department of economics
"S. Cognetti de Martiis"

Via Po, 53 – 10124 Torino (Italy)
Tel. (+39) 011 6704917 - Fax (+39) 011 6703895
URL: http//www.de.unito.it

WORKING PAPER SERIES

The Diffusion of Information Technology and the Increased Propensity of Teams to Transcend Institutional and National Borders

Anne E. Winkler, Wolfgang Glänzel, Sharon Levin and Paula Stephan

Dipartimento di Economia "S. Cognetti de Martiis"

LEI & BRICK - Laboratorio di economia dell'innovazione "Franco Momigliano"
Bureau of Research in Innovation, Complexity and Knowledge, Collegio Carlo Alberto

Working paper No. 10/2011

Università di Torino

The Diffusion of Information Technology and the Increased Propensity of Teams to Transcend Institutional and National Borders

Anne E. Winkler, University of Missouri St. Louis
Wolfgang Glänzel, Department of MSI, Centre for R&D Monitoring, K.U. Leuven, Belgium
Sharon Levin, University of Missouri St. Louis
Paula Stephan, Georgia State University and NBER

Draft: March 14, 2011

This research is funded in part by a grant from the Andrew W. Mellon Foundation titled "The Diffusion of Information Technology Across Institutions of Higher Education: Effects on Productivity by Type of Institution and Gender." The authors are grateful for expert research assistance from Erin Coffman and useful discussions with Michael T. Allison.

Corresponding author: Paula Stephan, Department of Economics, Georgia State University, Atlanta, GA 30302-3992, pstephan@gsu.edu.

Abstract

This study examines the relationship between the diffusion of IT and changes in collaboration patterns across institutional and national borders. To undertake the research, the authors match an explicit measure of institutional IT adoption (domain names, e.g. www.umsl.edu) with institutional data on all published papers indexed by ISI for over 1,200 U.S. four-year colleges, universities and medical schools for the years 1991-2007. The publication data examined cover the social sciences and natural sciences and narrower fields such as economics and biology. Two measures of institutional collaboration are examined: (1) percent of papers produced by a U.S. institution with one or more co-authors at another U.S. institution (US-US); and (2) percent of papers produced by a U.S. institution with one or more non-U.S. coauthors (US-INTL). We first describe collaboration patterns across universities and then use regression analysis to examine the impact of IT exposure on multi-institution collaboration. IT exposure is measured by the number of years elapsed since an institution's adoption of a domain name. Results indicate dramatic growth in the percentage of both US-US and US-INTL collaborations, as well as important differences by field. The study provides modest evidence that length of IT exposure has had a positive and significant effect on both US-US and US-INTL collaborations.

JEL Codes: A14, I23, O33
Key Words: Coauthorship; Collaboration; Information Technology, Diffusion, Higher Education

I. Introduction

It has been widely documented that the mean number of authors per paper in a range of fields including science and engineering and the social sciences has been increasing dramatically (Heffner, 1981; Braun et al., 2001; Cronin, 2001; Glänzel, 2002; Cronin et al., 2004; Glänzel & Schubert, 2004; Adams et al., 2005; Wuchty, Jones & Uzzi, 2007). The growth is due to both an increase in the number of authors working together at one institution as well as an increased propensity to work with individuals at different institutions. A variety of factors have contributed to these changing patterns, such as the increased financial support for science (Price & Beaver, 1966; Patel, 1973; Heffner, 1981) and more recently increased funding to foster collaboration (Stephan, 2010), the increased importance of interdisciplinary research, the increased mobility of scientists (Price & Beaver, 1966; Adams et al. 2005) and the narrowing of the expertise that individual researchers bring to a research problem (Beaver, 2001; Jones, 2005; Stephan, 2010). The rapid spread of connectivity has also contributed to the increased propensity of teams to transcend institutional boundaries (Agrawal & Goldfarb, 2008; Hamermesh & Oster, 2002; Kim, Morse & Zingales, 2009; Ding et al., 2010). This paper extends this research by examining the relationship between the diffusion of IT and changes in collaboration patterns across both institutional and national borders.

To investigate changing collaboration patterns and the role of IT, we match an explicit measure of institutional IT adoption (domain names, e.g. www.umsl.edu) with institutional data on all published papers in the natural science and social science fields indexed by ISI for over

1,200 four-year colleges, universities and medical schools for the years 1991-2007.[1] Information technology is measured by when/whether each of the 1,200-plus institutions adopted a domain name. The analysis is disaggregated by field (biology, chemistry, physics, and economics) given field differences in how research is produced. The unit of analysis is the institution. In addition, findings are reported by "tier" given the considerable variation in research mission and resources available to different institutions. Two measures of collaboration are examined: (1) percent of papers produced by a U.S. institution with one or more co-authors at another U.S. institution and (2) percent of papers produced by a U.S. institution with one or more non-U.S. coauthors.

From a policy perspective, collaborations that transcend institutions are important for at least two reasons. First, they have been shown to produce higher quality research as measured by citations (Hamermesh & Oster, 2002; Carayol & Matt, 2006; Jones et al., 2008; He et al., 2009). Second, the availability of IT has been shown to have a democratizing effect for *research active scientists* (Ding et al., 2010). It is not clear, however, whether the availability of IT has a democratizing effect at the *institutional* level. While previous research has shown that IT is particularly beneficial to active researchers working at lower-tier institutions, it does not follow that IT raises the productivity of lower-tier institutions, which are disproportionately staffed by research inactive scientists, relative to the productivity of higher-tier institutions. This is an empirical question.

The plan of this paper is as follows. In section two we provide a brief overview of the prior literature on trends in co-authorship and the likely explanations for these trends. Section

[1] In prior work (Levin et al. 2010) we identified 1,348 four-year colleges, universities and medical schools in the United States that have been in existence since 1980 and have not undergone a "substantial" change in structure such as a major acquisition or merger. The 1,348 total excludes specialized institutions such as engineering schools and religious institutions and represents the universe of institutions. In this study, the set of institutions is slightly smaller as discussed in the data section.

three extends this discussion by focusing exclusively on the likely impact of IT on collaboration patterns. In section four we discuss the publication data and in section five we discuss our measure of IT diffusion and the patterns that we observe. Finally, section six presents the regression results; the discussion and conclusion follow in section seven.

II. Prior Research on Collaboration Patterns

A. Collaboration Trends

A large and growing number of studies have identified a significant increase in the number of co-authors per paper ("team size") and in the number of coauthored papers.[2] For instance, in their analysis of approximately 13 million published papers in science and engineering from 1955 to 2000, Wuchty, Jones & Uzzi (2007) found an increase in team size in all but one of the 172 subfields studied. They also found that average team size nearly doubled, going from 1.9 to 3.5 authors per paper.[3] Adams et al. (2005) found similar results for the top-110 research universities in the United States, reporting that the average number of authors per paper in the sciences grew by 53.4%, rising from 2.77 to 4.24 over the period 1981-1999.

Notably, much of the observed change in collaboration is a result of increased collaboration across institutions, rather than solely among researchers located at the same physical location. For instance, Jones, Wuchy & Uzzi (2008) analyzed papers published by 662 universities in the United States which had received one or more NSF grants using data from ISI/Web of Science for the period 1975-2005. Publication patterns (sole-authored, multi-

[2] In this paper collaboration is defined as more than one institutional address on the same article. The data do not permit us to study intra-university collaboration. Katz and Martin (1997) duly caution that authorship-based measures such as this miss some collaborations and erroneously captures others. In the first case, individuals may collaborate but publish separately. In the second case, individuals decide to publish similar work together but did not do it jointly.

[3] Team size even increased in mathematics, generally seen as the domain of individuals working alone and the field least dependent on capital equipment.

authored within the same institution, multi-authored across U.S. institutions) were analyzed for three broad fields: Science and Engineering (S&E), Social Sciences, and Arts & Humanities. Among their findings, in S&E, multi-institution collaboration was rare in 1975 but was the fastest growing type of collaboration between 1975 and 2005. Indeed, by 2005, 32.8% of all publications had co-authors from more than one institution. Growth in multi-institution collaboration was slightly greater in the social sciences, amounting to 34.4% of all publications in 2005. When they broadened the set of possible collaborators to include other U.S. institutions (private, government labs) and international institutions, they find this pattern reversed: collaborations that span institutional boundaries were greater for S&E than for the social sciences.[4] A rise in US-US and US-INTL collaborations has been identified by the National Science Board. During the period 1988 to 2003, the number of addresses on an article in S&E with at least one U.S. address grew by 37% while the number of foreign addresses more than tripled (National Science Board, 2006, Table 5-18).[5]

B. Explanations for Observed Trends in Collaboration

Several factors have likely contributed to the increased role that collaboration plays in research, as outlined in reviews by Sonnenwald (2007) and Stephan (2010).[6] First, the importance of interdisciplinary research and the fact that major breakthroughs often occur in emerging disciplines, encourages collaboration. Systems biology, which involves the intersection

---

[4] See Wuchty, Johnes, & Uzzi (2008) Supplemental Online Material, Figure 1.

[5] During the same period, the number of names increased by approximately 50%, suggesting that lab size was growing slightly faster than institutional collaboration (National Science Board 2006, Table 5-18).

[6] Changing patterns in collaboration present certain challenges for organizations. For example, as the number of co-authors grows, it becomes increasingly difficult to evaluate curriculum vitas at tenure and promotion time. Historically, for example, individuals were penalized if they only published with their mentor after completing a postdoctoral appointment. In recent years, however, programs such as the Medical College at the University of Pennsylvania have relaxed this rule and now consider such individuals for promotion.

of biology, engineering and physical sciences, is a case in point. By definition, no one has all the requisite skills required to work in the area; researchers must rely on working with others.

Second, and related, researchers arguably are acquiring narrower expertise over time in order to compensate for the educational demands associated with the increase in knowledge (Jones, 2005). Narrower expertise, in turn, leads to an increased reliance on teamwork in research. This is true in the social sciences as well as the "hard" sciences.

A third factor that fosters collaboration is the vast amount of data that are becoming available. While the Human Genome Project (and the associated GenBank database) is perhaps the best known example, other large databases in the natural sciences have also become available, such as PubChem. In the social sciences, databases such as the Panel Study of Income Dynamics, and various U.S. Census data sets, also play a large role. The availability of such data sources promote collaboration by providing a common resource for authors to draw upon. Another factor that contributes to collaboration both in the sciences and the social sciences is the practice of sharing materials and data. Increased complexity of equipment also fosters collaboration. For example, at the very extreme, are the teams assembled to work at colliders. The four detectors associated with the Large Hadron Collider (LHC) that recently came on line at CERN outside Geneva, have combined team size of just under 6,000 (Overbye, 2007). Barnett, Ault & Kaserman (1988) suggest two other factors that lead persons to seek coauthors. One is the desire to minimize risk by diversifying one's research portfolio through collaboration; the other is the increased opportunity cost of time.

In addition, quality considerations may play a role in collaboration. The literature on scientific productivity suggests that scientists who collaborate produce "better" science than do individual investigators (Wuchty, Jones & Uzzi, 2007). Also, Bozeman & Corley (2004) suggest that some types of collaboration, such as between a senior faculty member and a junior researcher, might benefit the field as a whole, by enhancing technical skills and expanding networking ties (Bozeman & Corley, 2004). Finally, there is a relatively new factor which has contributed to the growth in team size: the rapid spread of connectivity, which began in the early 1980s with the adoption of BITNET by a number of universities and accelerated in the early 1990s with the spread of the INTERNET.[7] The specific relationship between IT and collaboration is discussed at greater length in the next section.

While many of the same factors explain collaboration patterns across fields, there are important field differences for a variety of institutional and cultural reasons. For instance, research in the biological and chemical sciences almost invariably requires a lab and thus has a strong local component. It may involve collaborators at another lab if materials or data have been exchanged or the research is extremely large in scope, such as mapping the human genome. By way of contrast, work in experimental high energy physics, which requires access to highly specialized and extremely expensive equipment, almost always occurs offsite and thus almost invariably involves collaboration among scientists from different institutions and countries. The same could be said of astronomy. Articles coming out of the IceCube Project, a neutrino observatory in Antartica, for example, lists all project members—256 on the most recent paper, coming from nine countries.[8] Research in economics is different: Except for experimental economics, labs are rarely part of economic research; neither is specialized equipment. But data and software can be readily shared and this encourages collaboration.

[7] Notably, Wuchy, Jones & Uzzi (2007) find that team size has grown in all but one of the 171 S&E fields studied during the past 45 years. This suggests that while technology has played a pivotal role in the recent period, many of the other factors encouraging collaboration have been at play for a number of years.

[8] The telescope is the brainchild of Francis Halzen, University of Wisconsin-Madison, and involves 67 faculty, 62 PhD research scientists and postdocs, and 95 students, drawn from 33 institutions, approximately half of which are located outside the United States.

Differences in how research is conducted by field also likely explain the difference in US-US and US-INTL collaboration patterns found in Jones, Wuchty & and Uzzi (2008) and found in the analysis here as well: US-INTL collaboration is more common in some science subfields than in the social sciences, while the reverse is true for US-US collaborations. Again, consider physics. Although theoretical papers often only have one to two authors, experimental papers can have 100's of authors because of the nature of the equipment. As found by Adams et al. (2005), because of the large-sized equipment needed—colliders for physics and telescopes for astronomy—international collaboration is especially common in these fields.

Moreover, the research conducted in fields may differ according to its universal appeal. While problems studied in the natural sciences are important to scientists world-wide, this is not necessarily so in the social sciences, where research may have more of a national focus. Thus, for example, US-INTL collaborations may be less common than US-US collaborations in economics. This tendency may be reinforced by the large investment researchers in the social sciences make in working with country-specific data sources. This is consistent with the findings of Jones, Wuchy & Uzzi (2008) when non-university and international collaborators are taken into account.

Collaboration patterns can also be affected by the fact that faculty often write with students while they are in graduate school and former students after they graduate (Black & Stephan, 2010 and Adams et al., 2005). Thus another factor that can contribute to differential patterns in international collaboration across fields is the likelihood that a PhD trained in the United States is a temporary resident coupled with the likelihood that the student leaves the country after graduation. To be more specific, in 2000, 34% of PhDs who received their degree that year from at U.S. institution in the natural sciences were temporary residents; 53% of those trained in economics were temporary residents and 27% of those trained in the social sciences were temporary residents.[9] These field differences are magnified by the fact that the stay rate (measured in this instance two years after receipt of the PhD) averages 70% in the natural sciences, 48% in economics and 45% in the social sciences.[10] Combining these two effects, one concludes that almost 28% of all newly trained PhDs in economics are temporary residents at the time they graduate and leave within two years of graduating. The comparable percent in the natural sciences is 10%; for the social sciences it is 14.6%.

III. IT, Research, and Collaboration

A. Effect of IT on Research

Without question, technology has played an important role in changing collaboration patterns among researchers. Regardless of field, researchers at different institutions can more easily collaborate as a result of reduced communication costs. Further, IT as noted earlier, has permitted the shared use of large databases such as GenBank.

Not surprisingly, given that how research is produced differs considerably by discipline, the impact of technology on distinct disciplines has differed as well (Walsh & Bayma, 1996; and Walsh, Kucker, Maloney & Gabbay, 2000). Regarding US-US collaboration, we might expect more of such collaboration in the social sciences because most research can be conducted virtually, while working in close proximity (in labs) may be more crucial to producing research in the natural sciences. On the other hand, in the case of fields like physics, especially high-energy physics, that rely on large-scale equipment (colliders), one would expect to find more

[9] National Science Foundation (2010), Table 2-28, Appendix. The natural sciences exclude medical/other life sciences.
[10] Stay rates are estimated by Finn (2007).

U.S.-International collaboration. What remains to be fully sorted out, however, is what role technology has played in multi-institution collaboration patterns and the degree to which its effect differs by field and by institutional tier.

B. Empirical Investigations of the Role of IT on Research and Collaboration

A number of studies have sought to examine the role of IT in explaining the observed increase in research productivity and collaboration. Due to field differences and data limitations, as discussed below, little research has systematically analyzed the impact of an explicit measure of IT by field. Also, most studies have focused on determinants of productivity and collaboration at the individual-level, but not on the drivers of collaboration across institutions. Yet, as noted in the introduction, from a policy perspective it is of interest to know whether the availability of IT has differential effects by tier of institution. Below we review related research and then point to the specific contributions made in this paper.

Much of the early research on the role of IT on research and collaboration has been descriptive in nature (see Friedlander & Bessette, 2003; Appendix B). For example, Hesse et al. (1993) surveyed the subset of oceanographers who used the electronic network SCIENCEnet and found a positive relationship between frequency of use and publication counts. Further, they found that geographically-disadvantaged scientists received a relatively higher productivity gain from IT. Subsequent research by Cohen (1996) and Walsh et al. (2000) expanded the number of disciplines surveyed to include philosophy, political science, and sociology, as well as math and a number of natural sciences and also found a relationship between IT usage and productivity. Notably, however, Cohen's (1996) survey of scientists from a broader set of disciplines found no support for the hypothesis of disproportionate benefits for scientists employed at lower-tier institutions.

A second generation of research has used nationally-representative data sets and more advanced research methods in an effort to identify a causal relationship between IT and productivity. "Second generation" studies in the fields of engineering and the life sciences include Agrawal and Goldfarb (2008), Ding et al. (2010) and Winkler et al. (2010). Agrawal and Goldfarb (2008) examined the impact of BITNET, as measured by date of institutional adoption, on collaboration (coauthorship) in engineering at the institutional level. In their study they used publication data from seven top journals in the field of electrical engineering for the period 1977 to 1991 and divided the institutional affiliations of authors into three groups: elite, medium, and lower tier. They found that faculty at medium-ranked research universities benefitted the most from the adoption of BITNET having increased collaboration with top-tier institutions and increased publishing productivity.

Winkler et al. (2010) appended information on date of adoption of BITNET and DNS to individual-level data on a cross-section of life scientists drawn from the Survey of Doctoral Recipients (SDR). They found some evidence, albeit modest, that individuals at lower-tier institutions benefitted relatively more from IT. Ding et al. (2010) appended these same institutional-level measures of IT to longitudinal individual-level data on research-active life scientists. They found that IT directly enhances research productivity as well as collaboration (as measured by gain in co-authors), and these effects are greater for those located at lower-tiered institutions as well as for women.

Research in the social sciences has largely, though not exclusively, focused on identifying the impact of IT by inferring it from time effects, rather than explicit measure of IT.

For example, Hamermesh & Oster (2002) compared publishing activity in three economics journals for the period 1970-1979 with that for the period 1992-1996. They found almost 20% of authors of jointly-produced articles to be located at distant locations in the more recent period compared to 5% in the earlier period. Rosenblat & Mobius (2004) looked at co-authorship patterns in economics from 1969 to 1999 based on papers published in 8 top economics journals. A novel feature of their study is they also look at the changing nature of the co-authorship--the degree of similarity of the author's research fields. Their analysis found that, at least in the field of economics, as communication costs fall, researchers seek to collaborate with more distant colleagues who share similar interests.

Kim, Morse & Zingales (2009) examined publishing productivity of faculty in economics and finance who were located at an elite institution at some point in time during the period 1970 to 2001. They found that the advantage to being located at an elite institution fell starting in 1970 and had in fact disappeared by the 1990s. Finally, Butler, Butler & Rich (2008) examined collaboration (measured as co-authorship) across universities in the fields of economics and political science using publication data from three top journals in each field. They inferred the time that IT became available based on a review of NBER working papers published during the 1990s. They found that prior to January 1997 an e-mail address was never included; since January 1999, however, almost all papers have an e-mail address. Using this indicator of IT, they found that co-authorship increased with IT, especially at lower-ranked institutions.

A much smaller body of research has examined multi-university collaborations and how they differ by field. These studies have only speculated about the role of IT. The prime example is Jones et al. (2008), which looks at multi-institution collaborations in S&E, Arts and Humanities (A&H) and the Social Sciences since 1975. Not only do they find that multi-institution collaborations are growing faster than within university partnerships, but such collaborations, most notably those that involve a top-ranked institution, produce the most highly-cited research. Sutter and Kocher (2004) examine publishing patterns in economics and similarly point to the rising share of multi-university collaborations as well as the preeminent role of elite institutions. What neither study explicitly investigates is the role of IT in explaining recent trends.

This study builds on and extends the extant research by using institution-level data to systematically examine collaborations across US-US and US-INTL institutions for several fields (biology, physics, chemistry, economics). To date, as seen in this review, far less attention has focused on U.S. collaborations that transcend national boundaries. Moreover, by appending these data to an explicit measure of information technology — measured by an institution's adoption of and length of time exposed to the domain name system, DNS—we are able to provide some preliminary insight into the role that IT plays in explaining recent trends in research collaboration across institutions.

IV. Institutional-Level Collaboration Patterns

A. Data

We utilize institutional data on publications from Web of Science/ ISI for 1,281 four-year colleges and universities for the years 1991-2007.[11] All bibliometric indicators are based on the Web of Science (WoS) volume year in order to avoid the problem that the last available year (in

[11] Initially, we identified 1,348 institutions, approximately the entire universe of institutions that grant baccalaureate degrees or above in the U.S. These data include all four-year colleges, universities and medical schools in the United States that have been in existence since 1980 and have not undergone a "substantial" change in structure such as a major acquisition or merger. Specialized institutions such as engineering schools and religious institutions were excluded. For the analysis at hand, free-standing medical institutions were dropped as well as cases where it was not possible to make a clean match between institution and publication data.

our case 2007) is incomplete because of delayed indexing.[12] A year therefore means WoS volume year, not publication year. This makes annual publication/citation counts more stable.

The analysis here focuses on publications in two broad fields: Natural Sciences and Social Sciences. Arts and Humanities are omitted from all analyses because the way in which research is conducted and published differs between the social and natural sciences. The disciplines are defined according to a classification scheme developed by Glänzel & Schubert (2003). In the case of core journals, assignment is straightforward. For instance, a core journal in economics is the *American Economic Review* and papers in that journal are assigned to economics. On the other hand, some journals cover a broader set of disciplinary topics. In these cases, journal articles may be assigned to more than one disciplinary field. Thus, in this paper, analysis is only done at the "all fields" level and then for select subfields; subfields are not aggregated together, however, to avoid duplication of publications. Three natural science subfields are examined, biology, chemistry, and physics, and one in the social sciences, namely economics.[13]

The institutional data are measured as whole counts, meaning that an article with authors at two institutions is counted once at each institution (and therefore twice, in total) while an article with two authors at the same institution is counted once.[14] Whole counts, also used in related research for economics by Sutter and Kocher (2004), are useful in understanding research production at each institution.

---

[12] This delay can cause publications to be undercounted by 10 to 20%.

[13] Per Table 1 of Glänzel & Schubert (2003), Biology is defined as Category 3, Biosciences (general, cellular & subcellular, genetics); Chemistry is category 8, Physics is category 9, and Economics (which also includes business & management) is subcategory 01 of category 14.

[14] Fractional counts (which counts each article once and then assigns shares to each author) are not possible with these data.

We examine two measures of multi-institution co-authorship patterns: “US-US” measures the number of publications produced at U.S. institutions where at least one co-author is located at another institution inside the U.S.; “US-INTL” measures the number of publications produced by U.S. institutions where at least one co-author is located at another institution outside the U.S.   By way of example, this paper when published will be counted in the publication count of both the University of Missouri-St. Louis and Georgia State University. It will contribute one count to US-US for both institutions since there is at least one co-author at another U.S. institution. It will also contribute one count to US-INTL at both institutions since one of the co-authors works at an institution outside the U.S.

While our measure of collaboration does not measure the percent of coauthors at another institution but rather the presence of one or more coauthors at another institution, we expect that our measure is positively correlated with the number of coauthors. Consider, for example, a paper with two authors and another paper with three authors. In each instance, at least one of the authors is at the University of X. Thus, assuming a uniform distribution for the location of coauthors, if a coauthor has a 50% chance of being at the same institution, then a paper with two authors has a 50% chance of having an author at another institution; the paper with three authors has a 75% chance. This is relevant for our work given that the average number of authors is field dependent. Wuchty, Jones & Uzzi (2007), for example, report that the average number of coauthors in chemistry was 3.69 the period 1996-2000; that in physics was 4.05 and that in economics was 1.71.

Institutions of higher education have very different teaching and research environments depending on their mission. Doctoral institutions devote many more resources to the research mission, including expenditures on IT. Indeed, a recent NSF brief (Christovich, 2010) points to the gap in IT between doctoral and non-doctoral institutions, whether measured in terms of networking (bandwith available) or the availability of supercomputing. Moreover, this gap appears to be widening. Given these differences by tier/mission, this study stratifies institutions into one of four groups: Top Research/Doctoral, Other Research/Doctoral, Master's Level, and Top Liberal Arts. Tiers are assigned based on the 1994 Carnegie Codes (*Carnegie Foundation for the Advancement of Teaching*, 1994. Top Research/Doctoral corresponds to code 11; Other Research/Doctoral corresponds to codes 12, 13, 14; Master's level corresponds to codes 21, 22. Top Liberal Arts institutions refer to the 80 institutions identified in 1996 by *US News and World Report* (1996). Other liberal arts institutions are not analyzed separately here since they tend to principally have a teaching rather than a research orientation. However, the designation "all tiers" includes these teaching institutions.

B. Findings Regarding Trends in US-US and US-INTL Institutional Collaborations

Table 1 provides information on total publications for three time periods (1991-1995, 1996-2000, 2001-2007) for all fields (Social and Natural Sciences combined) and for the four selected subfields: biology, chemistry, physics, and economics. Publication data are also presented separately by tier (excluding liberal arts teaching institutions). The most notable pattern is that regardless of field or tier, the average number of publications has increased substantially. For all fields, all tiers (1,281 institutions), the mean number of publications per institutions increased from 159 to 228 publications over the full period 1991 to 2007. In addition, the percent of institutions with zero publications declined from 24 to 19 percent. These

data also show that publication patterns are highly skewed; the median number of publications rose from 5 to 8. In other words, it is a small set of institutions—the most researchoriented—which are producing most of the publications. This is also apparent in comparing publication patterns for Top Research/Doctoral institutions with the other tiers shown in Table 1. The remainder of the analysis (apart from column 1, Table 1) excludes "other" liberal arts teaching institutions given their different institutional mission; only select liberal arts institutions are analyzed.

Figures 1-3 and Table 2 examine the percent of multi-institution publications for those institutions actively engaged in publishing, defined to be publishing a minimum of four articles in the respective field in a given year. For instance, if an institution produced 50 publications in a given year and 5 of these had at least one other U.S. address listed on the paper, then the percent of US-US collaborations for the institution is 10%. As shown in Figure 1, US-US institutional collaborations were much more frequent than US-INTL institutional collaborations both at the start and end of the study period, though they both experienced substantial increase. US-US collaborations rose from 46% to 70% for all research-active institutions and US-INTL collaborations more than doubled, from 9% to 23%.[15] The pattern is not unexpected given that the opportunity to meet potential coauthors who are domestic is generally greater and the costs of collaboration with domestic colleagues are generally lower.

Figures 2-3 and Table 2 also provide information on trends in US-US and US-INTL collaborations by field. Turning first to US-US institutional collaborations, Figure 2 and Table 2 indicate a higher percentage of US-US collaboration in the social sciences, reflected here by the subfield of economics, as compared with the natural sciences. For instance, in economics, among Top Research/Doctoral institutions, US-US collaborations increased from 57% of all publications for the period 1991-1995 to 70% for the period 2001-2007. By way of comparison,

[15] In interpreting these numbers, readers should keep in mind that the publication data are whole, not fractional, counts.

the figures for top Research/Doctoral institutions in biology were 40% and 55%, respectively. This is an interesting finding in the sense that based solely on the number of coauthors--as the above discussion indicates-- one would expect economics to have a lower percent, not a higher percent.[16]

Figure 2 and Table 2 further show that the pattern of multi-institution collaboration field is reversed for US-INTL collaborations: US-INTL collaborations are lower for economics than for the other natural science subfields examined. As noted earlier, Jones, Wuchty & Uzzi (2008) identified similar patterns. For instance, for top Research/Doctoral institutions in economics, %US-INTL increased from 12% in 1991-1995 to 21% in 2001-2007, while the comparable figures for biology were 16% and 27%. Notably the field with the highest %US-INTL collaborations in 2001-2007 was physics, with 44% of the total. This is not surprising given the important role that large scale equipment plays in research in physics and the importance of the role of international collaboration in facilitating this research (Adams et al., 2005). Furthermore, the findings with respect to economics are consistent with the hypothesis that economics research is more nationally focused than research in the natural sciences

V. Diffusion of IT

As discussed earlier, one explanation for recent trends in collaboration is the diffusion of IT in higher education. During the period under study, 1991 to 2007, a major innovation in IT that facilitated the growth of the INTERNET[17] was the introduction of the Domain Name System (DNS) (Griffiths, 1984). This system, which developed in 1984 and was fully diffused

[16] Laband & Tollison (2000) find that in 1950, over 30% of top articles in biology were co-authored, as compared with 5 % of top articles in economics. By 1994, co-authorship increased in both fields, but relatively more in economics; 80% of top articles were co-authored in biology as compared with 70% in economics.

[17] For a highly readable historical account, see Greenstein (forthcoming).

by the mid 2000s, classifies addresses according to whether the host computer connecting to the network was an educational (edu), commercial (com), governmental (gov), or international (org) institution; it also provided for a series of country codes. Prior to the invention of DNS, every host (computer workstation or server) on the Internet needed to know the exact name and IP-address of every other system on the network.

Given the importance of DNS in facilitating widespread use of the INTERNET, this study uses an institutions' date of adoption of the domain name system (DNS) as a proxy for the institution's early IT environment. Specifically, the date used indicates when universities formally registered their domain names on the INTERNET. Information on the adoption of domain names was obtained from the *ALLWHOIS* registry site available on-line. In cases where the university had more than one server registered, we examined the dates of all named servers and recorded the earliest date. Because branch campuses may have relied on a system-wide server before obtaining their own domain names, we collected both the earliest date of the domain name registered for the system, along with the earliest date that the branch campus registered its own domain name and used the earliest of the two.

Figure 4 depicts adoption of DNS by the study's full set of 1,281 institutions, as well as adoption stratified by institutional "tier," as defined earlier, for the years 1985-2007. As shown in Figure 4, DNS technology was first adopted by institutions in 1985. By 1991, the first year that the publication data used in this study are available, 33% of the 1,281 institutions had adopted DNS; by 2001, the figure was just above 97%, and by 2007, the technology had fully diffused among all institutions.[18] Prior research on the diffusion of DNS (Levin et al., forthcoming) as well as Figure 4 shows that the data exhibit the usual S-curve associated with

[18]Although publication data for the 1980s are also available, they could not be reliably matched with specific institutions and therefore are excluded from the present study.

diffusion patterns (for example, Geroski, 2000; Stoneman, 2002; Rogers, 2003): adoption first rises at an increasing rate and then levels off.

Not surprisingly, as shown in Figure 4, Top Research/Doctoral institutions were more likely to be early adopters and adopted at a much faster rate; indeed by 1991, nearly 97% of these institutions had already adopted domain names. In contrast, by 1991, only 78% of Other Research/Doctoral institutions, 52% of Top Liberal Arts institutions, and just 27% of Master's institutions had adopted. In the regression analysis, as discussed shortly, we look at the impact of *length* of exposure to this technology on multi-institution collaboration. Even among Top Research/Doctoral institutions, exposure to DNS measured in terms of length, varies, though less so than among other tiers.

The impact of exposure on institutional productivity is also expected to be positively affected by the size of the number of other users, or what is typically referred to as a "network effect" or "network externality" (Page & Lopatka, 1999). In the case of higher education, we expect that as another user around the world adopts DNS, this directly benefits all the other users, including the higher education institution under study. Network effects are measured here using data on the number of net total domains (DNS) registered worldwide. Specifically, domain count information for 1989 through 1997 are taken from Zakon (2005) and data for 1998-2007 are from Zooknic Internet Intelligence. These data (not shown here) indicate that there was a dramatic acceleration in the growth of worldwide DNS through the 1990s and early 2000s, followed by a virtual standstill following the technology bust in 2001. By the mid-2000s, growth in worldwide DNS had resumed, though at a reduced rate compared with the earlier period.

B. Regression Analysis: An Investigation of the Effect of IT Exposure on Collaboration

Here we analyze the role that IT has played in influencing collaboration patterns across institutions. This analysis is exploratory in nature for a number of reasons. First, there is no single measure of IT that reflects what has happened to information technology at a given institution. Rather, we must rely on adoption of DNS as a proxy. Second, the fields are (still) highly aggregated. For example, we do not differentiate between theoretic physics and experimental physics papers, or empirically-based economics papers rather than theoretical papers. Yet we would expect collaboration to be greater for the experimental and empirically-based work. Thus our findings should be regarded as suggestive at best.

Using the sample of institutions that regularly publish (those with more than 4 publications per year in the respective field), we estimate a "modified" difference equation to investigate US-US and US-INTL collaboration as a function of exposure to IT, using DNS as a proxy. The dependent variable is specified as the change in the number of US-US collaborations ($USUS\_change_{i,t} = USUS_t - USUS_{t-1}$). The virtue of focusing on *change* in number of multi-institution co-authored articles rather than the absolute number of articles is that this specification effectively "differences out" institutional fixed effects such as research intensity, grantsmanship, faculty size, and student quality. To account for scale effects, the change in total publications ($Pub\_change_{i,t} = pub_t - pub_{t-1}$) is included as a control variable. Exposure is defined as the amount of time (measured annually) that an institution has had access to DNS. Given an expected lag between the time of adoption of DNS and its impact on collaboration and publication, the exposure variable is lagged by one year. In other words, if an institution adopted DNS in 1991, then exposure for 1991 is coded as 0, and then coded 1 for 1992, 2 for 1993, and so on.

Preliminary work investigated alternative specifications for exposure including a linear function (in lieu of a quadratic) and dummy variables that reflect differing levels of exposure. The results presented here model IT exposure as a quadratic function to reflect the expectation that length of exposure to DNS is expected to have a positive impact but at a declining rate as the technology diffused over the period ($B_2 > 0$ and $B_3 < 0$). The model also includes a measure of the size of the IT network to reflect "network effects;" exposure is expected to have a greater impact as the number of DNS sites worldwide increases ($B_4 > 0$). Consider an institution that has 2 years of exposure to DNS in 1992 and another that has 2 years of exposure to DNS in 1999. While length of exposure is the same, the IT environment in these two years is very different; those adopting later encounter a larger network of users with more sophisticated tools for online applications compared to earlier entrants. One might expect that the effect of a year of exposure on collaboration would be greater for these institutions, all else equal.

In equation form the model is written as follows:

$$\text{USUS\_change}_{i,t} \text{ or } \text{USINTL\_change}_{i,t} = B_o + B_1 \text{ Pub\_change}_{i,t} + B_2 \text{ EXP}_{i,t-1}$$

$$+ B_3 \text{ EXPSQ}_{i,t-1} + B_4 \text{ EXP}_{i,t-1} * \text{change\_ln(IT WORLD)}_{t-1} + \varepsilon_{i,t}$$

where Pub_change = change in total number of publications at institution i in year j
USUS_change = year-to-year change in number of publications by institution i with at least one co-author from another institution
USINTL_change = year-to-year change in number of publications by institution i with at least one co-author from an international institution
EXP = years of institutional exposure to DNS
EXPSQ = squared years of institutional exposure to DNS
Change_ln(ITWORLD) = ln(worldwide DNS in period t) – ln(worldwide DNS in period t-1)

Given the earlier findings regarding differences in collaboration patterns by field, models are estimated separately for the four subfields studied here: biology, chemistry, physics, and economics. In addition, as a point of comparison, results are provided for "All Fields" (natural and social sciences combined). It is important to keep in mind that this broad designation includes the four subfields analyzed separately as well as a large number of other fields in the natural and social sciences as classified by Glänzel and Schubert (2003).

Each model is also estimated separately by major tier (Top Research/Doctoral, Other Research/Doctoral, Master's Level, Top Liberal Arts). One rationale for doing so is that research expectations and norms differ by tier. A second rationale, albeit an empirical one, is that estimation by tier reduces the substantial variation in publication rates and collaboration patterns observed. Nonetheless, and as seen in the earlier tables for the underlying *level* variables, Table 3 shows that there is still considerable variation in USUS_change and USINTL_change by tier. For instance, for Top Research/Doctoral (all fields), USUS_change has a mean of 75 and a median of 51, with a min of -312 and a maximum of 1,267. The analysis is conducted using OLS. Robust standard errors are reported alongside the OLS estimates given the tremendous variation in the dependent variable(s) as well as evidence of heteroskedastic errors uncovered by additional testing.

Turning to the results, Tables 4 and 5 provide results for the dependent variables, USUS_change and USINTL_change, respectively. As would be expected, in all model specifications, the change in the number of papers with more than one institutional co-author is a statistically significant function of the change in total publications. Most relevant to this study, however, is the impact of length exposure to IT on the change in the number of papers with more than one institutional coauthor. Thus, the remainder of the discussion focuses on this relationship. First and consistent with expectations, length of exposure is found to have a statistically and significant positive but diminishing effect on collaborations in All Fields at Top Research/Doctoral and Other Research Doctoral institutions, as measured by the USUS_change variable for All Fields. Also consistent with expectations, the impact of exposure increases with the size of the network (significant at the 10 percent level or better) for these tiers as well as Master's institutions. Results are far more modest by field. Length of exposure is found to have a positive and significant effect on US-US collaborations for Top Research/Doctoral in Biology and Master's Level in Economics only. Thus, this exploration reveals little evidence of a differential effect of IT exposure by field.

Table 5 presents results regarding US-INTL collaborations. In terms of statistical significance, findings regarding the relationship between IT and collaboration are much weaker compared with those for US-US collaborations. Length of exposure (entered by itself is not found to be statistically significant for any field/tier, though interestingly, for the Top Doctoral/Research tier for All Fields and for Physics, length of exposure is found to have a positive and significant effect on such collaborations as the size of the network increases. A significant coefficient on the interaction term is also found for Other Research/Doctoral for Economics. These results suggest that the size of the network plays a particularly important role in collaborations that transcend national boundaries. On the other hand, Table 5 also presents some anomalous results, including a positive significant coefficient on exposure squared for Top Research/Doctoral for All Fields and a negative coefficient on the interaction of exposure and size of network for Master's Level Biology.

VI. Discussion and Conclusion

This paper examines multi-institution collaborations, both US-US and US-INTL, using data from ISI/Web of Science for the period of 1991-2007 for the social sciences (as represented by economics) and the natural sciences (biology, physics, chemistry). Among the findings, for

Top Research/Doctoral institutions in economics, the percent US-US co-authored papers increased from 57% in the early 1990s to 70% in the mid 2000s. Comparable figures for biology were 40% and 55%. Growth in the propensity to co-author with an institution outside the U.S. grew even faster than growth in the propensity to co-author with an institution in the U.S. For instance, for Top Research/Doctoral institutions in economics, the percent of US-international co-authored papers increased from 12% for the early period to 21% for the later period. In the natural sciences, US-INTL co-authorship was greater at the start of the period and, for these fields as well, there was a considerable increase. For Top Research/Doctoral institutions in biology, the US-international share increased from 16% to 27%. Notably, of the fields considered, US-international collaboration was the highest in physics at the start and end of the period considered: 27% in the early 1990s and 44% in the early to mid 2000s. These figures point to important field differences in how research is conducted.

The paper next analyzes the impact of IT, as measured by exposure to DNS, on multi-institution co-authorship. Using a modified "first-difference" approach, the regression results provide preliminary evidence of a role for IT. IT exposure is found to have a significant effect on collaboration for Top (and Other) Research/Doctoral institutions, an effect that increases with the size of the network, for both US-US and US-INTL collaborations. Statistically significant effects are more often found for top tiers, suggesting that it is the most research-active institutions that benefit from the adoption and diffusion of IT.

Previous research has shown that the availability of IT has a democratizing effect, giving a particular boost to the productivity and collaboration patters of research active scientists at lower-tier institutions. (Ding et al. 2010). The current research suggests that IT does not have a democratizing effect at the institutional level. The two findings are not at odds but rather suggest that although IT is beneficial to active researchers working at lower-tier institutions it does not transform research-inactive scientists—who dominate at lower-tier institutions-- into research active scientists. To put it metaphorically, IT does not raise all ships—only those that are already launched—and there are few launched ships at lower-tier institutions.

Bibliography

Adams, J., G. Black, R. Clemmons & P. Stephan. 2005. Scientific Teams and Institution Collaborations: Evidence from U.S. Universities, 1981-1999. *Research Policy* 34 (3): 259-285.

Agrawal, A. & A. Goldfarb. 2008. Restructuring Research: Communication Costs and the Democratization of University Innovation. *American Economic Review* 98(4): 1578-1590.

ALLWHOIS. (http://www.allwhois.com/).
Atlas of cyberspaces. (http://www.cybergeography.org/atlas/bitnet_topology.txt).

Barnett, A., R. W. Ault & D. Kaserman. 1988. The Rising incidence of Co-authorship in Economics: Further Evidence. *Review of Economics and Statistics* 70(3): 539-543.

Beaver, D. 2001. Reflections on Scientific Collaborations (and its Study): Past, Present and Prospective, *Scientometrics* 52(3): 365–377.

Bozeman, B. & E. Corley. 2004. Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy* 33(4):599-616.

Braun, T., Schubert, A., Glänzel, W. 2001. Publication and Cooperation Patterns of the Authors of Neuroscience Journals, *Scientometrics* 51(3): 499–510.

Butler, D.M., R.J. Butler & J.T. Rich. 2008. The Equalizing Effect of the Internet on Access to Research Expertise in Political Science and Economics. *PS-Political Science & Politics* 41(3): 579-584.

Cameron, A. & P.K. Trivedi. 2009. *Microeconometrics for Stata*. College Station, TX: Stata Press.

Carayol, N. & M. Matt. 2006. Individual and Collective Determinants of Academic Scientists' Productivity. *Information Economics and Policy* 18(1): 55-72.

Carnegie Foundation for the Advancement of Teaching. 1994. *A Classification of Institutions of Higher Education—1994 Edition.* Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Christovich, L. 2010. Gap Between Doctorate and Nondoctorate Granting Institutions Widens. National Science Foundation *Info Brief* 10-328. http://www.nsf.gov/statistics/infbrief/nsf10328/

Cohen, J. 1996. Computer Mediated Communication and Publication Productivity Among Faculty. *Internet Research* 6(2/3):41–63.

Cronin, B., D. Shaw & K. La Barre. 2004. Visible, Less Visible, and Invisible Work: Patterns of Collaboration in 20th Century Chemistry Context Sensitive Links. *Journal of the American Society for Information Science and Technology* 55 (2): 160–168.

Ding, W.W., S.G. Levin, P.E. Stephan & A.E. Winkler. 2010. The Impact of Information Technology on Academic Scientists' Productivity and Collaboration Patterns. *Management Science.*

Finn, M. G. 2007. Stay Rates of Foreign Doctorate Recipients from U.S. Universities, 2007. *Oak Ridge Institute for Science and Education.* November. http://orise.orau.gov/files/sep/stay-rates-foreign-doctorate-recipients-2007.pdf

Friedlander, A. & R.S. Bessette. 2003. *The Implications of Information Technology for Scientific Journal Publishing: A Literature Review.* Arlington, Virginia: National Science Foundation. http://www.nsf.gov/statistics/nsf03323/htmstart.htm

Geroski, P. A. 2000. Models of Technology Diffusion. *Research Policy* 29 (4-5):603-625.

Glänzel, W. 2002. Co-authorship Patterns and Trends in the Sciences (1980–1998). A Bibliometric Study with Implications for Database Indexing and Search Strategies, *Library Trends* 50: 461–473.

Glänzel, W. & A. Schubert. 2003. A New Classification Scheme of Science Fields and Subfields Designed for Scientometric Evaluation Purposes. *Scientometrics* 56(3): 357-367.

Greenstein, S. forthcoming. Nurturing the Accumulation of Innovations: Lessons from the Internet. R. Henderson and R. Newell eds. Accelerating Innovation in Energy: Insights from Multiple Sectors. University of Chicago Press.

Griffiths, R. T. 2002. From ARPANET to World Wide Web, Chapter Two in *History of the Internet, Internet for Historians (and just about everyone else).* (http://www.let.leidenuniv.nl/history/ivh/chap2.htm).

Hamermesh, D. & S. Oster. 2002. Tools or Toys? The Impact of High Technology on Scholarly Productivity. *Economic Inquiry* 40 (4):539-555.

He, Z. , X. Geng & C. Campbell-Hunt. 2009. Research Collaboration and Research Output: A Longitudinal study of 65 Biomedical Scientists in a New Zealand University. *Research Policy*, 38 (2), 306-317.

Heffner, A. G. 1981. Funded Research, Multiple Authorship, and Subauthorship Collaboration in Four Disciplines. *Scientometrics* 3 (1): 5–12.

IPEDS (Integrated Postsecondary Education Data System). National Center for Education Statistics. (http://nces.ed.gov/).

Jones, B. 2009. The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *Review of Economic Studies* 76 (1): 283-317.

Jones, B.F., S. Wuchty & B. Uzzi. 2008. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science* 332 (21 November 2008): 1259-1039.

Katz, J.S. & B.R. Martin. 1997.What is Research Collaboration? Research Policy 26(1):1-18.
Kim E. H., A. Morse & L. Zingales. 2009. Are Elite Universities Losing Their Competitive Edge? *Journal of Financial Economics* 93(3): 353-381.

Laband, D.N. & R.D. Tollison. 2000. Intellectual Collaboration. *Journal of Political Economy* 108 (3): 632-662.

Levin, S. G., P. E. Stephan & A. E. Winkler. (forthcoming). Innovation in Academe: The Diffusion of Internet Technologies. *Applied Economics*.

Rosenblat, T.S. & M.M. Mobius. 2004. Getting Closer or Drifting Apart? *Quarterly Journal of Economics* 119(3): 971-1009.

National Science Board. 2006. *Science and Engineering Indicators*. Arlington, VA: National Science Foundation.

National Science Board. 2010. *Science and Engineering Indicators*. Arlington, VA: National Science Foundation.

Overbye, D. 2007. A Giant Takes on Physics' Biggest Questions. *New York Times* May 15, D1, D4-D5.

Page, W.H. & J. E. Lopatka. 1999. Network Externalities. In B. Bouckaert and G. De Geest , eds. *Encyclopedia of Law and Economics*. Edward Elgar Publishers. http://users.ugent.be/~gdegeest/

Patel, N. 1973. Collaboration in Professional Growth of American Sociology. *Social Science Information* 12(6): 77–92.

Price, D., J. deSolla &D. Beaver. 1966. Collaboration in an invisible college. *American Psychologist* 21 (11): 1011-1018.

Rogers, E. M. 2003. *Diffusion of Innovations*, 5th ed. New York: Free Press.

Schonfeld, R. C. 2003. *JSTOR: A History*. Princeton, NJ: Princeton University Press.

Sonnenwald, D. H. 2007. Scientific Collaboration. *Annual Review of Information Science and Technology* 41: 643–681.

Stephan, P. E. 2010. The Economics of Science. In B. H. Hall & N. Rosenberg, editors, *Handbook of Economics and Innovation, Vol. 1,* pp. 217-274, Elsevier.

Stoneman, P. A. 2002. *Economics of Technological Diffusion*. Oxford, UK and Malden, MA: Blackwell Publishing.

Sutter, M. & M. Kocher. 2004. Patterns of Co-Authorship Among Economics Departments in the USA. *Applied Economics* 36: 327-333.

Wagner, C.S. & L. Leydesdorff. 2005. Network Structure, Self-Organization, and the Growth of International Collaboration in Science. *Research Policy* 34(10): 1608-1618.

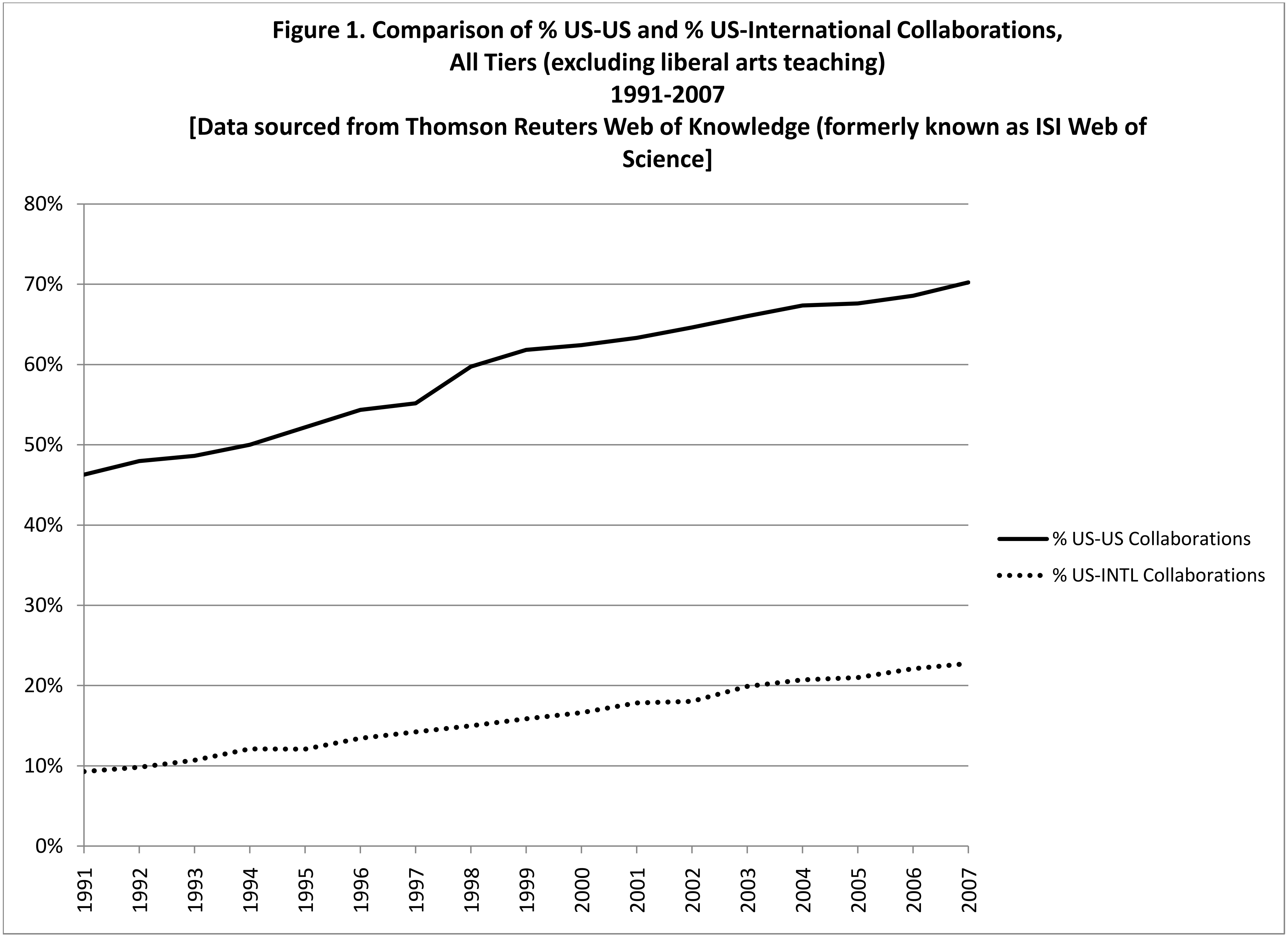
Walsh, J.P., S. Kucker, N.G. Maloney & S. Gabbay. 2000. Connecting Minds: Computer-Mediated Communications and Scientific Work. *Journal of the American Society of for Information Science* 51(4): 1295-1305.

Walsh, J.P. & T. Bayma. 1996. Computer Networks and Scientific Work. *Social Studies of Science* 26(3): 661-703.

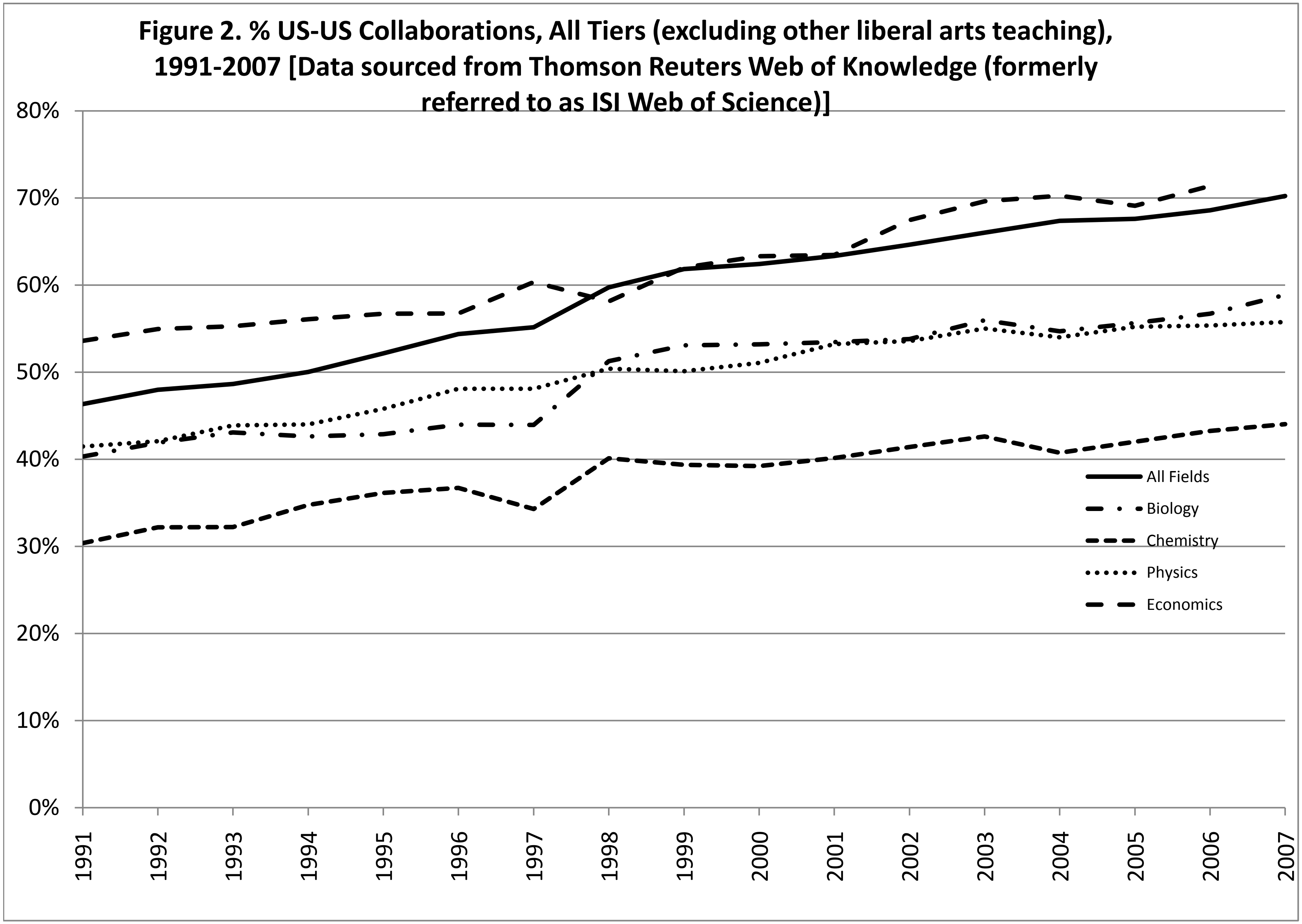
Wuchty, S., B. Jones & B. Uzzi. 2007. The Increasing Dominance of Teams in the Production of Knowledge. *Science* 316 (5827):1036-1039.

Zooknic Internet Intelligence. http://www.zooknic.com/Domains/counts.html\ [Accessed 21 January 2011].

Zakon, R. 2005. *Hobbes' Internet Timeline v8.1.*(http://www.zakon.org/robert/internet/timeline)



Restricted to institutions with > 4 publications per year.



Restricted to institutions with > 4 publications for each field for each year.

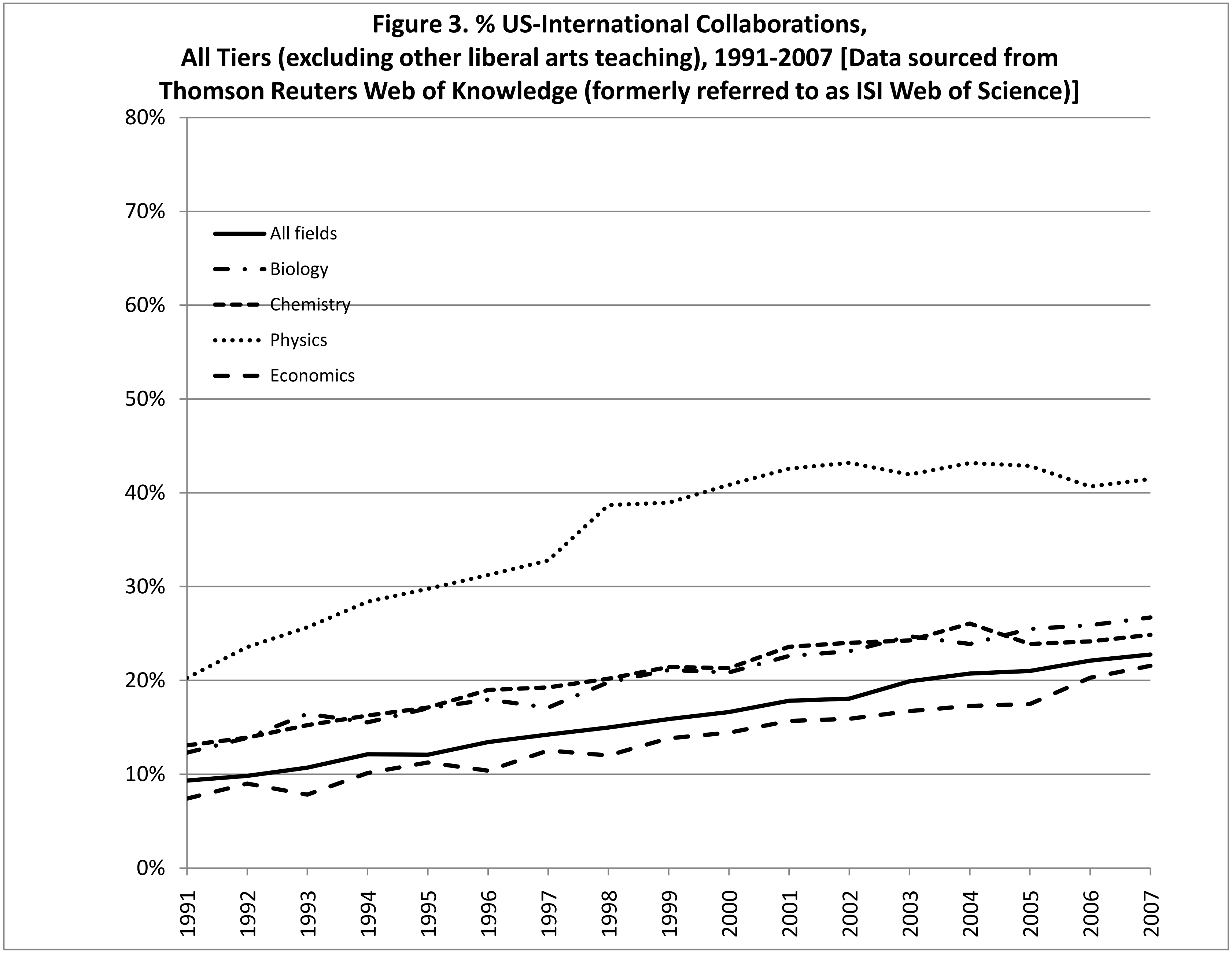


Figure 3. % US-International Collaborations,
All Tiers (excluding other liberal arts teaching), 1991-2007 [Data sourced from
Thomson Reuters Web of Knowledge (formerly referred to as ISI Web of Science)]
80%
70%
60%
50%
40%
30%
20%
10%
0%
All fields
Biology
Chemistry
Physics
Economics
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007

Figure 4. Adoption of DNS by U.S. Institutions by Year, Tier

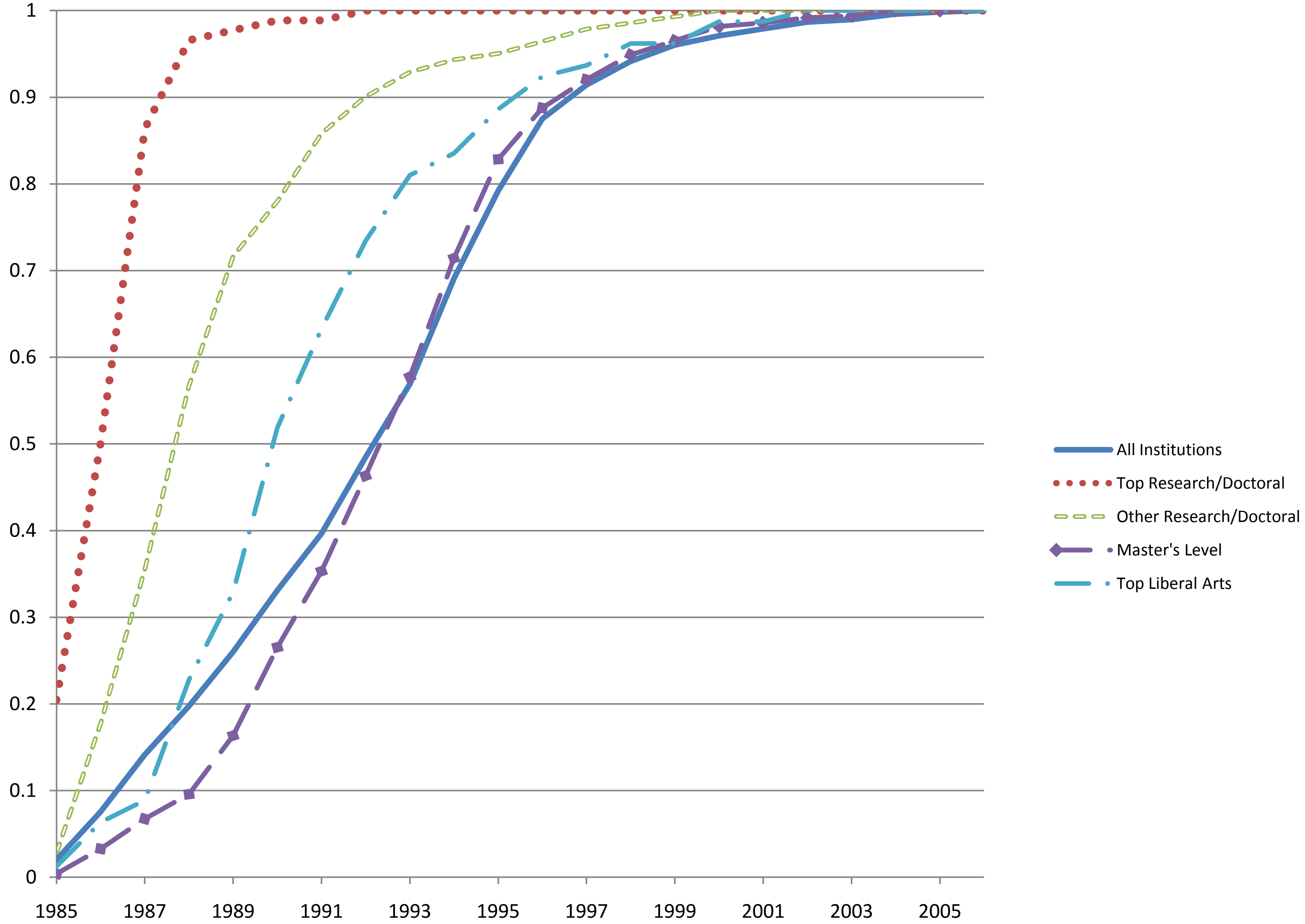


Table 1. Summary Statistics on Institutional Publication Data, by Tier and Field [Data sourced from Thomson Reuters Web of Knowledge (formerly referred to ISI Web of Science)]

	All Tiers			Top Research/Doctoral			Other Research/Doctoral			Master's			Top Liberal Arts		
Field	Mean Pubs	Median Pubs	% Zero Pubs	Mean Pubs	Median Pubs	% Zero Pubs	Mean Pubs	Median Pubs	% Zero Pubs	Mean Pubs	Median Pubs	% Zero Pubs	Mean Pubs	Median Pubs	% Zero Pubs
All Fields															
1991-1995	158.6	5.0	24.3%	1729.6	1412.5	0.0%	265.0	198.0	0.9%	21.9	7.0	16.7%	21.6	17.0	2.5%
1996-2000	186.5	7.0	20.1%	2029.0	1706.0	0.0%	311.5	238.0	0.4%	26.6	9.0	12.1%	25.9	21.0	2.8%
2001-2007	227.7	8.0	18.5%	2459.0	2041.5	0.0%	388.1	303.0	0.1%	33.1	11.0	10.9%	32.8	26.0	2.7%
Biology															
1991-1995	17.3	0.0	64.2%	204.2	163.0	0.0%	23.1	10.0	9.9%	1.6	0.0	68.4%	1.6	1.0	44.8%
1996-2000	22.5	0.0	58.3%	263.3	218.5	0.0%	30.3	14.0	7.5%	2.2	0.0	60.8%	2.1	1.0	34.2%
2001-2007	26.5	0.0	54.1%	307.8	245.0	0.0%	37.1	20.0	5.1%	2.7	0.0	53.8%	2.5	2.0	27.3%
Chemistry															
1991-1995	16.3	0.0	56.8%	166.4	141.5	0.0%	34.9	25.0	7.4%	2.0	0.0	56.4%	2.2	1.0	33.2%
1996-2000	19.3	0.0	53.5%	196.2	172.0	0.0%	41.2	30.0	5.7%	2.5	0.0	52.9%	2.6	2.0	26.6%
2001-2007	22.9	0.0	51.2%	229.1	197.0	0.0%	50.3	36.0	5.6%	3.4	1.0	49.7%	3.3	2.0	20.4%
Physics															
1991-1995	19.4	0.0	61.2%	215.9	181.0	0.0%	32.3	21.0	9.4%	2.0	0.0	64.3%	2.2	1.0	35.4%
1996-2000	22.1	0.0	58.3%	245.9	209.0	0.0%	36.8	25.0	7.5%	2.3	0.0	59.7%	2.5	2.0	32.9%
2001-2007	27.3	0.0	55.1%	292.4	253.5	0.0%	49.9	30.5	6.6%	3.5	0.0	54.7%	3.6	2.0	23.9%
Economics															
1991-1995	5.2	0.0	62.2%	49.8	42.0	1.1%	10.9	8.0	10.9%	1.2	0.0	62.0%	1.5	1.0	40.3%
1996-2000	5.6	0.0	58.8%	51.7	42.0	0.9%	12.1	10.0	8.7%	1.4	0.0	56.1%	1.6	1.0	38.0%
2001-2007	6.4	0.0	58.6%	59.1	46.0	1.3%	14.4	11.0	8.9%	1.6	0.0	55.5%	1.9	1.0	36.7%

Notes: All Fields include all fields in the Natural and Social Sciences (See Glanzel and Schubert, 2003). Arts and Humanities are excluded.
Total number of institutions (all tiers) = 1,281; n = 88 for Top Research/Doctoral; n =141 for Other Research/Doctoral; .
n=490 for Master's; n= 79 for Top Liberal Arts. Other liberal arts institutions (n=483) included in All Tiers but not shown separately.

Table 2: Multi-Institution Collaborations, Measured in % [Data sourced from Thomson Reuters Web of Knowledge (formerly referred to as ISI Web of Science)]

Panel A. % U.S. - U.S. Collaborations (calculated as USUS/Pubs)

	Top Research/Doctoral		Other Research/Doctoral		Master's Level		Top Liberal Arts	
	n of inst	%	n of inst	%	n of inst	%	n of inst	%
Biology								
1991-1995	437	40.1	485	40.7	138	51.8	24	54.1
1996-2000	440	48.3	553	46.9	203	55.2	56	58.5
2001-2007	616	54.7	814	53.9	395	59.9	97	59.3
Chemistry								
1991-1995	437	30.0	559	30.5	308	41.0	56	41.0
1996-2000	440	34.7	575	35.7	392	44.3	76	40.8
2001-2007	616	39.0	826	38.5	629	48.8	136	46.7
Physics								
1991-1995	434	41.0	530	41.2	265	50.1	67	52.1
1996-2000	440	47.7	554	46.3	321	55.5	74	58.7
2001-2007	616	52.0	805	50.2	620	60.7	132	65.7
Economics								
1991-1995	417	57.4	466	53.7	196	56.4	33	46.7
1996-2000	418	62.1	503	59.2	250	59.2	31	55.3
2001-2007	590	70.0	704	69.0	367	70.0	64	56.2

Panel B. % U.S. - International Collaborations (calculated as USINTL/Pubs)

	Top Research/Doctoral		Other Research/Doctoral		Master's Level		Top Liberal Arts	
	n of inst	%	n of inst	%	n of inst	%	n of inst	%
Biology								
1991-1995	437	16.3	483	13.7	136	16.0	24	17.2
1996-2000	440	21.4	552	18.0	203	18.7	55	20.2
2001-2007	616	26.7	814	24.6	394	23.1	96	19.2
Chemistry								
1991-1995	437	15.7	557	14.6	294	15.0	55	17.2
1996-2000	440	20.9	575	19.8	379	21.6	75	12.3
2001-2007	616	26.1	826	24.0	622	25.0	136	16.5
Physics								
1991-1995	434	27.0	530	24.4	260	25.6	64	26.7
1996-2000	440	38.3	554	34.5	320	38.1	73	33.9
2001-2007	616	44.4	805	39.8	620	44.7	132	35.3
Economics								
1991-1995	416	12.1	465	7.5	183	6.8	32	8.3
1996-2000	418	16.8	503	10.7	237	10.1	30	7.5
2001-2007	590	21.2	704	16.7	365	15.4	63	14.1

Notes: Restricted to >4 publications in each year in the given field.
For 1991, several observations were deleted due to missing data; thus the count for Top Research is 440 (all 88 institutions had > 4 pubs for each year) for 1996-2000, but 437 for 1991-1995.

Table 3. Summary Statistics for Variables Used in Regressions, 1992-2007 [Data sourced from Thomson Reuters Web of Knowledge (formerly referred to as Web of Science)]

	Top Research/Doctoral				Other Research/Dcotoral				Master's Level				Top Liberal Arts			
	Mean	Min	Median	Max	Mean	Min	Median	Max	Mean	Min	Median	Max	Mean	Min	Median	Max
All Fields	(n=1408)				(n=2207)				(n=5115)				(n=1150)			
Pub_change	70.0	-404.0	46.0	1226.0	11.7	-750.0	7.0	358.0	2.1	-94.0	1.0	193.0	1.1	-37.0	1.0	41.0
USUS_change	75.1	-312.0	51.0	1267.0	11.9	-447.0	6.0	716.0	1.8	-89.0	1.0	163.0	1.1	-27.0	1.0	39.0
USINTL_change	37.0	-150.0	27.0	419.0	5.7	-94.0	3.0	121.0	0.7	-32.0	0.0	76.0	0.4	-14.0	0.0	25.0
Exposure	13.0	0.0	13.0	22.0	10.8	0.0	11.0	22.0	8.0	0.0	8.0	22.0	9.0	0.0	9.0	22.0
Exposure Squared	191.4	0.0	169.0	484.0	144.2	0.0	121.0	484.0	92.2	0.0	64.0	484.0	108.7	0.0	81.0	484.0
Change_ln(ITWORLD)	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4
Exp*Change_ln(ITWORLD)	6.3	0.0	5.3	20.5	5.1	0.0	4.4	20.5	3.6	0.0	3.0	20.5	4.1	0.0	3.6	20.5
Biology	(n=1408)				(n=1757)				(n=713)				(n=173)			
Pub_change	10.3	-142.0	8.0	203.0	1.8	-189.0	1.0	57.0	1.7	-27.0	2.0	52.0	1.8	-8.0	2.0	15.0
USUS_change	75.1	-312.0	51.0	1267.0	14.3	-447.0	9.0	716.0	6.4	-89.0	5.0	163.0	2.6	-27.0	2.0	39.0
USINTL_change	37.0	-150.0	27.0	419.0	6.8	-94.0	4.0	121.0	2.9	-32.0	2.0	76.0	1.1	-11.0	1.0	22.0
Exposure	13.0	0.0	13.0	22.0	11.4	0.0	11.0	22.0	10.5	0.0	11.0	21.0	12.1	0.0	13.0	22.0
Exposure Squared	191.4	0.0	169.0	484.0	155.4	0.0	121.0	484.0	141.2	0.0	121.0	441.0	169.5	0.0	169.0	484.0
Change_ln(ITWORLD)	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4
Exp*Change_ln(ITWORLD)	6.3	0.0	5.3	20.5	5.3	0.0	4.6	20.5	4.4	0.0	3.9	19.1	5.3	0.0	4.5	20.5
Chemistry	(n=1408)				(n=1854)				(n=1282)				(n=257)			
Pub_change	5.8	-86.0	4.0	139.0	1.8	-58.0	1.0	64.0	1.6	-37.0	2.0	33.0	1.6	-11.0	2.0	18.0
USUS_change	75.1	-312.0	51.0	1267.0	13.9	-447.0	9.0	716.0	4.7	-89.0	3.0	163.0	2.8	-21.0	3.0	39.0
USINTL_change	37.0	-150.0	27.0	419.0	6.6	-94.0	4.0	121.0	2.0	-32.0	1.0	76.0	1.0	-11.0	1.0	25.0
Exposure	13.0	0.0	13.0	22.0	11.2	0.0	11.0	22.0	9.6	0.0	10.0	21.0	11.0	0.0	11.0	21.0
Exposure Squared	191.4	0.0	169.0	484.0	151.8	0.0	121.0	484.0	121.0	0.0	100.0	441.0	147.8	0.0	121.0	441.0
Change_ln(ITWORLD)	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4
Exp*Change_ln(ITWORLD)	6.3	0.0	5.3	20.5	5.3	0.0	4.5	20.5	4.2	0.0	3.6	20.5	4.8	0.0	4.3	19.1
Physics	(n=1407)				(n=1789)				(n=1163)				(n=261)			
Pub_change	7.4	-130.0	4.0	279.0	2.2	-79.0	1.0	90.0	1.6	-31.0	2.0	46.0	1.8	-19.0	2.0	21.0
USUS_change	75.1	-312.0	51.0	1267.0	14.3	-447.0	9.0	716.0	4.7	-89.0	4.0	163.0	2.8	-22.0	3.0	39.0
USINTL_change	37.1	-150.0	27.0	419.0	6.8	-94.0	4.0	121.0	2.3	-32.0	2.0	76.0	1.2	-12.0	1.0	25.0
Exposure	13.0	0.0	13.0	22.0	11.3	0.0	11.0	22.0	10.4	0.0	11.0	22.0	10.9	0.0	11.0	22.0
Exposure Squared	191.5	0.0	169.0	484.0	154.5	0.0	121.0	484.0	134.6	0.0	121.0	484.0	146.6	0.0	121.0	484.0
Change_ln(ITWORLD)	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4
Exp*Change_ln(ITWORLD)	6.3	0.0	5.3	20.5	5.4	0.0	4.6	20.5	4.4	0.0	3.9	20.5	4.7	0.0	4.2	19.1
Economics	(n=1341)				(n=1587)				(n=781)				(n=117)			
Pub_change	1.2	-61.0	1.0	51.0	0.9	-23.0	1.0	37.0	1.6	-20.0	2.0	18.0	2.3	-8.0	2.0	13.0
USUS_change	76.5	-312.0	52.0	1267.0	15.0	-447.0	9.0	716.0	5.2	-89.0	4.0	163.0	2.1	-22.0	2.0	27.0
USINTL_change	37.8	-150.0	29.0	419.0	7.0	-94.0	4.0	121.0	2.1	-32.0	1.0	76.0	0.8	-9.0	1.0	19.0
Exposure	13.0	0.0	13.0	22.0	11.4	0.0	11.0	22.0	9.9	0.0	10.0	22.0	12.5	1.0	13.0	22.0
Exposure Squared	192.7	0.0	169.0	484.0	155.4	0.0	121.0	484.0	124.6	0.0	100.0	484.0	179.0	1.0	169.0	484.0
Change_ln(ITWORLD)	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4	0.5	0.0	0.3	1.4
Exp*Change_ln(ITWORLD)	6.3	0.0	5.3	20.5	5.4	0.0	4.7	20.5	4.5	0.0	3.9	20.5	5.4	0.0	4.5	20.5

Notes: Restricted to institutions with > 4 publications for each field for each year. Years 1992-2007.

Table 4. OLS Regression Results, Dependent Variable is USUS_change [Data sourced from Thomson Reuters Web of Knowledge (formerly referred to as ISI Web of Science)]

	Pub_change			Exposure			Exposure Squared			Exp*Change_ln(ITWORLD)			Constant				
	Coeff.		(std err)	Coeff.		(std err)	Coeff.		(std err)	Coeff.		(std err)	Coeff.		(std err)	R-sq	n
All Fields																	
Top Research/Doctoral	0.700	***	(0.01)	7.0	***	(1.91)	-0.235	***	(0.08)	1.197	***	(0.27)	-27.368	***	(9.63)	0.72	1408
Other Research/Doctoral	0.672	***	(0.02)	0.5	*	(0.29)	-0.019		(0.01)	0.284	***	(0.09)	-0.476		(1.12)	0.66	2207
Master's Level	0.655	***	(0.02)	0.0		(0.05)	0.001		(0.00)	0.046	*	(0.02)	0.115		(0.16)	0.68	5115
Top Liberal Arts	0.618	***	(0.02)	0.0		(0.08)	0.000		(0.00)	-0.035		(0.04)	0.211		(0.24)	0.63	1150
Biology																	
Top Research/Doctoral	0.590	***	(0.05)	1.5	***	(0.55)	-0.054	***	(0.02)	-0.023		(0.09)	-7.229	**	(3.25)	0.51	1408
Other Research/Doctoral	0.620	***	(0.05)	0.0		(0.20)	-0.002		(0.01)	0.008		(0.03)	0.247		(1.00)	0.48	1756
Master's Level	0.601	***	(0.04)	-0.1		(0.09)	0.003		(0.00)	0.003		(0.03)	0.303		(0.39)	0.59	713
Top Liberal Arts	0.584	***	(0.07)	0.0		(0.13)	-0.002		(0.01)	-0.058		(0.04)	0.441		(0.68)	0.44	173
Chemistry																	
Top Research/Doctoral	0.350	***	(0.02)	0.1		(0.40)	-0.001		(0.02)	-0.023		(0.07)	1.192		(2.23)	0.38	1408
Other Research/Doctoral	0.318	***	(0.01)	0.2		(0.10)	-0.007		(0.00)	-0.017		(0.03)	-0.257		(0.39)	0.36	1853
Master's Level	0.432	***	(0.03)	0.0		(0.06)	0.000		(0.00)	-0.021		(0.02)	0.189		(0.19)	0.41	1282
Top Liberal Arts	0.399	***	(0.06)	0.1		(0.11)	-0.003		(0.01)	-0.043		(0.04)	0.250		(0.47)	0.33	257
Physics																	
Top Research/Doctoral	0.530	***	(0.02)	-0.3		(0.53)	0.003		(0.02)	0.095		(0.08)	5.076		(3.17)	0.60	1407
Other Research/Doctoral	0.504	***	(0.02)	0.0		(0.12)	-0.002		(0.01)	0.014		(0.03)	0.204		(0.52)	0.57	1788
Master's Level	0.526	***	(0.03)	0.0		(0.07)	0.001		(0.00)	0.009		(0.03)	0.355		(0.29)	0.53	1163
Top Liberal Arts	0.720	***	(0.04)	0.1		(0.11)	-0.006		(0.00)	-0.030		(0.03)	-0.396		(0.51)	0.72	261
Economics																	
Top Research/Doctoral	0.612	***	(0.02)	0.0		(0.18)	0.001		(0.01)	0.016		(0.03)	0.116		(0.97)	0.65	1341
Other Research/Doctoral	0.660	***	(0.02)	0.0		(0.07)	0.001		(0.00)	0.017		(0.02)	0.228		(0.34)	0.65	1586
Master's Level	0.652	***	(0.02)	0.1	*	(0.05)	-0.003		(0.00)	-0.022		(0.02)	-0.257		(0.21)	0.67	781
Top Liberal Arts	0.676	***	(0.08)	0.0		(0.15)	0.004		(0.01)	-0.011		(0.04)	-0.307		(0.89)	0.54	117

Notes: Restricted to institutions with > 4 publications in a given field in a given year. Sample is for years 1992-2007.
Robust standard errors in parentheses.
***p<.01, **p<.05, *p<.1

Table 5. OLS Regression Results, Dependent Variable is USINTL_change [Data sourced from Thomson Reuters Web of Knowledge (formerly referred to as ISI Web of Science)]

	Pub_change			Exposure		Exposure Squared			Exp*Change_ln(ITWORLD)			Constant				
	Coeff.		(std err)	Coeff.	(std err)	Coeff.		(std err)	Coeff.		(std err)	Coeff.		(std err)	R-sq	n
All Fields																
Top Research/Doctoral	0.261	***	(0.01)	-1.5	(1.45)	0.102	*	(0.06)	0.755	***	(0.22)	14.217	**	(6.93)	0.54	1408
Other Research/Doctoral	0.236	***	(0.02)	0.1	(0.21)	0.002		(0.01)	0.053		(0.07)	1.067		(0.78)	0.42	2207
Master's Level	0.202	***	(0.01)	0.0	(0.04)	0.002		(0.00)	0.023		(0.02)	-0.033		(0.10)	0.29	5115
Top Liberal Arts	0.167	***	(0.02)	0.0	(0.07)	0.004		(0.00)	-0.003		(0.03)	0.227		(0.19)	0.17	1150
Biology																
Top Research/Doctoral	0.272	***	(0.01)	0.0	(0.37)	0.007		(0.01)	-0.011		(0.06)	0.190		(1.78)	0.36	1408
Other Research/Doctoral	0.232	***	(0.02)	0.0	(0.08)	0.002		(0.00)	0.000		(0.02)	0.122		(0.29)	0.30	1756
Master's Level	0.205	***	(0.03)	0.0	(0.08)	-0.002		(0.00)	-0.043	*	(0.03)	0.038		(0.32)	0.23	713
Top Liberal Arts	0.184	***	(0.04)	-0.1	(0.09)	0.006		(0.00)	0.010		(0.03)	0.468		(0.52)	0.11	173
Chemistry																
Top Research/Doctoral	0.231	***	(0.01)	-0.1	(0.37)	0.007		(0.02)	0.098		(0.06)	1.595		(1.81)	0.31	1408
Other Research/Doctoral	0.252	***	(0.01)	0.1	(0.08)	-0.002		(0.00)	-0.020		(0.03)	0.165		(0.34)	0.34	1853
Master's Level	0.241	***	(0.02)	0.0	(0.04)	-0.001		(0.00)	0.003		(0.02)	-0.120		(0.15)	0.27	1282
Top Liberal Arts	0.199	***	(0.02)	0.0	(0.07)	0.003		(0.00)	-0.033		(0.02)	0.168		(0.33)	0.27	257
Physics																
Top Research/Doctoral	0.453	***	(0.01)	-0.1	(0.46)	-0.003		(0.02)	0.306	***	(0.09)	3.458		(2.33)	0.60	1407
Other Research/Doctoral	0.476	***	(0.02)	0.0	(0.12)	-0.001		(0.01)	0.047		(0.03)	0.872		(0.55)	0.53	1788
Master's Level	0.423	***	(0.02)	0.0	(0.06)	-0.001		(0.00)	0.038		(0.03)	-0.031		(0.24)	0.48	1163
Top Liberal Arts	0.604	***	(0.07)	0.1	(0.13)	-0.003		(0.01)	-0.010		(0.04)	-0.822		(0.63)	0.56	261
Economics																
Top Research/Doctoral	0.172	***	(0.01)	-0.1	(0.15)	0.005		(0.01)	0.016		(0.02)	0.889		(0.75)	0.20	1341
Other Research/Doctoral	0.150	***	(0.01)	-0.1	(0.05)	0.004	**	(0.00)	0.021	*	(0.01)	0.264		(0.19)	0.17	1586
Master's Level	0.103	***	(0.01)	0.0	(0.03)	0.001		(0.00)	0.004		(0.01)	-0.048		(0.12)	0.10	781
Top Liberal Arts	0.094	***	(0.04)	0.1	(0.10)	-0.003		(0.00)	-0.019		(0.02)	-0.365		(0.49)	0.07	117

Notes: Restricted to institutions with > 4 publications in a given field in a given year. Sample is for years 1992-2007.
Robust standard errors in parentheses.
***p<.01, **p<.05, *p<.1