“Sited, Sighted, and Cited:

The Effect of JSTOR in Economic Research”

I. Introduction

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

Specifically, we examine the impact of one particular Internet tool, the JSTOR journal archive, on one particular discipline, economics. JSTOR is the first large scale Internet-based journal article storage, search, and retrieval service.[1] Scholars at research institutions that subscribe to JSTOR can easily find and read at their desktops the archived articles published in

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

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 economists refer to these journals more often and refer to excluded, i.e., non-JSTOR, journals less often. Thus JSTOR appears to have lowered the relative cost of the former causing a substitution away from the later. Moreover, we find evidence that JSTOR led to an increase the research productivity of these economists as measured by the rate at which they publish but that this increase was experienced only at lesser ranked institutions.

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

[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.

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).

There have been investigations of the effect of the Internet on university research as well. 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. A minimum requirement for the Internet to have increased productivity is that a research tool that exists only because of the Internet, such as JSTOR, has increased productivity. We can say nothing about the net effect of ICT that would also include greater access to Internet toys that could decrease productivity.

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

As a panacea to the binding space constraints, JSTOR appears to have failed, 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 January, 2013, the archive contained over 50 million pages from nearly 1,700 academic journals with 800 participating publishers and more than 10,000 participating subscribers from 160 countries. Usage has steadily grown to the point that users downloaded more than 74 million articles in 2010. If traffic to the web site is any indication, it appears evident that increasing numbers of publishers, subscribing institutions, and scholars have benefited from the development of the JSTOR archive.

The first journals and institutions included in JSTOR tended to be more research oriented. JSTOR management consciously decided to first archive the journals that were most widely read and had the largest number of older volumes so to maximize the physical amount of shelf space released. Similarly, subscribership diffused from the leading research institutions to progressively weaker research institutions. Many of the leading US institutions were charter members at the time of JSTOR's launch and there were almost 200 US subscribers by the end of 1997. Some non-US institutions obtained access during 1997 but non-US subscribership only accelerated in 1999-2000. Among the non-US subscribers too, the leading institutions tended to be earlier adopters.

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 or intertemporal changes so that our estimates reflect only the increased referencing and publishing due to JSTOR for a given institution.

III. A Simple Model of Research Production

Notwithstanding the obvious metric of web traffic, it is not immediately clear whether JSTOR or other online "tools" actually enhance research output, either in quality or quantity. We adopt a simple model of the academic research production process using standard neo-classical theory. The model provides a framework in which to develop testable hypotheses regarding the 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 *p* is positive and that researchers maximize a shadow profit function:

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

That is, researchers face an optimization problem analogous to the optimization problem facing any neo-classical firm. Quite generally, researchers equate the marginal rate of transformation, -MP_i/MP_j, with the ratio of factor input prices, -w_i/w_j, for $i \neq j, i = 1 \dots N, j = 1 \dots N.$. Let x_1 and x_2 be the processes of searching for, reading, and incorporating an article from journals 1 and 2 into one's research. These searches have marginal benefits of MP_1 and MP_2 and costs of w_1 and w_2. Access to journal 1 through JSTOR, but not journal 2, is assumed to reduce w_1, but not change w_2. As a consequence, we expect the researcher to make more use of articles in journal 1. This will involve a substitution effect away from articles found in journal 2 (see Figure 1). Since the costs of production for any level of research will have declined, we expect a scale effect as researchers produce more and/or better research. This could be the case either if a more thorough understanding of past research improves the quality of current research or if time saved through easier access to past research allows a researcher to work on more projects. The net effect on substitute inputs is ambiguous but the direct effect on JSTOR accessible journals is unambiguously toward greater usage. The goal of our analysis is to determine if we can detect 1) an increase in usage of JSTOR accessible journals (the direct effect), 2) a decrease in usage of journals not accessible from JSTOR (an indirect effect), and 3) an increase in research output and/or quality (a scale effect).

IV. Journal and JSTOR Data

The data for the analysis come from JSTOR's own records of journals archived and institutions' access arrangements and from ISI's Social Science Citation Index (SSCI) database for the economics discipline from 1985 through 2006. After matching these data sources by institution, journal and year, 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 over 2,000 institutions worldwide 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.[3] 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

[3] 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 three 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. We confine ourselves to the sample years 1991-2006. This provides some years prior to the inception of JSTOR in 1998 and ends before competing journal searching mechanisms became popular. While JSTOR was the clear first-mover for this service, by 2006 many journals were also offering access to their archived past issues.[4] In general, the journals generally regarded as the most important 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 the top 100 journal titles as determined by the ratio of 'in' citations to 'out' citations over the entire sample.. 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*. Out of these titles, 31 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,

[4] Still, many journals implementation of archival information was through 'outsourcing' to JSTOR.

government agencies, non-governmental agencies, private consultancies, and many high schools. Since our focus is on the ‘production’ of journal articles, most of those entities that are consumers of 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. For our broad view of what constitutes a research institution, our sample included the top 500 institutions in the world in terms of publication output over the sample.[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. Figure 2 indicates that research institutions had access to more journals over time, more so for higher ranked institutions. At the same time, Figure 3 indicates that as late as 2006, there was considerable variation in JSTOR availability within a tier of research institution.

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

---

[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, we rarely observe zero publications for an institution and year.

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.[7]

We distinguish "elite" institutions from other researching institutions based on within sample calculations of publishing production. We suspect that access to journals through JSTOR will have a larger impact on researchers at institutions less well-known for their research. These institutions tend not have as extensive of library facilities, will tend not to have as many researcher colleagues and will host fewer research presentations. In a sense, JSTOR may serve to reduce the comparative advantage of elite research institutions. For some of our analyses, we divide the sample of 500 top institutions into the top 100 and the 101st through 500th institutions. We are not interested in generating a ranking of departments but institutions around the 100th by our ranking include: George Mason University, Georgia State University, Southern Methodist University, University of Vienna, and University of Alberta. By comparison, institutions around the 500th include: California State University at Northridge, University of Denver, Kent State University, Middle East Technical University, and University of Göttingen.

V. The Effect of JSTOR on Referencing Patterns

We first document that JSTOR has affected the referencing patterns of scholars writing in peer reviewed economics journals. We construct a balanced panel measuring the number of references made by scholars at institution *i* to journal *j* in year *t*, $Referenes_{ijt}$. Because of the "moving wall" in which the most recent articles, usually those published in the previous three

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

[7] These are citations as of August, 2007 when the data were collected and so represent a data truncation issue.

years, are not available through JSTOR, our count measure omits these references. Our sample includes the top 500 institutions, the top 100 journals and the years 1991 through 2006 for a total of 800,000 observations. Likewise, we construct a dummy variable indicating weather researchers at institution i had JSTOR access to journal j at time t, $JSTOR_{ijt}$. We expect that, all else equal, JSTOR access will increase referencing to a journal. In our sample, no institutions had access to any journals in at the beginning of the sample and some did not at the end. Moreover, some institutions gained to different sets of journals and did so in different years. Thus, we have three-dimensional variation in our variables of interest.

We also hypothesize that a researcher will be less likely to reference any particular journal when he has access to an increasing number of other journals via JSTOR. In this sense, references "compete" with each other across journals. To test this hypothesis, we construct a variable equal to the number of other journals that researchers at institution i have access to in year t, $NumJSTOR_{ijt}$. If references compete with each other, then the number of references to journal j should fall as $NumJSTOR_{ijt}$ rises.

Our specification relates the count of references to our measures of JSTOR access using the negative binomial estimator.[8] The count of references differs across journals as some are generally more highly cited, across institutions as some exhibit are more research productivity than others, and over time as there is a gradual secular increase in the number of references per article. We are not interested in modeling the characteristics of journals or institutions that are the cause in of this variation in references. Instead, we adopt three way fixed effects that proxy for the specific bundle of characteristics of a journal, institution and year. Implicitly, we are

[8] Since over ninety percent of all institution, journal, and year cells have zero references, we also estimated a logit model for any reference. These are not reported but are similar to the negative binomial specification.

assuming that these bundles of characteristics do not change over the 16 years, or do not do so in a way correlated with JSTOR access. Our estimating equation assumes becomes:

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

where we expect $\beta_0 > 0$ and $\beta_1 < 0$ and $\beta_0 > |\beta_1|$.

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

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

The variables of interest are those relating to JSTOR access. For all three samples, JSTOR access to a journal significantly increases references to that journal and an increase in the number of other journals an institution has access to significantly reduces references to the journal. Moreover, the magnitude of the first effect is substantially larger than the latter effect.[9] As hypothesized, obtaining JSTOR access to a journal increases references to that journal by almost 20% and decreases references to another journal by about three-quarters of a percent on average. In addition, it appears that the referencing patterns at top 100 institutions are affected to a larger extent than at non-top 100 institutions.

VI. The Effect of JSTOR on Research Productivity

While the findings thus far suggest that researchers are using JSTOR in increased levels, they do not indicate that research benefits from increased JSTOR use. We now turn to the possible effect of JSTOR access on the research output produced by institutions. Since our raw data include every article published in these 100 journals, we aggregate all articles authored by researchers affiliated with the top 500 institutions for each of the 16 years. Output is measured three ways: 1) the number of publications by researchers at the institution and published in the year in question, 2) the number of publications weighted by a quality measure for the journal, and 3) the number of forward citations the these published articles subsequently generate. The journal quality measure is simply the ratio of forward citations to the journal to the number of backward citations from the journal over the whole sample period. While the determination of the actual quality of journals may be more nuanced, this measure has the virtue of being simply

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

and consistently applied. Again, our purpose is not to generate another journal ranking but to create an objective measure of journal quality. The third measure examines the quality of each article, as defined by the citations it later garners, rather than imputing quality from the journal. This provides us with three possible output measures for a balanced panel of 8,000 observations Summary statistics for this sample are reported in Table 4. The average institution generates seven publications per year which are cited an average of 65 times.

We implement a dynamic panel estimator in which an institution's research output in a year is a function of previous values of its annual research output and JSTOR access when the article was written:

$$Res.Output_{i,t} = \sum_{\tau=1}^{T} \rho_\tau Res.Output_{i,t-\tau} + \beta CountJSTOR_{i,t-3} + \mu_i + v_{i,t}. \quad (2)$$

The count of the 100 journals that a researcher has JSTOR access to, *CountJSTOR*, is lagged three years to account for publication lags.[10] Since our sample ends in 2006, the latest value for *CountJSTOR* is 2003, which largely precedes widespread adoptions of alternatives to JSTOR. The lagged values of journal article production represent persistence in the publication propensity of an institution over time. The number of lags is typically chosen in each application as required to reduce the likelihood of bias in the coefficient values.

We implement a two-step GMM dynamic panel estimator of equation (2). Because of the inclusion of lagged dependent variables, the error term is not independent of the regressors. A strategy to avoid this problem is including instrumental variables drawn from further lagged values and differences. Roodman (2006) provides an overview of the methods that have been developed for this purpose. In our case, we include year dummy variables that are treated as

[10] Results are only mildly sensitive to different publication lag assumptions.

exogenous. However, both the dependent research output variables and the JSTOR count variable are treated as endogenous. In practice, three lags of the dependent variables are needed to eliminate the confounding effects of lags. Thus, we include lags from four years before and earlier. However, to avoid the problem of too many instruments we limit instruments to lags of eight years. Still, these specifications include a total of 134 instruments. The specifications reported in Table 5 fail to reject the second order autocorrelation in an Arellano-Bond test with three lags of the dependent variable. Also, with these specifications, both the Sargan and Hansen tests fail to reject the exogeneity of the instrument set.

As shown in Table 5, all three measures of research output are persistent over time. These effects fade with lag length but three lags are often significant. The JSTOR results are remarkably consistent across output measures. Over this sample period, these measures of research output increased about 1% in the short-run with each additional journal available through JSTOR. However, these effects get amplified due to the estimated persistence of research output over the previous three years. The half-life of a JSTOR "shock" is about five to eight years by which time an additional journal available through JSTOR would have increased output by 4% to 6% depending on the measure of research output. These results suggest that specific Internet research tools do have the desired effect of increasing research productivity.

A concern is that these results may not reflect a causal link between JSTOR and research productivity because of reverse causality. Since an institutions' research productivity is not constant over time, it is possible that institutions with rising research potential, and expectations of greater future research output, obtained earlier and more extensive access to research related resources, including JSTOR, than otherwise similar institutions. This would be the case, for example, if administrators invested in greater research-related resources because they foresaw they would be complementary with the higher expected future research productivity. In this way, JSTOR access could be endogenous to future research productivity.

A robustness check against this form of JSTOR endogeneity is to exclude from the sample institutions with the greatest increases in research productivity over the sample period. We measure increased research productivity by comparing the average number of quality-weighted Economics publications for the 1986-1990 period to the same measure for the 2001-2006 period. The top 20 percent, or 80 institutions, with the highest percent growth rate were then excluded from the sample and the models were re-estimated; the results are reported in Table 6. As can be seen, the results are not qualitatively different from those using the entire sample. The estimated coefficients on the number of JSTOR journals are slightly smaller and the standard errors are slightly higher than those obtained using the full sample. However, only the measure for the number of backward citations loses significance at standard levels. This suggests that the measured JSTOR effects are not solely due to reverse causality.

VII. Conclusions

JSTOR represents a single, though important, new tool available to academic researchers. New tools are emerging, such as the Research Papers in Economics (RepEc) and Social Science Research Network (SSRN) working paper archive for economics, improvements to tools are becoming more common, 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.

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

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

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
The Effects of Lower Input Costs Due to JSTOR

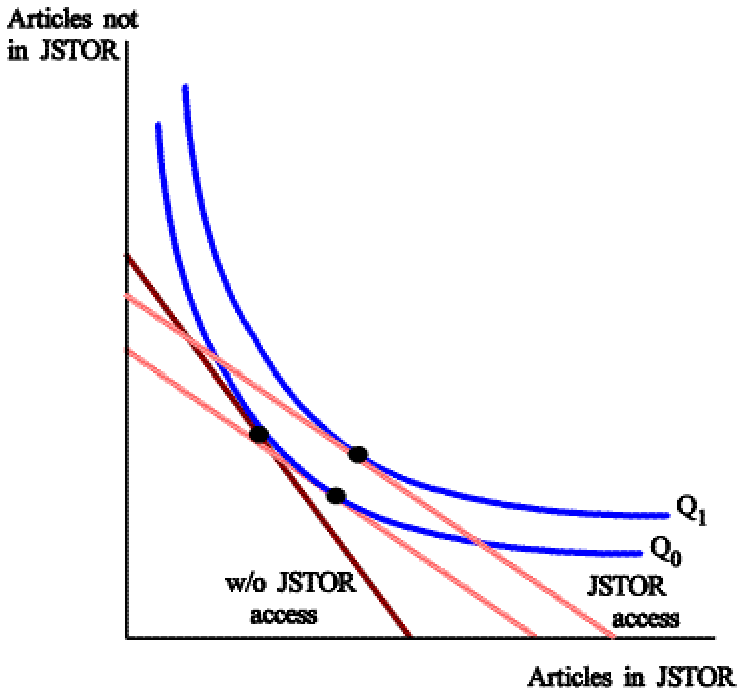

Figure 2

Growth in Average Number Journals Available through JSTOR over Time

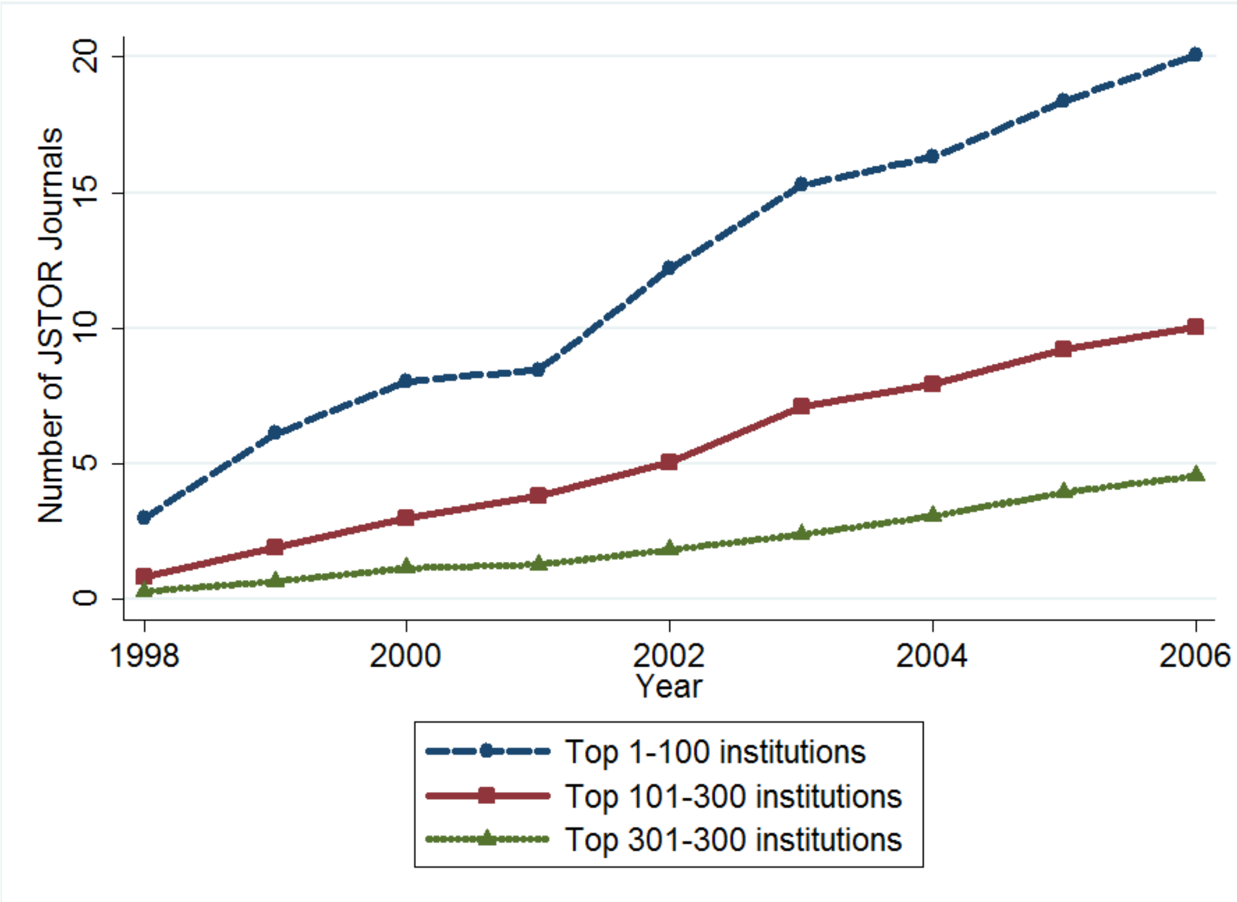

Figure 3

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

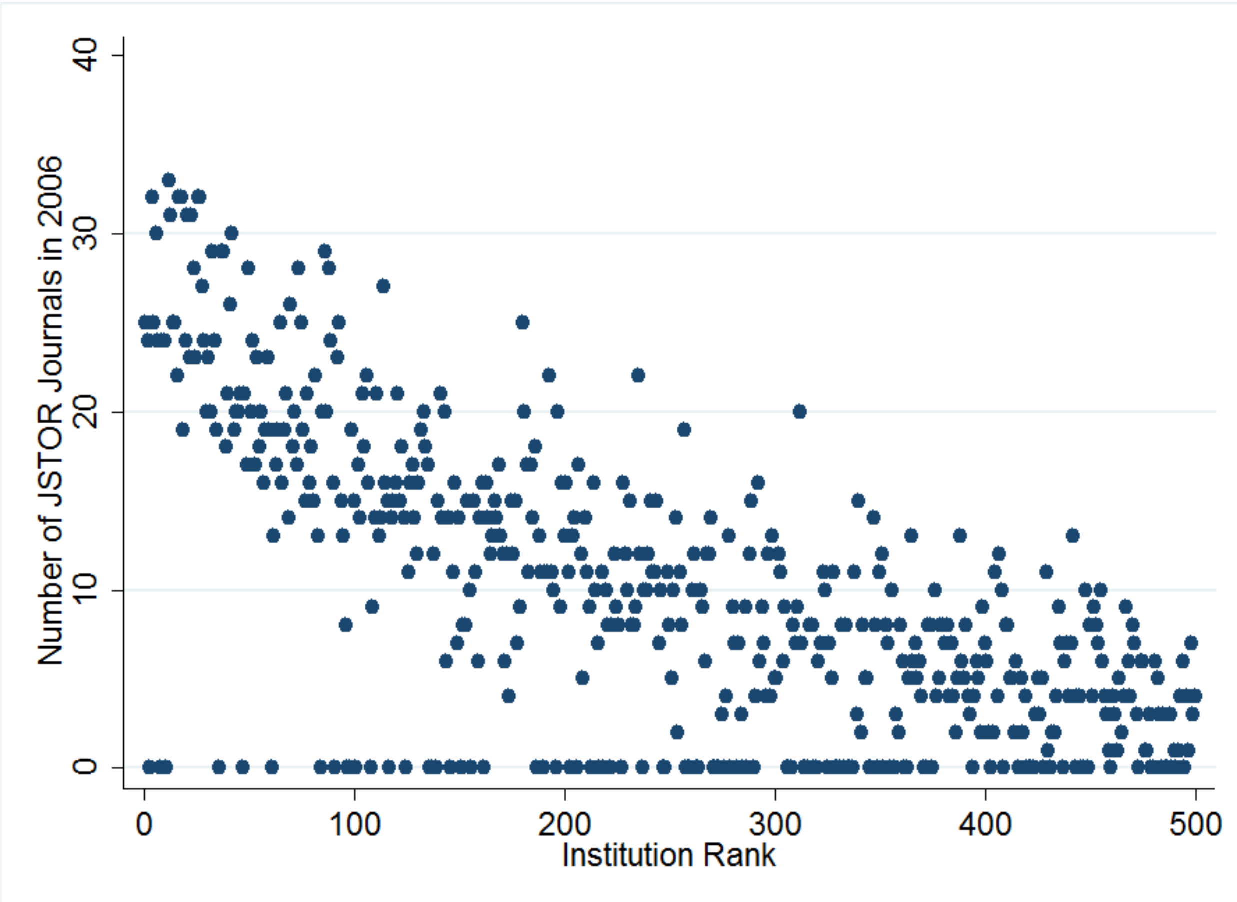

Table 1
Sample Journals and Year First Available in JSTOR

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

Table 2

Mean Values for Referencing Sample

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

Table 3

The Effect of JSTOR Access on Referencing Patterns

	All 500 Institutions	Top 100 Institutions	101-500 Institutions
JSTOR Access	0.184**	0.153**	0.088**
	(0.012)	(0.054)	(0.007)
Number of other journals with JSTOR Access (×100)	-0.752**	-2.462**	-0.332**
	(0.090)	(0.542)	(0.062)
Journal	sign.	sign.	sign.
Year	sign.	sign.	sign.
Institution	sign.	sign.	sign.
Observations	800,000	150,000	640,000

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

Table 4

Mean Values for Publishing Sample

	Mean	Standard Deviation
Number of Publications	7.179	9.669
Quality Weighted Publications	3.506	6.398
Forward Citations	65.635	176.957
Ln Number of Publications	1.183	1.494
Ln Quality Weighted Publications	0.332	1.436
Ln Forward Citations	2.365	2.333
Number of JSTOR Journals Available Lagged Two Years	3.055	5.703
8,000 Observations over 16 years and 500 institutions.		

Table 5

The Effect of JSTOR Access on Research Productivity

	Ln Total Publications	Ln Quality Adjusted Publications	Ln Forward Citations
JSTOR Journals Lagged 3 Years	0.010**	0.009**	0.010**
	(0.003)	(0.002)	(0.003)
Dependent Variable Lagged 1 year	0.473**	0.477**	0.544**
	(0.105)	(0.116)	(0.124)
Dependent Variable Lagged 2 years	0.359**	0.391**	0.326**
	(0.097)	(0.108)	(0.112)
Dependent Variable Lagged 3 years	0.041+	0.048*	0.023
	(0.021)	(0.020)	(0.019)
Year Dummies	Sign.	Sign.	Sign.
Arellano-Bond for AR(2) in 1st Diff.	-1.32	-1.38	-0.64
	[0.19]	[0.17]	[0.53]
Sargan test of over-identification $\chi^2(113)$	113.1	103.1	107.5
	[0.48]	[0.74]	[0.63]
Hansen test of over-identification $\chi^2(113)$	126.80	117.3	119.7
	[0.17]	[0.37]	[0.32]

The sample is a strongly balanced panel of 500 institutions and 16 years. Year dummy estimates are suppressed. Both JSTOR and the dependent variables are treated as endogenous with lags from 4 to 8 years generating 134 instruments. Standard errors in parentheses. P-values are in brackets. ** p<0.01, * p<0.05, + p<0.1

Table 6

The Effect of JSTOR Access on Research Productivity

Excluding 80 Institutions with the Highest Percent Growth in Research Productivity

	Ln Total Publications	Ln Quality Adjusted Publications	Ln Forward Citations
JSTOR Journals Lagged 3 Years	0.009**	0.006*	0.006
	(0.003)	(0.003)	(0.004)
Dependent Variable Lagged 1 year	0.526**	0.549**	0.660**
	(0.120)	(0.120)	(0.130)
Dependent Variable Lagged 2 years	0.346**	0.373**	0.264*
	(0.110)	(0.120)	(0.120)
Dependent Variable Lagged 3 years	0.033	0.035+	0.014
	(0.023)	(0.021)	(0.021)
Year Dummies	Sign.	Sign.	Sign.
Arellano-Bond for AR(2) in 1st Diff.	-0.84	-1.01	0.02
	[0.40]	[0.31]	[0.98]
Sargan test of over-identification $\chi^2(113)$	97.5	89.4	114.7
	[0.85]	[0.95]	[0.44]
Hansen test of over-identification $\chi^2(113)$	113.7	103.4	120.0
	[0.47]	[0.73]	[0.31]

The sample is a strongly balanced panel of 420 institutions and 16 years. Year dummy estimates are suppressed. Both JSTOR and the dependent variables are treated as endogenous with lags from 4 to 8 years generating 134 instruments. Standard errors in parentheses. P-values are in brackets. ** p<0.01, * p<0.05, + p<0.1