The Impact of an Online Journal Database on Academic Scholarship:

JSTOR and Economics

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

Michael R. Ward, *University of Texas at Arlington*

March, 2009

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 of scholars 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 both the quantity and quality of the 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

This research 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.

The Impact of an Online Journal Database on Academic Scholarship:

JSTOR and Economics

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

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

subscription for institutions and the number of journals available to scholars at those institutions from JSTOR has varied since the service began in 1997. 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 that JSTOR increased the research productivity of these economists as measured by both the rate at which they publish and the citations that these publications receive.

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]

[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, of course, the game Monopoly.
[3] The investigation of these hypotheses is planned for our future research in this area.

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. 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 has 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) merely serves to add to the consumptive value of being an academic without enhancing research productivity.

JSTOR is one such 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. There are currently 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 pattern of journal incorporation into 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 both 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.[4]

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. For instance, some professors might download JSTOR articles and print them on a network printer to appear learned or actively researching to their colleagues. Alternatively, some

[4] It is still possible that an institution with expectations of increasing research output might subscribe to JSTOR sooner than an otherwise similar institution without such an expectation. In this case, our results would reflect some reverse causality.

researchers might download and print JSTOR articles only to later discard them or otherwise fail to utilize the tool to enhance research.

To address this ambiguity, we adopt a simple model of 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, existing human capital, human capital they may acquire, 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 and 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 may be 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, is separable from the other inputs used in research production.

We assume researchers face input costs, *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, $-MP_1/MP_2$, with the ratio of factor input prices, $-w_1/w_2$. 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 journal articles that were continuously indexed by ISI, 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 us to develop a sample of over 40,000 articles in 79 journals written by authors at one of 444 institutions over 23 years making a total of about 500,000 references to these journals.

Information about research institutions' access to JSTOR economics and business collections was made available by JSTOR.[5] Institutions can 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

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

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 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 73 journal titles archived by JSTOR are also among the more than 160 journal titles indexed by ISI. Our final sample will include 79 journal titles that were indexed by ISI continuously from 1985-mid-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 be 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 these collections that will eventually include 29 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 these primarily 'consumers' of journal articles will not be included in our analyses. Our sample includes both top research institutions and those not as well known for their research output. Our broad view of

what constitutes a research institution yields a total of 444 institutions whose scholars published 25 or more articles in the sample of 79 journals from 1985 through 2007. 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 reports the number of institutions by degree of JSTOR access by year.

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 444 institutions authored 42,181 articles in the 79 journals and collectively made close to half a million citations.[6] For the purpose of this study, variables of interest for an article include the journal title, date of publication, the authors' 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 to date.

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

[6] That is the affiliation of the first-author was one of the 444 institutions. See more on this below.
[7] In fact, we include only the first 200 citations made by an article, Fewer than 10 articles include more than 200 citations.

available to her in the archive.[8] It does not change the 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. Articles by Reference Sample

We first examine the effect of JSTOR on the individual references made by an article. Even 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 an article that should be cited 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 discover and incorporate more 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

[8] 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 bur are *cited* by articles subsequently published at a later date.

topic, thereby likely reducing search costs.[9] 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 would have no effect on the researcher's propensity to refer to articles available through JSTOR.

This simple analysis suggests a test of the hypothesis. If JSTOR increases research productivity, having access to a journal through JSTOR should lead a researcher to refer more often to articles in that journal. 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 444 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-reference. The sample was limited to articles written by an author affiliated with one of these 444 institutions.[10] The sample was further limited to 42,181 articles published in one of these 79 journals and potential references to one of these 79 journals. Thus, each article generates exactly 79 observations indicating the number of references to each of the 79 journals, including the instances of zero references.

[9] This assumes *certis 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 the unaffected and concentrate on JSTOR's search cost reducing effects.

[10] In fact, the first author must have been affiliated with one of the 444 institutions. Focusing on the first author may omit some publications of an institution. However, there is no reason to believe that this omission is related to JSTOR access and thus no reason to believe that this selection criterion will bias our estimates.

Most of the sample institutions became subscribers to JSTOR sometime after it was launched in 1997 but many did not. 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.[11] We also create a variable measuring the number of JSTOR journals available to the authors as well as the total number of references made by the article.

With this sample, our tests of hypotheses are implemented with a simple Ordinary Least Squares (OLS) regression model describing the number of references made to a journal. Because the dependent variable can only take on positive integer values, the assumption of normally distributed errors is invalid. This will tend to bias estimates of standard errors but not of coefficient estimates. Since our estimated standard errors are often a fraction of the value necessary for hypothesis testing and traditional levels, for reasons described below, we proceed with OLS assuming that the bias in standard errors is small enough that inference testing will not be problematic.

We fit the data to the regression equation:

$$NRef_{aj} = \beta_0 JSTOR_{aj} + \beta_1 NJSTOR_a + +\beta_2 NJSTOR_a^2 + +\beta_3 NRef_a + \gamma X + \varepsilon_{aj},$$

where a indexes an article and j indexes a potential reference. The variable $NRef_{aj}$ is the number of references to journal j by article a. $JSTOR_{aj}$ is a dummy variable indicating whether any authors have access to the journal through JSTOR. Our direct test is whether or not β_0 is positive.

[11] 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.

The variables $NJSTOR_a$ and $NJSTOR_a^2$ are the number of JSTOR journals available to the authors and its square. Our indirect test of a JSTOR effect is whether or not β_1 is positive. Because the increase in the number of journals that became available to researchers over time was primarily due to the inclusion increasingly less well-known journals, we expect to observe a diminishing marginal effect of the number of JSOR available journals, or that β_2 is negative. An article that refers to more of the previous literature will have a greater likelihood of referring to any particular journal, implying that β_4 is positive. Finally, X represents different vectors of potential control variables allowing for fixed year, referring journal and referred to journal effects.

In practice, inclusion of a high dimensional X vector poses an estimation challenge. Our sample includes 3,332,299 observations and X can include over 100,000 dummy variables. Our controls will variously include 23 year dummies, 79 referring journal dummies, 79 referred to journal dummies and interactions of these sets of dummies. Instead of estimating these parameters directly, we implement these fixed effects by differencing data relative to the mean value for the cell defined by the dummy variables. With an OLS estimator, differencing the data as above will yield the same coefficient estimates as would be obtained by including the dummy variables (Greene, 2006).

Summary statistics for this sample are provided in Table 3. The mean number of references to any of these journals is 0.124. At least one journal was referenced 51 times in by single article but over 90% of all article-reference observations result in zero references.[12] By coincidence the mean of the JSTOR availability dummy variable is also 0.124. This is always

[12] The highest number of references to a single journal was the *Jounal of Accounting and Economics* referrenced 51 times in Fields, et al (2001).

zero prior to 1997 but reaches 0.31 by 2007. This implies that by 2007 the average number of JSTOR journals available to any particular institution is less than one, yet for those institutions that subscribed to JSTOR the average number of journals available during the sample period was 0.58. Likewise, the number of the 79 journals available through JSTOR is zero before 1997, grows to a mean of 24 by 2007, and has a sample mean of just fewer than 10.

Estimation results are reported in Table 4. Different columns refer to different sets of controls included in *X*. Because the most often cited journals (e.g., the *American Economic Review*, the *Journal of Political Economy*, and the *Quarterly Journal of Economics*) also were included in JSTOR earliest, we include referred to and year dummies in all specifications. Column (2) also accounts for referring journal fixed effects so that any differences in researcher propensity to cite to JSTOR available journals across referring journal will not be captured by JSTOR variables. Column (3) includes the interactions between the referring and referred to journals to capture a greater likelihood of referring within a field. Finally, column (4) includes three-way interactions between referring journal, referred to journal and year to capture any changes over time in the likelihood of within field reference patterns.

Across all sets of controls, our hypotheses are generally confirmed – JSTOR access is estimated to affect reference patterns as predicted. The direct effect, the estimated coefficient of $JSTOR_{ai}$, is about 0.010 to 0.012 in columns (1-3) and is one-quarter the size in column (4). This represents about a 10% average increase in the number of references to a journal once it is accessible from JSTOR. The indirect effect, the estimated coefficients of $NJSTOR_a$ and its square indicate an increase in the likelihood of referring to a journal in column (1) but indicate a decrease in specifications with more controls. At the sample mean of 9.76 JSTOR available journals, columns (2-4) indicate a decrease in the average number of references to a journal of -

0.003 to -0.008 or 3-7%. The diminishing marginal effect of journals available to a researcher is only significant in column (4).

b. The Articles Sample

Aggregating the data to the article level provides the opportunity for alternative tests of the impact of JSTOR on reference patterns, but also allows for a test of any JSTOR effect on research productivity. An alternative, but related, test to that described above is to examine if the number of references in (or out) of the subset of journals that an author will have JSTOR access to increases (or decreases) once she does obtain JSTOR access to more journals. Thus, for the authors of an article, we define the set of references to articles in our set of 79 journals that any author at these institutions would eventually obtain JSTOR access to. We likewise define the set of references to articles in the 79 journals that these institutions did not have JSTOR access to by the end of our sample period. Our test of this hypothesis is to see if the size of either set of references is affected by improved JSTOR access.

One measure of enhanced research productivity is the number of citations an article attracts in the future. JSTOR access allows authors to become more aware of the prior literature and to take this knowledge into account in their research. If improved access to the accumulated knowledge base improves the quality of analysis, then the analysis associated with improved access should have a greater subsequent impact on the discipline, all else equal. We measure the impact of a research article by the number of subsequent citations made to an article. Our hypothesis is that, all else equal, an article will be cited more often in the future if its authors had JSTOR access to more journals.

Formally, our estimating equations are:

$$Ln\left(NRef_a^{in} + 1\right) = \beta_0 NJSTOR_a + \beta_0 NJSTOR_a^2 + \Gamma X_a + \nu_a^{in}$$
$$Ln(NRef_a^{out} + 1) = \beta_0 NJSTOR_a + \beta_0 NJSTOR_a^2 + \Gamma X_a + \nu_a^{out}$$
$$Ln\left(NCite_a + 1\right) = \beta_0 NJSTOR_a + \beta_0 NJSTOR_a^2 + \Gamma X_a + \nu_a^{cites}$$

where $NJSTOR_a$ is defined as above. $NRef_a^{in}$ and $NRef_a^{out}$ are the numbers of references to our set of 79 journals that any coauthor of article *a* will and will not have access to through JSTOR. $NCite_a^{out}$ is the numbers of future citations that an article will eventually receive. As before, we allow for diminishing marginal productivity of the number of journals archived within JSTOR by including a quadratic term and we allow for various controls, *X*, representing different sets of dummy variables. Here, we assume exponential distributions for what are integer count data.

Table 5 provides summary statistics for the data set aggregated to the article level. On average, there are 5.6 references to journals that will be included in an article authors' set of JSTOR available journals. There are 4.2 references to journals that will not be included in an article authors' set of JSTOR available journals. The average article makes 27 references in total. Because we are counting references only to our set of 79 journals, references to other journals, or to books, reports, working papers, and the popular press are excluded from the analysis. The average article will be cited just over 10 times in subsequent ISI indexed articles, but Newey and West (1987) had garnered nearly 1,700. Articles published earlier in the sample have had more time over which to be cited regardless of their future impact. To account for this truncation, it is important to include year dummies in *X*. As before, the average number of JSTOR available journals is just under 10, increasing from zero before 1997 to 24 in 2007.

Table 6 reports results for the effect of JSTOR availability on references to JSTOR available journals. Year dummies are included in all specifications, referring journal dummies are included in columns (2) and (4) and author institution dummies are included in columns (3)

and (4). The journals included in JSTOR journals earlier and the institutions subscribing to JSTOR earlier both tend to be relatively higher quality, which could be correlated with referencing patterns. Nevertheless, even after accounting for these effects in column (4), we observe a positive effect of JSTOR availability on referencing to journals that will be in JSTOR. The linear term is always positive and the quadratic term is always negative. At the mean number of JSTOR available journals, the column (4) estimates imply a 30% increase in references (or 1.7 more references).

Table 7 reports results for the effect of JSTOR availability on references to journal articles not available through JSTOR available. As before, year dummies are included in all specifications, with journal dummies and author institution dummies variously included in columns (2) through (4). The journals included in JSTOR journals earlier and the institutions subscribing to JSTOR earlier both tend to be of relatively higher quality. As before, the specification in column (4) is the most conservative, but all specifications indicate that JSTOR access decreased references to journals not available through JSTOR. The linear term is always negative and the quadratic term is always positive, though not always significant. At the mean number of JSTOR available journals, the column (4) estimates imply an 11% decrease in references (or 0.5 fewer references).

Table 8 reports our examination of the effect of JSTOR access on one measure of article quality. The results reported above indicate that JSTOR access has had an effect on research inputs: researchers substitute away from journals not available through JSTOR toward those that are. We hypothesize that lower costs for this input caused researchers to move to a higher isoquant. Future citations to an article are one measure of article quality. As before, year dummies are included in all specifications, and are meant to control for differences in citation

patterns attributed to the age of the article. As before, journal dummies and author institution dummies are variously included in columns (2) through (4) and, as before, the specification in column (4) is the most conservative. All specifications are consistent with increased access to JSTOR increasing the number of citations to an article. Moreover, all specifications are consistent with diminishing marginal productivity of additional journals included in JSTOR. At the mean number of JSTOR available journals, the column (4) estimates imply a 7% increase in citations (or 0.7 more citations).

c. The Institution by Year Sample

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 make the production of a given article more efficient implying less time spent on each article, or it could allow a researcher to more thoroughly investigate the prior literature relevant to the article implying a more insightful analysis. The former would imply a quantity improvement while the latter would imply a quality improvement. We identify these effects with the number of publications generated by authors affiliated with an institution in a year and the sum of all subsequent citations that these articles will generate.

Assuming an exponential distribution, this implies the research quantity and quality relationships:

$$Ln(Pubs_{it}+1) = \beta_0 JSTOR_{it} + \beta_1 JSTOR_{it}^2 + \Gamma X_{it} + \omega_{it}^{Pubs}$$
$$Ln(Cites_{it}+1) = \beta_0 JSTOR_{it} + \beta_1 JSTOR_{it}^2 + \Gamma X_{it} + \omega_{it}^{Cites}$$

where *i* refers to institution and *t* refers to year. This is augmented with zero values for the year-institution combinations in which no articles are published. Additionally, for multiple authored

articles, we aggregated to the institution of the first author. This is likely to introduce some measurement error into the research output for an institution, but this error is not likely to be correlated with our JSTOR variables.[13]

Table 9 reports summary statistics for this aggregation. The average institution in our sample generated just over four articles per year among our sample of 79 journals, though Harvard University generated 70 in a year. These articles generated an average of 42.5 citations in a year though Harvard University generated nearly 3,000 citations to articles published in a single year. On average, institutions had 6.7 of the 79 journals available to it through JSTOR in a given year. Of course, this was initially zero, but rose to 29 journals for many institutions by 2007.

Tables 10 and 11 report regression results for the number of publications and the number of citations using various specifications of X. As before, all specifications include year dummies. Columns (2) and (4) include journal effects and columns (3) and (4) include institution fixed effects. While institution fixed effects are implemented with a series of dummy variables, journal effects are implemented as the number of publications in journal j in a year divided by the total number of journals published in the year so as to not artificially inflate estimates.[14] It may be possible that to increase the number of publications for an institution if articles are placed into journals with lesser publication standards. These publications may then be cited less often as they are in less widely cited journals Similarly, better institutions are better able to place articles

[13] Alternatively, we could assign each of N institutions for N coauthors some weighted output level f(N) (e.g. 1/N). While this might reduce this measurement error, it is not clear that it would affect the estimates of our variables of interest.

[14] Consider an institution that places 10 publications into 10 journals and another that places 10 publications into 5 journals. In the first case, each of the 10 journal variables where the institution placed articles have a value of 0.1 and journal variables where the institution did not place articles have value 0. In the second case, journal variable values are 0.2 and 0 respectively. In either case, the sum of the dummies is not affected by the number of articles published.

in journals that are more heavily cited. Inclusion of journal and institution variables attempts to control for these effects.

In practice, journal and institution effects appear to be important to the estimated JSTOR effect for both publication and citations. Across all columns, greater JSTOR access is associated with more articles and more citations to articles. Moreover, both effects appear to exhibit diminishing marginal productivity. At the sample mean column (4) of both tables indicate that JSTOR access is associated with 9% more publications and 14% more citations.

Combined, the empirical evidence suggests that JSTOR access has a number of different, if predictable, influences on the pattern of economic research over the past decade. First, we document that articles written by authors at institutions with access to the JSTOR archive of economics-related journals are more likely to refer to a journal after it is accessible through JSTOR and, holding this effect constant, are less likely to refer to any one journal when more journals are accessible through JSTOR. This is consistent with the idea that the JSTOR archive reduces search costs and thereby increases the likelihood that a JSTOR archived journal will be cited, *ceteris paribus*. Next we document that articles published by authors with access to more journals in the JSTOR archive tend to refer to more JSTOR archived articles after access is obtained, at the expense of non-JSTOR archived journals. This indicates a substitution effect induced by JSTOR. To the extent that JSTOR archived the highest quality economics journals first, this might suggest that the research referring to these high quality journals is itself of higher quality. We find that articles published by scholars at institutions with JSTOR access tend to receive more citations in the future, consistent with the hypothesis that JSTOR access increases the quality of economic research at the article level. Finally, we document that economics research developed at institutions with JSTOR access experience both an increase in the quantity

of research published in JSTOR archived journals, as measured by total publications in a rather broad sample of 79 economics journals, but also an increase in the quality of economic research, as measured by the total citations to papers published after JSTOR access is granted. Combined, these findings suggest that JSTOR alters the relative price of at least one input to the economic research production function, inducing both substitution and scale effects. We interpret this as suggesting that the public-good nature of JSTOR, from the point of view of individual researchers, might be justified by the impact of the generation of new, arguably higher quality, economic research on social welfare.

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 linkable citations within 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 science, 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

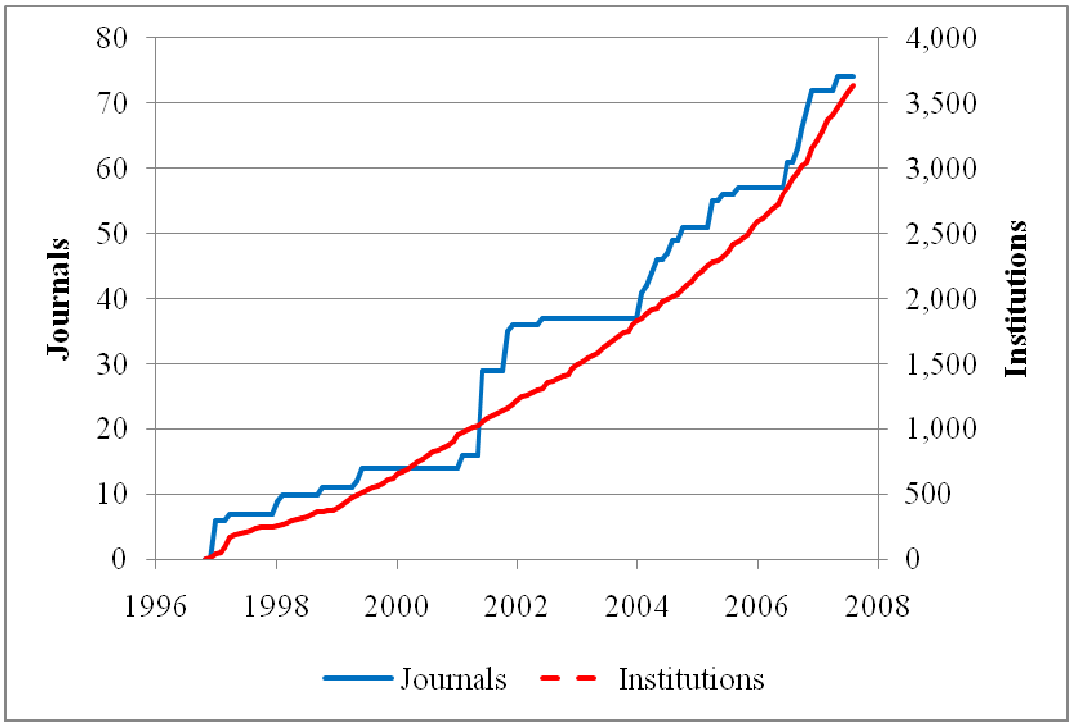

Figure 2
The Effects of Lower Input Costs Due to JSTOR

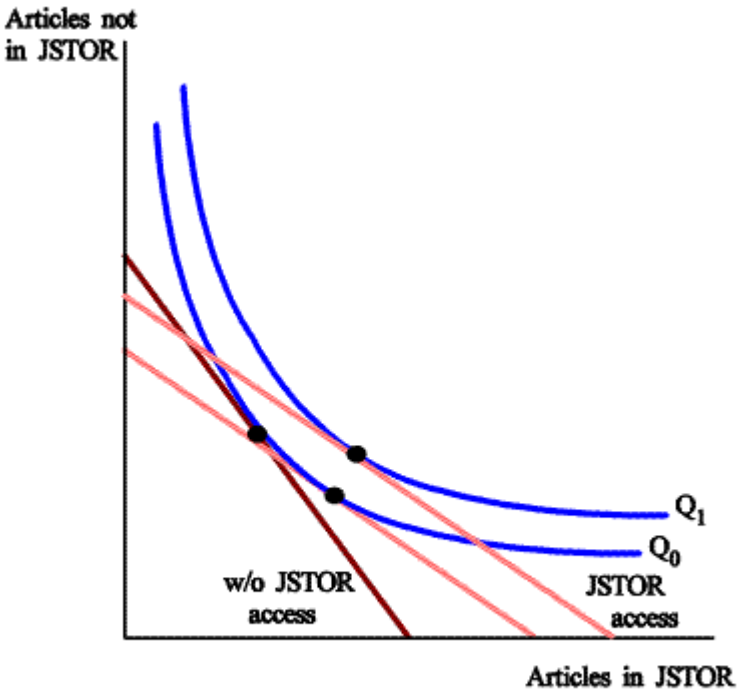

TABLE 1
Sample Journals and Year First Available in JSTOR

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
Pre 1997	444	0	0	0
1997	294	150	0	0
1998	194	246	3	1
1999	158	282	2	2
2000	140	299	2	3
2001	107	86	248	3
2002	68	24	351	1
2003	59	13	370	2
2004	48	10	125	261
2005	50	3	91	300
2006	36	10	42	356
2007	62	6	47	329

Table 3

Descriptive Statistics for Article x Reference Sample

	Mean	Std. Dev.	Min.	Max.
Number of References	0.124	0.623	0.0	51.0
Dummy if Journal Available Through JSTOR	0.124	0.329	0.0	1.0
Number of JSTOR Available Journals	9.761	10.312	0.0	29.0
Number of JSTOR Available Journals Squared	201.608	262.185	0.0	841.0
Total Number of References	27.099	16.928	0.0	554.0
Sample includes 3,332,299 observations (42,181 articles by 79 journals).				

Table 4

The Estimated Effects of JSTOR Journal Access on Referencing Patterns

	(1)	(2)	(3)	(4)
Dummy if Journal Available Through JSTOR	0.01292 (0.00106)**	0.01292 (0.00106)**	0.01035 (0.00094)**	0.00269 (0.00091)**
Number of JSTOR Available Journals	0.00067 (0.00014)**	-0.00046 (0.00014)**	-0.00043 (0.00013)**	-0.00106 (0.00009)**
Number of JSTOR Available Journals Squared	-0.00001 (0.00001)	0.00001 (0.00001)	0.00001 (0.00001)	0.00002 (0.00000)**
Total Number of References	0.00303 (0.00002)**	0.00283 (0.00002)**	0.00283 (0.00002)**	0.00261 (0.00002)**
Other Controls:				
Years (23 Dummies)	X	X	X	
Referred to Journal (79 Dummies)	X	X		
Referring from Journal (79 Dummies)		X		
Referred from x Referred to (6,241 Dummies)			X	
Referred from x Referred to x Year (143,543 Dummies)				X
Standard errors in parentheses. Sample includes 3,332,299 article by reference observations.				
* significant at 5%; ** significant at 1%				

Table 5

Descriptive Statistics for Article Sample

	Mean	Std. Dev.	Min.	Max.
Number of References to Journals that Will Become Available Through JSTOR	5.615	5.350	0.0	81.0
Number of References to Journals that Will Not Become Available Through JSTOR	4.206	4.324	0.0	64.0
Number of Forward Citations	10.304	31.318	0.0	1698.0
Number of JSTOR Available Journals	9.761	10.312	0.0	29.0
Number of JSTOR Available Journals Squared	201.608	262.189	0.0	841.0

Sample includes 42,181 observations.

Table 6

The Estimated Effect of JSTOR Availability on Log References to JSTOR Available Journals

	(1)	(2)	(3)	(4)
Number of JSTOR Available Journals	0.07263**	0.05955**	0.03628**	0.03098**
	(0.00185)	(0.00169)	(0.00193)	(0.00175)
Number of JSTOR Available Journals Squared	-0.00093**	-0.00075**	-0.00051**	-0.00043**
	(0.00007)	(0.00007)	(0.00007)	(0.00007)
Other Controls:				
Year (23 Dummies)	X	X	X	X
Journal (79 Dummies)		X		X
Institution (444 Dummies)			X	X
R Squared	0.12	0.28	0.03	0.20

Standard errors in parentheses. Sample includes 42,181 articles.

* significant at 5%; ** significant at 1%

Table 7

The Estimated Effect of JSTOR Availability on Log References to

Sources Not Available Through JSTOR

	(1)	(2)	(3)	(4)
Number of JSTOR Available Journals	-0.02435**	-0.02774**	-0.01139**	-0.01241**
	(0.00167)	(0.00156)	(0.00183)	(0.00169)
Number of JSTOR Available Journals Squared	0.00016*	0.00018**	0.00007	0.00006
	(0.00007)	(0.00006)	(0.00007)	(0.00006)
Other Controls:				
Year (23 Dummies)	X	X	X	X
Journal (79 Dummies)		X		X
Institution (444 Dummies)			X	X
R Squared	0.04	0.19	0.02	0.17

Standard errors in parentheses. Sample includes 42,181 articles.

* significant at 5%; ** significant at 1%

Table 8

The Estimated Effect of JSTOR Availability on Log Citations to an Article

	(1)	(2)	(3)	(4)
Number of JSTOR Available Journals	0.03719**	0.01435**	0.01869**	0.01273**
	(0.00228)	(0.00199)	(0.00243)	(0.00220)
Number of JSTOR Available Journals Squared	-0.00087**	-0.00062**	-0.00067**	-0.00060**
	(0.00009)	(0.00008)	(0.00009)	(0.00008)
Other Controls:				
Year (23 Dummies)	X	X	X	X
Journal (79 Dummies)		X		X
Institution (444 Dummies)			X	X
R Squared	0.29	0.47	0.29	0.42

Standard errors in parentheses. Sample includes 42,181 articles.

* significant at 5%; ** significant at 1%

Table 9

Descriptive Statistics for Institution by Year Sample

	Mean	Std. Dev.	Min.	Max.
Number of Publications	4.131	5.819	0.0	70.0
Number of Forward Citations to these Publications	42.562	123.404	0.0	2,956.0
Number of JSTOR Available Journals	6.692	9.956	0.0	29.0
Number of JSTOR Available Journals Squared	143.896	251.730	0.0	841.0
Sample includes 10,212 observations (444 Institutions by 23 Years).				

Table 10

The Estimated Effect of JSTOR Availability on Log of Articles Published

	(1)	(2)	(3)	(4)
Number of JSTOR Available Journals	0.06723**	0.03720**	0.03029**	0.02726**
	(0.00428)	(0.00314)	(0.00300)	(0.00219)
Number of JSTOR Available Journals Squared	-0.00071**	-0.00037**	-0.00038**	-0.00039**
	(0.00017)	(0.00012)	(0.00012)	(0.00008)
Other Controls:				
Year (23 Dummies)	X	X	X	X
Journal (79 Variables)		X		X
Institution (444 Dummies)			X	X
R Squared	0.29	0.64	0.43	0.70

Standard errors in parentheses. Sample includes 10,212 articles.

* significant at 5%; ** significant at 1%

Table 11

The Estimated Effect of JSTOR Availability on Log of Forward Citations

	(1)	(2)	(3)	(4)
Number of JSTOR Available Journals	0.16495**	0.07215**	0.07062**	0.05413**
	(0.00945)	(0.00623)	(0.00708)	(0.00517)
Number of JSTOR Available Journals Squared	-0.00396**	-0.00280**	-0.00293**	-0.00283**
	(0.00037)	(0.00024)	(0.00027)	(0.00020)
Other Controls:				
Year (23 Dummies)	X	X	X	X
Journal (79 Variables)		X		X
Institution (444 Dummies)			X	X
R Squared	0.18	0.66	0.26	0.61

Standard errors in parentheses. Sample includes 10,212 articles.

* significant at 5%; ** significant at 1%