Sited, Sighted, and Cited:

The Effect of JSTOR in Economic Research

Craig A. Depken, II,
Department of Economics
University of North Carolina at Charlotte
cdepken@uncc.edu

Michael R. Ward
Department of Economics
University of Texas at Arlington
mikeward@uta.edu

August, 2009

Sited, Sighted, and Cited:

The Effect of JSTOR in Economic Research

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 is hypothesized to 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 available 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, JSTOR access increases the quantity, if not quality, of economic research generated at an institution. From this accumulated evidence, we deduce that Information and Communication Technology has the potential to not only increase productivity, but by increasing research productivity, can also increase the rate of economic growth.

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

Keywords: Internet, Scholarly Communication, Academic Research

We wish to thank Mahmut Yasar, Jeff Desimone, Babur De Los Santos, and workshop participants at 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.

I. Introduction

Academic scholars who earned their degrees within the past decade may be dumbfounded upon learning how, before the Internet, their elders stayed current with the work in their field. They may 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, 2007), concerts, (Krueger, 2005), and the insurance industry (Brown and Goolsbee, 2002), it appears the Internet has had a major effect on the "research industry." Clearly, 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. Has the Internet 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 the 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 has varied across institutions since the service began in 1997. We find that once a journal's previously published articles become available to economists at an institution, these

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

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 that JSTOR increased the research productivity of these economists as measured by the rate at which they publish but did not significantly affect the citations to these economists' work.

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 in 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.[3]

II. The Internet, the Academy and JSTOR

The academy was instrumental in the development, use and popularization of the Internet. Research universities were among the first to develop applications for the Internet.

[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.
[3] The investigation of these hypotheses is planned for our future research in this area.

Many of the pioneering applications were developed on university campuses 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).

The effect of the Internet on university research is being documented. There is evidence that the Internet has broken down many geographical and international barriers that hampered economics and finance researchers outside of elite universities (Kim, Morse and Zingales, 2006). Early 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 opened up alternative venues to peer-reviewed journal articles for the dissemination of research for the top researchers (Ellison, 2007). However, Hamermesh and Oster (2002) provide evidence suggesting that Information and Communication Technology (ICT) provides "toys" as well as "tools" and may merely serve to add to the consumptive value of being an academic without enhancing research productivity. Since we focus on the effects of a specific tool, we can say nothing about the net effect of ICT that would include greater access to Internet toys that could decrease productivity.

JSTOR is an Internet application believed to have enhanced research productivity at universities. 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 met 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, although many research libraries have reduced their possession of physical copies of many of JSTOR archived journals. 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 March, 2009, the archive contained nearly 5 million articles archived from nearly 1,000 journals. As of March 2009, there were 568 participating publishers and more than 5,400 participating institutions, more than half of which are outside of the US. Figure 1 depicts the growth in the number of economics journals archived in JSTOR and subscribing institutions over time. Usage has steadily grown to the point that JSTOR is currently averaging 12 million searches and 20 million page views per month. 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.

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 is not random. Thus, it is possible that JSTOR effects will be biased since early adopters of JSTOR are more research intensive. Again, we will generally include institution fixed effects so that our estimates reflect only the increased referencing and publishing due to JSTOR for a given institution. As an additional robustness check, we include for institution-specific time trends in some specifications to account for the possibility that institutions with increasing research missions adopt JSTOR earlier, all else equal.

III. A Simple Model of Research Production

Notwithstanding the obvious metric of web traffic, it is not immediately clear whether JSTOR or other online "tools" actually enhances 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 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, the 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 optimal mix of inputs 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 that researchers' objective functions are to 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, $-w_i/w_j$, i $i \neq j, i = 1...N, j = 1...N$.. Let x_1 be the process of searching for, reading, and incorporating an article from Journal 1 into one's research. This search has a marginal benefit of MP_1 and a cost of w_1. Let x_2 be the process of searching for, reading, and incorporating an article from Journal 2 into one's research. This search has a marginal benefit of MP_2 and a cost of w_2. If access to Journal 1 through JSTOR reduces w_1, but does not change w_2, we expect the researcher to make more use of articles in Journal 1. This will involve a substitution effect away from articles found in another journal (see Figure 2). 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. In this case, 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 mid-2007. After matching these data sources by institution, journal and year, a usable sample was created of articles from journals that were continuously indexed by ISI over that time period, articles in these journals authored by economists at research institutions worldwide, and references made by these articles to this same set of journals. This led to a sample of over 40,000 articles in 79 journals written by authors at one of 542 institutions during a 22 year period entailing more than 400,000 references to these journals.

Information about research institutions' access to JSTOR economics and business collections was made available by JSTOR.[4] Institutions could subscribe to any of seven different collections that include economics related journals are archived by JSTOR (Arts & Sciences I, II, III, IV, and Complement, and Business I and II). Each collection includes a set of specified journal titles that has grown in number over time and are not necessarily mutually exclusive. Scholars at these institutions have access to a covered journal's archive except for a few years prior to the present as dictated by the journal's 'moving wall.' Most journals have opted to hold back the most current issues, usually 3-5 years' worth, from JSTOR to avoid cannibalizing

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

journal 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 and their moving walls. In general, the most important journals were 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 archived by JSTOR are also among the more than 160 journal titles indexed by ISI. The sample of articles we use includes 79 journal titles indexed by ISI continuously from 1985 through 2007. These include all of the most important general 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*. Journals that either began publication after 1985 or were first indexed by ISI after 1985 were omitted from the sample. Out of these 79 titles, 29 will have been 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 institutional subscribers to any of those collections that will eventually include the different economics-related journal titles. These subscribers include most of the research universities worldwide but also include lesser-known colleges, government agencies, non-governmental agencies, private consultancies, and even some high schools. Since our focus is on the 'production' of journal articles, most of those entities that 'consumer' journal articles are not included in this analyses. Ultimately, the sample includes top research institutions and institutions that are not as well known for their research output. Our broad view of what constitutes a research institution yields a total of 542 institutions whose

scholars published seven or more weighted publications in the sample of 79 journals from 1985 through 2007.[5] 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. Table 2 describes the evolution of JSTOR access across these institutions over time.

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 59,097 articles over the sample period of which authors of the top 542 institutions authored 43,111 articles in the 79 journals and 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.[6] Also available for each article is a variable indicating the number of citations it has received.

V. Empirical Results

We analyze the effects of JSTOR access on both inputs and outputs. JSTOR access lowers researcher costs to finding, reading, benefiting from and ultimately referring to papers

[5] 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, almost all institutions are associated with one or more publication per year.

[6] In fact, we include only the first 200 citations made by an article, Fewer than 10 articles, usually survey articles, include more than 200 citations.

available to her in the archive.[7] It does not change the nominal costs of using and referring to papers found by traditional methods. This is analogous to a shift in the isocost curve facing a researcher. Standard production theory predicts that researchers will refer to JSTOR available articles more often. However, the net effect on articles not available via JSTOR is theoretically ambiguous as scale and substitution effects push in opposite directions and their magnitudes are uncertain. However, while it is not clear if there is an increase in the quality of this research, there should be an unambiguous increase in output as researchers move to a higher isoquant.

a. The Effect of JSTOR on Referencing Patterns

We first examine the effect of JSTOR on the references made by the authors of an article. Above, we describe how JSTOR, by lowering the cost of referencing some articles, would increase the number of references made to these articles and, possibly, reduce the number of references to articles not available through JSTOR. This presupposes that some references are substitutes for others. Before the advent of online article archives, researchers would draw on both their own knowledge of the extant literature and on a directed 'manual' search of the 'state-of-the-art' related to their paper. Ideally, the directed search would uncover all of the literature relevant to a current topic, but if search costs are convex and total benefits of the search are quasi-concave, the researcher would equate the expected marginal benefit of search with its marginal costs. Therefore, it is possible that a directed search could miss some relevant articles because the optimal search is not completely exhaustive. A decrease in search costs would induce researchers to extend the directed search of the literature making it more likely to

[7] To facilitate exposition, we will adopt a directional reference/cited terminology from the point of view of the authors of an article. Authors *refer* to previously published articles but are *cited* by articles subsequently published at a later date.

discover and incorporate a greater number of relevant papers. Searchable databases of journal articles, such as ECONLIT, JSTOR, SSCI, and publisher archives greatly facilitate the search for papers related to a specific topic, thereby likely reducing search costs.[8] If so, researchers should be able to find more articles relating to a topic and refer to more of the previous literature. Alternatively, if an online index, such as JSTOR, had no effect on a researcher's ability to find and refer to related papers, then we would expect having access to a journal through JSTOR to have no effect on the researcher's propensity to refer to articles available through JSTOR.

This simple analysis suggests an empirical test of the hypothesis. If JSTOR increases research productivity, having access to more journals through JSTOR should lead a researcher to refer more often to articles in those journals. As a corollary, having access to more, possibly substitute, journals through JSTOR could lead a researcher to refer to non-JSTOR journals less often. To test these hypotheses, we focus on a set of 79 commonly referenced economics journals published and also indexed by ISI continuously from 1985 through 2007. We examine only references from this set of journals to this set of journals. While the references we measure were made after 1985 the referred to article may well have been published earlier. We focus on a set of 542 institutions including those with the highest economic research output as well as those with more moderate levels of economic research output. Our unit of observation for this sample is an article. The sample was limited to the 43,111 articles published in these 79 journals written by an author affiliated with one of these institutions.

[8] This assumes *ceteris paribus*. If one rather quickly discovers additional articles via the Internet this will also require time to read, digest, and integrate the papers into the current project. Here, we assume that other inputs remain unaffected and concentrate on JSTOR's search cost reducing effects.

Most of the sample institutions became subscribers to JSTOR at some point after its launch in 1997 but many did not subscribe to JSTOR during the sample period.[9] . Even when they did subscribe, they often chose subscriptions to different collections of journals and the collections themselves grew to include more journals. Thus, for a stable set of journals and institutions, we observe substantial variation in the timing and level of JSTOR access. We measure whether any of the coauthors of the article had access to each potential reference journal through JSTOR at the time of publication.[10]

Our first tests relate the number and type of references authors make to journals to the number of journals available to the authors through JSTOR. For each article, we identify the set of journals available to any author at the time of publication. From this, we calculate the number of JSTOR available journals to the authors as well as that the number of article references to this set of JSTOR available journals and to journals not in this set. Our estimating equations are:

$$\ln(Ref_i^{in} + 1) = \beta_0^{in} NumJSTOR_i + \beta_1^{in} NumJSTOR_i^2 + \Theta^{in} X_i + \varepsilon_i^{in}$$
$$\ln(Ref_i^{out} + 1) = \beta_0^{out} NumJSTOR_i + \beta_1^{out} NumJSTOR_i^2 + \Theta^{out} X_i + \varepsilon_i^{out}$$

where *in* and *out* refer to the set of JSTOR available journals or not, respectively. The square of the number of JSTOR available journals is included to test for possible diminishing marginal value of additional journal inclusion. Our control variables, X_i, will include sets of year dummy variables, to account for a slight increase in number of references per paper during the sample period, and citing journal dummy variables, to account for differences in referencing patterns across journals.

[9] We do not investigate why an institution did not subscribe to JSTOR. Perhaps the institution did not have the resources to subscribe to the sections of JSTOR we investigate here. Perhaps the institution was subscribing to other electronic resources. Our institution fixed effects and institution time trends hopefully capture some of this idiosyncratic yet unobservable reasons for not subscribing (or subscribing) to JSTOR.

[10] Our measure of availability takes into account a standard a three year 'moving wall' during which the articles in the three most recent years of a journal are not available to JSTOR subscribers.

Table 3 provides summary statistics for this sample. On average, these articles make 5.6 references to journals that will be archived by JSTOR and 4.2 references to journals that will not. These articles make an average of 23 total references with just over half going to journals not in this set, government reports, working papers, the popular press, and other sources. On average, the authors of an article had access to 9.3 journals through JSTOR. Of course this was zero for all years prior to 1997 and rose to 23.7 by 2006.

Evidence of a change in referencing behavior is found in Table 4. The first two columns report coefficient estimates from a regression of references to JSTOR available journals against the number of JSTOR journals available to authors. The next two columns report coefficient estimates from the same specifications applied to the number of references to journals not available through JSTOR. In columns one and three, a set of year dummy variables are included and in columns two and four, sets of year and journal dummy variables are included. The specifications indicate that access to more JSTOR journals is associated with more references in JSTOR accessible journals and with fewer references to journals not accessible through JSTOR. Both of these effects exhibit a diminished marginal effect as access to more JSTOR journals is available. This could be due to either the decreased "productivity" from any additional journal archived by JSTOR or due to the earlier inclusion in JSTOR of more oft cited journals. At the sample average for *NumJSTOR*, these estimates imply 2.5 more references to JSTOR journals and 0.9 fewer references to non-JSTOR journals, on average.

An additional test of a behavioral change by researchers is to examine the characteristics of the cited material. In particular, we conjecture that JSTOR allows researchers to find and refer to older articles. If so, all else equal, references made with the assistance of JSTOR should be older, on average. To test this, for our sample of articles, we calculate the age to the nearest year

of each reference and indicate whether this reference was available to the authors via JSTOR. This generates a sample of 423,861 references to the set of 79 journals indexed by ISI. Our estimating equation is:

$$age_{ij} = \beta_0^{age} DumJSTOR_{ij} + \Theta^{age} X_{ij} + \varepsilon_{ij}^{age}$$

where *i* refers to the article and *j* refers to the reference within the article, and $DumJSTOR_{ij}$ is a dummy variable that takes a value of one if journal *j* is available on JSTOR when article *i* is published and zero otherwise. Our control variables, X_i, will include sets of year, citing journal, and cited journal dummy variables.

Age of reference regression results are presented in table 5. Column (1) indicates that the unconditional average age of a referenced work is 3.8 years older when made to a JSTOR available journal. Columns (2)-(4) successively add sets of dummy variables for the year of the published article, the citing journal and the cited journal. With the full set of controls, the conditional average age difference falls to 0.13 years but remains statistically significantly different from zero.

We take the above results as evidence consistent with JSTOR affecting the way in economic scholars' conduct research. The evidence presented above suggests that scholars refer to JSTOR articles, and refer to otherwise harder to find older articles, more often. It is plausible that JSTOR also led to scholars to read additional published work that affected the content of their work even if it was not cited. Publishing productivity, as one of these outcomes, is investigated next.

b. The Effect of JSTOR on Research Productivity

Aggregating the articles sample to the institution level for each publication year allows for tests of increased research output along both the quantity and quality dimensions. As described above, JSTOR access could affect output both by reducing the costs of producing an article and by increasing the amount of prior knowledge incorporated into an article. Holding article quality constant, the first effect would imply that less time is required to complete each article implying a possible increase in the quantity of research output. Alternatively, researchers may spend the same amount of time on a project, but are able to perform a more thorough review of past work that fosters a more thorough and rigorous analysis implying a possible increase in the quality of research. We attempt to uncover these effects with analyses of the number and placement of, and citations to, publications generated by authors affiliated with an institution in a year.

We construct three measures of research output – the number of publications, the number of citations, and an impact factor weighted number of publications. For each article, we identify the institutional affiliation of each author. For an article with N authors, we attribute to each author's institution an authorship weighting of $1/N$ of the article. For each institution and publication year we sum the weighted number of published articles, of article citations[11], and impact factor weighted number of publications. Citations to an individual article may be a noisy measure of quality. An alternative is to infer quality from editorial decisions to accept or reject a submission from different qualities of journals. We construct an impact factor as a quality-adjusting measure by weighting publications by journal impact factors constructed by dividing the sum of the total citations to a journal by the sum of references made by the journal over the

[11] These are citations as of August, 2007, when our data were collected.

entire sample period. That is, by construction, our impact factor measures do not vary over time. Changes in the quality of our sample of economics journals over our time period is not the focus of this study and are not likely to be large or correlated with our variables of interest.

These data are used first to test a parsimonious production function specification and then to attempt some robustness checks. The number of JSTOR available journals does not vary across researchers affiliated with an institution within a year. For each of our three outcome measures we estimate:

$$\ln(Outcome_{jt} + 1) = \gamma_0 NumJSTOR_{jt} + \gamma_1 NumJSTOR_{jt}^2 + \Gamma \mathbf{X}_{it} + \upsilon_{it}$$

where *j* refers to institution and *t* refers to year. Since the productivity of JSTOR could be a result of any journal archived by JSTOR, we include the full count of JSTOR available journals and not just those that were continuously indexed by ISI. The vector *X* will include sets of dummy variables for years and institutions. Year fixed effects attempt to account for a slight increase in the number of articles that these journals published per year. Citations are also affected by this trend and by a truncation issue toward the end of the sample. Institution fixed effects attempt to address the considerable variation in research output across institutions. Additionally, our results above suggest that an article published in a JSTOR archived journal could be cited more often because more subsequent authors will have access to JSTOR and not necessarily because the paper's authors benefited from using JSTOR. To address this concern, we constructed a time and institution varying measure of the likelihood that a future potential citer will have JSTOR access to an article. For each journal and year, we calculate the fraction of future citations where at least one author had JSTOR access to the journal. This is aggregated

across the articles published by all authors affiliated with an institution as a proxy for the additional ease with which these publications can be cited due to future JSTOR availability.

Table 6 provides summary statistics for this sample. The average institution generates 2.7 publications, 31 citations and 1.6 impact factor weighted publications. Over the entire sample, authors have access to an average of 3.9 journals through JSTOR, although, again this is zero for all institutions until 1997. Finally, over the sample, 19% of these institutions' publications had potential citers with JSTOR access to the publications.

Estimation results from our parsimonious specification are reported in Table 7. The number of publications produced by an institution in a year, and Impact Factor weighted publications, both increase with the number of JSTOR journals available to researchers and does so at a declining rate. These estimated magnitudes are both statistically and economically significant. At the sample average of 12.4 journals available in JSTOR after 1997, the mean effect is 11% more articles published per year and 10%-12% more impact factor weighted publications. The number of citations to these articles is not estimated to be affected greatly. Only the squared term is statistically significant and, calculated at the mean as above, JSTOR is associated with a statistically insignificant 1% to 3% fewer citations. The measure of potential future citers with JSTOR significantly increases citations and impact factor weighted publications. Its inclusion in the specification slightly reduces the coefficient estimates for the number of JSTOR journals available to authors.

Two potential problems with this specification make inferences about the productivity effects of JSTOR problematic. First, institutions with a greater publication record tended to obtain JSTOR access earlier and obtained access to more journals. It is likely that more research

oriented institutions expected a greater effect causing them to adopt JSTOR sooner. The fixed effect for each institution will account for the time invariant differences in institutions research productivity. However, institutions research orientation may not be time invariant. It is possible that institutions that were experiencing an upward trend in research output adopted JSTOR sooner than similar institutions that were experiencing a decline in research output. If so, some of the correlation between JSTOR and research output could reflect reverse causality. We attempt to address this concern by adding to the specification institution specific time trends. This way, the JSTOR variables would measure in increase in research output beyond that accounted for by the institution's trend in research output.

Second, greater access to JSTOR may affect research output with a lag. Because of publication lags, much research completed in one year will not be published until a later year. If so, the research will be completed with JSTOR availability as of a date prior to the publication date. Because there are opportunities to amend papers during the editorial process with new references as they become known, the lag may not be too long. To address this concern, we add lags and leads of our JSTOR variables to the specifications. Current and lagged JSTOR availability affecting research output is consistent with an increase in research productivity. Leads of JSTOR availability provide a falsification test since they should have no effect on productivity.

Table 8 reports regression results that add institution-specific time trends and lags and leads. For all three outcome measures, JSTOR availability increases productivity when these time trends are included. Focusing on the estimates without lags and leads first, we see a statistically significant increase with all three output measures; however, the publications and citations measures are statistically significant only at the 10% level. Moreover, the magnitudes

appear to be more reasonable. These estimates suggest an average increase of 5.7%, 7.7% and 5.2% in publications, citations and impact factor weighted publications during the JSTOR era. Next, two lags and leads are introduced along with the institution-specific time trends. While few individual coefficient estimates are statistically significant, those that are significant are among the lags and current value and not among the leads and are of the predicted sign. These tend to have large magnitudes as well. Finally, F-tests indicate that we can reject that the sum of lags and current coefficients are negative at the 5% level for the publications and impact factor weighted publications measures but not for the citations measure. The estimated research productivity effects from JSTOR appear to be robust to these concerns. In sum, our evidence is more consistent with JSTOR improving the quantity of research output rather than its quality.

VI. Conclusions

JSTOR represents a single, though important, new tool available to academic researchers. New tools are emerging, such as the Social Science Research Network (SSRN) working paper archive, improvements to tools are being developed, such as linkable citations within on-line papers, and academics are embracing new methods of discourse, such as blogging. The continuous development of the Internet is likely to continue to enhance the quality and quantity of academic research.

It is not clear how valuable enhanced academic research is to society. There is evidence from the sciences that industrial innovation is enhanced by academic research, (Ward and Dranove, 1995 and Toole, 2007). Unlike the sciences, it would be difficult to determine any specific benefits emanating from economic research. For the sciences, it is conceivable that one

could trace a connection from peer-reviewed articles through, for example, patent grants to product commercialization. The effects found for JSTOR in economics may also be at work with other Internet accessible bibliographic information applications in the sciences. If so, these applications could lead to economic growth far in excess to their costs.

If such a link between academic research productivity and economic growth does exist, it is not clear if each new innovation represents a change in the level of economic production or a change in economic growth rates. The recent growth literature has focused on economic growth emanating from the generation and exploitation of ideas (Kortum, 1997, Alvarez, et al., 2007, Lucas, 2009). In these models, sustained increases in the rate of economic growth require alterations to the way ideas are generated, disseminated, and exploited. The literature to which this paper contributes could be viewed as a contribution to the "micro-foundations" of this macro-oriented growth literature.

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), 1578-90.

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 (2007). "Is Peer Review in Decline?" NBER Working Paper No. 13272.

Fields, Thomas D., Thomas Z. Lys, 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.

Kim, E. Han, Adair Morse and Luigi Zingales (2006), "Are Elite Universities Losing Their Competitive Edge?," NBER Working Paper No. 12245.

Kortum, S. (1997). "Research, patenting, and technological change." *Econometrica*, 65, 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).

Lucas, E. R., Jr. (1988). "On the mechanics of economic development." *Journal of Monetary Economics*, 22, 3–42.

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

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.

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

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, "Online Sales, File Sharing, and the Decline of Retail Music Specialty," 2007 <papers.ssrn.com/sol3/papers.cfm?abstract_id=985115>.

Figure 1

JSTOR Institutions and Economics Journals over Time

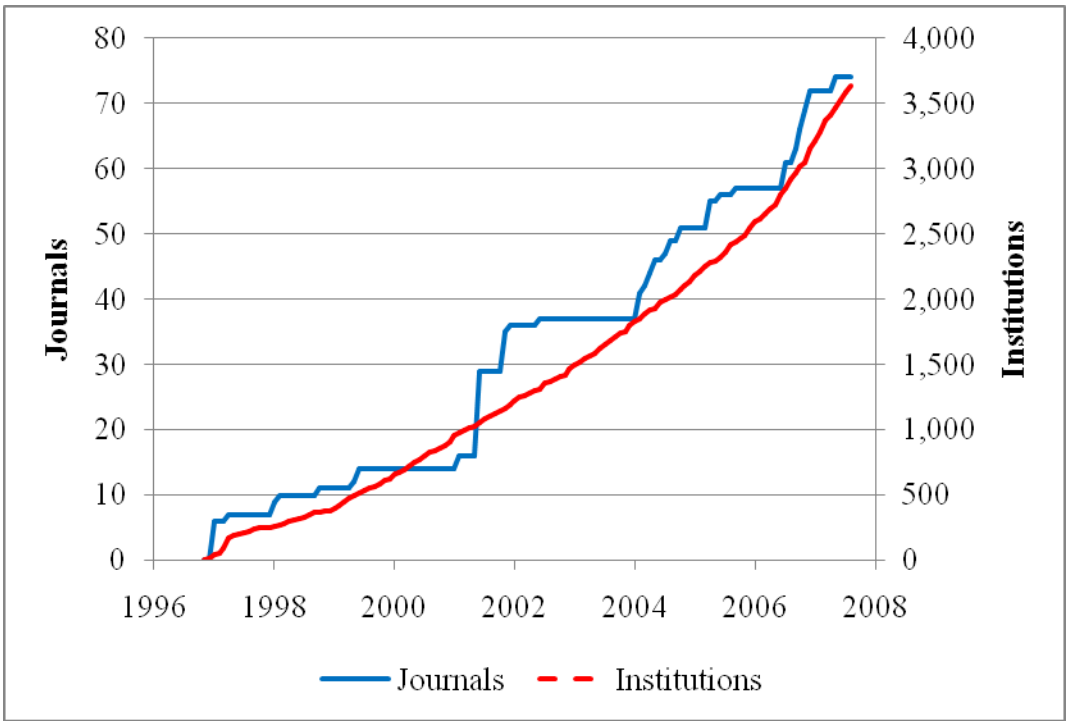

Figure 2
The Effects of Lower Input Costs Due to JSTOR

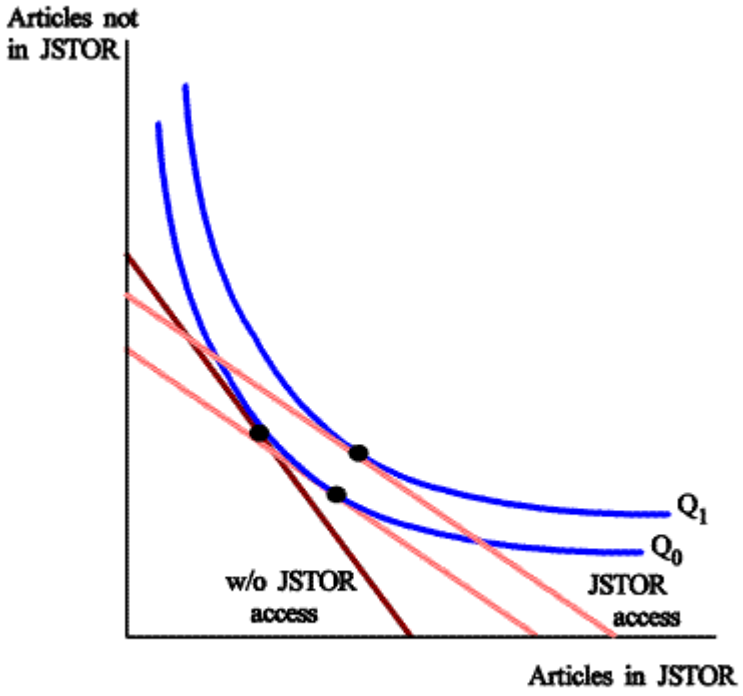

Table 1
Sample Journals and Year First Available in JSTOR

Journal	Year	Journal	Year	Journal	Year
AMER ECON REV	1997	J BANK FINAN		J POST KEYNES ECON	
AMER J AGR ECON	2004	J COMP ECON		J PUBLIC ECON	
AMER J ECON SOCIOL	2006	J DEVELOP ECON		J RISK INS	2001
APPL ECON		J DEVELOP STUD		J TRANSP ECON POLICY	
BROOKINGS PAPER	2001	J ECON BEHAV ORGAN		J URBAN ECON	
CAMB J ECON		J ECON DYN CONTROL		JAHRB NAT STATIST	
CAN J ECON	2001	J ECON EDUC	2004	KYKLOS	
DEVELOP ECON		J ECON HIST	1998	LAND ECON	2004
ECON DEVL CULT CHG	2004	J ECON ISSUE		NAT TAX J	
ECON GEOGR	2001	J ECON PSYCH		OXFORD BL ECON STAT	
ECON HIST REV	2001	J ECON THEOR		OXFORD ECON PAPERS	2002
ECON INQ		J ECONOMETRICS		POLIT EKON	
ECON J	1998	J ENVIR ECON MANG		PUBLIC CHOICE	
ECON LETT		J FINAN ECON		QUART J ECON	1997
ECON MODEL		J FINAN QUANT ANAL		RAND J ECON	2001
ECON REC		J HEALTH ECON		REG SCI URBAN ECON	
ECON SOC		J HUM RESOUR	2001	REV ECON STATIST	1997
ECONOMETRICA	1997	J IND ECON	1998	REV ECON STUD	1999
ECONOMICA	2001	J INT ECON		S AFR J ECON	
EKON CAS		J LABOR ECON	2001	SCAND J ECON	2006
EXPLOR ECON HIST		J LAW ECON	2004	SCOT J POLIT ECON	
FOOD POLICY		J MACROECONOMICS		SOC CHOICE WELFARE	
FUTURES		J MATH ECON		SOUTHERN ECON J	2004
INSUR MATH ECON		J MONETARY ECON		THEOR DECIS	
INT ECON REV	2001	J MONEY CREDIT BANK	1997	TIJD ECON SOC GEOG	
J ACCOUNT ECON		J POLICY MODELING		WORLD DEVELOP	
J AGR ECON		J POLIT ECON	1997	WORLD ECON	

Table 2
Number of Institutions by Number of Journals Available Through JSTOR

	Number of Journals Available Through JSTOR					
Year	0	1-10	11-20	21-30	31-40	41-50
Pre-1997	542	0	0	0	0	0
1997	446	96	0	0	0	0
1998	399	6	137	0	0	0
1999	340	0	202	0	0	0
2000	306	0	236	0	0	0
2001	267	6	119	150	0	0
2002	220	5	60	257	0	0
2003	180	3	44	315	0	0
2004	159	2	25	229	127	0
2005	144	2	19	189	188	0
2006	130	1	15	139	61	196

Table 3
Descriptive Statistics for Article Sample

	Mean	Std. Dev.	Min.	Max.
Number of References to JSTOR Journals	5.64	5.32	0	81
Number of References to non-JSTOR Journals	4.20	4.31	0	64
Number of JSTOR Available Journals	9.26	9.91	0	29
Number of JSTOR Available Journals Squared	183.98	241.74	0	841
Sample includes 43,111 articles.				

Table 4
The Estimated Effect of JSTOR Availability on Referencing Behavior

	JSTOR Journals		Non-JSTOR Journals	
Num. of JSTOR Journals	0.06523**	0.05759**	-0.02826**	-0.02967**
	(0.00188)	(0.00172)	(0.00169)	(0.00157)
Num. of JSTOR Journals Squared	-0.00079**	-0.00069**	0.00028**	0.00027**
	(0.00008)	(0.00007)	(0.00007)	(0.00007)
Other Controls:				
Year (32 Dummies)	X	X	X	X
Citing Jour. (79 Dummies)		X		X
R Squared	0.09	0.25	0.04	0.19

Robust standard errors in parentheses. Sample includes 43,111 articles.
* significant at 5%; ** significant at 1%

Table 5
The Estimated Effect of JSTOR Availability on Age of Reference

	(1)	(2)	(3)	(4)
JSTOR Available Dummy	3.808**	3.198**	3.370**	0.127**
	(0.033)	(0.037)	(0.038)	(0.047)
Other Controls:				
Year (32 Dummies)		X	X	X
Citing Jour. (79 Dummies)			X	X
Cited Jour. (79 Dummies)				X
R Squared	0.03	0.03	0.05	0.12

Robust standard errors in parentheses. Sample includes 423,861 references.
* significant at 5%; ** significant at 1%

Table 6
Descriptive Statistics for Institution by Year Sample

	Mean	Std. Dev.	Min.	Max.
Number of Publications	2.74	4.98	0	72.34
Number of Citations to these Publications	30.86	105.84	0	3167.17
Impact Factor weighted Number of Publications	1.64	4.02	0	78.92
Number of JSTOR Available Journals	3.87	9.74	0	48.00
Number of JSTOR Available Journals Squared	109.90	340.38	0	2304.00
Fraction of Potential Citers with JSTOR Access	0.19	0.18	0	1.00
Sample includes 17,344 observations (542 Institutions by 32 Years).				

Table 7
The Estimated Effect of JSTOR Availability on Research Productivity

	Pubs.	Citations		Impact Pubs	
Num. of JSTOR Journals	0.01056**	0.00131	-0.00045	0.01159**	0.01005**
	(0.00164)	(0.00315)	(0.00317)	(0.00140)	(0.00135)
Num. of JSTOR Journals Squared	-0.00014**	-0.00020*	-0.00017*	-0.00015**	-0.00013**
	(0.00004)	(0.00008)	(0.00008)	(0.00004)	(0.00004)
Fraction of Potential Citers with JSTOR			0.79815**		0.69976**
			(0.05405)		(0.01998)
Other Controls:					
Year (32 Dummies)	X	X	X	X	X
Inst. (542 Dummies)	X	X	X	X	X
R Squared	0.48	0.21	0.22	0.38	0.43

Robust standard errors in parentheses. Sample includes 17,344 institution years.
* significant at 5%; ** significant at 1%

Table 8
Robustness of JSTOR on Research Productivity Results

	Pubs.		Citations		Impact Pubs	
JSTOR Journals						
Lag 2 Years		0.00684		-0.00414		0.00230
		(0.00458)		(0.00963)		(0.00427)
Lag 1 Year		-0.00009		0.01010		0.00721^{+}
		(0.00424)		(0.00903)		(0.00388)
Current Year	0.00470^{+}	0.00597^{+}	0.00837^{+}	0.00520	0.00496*	0.00376
	(0.00246)	(0.00345)	(0.00484)	(0.00715)	(0.00204)	(0.00292)
Lead 1 Year		-0.00116		0.00205		0.00025
		(0.00319)		(0.00708)		(0.00265)
Lead 2 Years		0.00040		-0.00836		-0.00029
		(0.00255)		(0.00596)		(0.00212)
JSTOR Journals Squared						
Lag 2 Years		-0.00020		0.00033		-0.00011
		(0.00017)		(0.00037)		(0.00017)
Lag 1 Year		-0.00006		-0.00062^{+}		-0.00027^{+}
		(0.00016)		(0.00034)		(0.00015)
Current Year	-0.00001	-0.00007	-0.00017^{+}	-0.00006	-0.00006	-0.00001
	(0.00005)	(0.00011)	(0.00010)	(0.00022)	(0.00004)	(0.00009)
Lead 1 Year		0.00001		-0.00010		0.00002
		(0.00009)		(0.00019)		(0.00008)
Lead 2 Years		0.00001		0.00014		0.00003
		(0.00006)		(0.00012)		(0.00005)
Other Controls:						
Year	X	X	X	X	X	X
Inst.	X	X	X	X	X	X
Inst. Trend	X	X	X	X	X	X
R Squared	0.09	0.10	0.05	0.04	0.07	0.08
Observations	16,802	15,718	16,802	15,718	16,802	15,718

Robust standard errors in parentheses. $^{+}$ significant at 10%; * significant at 5%; ** significant at 1%. Specifications include year and institution fixed effects as well as institution specific trends.