-based journal article archives. However, these results do not indicate that their research benefited from increased JSTOR use although the archive both dramatically decreased the time spent performing a literature review and increased a researcher's access to the current state of the art. As these tasks represent a small portion of inputs to a research article it is not clear that even drastic reductions in the costs of this narrow aspect of conducting research would translate into observable productivity.

We measure economic research output using various counts of journal articles. Since our raw data include every article published in these 100 journals, we aggregate all articles authored by researchers affiliated with the top 500 institutions for each of the 16 years. Output is measured three ways: 1) the number of publications by researchers at the institution and published in the

[11] For ease of presentation, the coefficient values on the Number of JSTOR journals is multiplied by 100.

year in question, 2) the number of publications weighted by a quality measure for the journal, and 3) the number of forward citations these published articles subsequently generated.

The journal quality measure is simply the ratio of forward citations to the journal to the number of backward citations from the journal over the sample period. While the determination of the actual quality of journals may be more nuanced, this measure has the virtue of being simple and consistently applied. Again, our purpose is not to generate another journal ranking but to create an objective measure of journal quality. The third measure examines the quality of each article, as defined by the citations it later garnered, rather than imputing quality from the journal. This provides us with three possible output measures for 500 institutions over 16 years. Summary statistics for this sample are reported in Table 4. The average institution generated seven publications per year which were cited an average of 65 times.

In search of preliminary evidence that JSTOR might have increased research productivity, we compare the annual change in the number of publications for the research institutions that gained access to more journals through JSTOR to the equivalent group that had no change in JSTOR access. One concern is how long it takes for JSTOR to have an impact on research output. Figure 5 reports the annual change in the number of publications for institutions with increased JSTOR access in the previous year to institutions with no change in JSTOR access in the previous year. The subsequent bars repeat this exercise for changes in JSTOR access two, three, and four years in the past. This figure suggests that increased JSTOR access may have increased publications three years later. This roughly corresponds to the typical time span in economics from a project's inception to publication.

While Figure 5 is suggestive, it does not constitute a formal test. To test for productivity effects, we econometrically relate each institution's research output in a given year to that

institution's prior access to JSTOR. However, our measures of institution research productivity seem highly auto-correlated suggesting possible non-stationarity. Harris and Tzavalis (1999) provide an appropriate test for co-integration for these data with a large N in the cross-section and small T in the time series. The results of these tests for each of the three output variables are reported in Table 5. The test results indicate that we cannot reject the existence of unit roots in the levels but can reject it in the first differences. We conclude that estimation using the levels of the data is susceptible to estimation bias. In light of this, we proceed to analyzing the data in first differences.

Thus, we estimate:

$$\Delta\, Res.Output_{i,t} = \sum_{\tau=2}^{4} \beta_\tau \Delta CountJSTOR_{i,t-\tau} + \mu_i + \mu_t + \upsilon_{i,t}\,, \qquad (2)$$

where μ_i is a fixed effect for institution *i*, μ_t is a fixed effect for time period *t*, and $\upsilon_{i,t}$ is a zero-mean error term.[12] The number of the 100 journals that a researcher can access through JSTOR, *CountJSTOR*, is lagged two to four years to account for publication lags. Since our sample ends in 2006, the latest value for *CountJSTOR* is 2004, which largely preceded widespread adoptions of alternatives to JSTOR.

Results of estimating equation (2) are reported in Table 6. It appears that the strong three year lag effect from Figure 5 is confirmed in these estimates. Effects for lags of two or four years are essentially zero in all specifications. For all three measures of output, obtaining JSTOR access to more journals three years prior is associated with an increase in publication productivity. This effect is significant for publications and quality-adjusted publications but is not statistically significant for backward citations, perhaps because the number of backward citations was

[12] Since equation (2) is in first-differences, the institution fixed effect is actually an institution-specific growth rate in research output. Similarly, the time fixed effect is a growth rate in overall research output. Excluding the institution fixed-effects from equation (2) does not change the results, which are available from the authors upon request. Here, we present the more conservative results.

truncated at the time of data collection. The coefficient from column two indicates that access to eleven additional journals through JSTOR is associated with one additional publication for an institution. The coefficient from column four indicates that obtaining JSTOR access is associated with 1.7% more quality-adjusted publications for the average institution. This would be consistent with, for example, JSTOR reducing literature review costs by 50% and literature review representing about a 4% "factor share" of research inputs, not unreasonable assumptions.

A concern is that these results may not reflect a causal link between JSTOR and research productivity because of potential reverse causality. Since the research productivity of some institutions was growing during the sample period, it is possible that institutions with rising research potential, and expectations of greater future research output, obtained earlier and more extensive access to research related resources, including JSTOR, than an otherwise similar institution. This would be the case, for example, if administrators invested in greater research-related resources because they foresaw these would be complementary with the higher expected future research productivity. In this way, JSTOR access could be capturing the effects of broad-based increases in research related inputs that would also affect future research productivity.

As a robustness check, we re-estimate equation (2) for a sub-sample of 172 universities for which university-wide research expenditures are available from the Integrated Postsecondary Education Data System (IPEDS). This reduces the sample size by two-thirds. While most of these expenditures are not directly related to economics research, they indicate the general research focus of a university. To the extent that universities generally invest across a broad spectrum of disciples simultaneously, these will be indicative of the resources available for economics research.

Results that include this measure are reported in Table 7. We compare the JSTOR coefficient when research expenditures are included and excluded for this sub-sample. With a sample-size about one-third as large, few coefficient estimates are significantly different from zero. Still, both research expenditures and JSTOR access continue to have positive effects on economic research productivity. In particular, the coefficient estimates for JSTOR access do not appreciably differ from those in Table 6. We infer from this that the results for JSTOR access in Table 6 are not likely biased by the omission of measures of other research related inputs.

VII. Conclusions

The creation of JSTOR introduced a new and potentially important, tool to academic researchers. We use a natural experiment in the availability of journals on the JSTOR archive and the access of JSTOR across institutions from 1996-2006. We estimate the impact of JSTOR access on referencing patterns and economic research output and find that access to JSTOR journals increased citations to JSTOR archived journals and decreased citations to non-JSTOR archived journals. We also find that the quantity of economic research from an institution increased after JSTOR access was obtained at that institution. These partial equilibrium results appear to be robust to potential omitted variables and subsamples.

Given the available data, we are not able to answer the general equilibrium impact of the increased economic research after JSTOR access. While there is evidence from the sciences that industrial innovation is enhanced by academic research, (Ward and Dranove, 1995 and Toole, 2007), it would be difficult to determine any specific benefits emanating from economic research. In the sciences, it is conceivable that one could trace a connection from peer-reviewed articles through, for example, patent grants to product commercialization. In the case of economics, such

connections would be difficult to establish. Therefore, we cannot, at this point, judge whether the costs to access the JSTOR archive are offset by the benefits generated by the increased economic research. If the effects found for JSTOR in economics are also at work with other Internet accessible bibliographic information applications in the sciences. These applications could lead to economic growth far in excess to their costs.

The recent growth literature focused on economic growth emanating from the generation and exploitation of ideas (Kortum, 1997, Alvarez, et al., 2007, Lucas, 2009) suggests that sustained increases in the rate of economic growth require alterations to the way ideas are generated, disseminated, and exploited. Clearly journal article archives that can be easily searched and filtered provide such a sustained change in the access to ideas. Therefore, it is possible that JSTOR, and similar Internet sites, might have a more widespread benefit than those generated in a single field of investigation. Future research might focus on how JSTOR and similar sites has influenced the geographic dissemination of ideas and collaboration, how sustained accessed has varied across national borders and influenced economic growth, economic freedom, human development, and political economies.

References

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas, Jr. (2008). "Models of Idea Flows." NBER Working Paper No. 14135.

Agrawal, Ajay and Avi Goldfarb (2008). "Restructuring Research: Communication Costs and the Democratization of University Innovation," *American Economic Review*, 99(4), 1578-1590.

Belo, Rodrigo, Pedro Ferreira, and Rahul Telang (2014). "Broadband in School: Impact on Student Performance", *Management Science*, 60(2), 265-282.

Brown, Jeffrey R. and Austan Goolsbee (2002). "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy*, 110(3), 481-507.

Brynjolfsson, Erik and Michael D. Smith (2000). "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 6(4), 563-585

Ellison, Glenn (2011). "Is Peer Review in Decline?" *Economic Inquiry,* 49(3), 635-657.

Fields, Thomas D., Thomas Z. Lys, and Linda Vincent (2001). "Empirical research on accounting choice," *Journal of Accounting and Economics*, 31(1-3), 255-307.

Goldfarb, Avi (2006). "The (Teaching) Role of Universities in the Diffusion of the Internet". *International Journal of Industrial Organization* 24(2), 203-225.

Greene, William (2006). *Econometric Analysis,* 6th Ed. Prentice Hall, New York, NY.

Hamermesh, Daniel S. and Sharon M. Oster (2002). "Tools or Toys? The Impact of High Technology on Scholarly Productivity," *Economic Inquiry*, 40(4), 539-55.

Harris, Richard and Elias Tzavalis (1999). "Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed," *Journal of Econometrics,* 91(2), 201-226.

Kim, E. Han, Adair Morse, and Luigi Zingales (2006). "Are Elite Universities Losing Their Competitive Edge?" *Journal of Financial Economics,* 93(3), 353-381.

Kortum, Samuel S. (1997). "Research, Patenting, and Technological Change." *Econometrica*, 65(6), 1389–419.

Krueger, Alan (2005). "The Economics of Real Superstars: The Market for Rock Concerts in the Material World," *Journal of Labor Economics*, 23(1), 1-30.

Lucas, Robert E., Jr. (1988). "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1), 3–42.

Lucas, Robert E., Jr. (2009). "Ideas and Growth." *Economica,* 76(301), 1–19.

McCabe, Mark J. and Christopher M. Snyder (2015). "Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals." *Review of Economics and Statistics*, 97(1), 144-165.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703-708.

Roodman, David (2009). "How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata" *The Stata Journal*, 9(1), 86-136.

Sauer, Raymond D. (1988). "Estimates of the Returns to Quality and Coauthorship in Economic Academia," *The Journal of Political Economy*, 96(4), 855-866.

Schonfield, Roger C. (2003). *JSTOR: A History*. Princeton University Press: Princeton, N.J.

Toole, Andrew A. (2007). "Does Public Scientific Research Complement Private Investment in Research and Development in the Pharmaceutical Industry?," *Journal of Law & Economic*s, 50(1), 81-104.

Ward, Michael R. and David Dranove (1995). "The Vertical Chain of Research and Development in the Pharmaceutical Industry," *Economic Inquiry*, 33(1), 70-81.

Zentner, Alejandro (2008). "Online Sales, File Sharing, and the Decline of Retail Music Specialty," *Information Economics and Policy,* 20(3), 288-300.

Figure 1
The Effects of Lower Input Costs Due to JSTOR

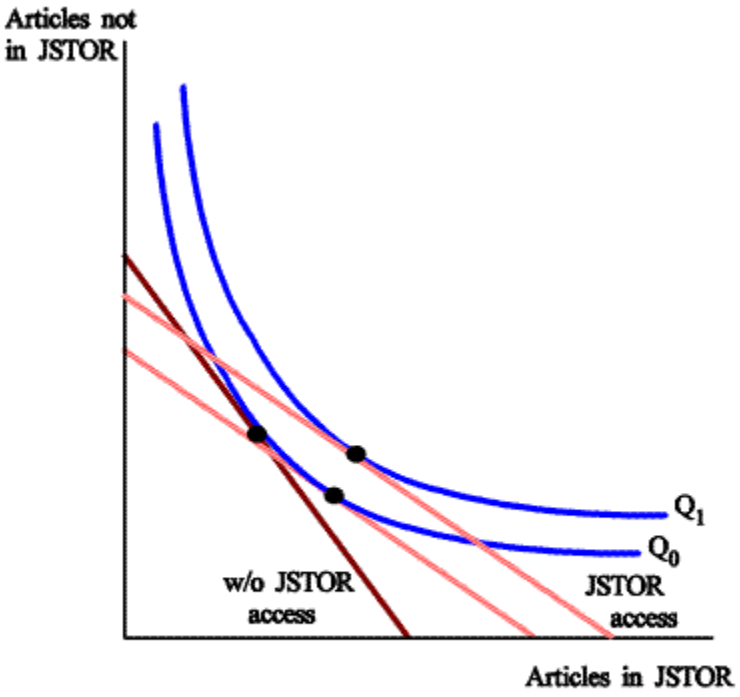

Figure 2

Growth in Average Number Journals Available through JSTOR over Time

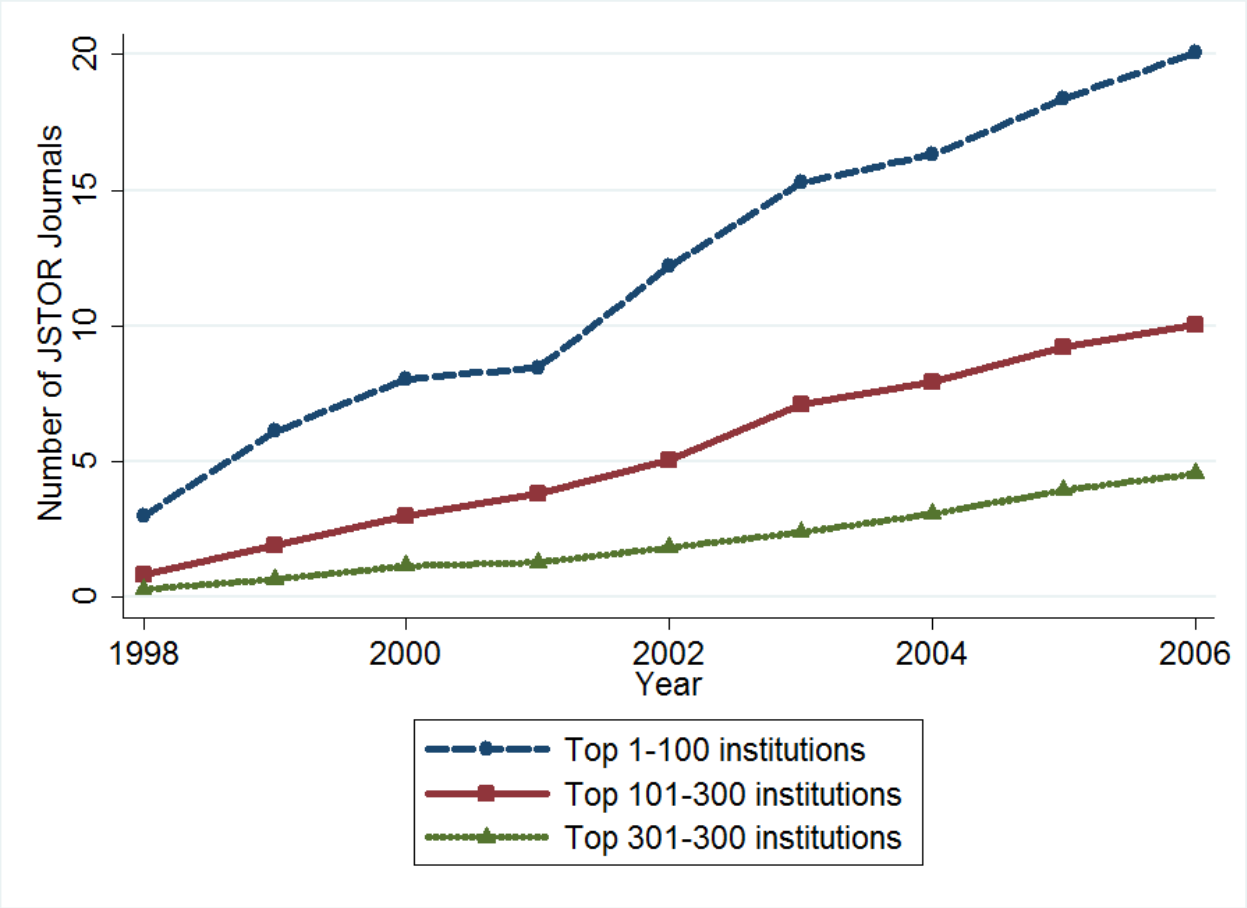

Figure 3

Number of top 100 Journals Available through JSTOR in 2006 by Institution Rank

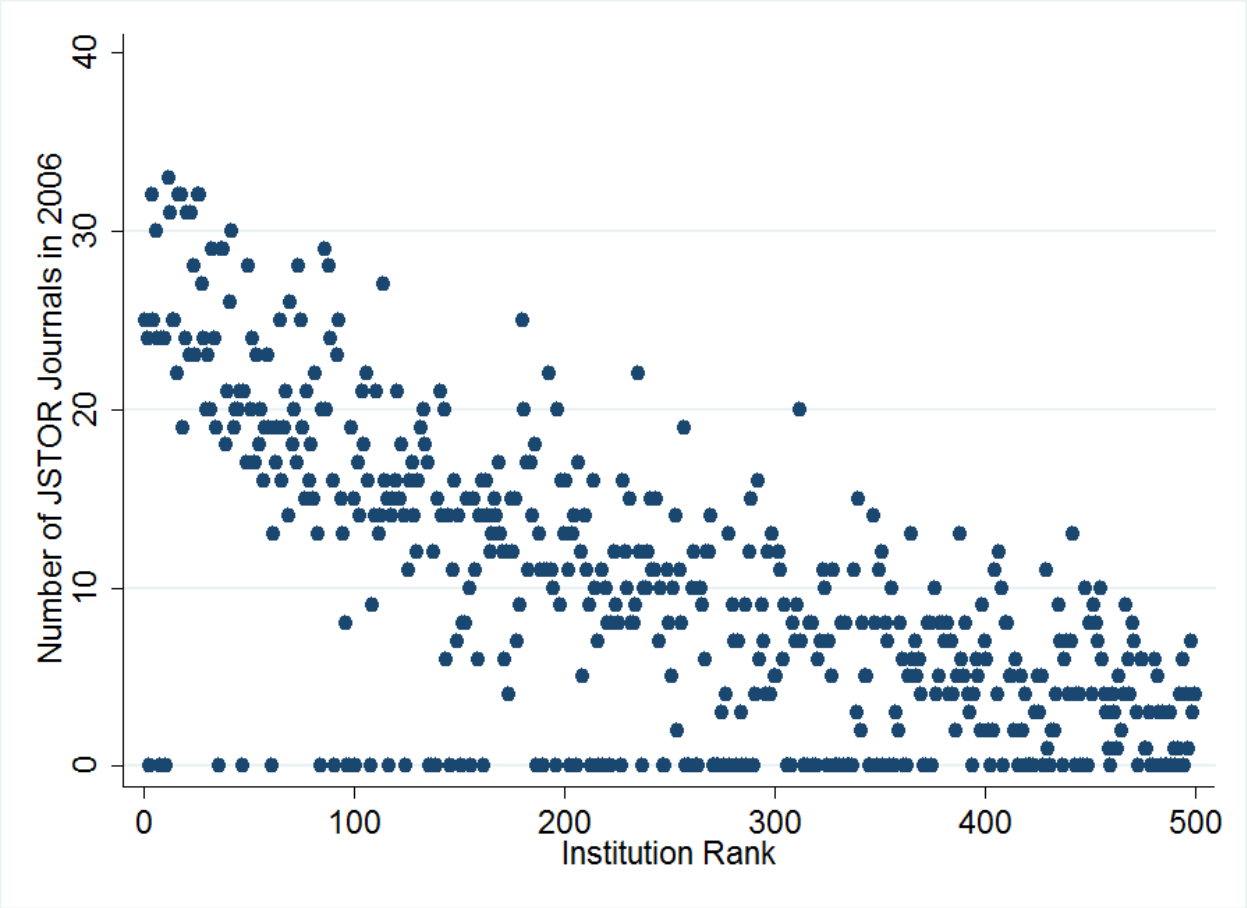

Figure 4

Referencing Patterns Before and After JSTOR Access to a Journal

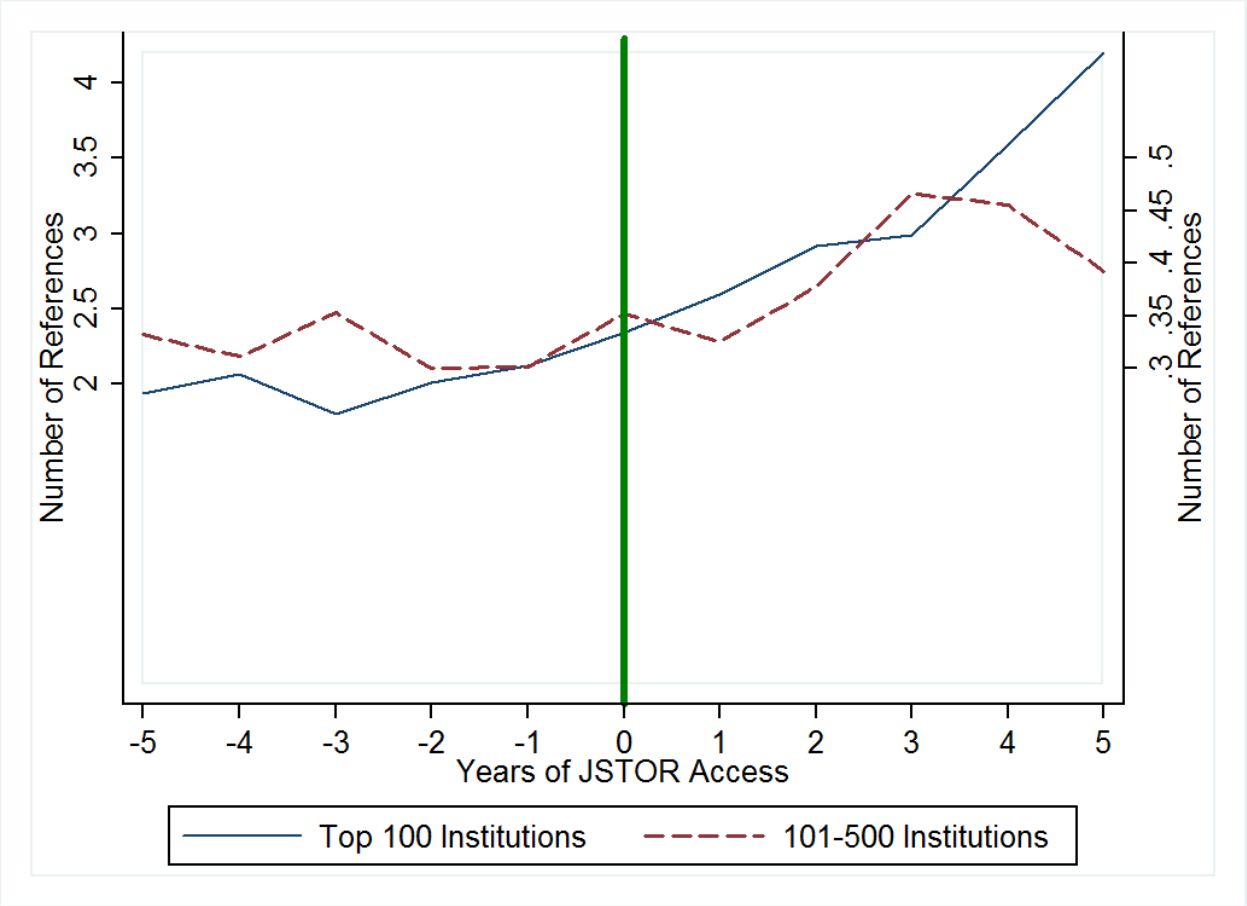

Figure 5

Change in Publications by When Change in JSTOR Access Occurred

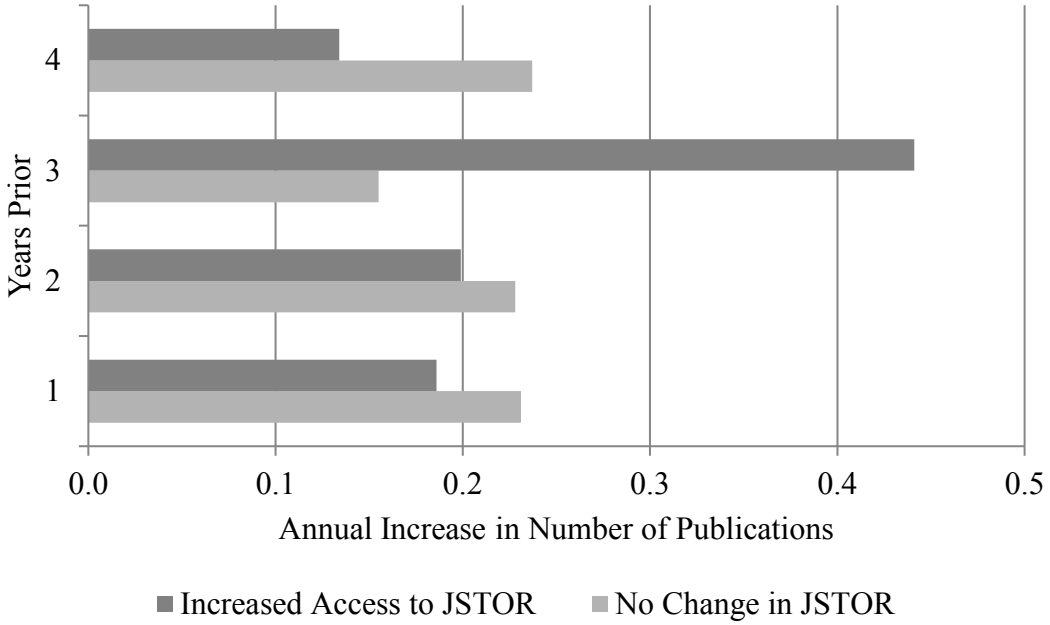

Table 1
Sample Journals and Year First Available in JSTOR

Journal	Year	Journal	Year	Journal	Year
AMER ECON REV	1997	J ACCOUNT ECON		J MONEY CRED BANK	1997
AMER J AGR ECON	2004	J AGR ECON		J POLIT ECON	1997
APPL ECON		J AGR RESOUR ECON		J PROD ANAL	
APPL ECON LETTERS		J APPL ECONOM	1998	J PUBLIC ECON	
BROOKS PAPERS	2001	J BANK FINANCE		J REAL EST FIN ECON	
CAMBRIDGE J ECON		J BUS ECON STAT	2005	J REGUL ECON	
CAN J ECON	2001	J COMP ECON		J RISK INSUR	2001
ECOL ECON		J DEVELOP ECON		J RISK UNCERTAINTY	
ECON DEV CULT CHG	2004	J DEVELOP STUD		J TRANS ECON POL	
ECON EDUC REV		J ECON BEH ORG		J URBAN ECON	
ECON GEOGR	2001	J ECON DYN CTL		KYKLOS	
ECON HIST REV	2001	J ECON EDUC	2004	LAND ECON	2004
ECON INQUIRY		J ECON HIST	1998	MATH FINANC	
ECON J	1998	J ECON ISSUES		NATL TAX J	
ECON LETT		J ECON LIT	1999	N E ECON REV	
ECON REC		J ECON MANG STRAT		OX BULL ECON STAT	
ECON THEORY	2006	J ECON PERSPECT	1997	OX ECON PAP	2002
				OX REV ECON	
ECONOMET THEORY		J ECON PSYCHOL		POLICY	
ECONOMETRICA	1997	J ECON THEORY		PUBLIC CHOICE	
ECONOMICA	2001	J ECONOMETRICS		QUART J ECON	1997
ENERGY ECON		J ENVR ECON MANG		RAND J ECON	2001
ENERGY J		J FIN ECON		REG SCI URB ECON	
ENV RES ECON		J FIN QUANT ANAL		REV ECON STAT	1997
EUR ECON REV		J HEALTH ECON		REV ECON STUD	1999
EXPLOR ECON HIST		J HUM RESOUR	2001	REV INC WEALTH	
FUTURES		J IND ECON	1998	SCAND J ECON	2006
GAME ECON BEHAV		J INST THEOR ECON		SCOT J POL ECON	
HEALTH ECONOMICS		J INT ECON		SMALL BUS ECON	
INSUR MATH ECON		J LABOR ECON	2001	SOC CHOICE WELF	
INT ECON REV	2001	J LAW ECON	2004	SOUTH ECON J	2004
INT J FORECASTING		J LAW ECON ORGAN	2004	WORK EMPL SOC	
INT J GAME THEORY		J MATH ECON		W B ECON REV	
INT J IND ORGAN		J MONETARY ECON		WORLD DEV	
				WORLD ECON	

Table 2

Mean Values for Referencing Sample

	All 500 Institutions	Top 100 Institutions	101-500 Institutions
References to Articles Published Three Years Back	0.276	1.006	0.093
Dummy Variable for JSTOR Access	0.031	0.067	0.021
Number of other journals with JSTOR Access	3.024	6.676	2.111
Observations	800,000	160,000	640,000

Table 3

The Effect of JSTOR Access on Referencing Patterns

	All 500 Institutions	Top 100 Institutions	101-500 Institutions
JSTOR Access	0.184**	0.153**	0.088**
	(0.012)	(0.054)	(0.007)
Number of other journals with JSTOR Access (×100)	-0.752**	-2.462**	-0.332**
	(0.090)	(0.542)	(0.062)
Journal	sign.	sign.	sign.
Year	sign.	sign.	sign.
Institution	sign.	sign.	sign.
Observations	800,000	160,000	640,000

Negative Binomial regressions include dummy variables for 16 years, 100 journals and each institution. This table reports marginal effects evaluated at the mean values rather than coefficient values.
** $p<0.01$

Table 4

Mean Values for Publishing Sample

	Mean	Standard Deviation
Number of Publications	7.054	9.563
Quality Weighted Publications	3.476	6.360
Forward Citations	69.756	182.007
Number of JSTOR Journals Available Lagged Three Years	1.037	3.091
Δ Number of Publications	0.217	3.894
Δ Ln Quality Weighted Publications	0.032	2.484
Δ Ln Forward Citations	-5.881	99.778
Δ Number of JSTOR Journals Available Lagged Three Years	0.347	1.457
7,500 Observations over 15 years and 500 institutions.		

Table 5

Harris-Tzavalis Unit-root Tests

	Publications	Quality Adjusted Publications	Backward Citations
Levels			
rho	0.929	0.927	0.821
z	31.16	31.02	22.04
p-value	1.0	1.0	1.0
First Differences			
rho	-0.451	-0.497	-0.439
z	-85.01	-88.85	-84.01
p-value	0.00	0.00	0.00
Null Hypothesis is that the panel contains unit roots. Test implemented with cross-sectional means removed and with time trend.			

Table 6

The Effect of JSTOR Access on Research Productivity

	Δ Publications		Δ Quality Adjusted Publications		Δ Backward Citations	
Δ Num. Journals via JSTOR 2 Years Lagged	-0.035		-0.023		-0.827	
	(0.030)		(0.019)		(0.750)	
Δ Num. Journals via JSTOR 3 Years Lagged	0.083*	0.088*	0.053*	0.056*	0.512	0.639
	(0.036)	(0.036)	(0.023)	(0.023)	(0.910)	(0.905)
Δ Num. Journals via JSTOR 4 Years Lagged	-0.037		-0.003		-0.963	
	(0.041)		(0.027)		(1.050)	
Year Fixed Effects	X	X	X	X	X	X
Inst. Fixed Effects	X	X	X	X	X	X

Each column represents a regression with 500 institutions from 1991-2006 for 7,500 observations. Standard errors in parentheses. ** $p<0.01$, * $p<0.05$, + $p<0.1$

Table 7

The Effect of JSTOR Access on Research Productivity

Robustness Check

	Δ Publications		Δ Quality Adjusted Publications		Δ Backward Citations	
Δ Num. Journals via JSTOR 3 Years Lagged	0.058	0.055	0.078+	0.073	0.727	0.676
	(0.062)	(0.062)	(0.047)	(0.047)	(2.003)	(2.006)
Δ Research Expenditures 3 Years Lagged		0.006		0.011+		0.117
		(0.008)		(0.006)		(0.254)
Year Fixed Effects	X	X	X	X	X	X
Inst. Fixed Effects	X	X	X	X	X	X

Each column represents a regression with 172 institutions from 1991-2006 for 2,537 observations. Standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1