20 November 2002

Patterns of Research Collaboration in U.S. Universities, 1981-1999

By

James D. Adams, University of Florida and NBER
Grant C. Black, Georgia State University
Roger Clemmons, University of Florida
Paula E. Stephan, Georgia State University

November 2002

Preliminary—Comments Welcome

JEL Codes: L3, O3

Keywords: Academic Science, R&D, Collaboration, Scientific Teams

Corresponding Author: James D. Adams, Department of Economics, University of Florida, 224 Matherly Hall, Gainesville, FL 32611-7140. Telephone: 1-352-392-0124, fax: 1-352-392-7860, e-mail: **James.D.Adams@notes.cba.ufl.edu**

*The Andrew W. Mellon Foundation provided generous support for this research. We thank Nancy Bayers and Henry Small of the Institute for Scientific Information (ISI) for suggestions and clarifications concerning the data, and Cherisa Yarkin of the UC-Berkeley Office of Technology Licensing for comments on an earlier draft.

Abstract

This paper explores recent time trends as well as cross-sectional patterns in the size of scientific teams, and in collaboration between scientific institutions. The data derive from 2.4 million scientific papers written in 110 leading U.S. universities over the period 1981-1999.

Our measure of team size is the number of authors on a scientific paper. By this measure the size of scientific teams increases by 50% over the 19-year period. Much of the increase takes place during 1991-1995, when the Internet was rapidly commercialized. Cross-sectional patterns indicate that sciences that are intensive in instruments or research assistants employ larger teams, as do top departments that receive large amounts of federal R&D and employ faculty who have received prestigious prizes and awards.

We also find evidence of rapid growth in institutional collaboration, especially international collaboration. Since these two factors determine the location of team members, we conclude that geographic dispersion of scientific teams has increased over time. Finally, collaboration in its different dimensions contributes generally positively to papers and citations received. Since collaboration implies an increase in the division of labor, these results are consistent with the notion that the division of labor increases scientific productivity, consistent with Smith's famous dictum of 1776.

I. Introduction

Over the past century teams of scientific specialists have largely replaced the independent scientist, just as corporate R&D laboratories have replaced the independent inventor. The trend towards larger teams is evident; our own data confirm its progress[1]. Advancing instrumentation and the sheer quantity of what there is to know have pushed it, while improving communications have pulled it, to its present state. A diminishing sphere of pencil-and-paper research may be only the last to withstand these changes, rather than an exception to them.

One would think that collaboration in science, especially in capital-intensive empirical fields, could contribute to effectiveness, just as specialization in general is said to increase productive efficiency. According to this view, growth in equipment and structures capital, knowledge, and enhanced communications has mandated specialization, so that going it alone in science does not make sense. The evidence on these points is suggestive, but somewhat mixed up with other considerations. Collaborative research is more highly cited (Presser, 1980), but more capable researchers also attract collaboration (Merton and Zuckerman, 1973), so that separating efficiency from talent is not easy. But the size of scientific teams has increased steadily with time, whereas mean ability has probably changed very little, and this suggests that the division of labor and its advantages account for the rising propensity to collaborate.

Besides its contributions to efficiency, scientific collaboration could be important as a channel of knowledge transfer between scientists. And since collaborators are increasingly found in different institutions and countries, the entire subject is relevant to the tendency for knowledge to flow more readily and over greater distances than ever before.

In this paper we assemble findings on the size of scientific teams, inter-institutional collaborations, and their influence on scientific productivity, making use of a large database that covers most of academic science in the U.S. in the years 1981-1999. We confirm that team size and specialization have increased,

[1] See Merton and Zuckerman (1973), De Solla Price (1986), and Hicks and Katz (1996), among others, for trends in the size of scientific teams since 1900. Weiner (1994), writing in the mid-20th century, strongly disapproved of these trends, including the notion that corporate R&D laboratories might one day supplant individual university researchers and independent inventors.

especially 1990. In addition we empirically address some of the factors that promote or deter teams and collaborations, where the analysis is conducted at the level of individual science fields in U.S. universities.

The rest of the paper consists of four sections. Section II reviews the literature on scientific teams and collaborations and some literature that is specifically from economics, on teams and specialization. Section III describes the database and presents descriptive findings on trends in team size and collaboration over numerous dimensions of science. Section IV provides regression evidence on the determinants of team size, collaboration, and research quantity and quality. Section V concludes.

II. The Literature on Collaborations and Teams

A. Academic Research Collaborations

The size of academic teams is one preoccupation of bibliometrics. De Solla Price (1963, revised ed. 1986) examines multiple authorships in chemistry from 1910-1960. His main finding is that papers with three or more authors gain share over time, while papers written by one or two authors steadily lose share, so that by this measure team size unambiguously increases in chemistry over the first half of the 20^{th} century. Clarke (1964) finds that team size increases in biomedicine from 1934-1946 but then flattens through 1963. However, this result is an anomaly. Merton and Zuckerman (1973) find that the percentage of multiple authored papers rises steadily from 1900-1959 in biology as well as in chemistry and physics. In addition Merton and Zuckerman compare a sample of Nobel Prize winners with a sample of other scientists of similar age. They find that collaboration is more common among top researchers than others, suggesting that the talent of principal investigators attracts coauthors and drives team formation.

Other studies examine inter-institutional collaborations. Katz (1994) focuses on the role of proximity in inter-institutional collaborations from the 1980s in the United Kingdom, Canada, and Australia. He counts collaborative pairings between institutions, calculates distances between pairs normalized by the maximum distance within a country, and computes their frequency distributions. In all three countries, collaboration declines sharply with distance. Hicks and Katz (1996) examine team size and inter-institutional collaboration in the United Kingdom from 1981-1991. Collaboration increases in every dimension: authors per paper increase from 2.63 to 3.34, the number of institutions rises from 1.19 to 1.28, and the share of international collaborations increases from 14 to 23 percent.

B. Industrial Research Collaborations

Other research has examined industrial research collaborations, including joint research with universities and federal laboratories. Arora and Gambardella (1990), Mowery (1992), Powell (1996), Adams, Chiang and Starkey (2001), Adams (2002), Adams, Chiang and Jensen (forthcoming) and Zucker, Darby, and Armstrong (2001) are examples from a large literature.

Arora and Gambardella (1990) explore collaborations by biotechnology firms. In this industry innovations depend on contributions by universities, small research-intensive New Biotechnology Firms (NBFs), and established chemical and pharmaceutical firms. Arora and Gambardella point to four types of external linkages: (1) agreements with other firms; (2) research agreements with universities; (3) investments in the capital stocks of the NBFs; and (4) acquisitions of the NBFs. The use of external R&D resources appears to be driven by the increasingly inter-disciplinary character of knowledge. The four linkages are found to occur jointly, suggesting that strong complementarities are at work.

Mowery (1992) studies International Collaborative Research Ventures (ICRVs) and their success and failure in commercializing new technologies. He uncovers six factors that are implicated in the rise of ICRVs: (1) Growth in research capabilities of foreign firms has made them more attractive partners. (2) Increases in the cost of R&D performed wholly within the firm have made collaboration more appealing. (3) The growing importance of technical standards makes rapid building of an installed base more important, including overseas markets. (4) The apparent shortening of product cycles reinforces the importance of rapid market penetration. (5) The phenomenon of "technological convergence," which consists of a explosion of technologies coming together and required for successful R&D, spurs the search for partners with complementary capabilities. (6) The growth of non-tariff trade barriers such as public procurement policies that favor domestic firms and require technology transfer to these firms, also make these firms more useful as partners. Of the six, (1), (2) and (5) suggest that an expansion of complementary research inputs lies behind the growth of collaborations[2].

[2] Compare these findings on growth in research collaboration and research mergers due to expanding varieties of technology, with the following analogous statement on the role of increasing equipment variety in growth of firm size. Marshall (1920), p. 233 remarks: "In spite of the aid which subsidiary industries can give to small manufacturers, where many in the same branch of trade are collected in one neighborhood, they are still placed under a great disadvantage by the growing variety and expensiveness of machinery. For in a large establishment there are often many expensive machines each made specially for

Powell (1996) examines inter-firm collaborations in biotechnology. As in Mowery (1992) and Arora and Gambardella (1990) he finds that the collaborations are driven by research breakthroughs that are too diverse for any one firm to efficiently commercialize the R&D. However, Powell points to a further benefit of collaborations, namely their value in opening up information networks to member firms.

Adams, Chiang, and Starkey (2001) examine industry-university collaborative research centers (IUCRCs). They find that membership in IUCRCs is correlated with faculty consulting, co-authorship with faculty, and hiring of graduate students. Moreover, IUCRC membership is associated with small increases in patenting by the laboratories. IUCRCs appear to have succeeded in encouraging faculty to work more closely with companies in ways that are evidently valued by the companies.

Adams (2002) studies the geographic dimensions of firm-to-university and firm-to-firm research collaborations using a sample of industrial R&D laboratories. He finds that partnerships with universities are more localized than partnerships with firms, with the exception of top research universities. The results suggest that academic knowledge spillovers have a strongly regional component.

Adams, Chiang, and Jensen (forthcoming) examine interactions between the sample of R&D laboratories noted above and federal laboratories. Their principal finding is that Cooperative Research Agreements or CRADAs dominate other interactions between private and federal laboratories, and appear to stimulate patenting and company funded R&D in industry. The results suggest that CRADAs, which in principle could capitalize on complementarities between public and private R&D, may be beneficial to firms and public sector laboratories.

C. Economic Analysis of Teams

The economic analysis of team production examines incentive mechanisms for team members and the design of teams. Holmstrom (1982) studies the problem of moral hazard when there is group production, where the problem is one of inducing team members to provide the appropriate inputs when their actions cannot be observed and free riding occurs. His incentive scheme taxes each team member and

---

one small use. Each of them requires space in a good light, and thus stands for something considerable in the rent and general expenses of the factory; and independently of interest and the expense of keeping it in repair, a heavy allowance must be made for depreciation in consequence of its being improved upon before long. A small manufacturer must therefore have many things done by hand or by imperfect machinery, though he knows how to have them done by special machinery, if only he could find constant employment for it."

pays the team less than its output, if the output falls short of the Pareto optimal amount. By the same token, if output is optimal then his scheme pays team members amounts that satisfy the constraints of adding up and workers' opportunity costs[3]. Supervision plays no role in this analysis since the control variable, output, is observable. This theory suggests that management may be needed to handle problems of team production, if self-enforced penalties in labor-managed teams are not credible. Kandel and Lazear (1992) disagree with this conclusion. They suggest that peer pressure can replace management, so that worker managed teams need not be less efficient than teams managed by executives. The peer pressure function rewards each team member for greater effort, to an extent countering the free rider effects brought about by sharing of output. This argument suggests that free rider problems in teams can be solved as the case may be, either by means of a separate managerial input (Holmstrom) or by a group of unmanaged workers (Kandel and Lazear). But incentives must be available from somewhere: from the notching of rewards by management, or from sanctions imposed by peers.

Rosen (1982) examines the optimization problem of a firm that includes managerial input in the firm's production function. Quality of management has an environmental effect that is a local public good within the firm. In addition management improves supervision, though this effect is limited by the manager's time. The theory implies that superior talent sort to top management positions, to profit from their own environment and supervision. Another implication is that more able persons manage more profitable firms and firms that are more than proportionately larger, so that the span of control increases with talent. By analogy with scientific teams, the Rosen model suggests that star researchers who are the equivalent of managers are able to build larger laboratories and research groups that are more profitable, so that faculty wages and grants increase more than proportionately with skill of star researchers.

---

[3] Where $s_i(x)$ is the payment to team member i based on output x, the rule is, set

$$s_i(x) = \begin{cases} b_i & if \;\; x \geq x(a^*) \\ 0 & if \;\; x < x(a^*) \end{cases}$$

Where a^* is the vector of optimal inputs, $\sum_i b_i = x(a^*)$ so that the budget balances at a*, and $b_i > v_i(a_i^*)$, $x(a^*) > \sum_i v_i(a_i^*)$, so that payments and output exceed opportunity cost $v_i(a_i^*)$ of each worker. The employer "notches" the budget line and pays the team less than its output when output is less than a target value. But team payments automatically sum to output and satisfy opportunity cost if the target output is met.

Boning, Ichniowski and Shaw (2001) explore joint adoption of group incentives (I) and problem solving teams (T) in steel mini-mills. Modeling team output as an increasing and interactive function of T and I, they study the pattern of adoption of the two practices. They find that problem-solving teams are adopted only in the presence of incentive pay, that incentives raise productivity, and that incremental adoption of problem solving teams raises productivity even further. Applied to scientific teams the insights of Boning et al. suggest that successful work on more challenging problems requires enhanced group incentives compared to more routine investigations.

Theories of team production focus on the inadequacy of incentives when output is shared and the presence of free rider problems. In the case of scientific collaborations peer pressure and repeat collaboration may deter free riding for teams of modest size. The analogue to group incentives is less obvious. One clue is suggested by the winner-take-all nature of science, which rewards priority in discovery. Delay and poor quality of the work punish all members of dysfunctional teams with few publications or prospects, whereas influential publications and higher earnings reward successful teams.

Becker and Murphy (1991) explore the implications of specialization, the division of labor, and team size for economic growth. In their model an infinite number of tasks are rigidly combined to produce a unit of output. Output on each task depends negatively on the breadth of tasks but positively on knowledge. One implication is that output increases with team size since breadth of tasks decreases. This element is responsible for increasing returns to team size. And yet Becker and Murphy show that this view is too narrow. For one thing, the setup ignores costs of coordination, which tend to grow with team size. Rising costs in turn are attributed to moral hazard problems, hold ups, and breakdowns in communication. Given the element of increasing cost, they show that there is an equilibrium team size, and that equilibrium team size increases with knowledge.

III. Data and Descriptive Findings

A. Database of Scientific Papers

Our primary data set consists of 2.4 million scientific papers published during 1981-1999 by 110 universities that account for the majority of U.S. academic research. The Institute for Scientific Information (ISI) in Philadelphia is the source of the data on scientific papers, including the selection of the top 110 universities. Scientific papers consist of a standard set of communications—articles, reviews, notes and proceedings—that are included in ISI 's Current Contents publication from 1981 and 1999[4].

The papers are defined by journal classifications as belonging to one of 88 academic fields. We assign the 88 detailed fields to one of 12 major sciences in the National Science Foundation (NSF) CASPAR database, for purposes of linking with that database[5]. The Appendix lists the 110 universities, the 88 detailed fields, and the 12 major sciences in the data.

We make use of this data both at the paper level and the level of university-fields. At the paper level we compile time trends and cross-sectional patterns based on ISI assignments to universities, fields, and years. The data record date of publication, scientific fields of journals in which papers appear, institutional affiliation of authors, their address, and the number of authors[6].

The ISI data link papers to author's addresses. Besides the top 110 universities, the addresses identify U.S. and foreign institutions consisting of other universities and colleges; governments and government research institutes; medical centers; corporations; and all other institutions (plus non-assigned)[7]. We used the addresses to construct fractions of scientific papers written in one or more of the top 110, and in a number of cells. Within the U.S. we define cells consisting of (a) U.S. Government, (b) Other U.S. Universities, (c) U.S. Corporations, (d) U.S. Medical Centers, and (e) All Other U.S. Institutions. Outside the U.S. we define cells consisting of (a) Foreign Governments, (b) Foreign Universities, and (c) All Other Foreign Institutions. The cell information allows us to assign fractions of papers to different classes of institutions and provides an indication of proportional contributions by class of institutions[8].

---

[4] The journal set consists of 5500 journals that were active in 1999 and 1600 inactive journals that were cited in currently active periodicals.

[5] The 12 fields are: agriculture, astronomy, biology, chemistry, computer science, earth sciences, economics and business, engineering, mathematics and statistics, medicine, physics, and psychology.

[6] There is no limit at this time on the number of authors in the ISI data. The maximum number is 210, while the mean number is 2.36. Number of authors underestimates number of team members when it excludes contributors, such as research assistants. It is an overestimate when it includes honorific authors. In short, the number of authors measures the size of scientific teams with error.

[7] About 5% of the addresses could not be assigned.

[8] The fractions are ½, ½ in the case of two institutions, 1/3, 1/3,1/3 for three institutions, and so on. The cumulative distribution of the number of top 110 institutions per paper is as follows, with number of institutions in parentheses: 79.6% (1 institution), 96.8% (2 institutions or less), 98.3% (3 institutions or less), 99.5% (4 institutions or less). These are extremely crude indicators of contributions because they do not include time and effort by team members, nor do they differentiate among types of effort. In short, the institutional address fractions do not measure labor input, even though we use them to attribute output to scientific institutions.

Besides the use of the data to compile paper level statistics, we construct a panel at the level of universities, fields and years in order to undertake regression analysis. We combine the ISI data with information on university R&D and characteristics of doctoral programs. The NSF CASPAR database of universities provides university R&D. The National Research Council 1993 Survey of Doctoral Programs (NRC, 1995) provides characteristics of graduate programs, including counts of Nobel prizes and other prestigious awards as well as rankings of PhD programs as of 1993. Finally, microdata from the NSF Survey of Earned Doctorates (SED) estimates the migration of PhD students to particular universities, sectors, and countries—for those who have definite plans.

One other constraint is that we consider only leading departments out of the top 110. All other schools form a remainder cell within each field. We include the top 25 universities in astronomy, the top 50 universities in agriculture, chemistry, computer science, economics and business, earth sciences, mathematics and statistics, physics, and psychology, and the top 75 universities in biology, medicine, and engineering. Our purpose in breaking out fewer individual schools in smaller fields, and more in larger fields, is to avoid large numbers of empty cells for universities in which fields are small or non-existent[9]. The result is a panel that approximates broadly defined university "departments." This panel includes an array of variables that are likely to drive teams, collaborations, and research output. Table 1 describes the major variables in the data set and the sources of these variables.

B. Descriptive Findings Covering All of Science

Figures 1-8 provide an overview of the data. The figures cover all 12 main sciences and are at the paper level. Figure 1 displays aspects of the time series of scientific papers. The upper line consists of whole paper counts. The lower line consists of the contribution of the top 110 to the papers: the fraction written within the top 110 times the whole counts on the upper line. The increase in whole counts is 36% over the period 1981-1991 but just 16% from 1991-1999. Clearly there is a substantial slowdown in the growth of research during the 1990s; this is especially evident starting with 1995. Second, output measured by the internal contribution of the top 110 grows more slowly than whole counts, particularly in the 1990s.

[9] The size of the remainder of the top 110 equals an average "department" along the individual top 25, 50, or 75 schools in a field. This finding reflects the positive skew of academic programs. For more on this issue see Adams and Griliches (1998).

The internal contribution peaks in 1995 and declines by 2.5% from the peak as of 1999. All this reflects the increasing contribution of foreign institutions. One could see this as beneficial: foreign institutions shoulder more of the load of producing the research and transfer more of their stock of knowledge to U.S universities. Very likely international collaboration is also more feasible because of growth of Internet services after the early 1990s and because of other factors serving to promote such research. But the data also say that just staying even in research after 1995 seems to have required an increase in the foreign contribution. This suggests a shortfall in U.S. universities' research output, a more ominous interpretation.

Figure 2 displays contributions of "paper equivalents" by the top 10 collaborating universities to other schools in the top 110. By this we mean the sum of fractional papers contributed to papers of other institutions. Harvard leads with 11,000 papers; the remaining nine contribute about 6,000 papers each. The list of top 10 collaborators partly indicates scale of research programs, and yet a number of highly rated, research universities such as Chicago, Yale, Princeton, Pennsylvania, and Columbia are missing from the list. To explain this omission, we suggest that research styles vary among universities. Differences in capital intensity of research programs are partly responsible, and these are to a degree implied by differences in field mix. Subtle variations in vintage and quality of faculty and in the ability to attract able students for less are another possible cause of the differences.

Figure 3 shows the top 10 collaborating countries. The "paper equivalents" of these 10 countries account for 40% of the total foreign contribution, and not surprisingly, are concentrated in a familiar group of top R&D countries. The country contributions tail off fast, from over 18,000 paper equivalents contributed by Canada to 5000 papers contributed by the Netherlands.

Figure 4 displays the time series of authors per paper. This increases by 50 percent from 1981-1999. The increase accelerates during the 1990s, even as the growth of papers decelerates during the same period (see Figure 1), and the two graphs are almost mirror images. This could mean that rising team sizes are a behavioral response to rising limitations on resources, or that resources coincidentally grow at a faster pace elsewhere.

Figure 5 examines the trend in the number of top 110 universities per paper. According to this measure institutional collaboration in the U.S. increases by 10 percent from 1981-1999. This is of course an understatement since it ignores other institutional collaborations. To show this Figure 6 graphs the fractions of papers written inside and outside the top 110. The internal proportion falls from 78 to 65 percent, indicating the more rapid increase in the external contribution.

Figures 7 and 8 proceed further by decomposing trends in the external proportion of top 110 papers. Figure 7 breaks down the external U.S. proportion. And yet there is no clear trend in the overall, external U.S. contribution to papers of the top 110. In contrast Figure 8 shows that the foreign contribution increases rapidly, especially the contribution of foreign universities[10]. On the whole the foreign proportion of top 110 papers rises from five percent in 1981 to 15 percent in 1999.

C. Descriptive Findings By 12 Major Sciences

Tables 2 to 4 and figures 9 and 10 below are again at the paper level. Table 2 shows the number of papers written in the 12 major sciences consisting of agriculture, astronomy, biology, chemistry, computer science, economics and business, mathematics and statistics, medicine, physics, psychology, engineering, and earth sciences. The table presents whole counts and equivalent numbers of papers written by the top 110. The size of fields differs markedly: biology and medicine alone account for 60 percent of all papers, while astronomy, computer science and economics and business together account for four percent of all papers.

Figures 9 and 10 examine cross-field differences in authors and top 110 universities per paper. At the field level Figure 9 reveals a 3-1 variation in authors per paper. By this measure the largest teams are in astronomy, biology, medicine, and physics. All four are thought to be equipment-intensive or intensive in research assistance. Large telescopes and particle accelerators suggest a division of labor between theorists and experimentalists as well as a sharing of scarce equipment time. In medicine, large clinical trials imply teams that consist of principal investigators, managers of the trials, and laboratory assistants. The two fields that are the least equipment intensive and the least intensive in research assistance are economics and business and mathematics and statistics. They exhibit the smallest team size of any field.

Figure 10 displays top 110 institutions per paper by field. It would appear that rather different factors drive variations in team size and institutional collaboration (see Figure 9). The variation across fields is almost 2-1, less than in Figure 9, and the rankings of fields differ between the two figures. Consider again

[10] The task of separating foreign universities from foreign research institutes is not easy because government institutes that are in part university laboratories are often housed and addressed separately.

economics and business, and mathematics and statistics. These have the smallest team size but are at least average in institutional collaboration. While equipment intensity and research assistance help to drive team size, institutional collaboration may depend on the size of academic departments and hence the pool of local collaborators relative to ideal team size. If "technological convergence" indicates the use of large teams but university constraints limit the size of departments, then this would contribute to institutional collaboration. Moreover, this pattern is not unrelated to the availability of grant money, since department size is likely to increase with external funding.

Table 3 considers the composition of U.S. collaborators outside of the top 110. Astronomy and Earth Sciences are most prone to work with government, Economics and Psychology are the most likely to collaborate with other U.S. colleges and universities, Computer Science and Engineering are most likely to partner with U.S. Corporations, and Biology and Medicine more often with hospitals and medical centers. These patterns are as expected given the mix of sciences in non-academic institutions.

Table 4 reports proportions of papers written with different foreign institutions. The table is like Table 3 except that we combine foreign corporations, medical centers, and unassigned foreign addresses into a residual owing to their weak representation in the data. Co-authorship with foreign governments is most likely in Physics and Astronomy, while collaboration with foreign universities is most common in Physics, Mathematics, and Astronomy. Perhaps a common mathematical language facilitates international collaboration, since these fields use mathematics more intensively than others.

We turn now to descriptive findings at the level of universities, fields, and years. Table 5 reports means and standard deviations of the principal variables used in the regressions. The mean of authors per paper is 2.4, while the mean of top 110 universities per paper is 1.4. The foreign proportion is eight percent. Citations increase from 3.6 to 8.4 as the length of citation "window" increases. The stock of federally funded R&D in the average university and field is about 46 million dollars, private universities account for slightly more than a third of the observations; top 10 universities contribute one sixth. The table also reports data on the number of Nobel prizes and on the number of other lifetime achievement awards[11]. These data are not available for agriculture and medicine. Nobel prizes occur once every 33 university-

[11] The number of lifetime achievement awards is the sum of the number of Fields Medals (in Mathematics), McArthur Awards, National Medals of Science, National Medals of Technology, and election to the National Academies of Science and Engineering.

fields (“departments”); other lifetime achievement awards occur once every seven departments. The average department retains 1.4 PhD students after graduation per year.

IV. Regression Findings

Table 6 reports regressions that explain our measure of team size, the logarithm of authors per paper. The data are pooled across the 12 main sciences for the period 1981-1999, or across 10 fields that exclude agriculture and medicine. Allowing for missing values, the full data set includes 12,127 observations at the university, field and year level. All equations include dummy variables for year and field in which 1981 and chemistry are the omitted categories. The dummies absorb trend and field effects, which are similar to those depicted in the preceding figures and highly significant. We omit them and provide a brief interpretation instead. Year effects increase monotonically, but especially during the early 1990s, when sharp jumps are exhibited that amount to half the total increase in team size during 1991-1995 alone[12]. This is the period of rapid expansion of the commercial Internet, suggesting that some of the expansion of scientific teams reflects “technological pull” from this major invention[13]. Figure 11 is a graph of the time effects in regression 6.1. Turning to field effects, which are measured relative to agriculture, team size is largest in physics, followed by medicine, chemistry and astronomy. Team size is smallest in economics and mathematics. These effects follow the same order as the descriptive statistics.

The table includes five regressions. Equation 6.1is a baseline regression that includes the logarithm of the stock of federally funded R&D in thousands of 1992 dollars, indicators of private control, top 10 ranking of the university, and top 10 ranking in a field[14]. Consistent with the idea that larger projects entail greater specialization, the stock of R&D increases team size. Larger teams prevail in private universities, perhaps reflecting internal funding, but teams are smaller in top 10 universities, holding the size of R&D

[12] Divide the graph into four (almost) sized equal parts covering 1981-1985, 1986-1990, 1991-1995, and 1996-1999. The third of these periods accounts for 25% of the time but 50% of the time effects in the regression.

[13] A recent NRC report suggests a time frame for the Internet’s expansion: “As the NSFNET expanded, opportunities for privatization grew. … By the early 1990s, the Internet was international in scope, and its operation had largely been transferred from the NSF to commercial providers. … In April 1995, all commercial restrictions on the Internet were lifted.” NRC (1999), p. 179.

[14] The stock of R&D is lagged one year and is the discounted sum of R&D in that field and university conducted over the previous eight years, where the discount rate is fifteen percent. Thus the 1981 stock is the discounted sum over 1973-1980, and so on. The incomplete stock of R&D is dictated by the initial date of collection of R&D by university and field, 1973.

and the nature of control constant. One might think that this reflects differentially talent of faculty in top 10 schools, except that this effect is not replicated in top 10 departments.

Equations 6.2 and 6.5 successively add PhD student variables, numbers of faculty in the 1993 NRC survey, and counts of Nobel prizes as well as other lifetime achievement awards, which are not available for agriculture or medicine. The PhD and faculty count variables are included to capture the local pool of co-authors. However, this effect is largely subsumed in the R&D and school rank variables, and the pool variables are generally insignificant or actually negative. The prize variables typically increase team size, but not in the predicted order, since the coefficient on the Nobel prize effect is insignificant and less than that of other, less prestigious awards.

The dependent variable of table 7 is the logarithm of the number of top 110 universities per paper and is one measure of collaboration between institutions. The plan of table 7 is the same as table 6 and includes the same explanatory variables. As in Table 6 we omit a detailed reporting of the dummy indicators, offering a brief summary instead. We find that collaboration between the top 110 institutions increases almost monotonically with time. About half the increase occurs during 1991-1995[15]. The rise of collaboration between institutions is similar to that of team size. As there it is probably linked to growth of the Internet. Figure 11 includes a graph of the time effects in equation 7.1.

Field effects conform reasonably closely to the descriptive findings in table 3. Relative to the omitted category (chemistry), astronomy, medicine, and physics are the most prone to collaborate with other top U.S. universities. Agriculture, chemistry and engineering are the least prone to this kind of inter-institutional collaboration.

Equation 7.1 includes a basic set of explanatory variables. The logarithm of the stock of federally funded R&D discourages collaboration with other top 110 institutions. In combination with the positive effect of R&D on team size, this suggests that larger research programs tend to form more local teams. This interpretation receives support in equation 7.3, which includes numbers of faculty in the 1993 NRC survey. This variable strongly detracts from collaboration with other top 110 institutions, and steals from the effect of R&D in a university and field. Private control of universities increases institutional collaboration; this is very likely the result of the smaller size of private universities, suggesting that on the

[15] Again breaking the time periods almost into fourths as in fn. 12, we find that the period 1991-1995 contributes about half the total effects over the 19 years.

whole other schools satisfy an excess demand for coworkers. The top 10 indicators are insignificant, while the prize variables on the whole are consistent with greater collaboration among the top 110. The array of PhD student variables, which have no effect on team size, have a nice interpretation in Table 7. Retained PhDs reduce excess demand for inter-institutional team members; removal of PhDs from the same university-field to a distance increases the excess demand. This suggests that PhD students are preferred to other collaborators and the carriers of team arrangements, so that PhD retention and placement at a distance alternatively discourage and encourage institutional collaboration.

Table 8 employs grouped Logit to explain the relative proportions of papers that are contributed by foreign coworkers. We begin with an overview of year and field effects. Similar to the earlier tables, the foreign contribution increases especially rapidly during the early 1990s[16]. Figure 12 is a graph of the time effects in equation 8.1. Field effects follow the descriptive statistics noted in table 4. As there, the fields of astronomy, mathematics and physics are the most likely to engage in international collaboration.

The partial effect of the logarithm of federally funded R&D brings down international collaboration throughout the table. This is quite unlike its role in the other regression tables and actually suggests that university-fields that have *less* funding are the most likely to seek collaboration at a distance. But top 10 universities and top 10 fields, which have larger amounts of funding, are more likely to engage in international collaboration. Thus the total effect of federally funded R&D could increase the foreign contribution. There is evidence in the table that U.S. based PhD students substitute for foreign collaboration and that foreign-based graduates are complementary to it, but this evidence is rarely significant.

Equation 8.2 includes the logarithm of federally funded stocks of R&D in other U.S. schools within and beyond 500 miles. One possible interpretation of the results is that universities that are close to a large pool of federal R&D (within 500 miles) are less likely to engage in foreign collaboration while those that are distant (beyond 500 miles) are more likely to work with foreign co-authors. But more thought is needed before these results can be assessed. Nobel and other prizes in 8.3 also contribute to international collaboration, perhaps through matching with distinguished faculty in leading universities abroad.

[16] Following the pattern of fn. 10 and 12 of dividing time into quartiles, we find that 40 percent of the increase in the foreign share occurs during 1991-1995.

Table 9 concludes the empirical work with some preliminary findings on research "output" consisting of papers and citations received. We include all three collaborative indicators from Tables 6 to 8 in the regressions: authors per paper, top 110 universities per paper, and the relative foreign contribution, as well as field and university characteristics. All equations include year and field effects, as well as the stock of federally funded R&D and sometimes counts of faculty in the 1993 NRC data to control for size of a university-field.

Equations 9.1 and 9.2 have as their dependent variable the logarithm of papers. Equations 9.3 and 9.4 report findings for the logarithm of citations received over all remaining years. The citations exclude self-citations from the same university-field, and given the blurring of field boundaries in the data, citations from other fields in the same university. Equations 9.5 and 9.6 report findings for citations per paper over a fixed window consisting of the year of publication and the next four years[17]. In all cases we employ a fractional assignment of papers and citations. Thus 1/2 of the papers and citations received by papers written with another scientific institution are attributed to a given university-field, 1/3 of the papers and citations received are attributed in the case of three institutions, and so on. In this way we avoid multiple counting of scientific research that is produced cooperatively by several institutions and automatic "productivity" effects of institutional collaboration.

Throughout the table federally funded R&D as well as top 10 universities and fields produce more papers and receive more citations. Counts of faculty and PhD students retained also increase output where they appear. As before Nobel prizes and other career achievement awards typically increase output, though the effects of the different prizes may be difficult to separate.

We now turn to a discussion of the role of the collaboration indicators. In the papers equations, 9.1 and 9.2, authors per paper (team size) and the relative share of foreign coworkers contribute strongly to papers produced[18]. Top 110 universities per paper, which captures most of inter-institutional collaboration within the U.S., decreases the number of papers. That is to say, the net effect of dispersion of a team across the country is to make publication more difficult. Why this should be the case, when it does not hold true

[17] The five-year window excludes papers published after 1995, so the time period of equations 9.5-9.6 is 1981-1995.

[18] Strictly speaking, authors per paper as well as top 110 universities per paper should be lagged to avoid division error bias. We plan to lag these variables in our next draft.

for the relative foreign share, is puzzling. Division error bias as well as the comparative quality of domestic versus international research teams may be potential explanations.

In this respect the findings of equations 9.3-9.6 are reassuring. All three indicators of collaboration contribute to citations received, suggesting that collaboration consistently increases the quality of research. Of course these findings are an upper bound since we are unable to control for "hidden" self-citations of team members located in different institutions. Nevertheless, the findings as a group are consistent with the idea that specialization and the division of labor increase output, in academics as well as industry.

V. Conclusion

This paper has presented evidence on patterns of research collaboration in U.S. universities over the final two decades of the 20th century. Our evidence on the size of scientific teams, measured by authors per scientific paper, shows that specialization and the division of labor have increased markedly over this period. This finding confirms trends noted earlier in the century by other writers, as noted in Section II. We have several new findings to report. First, the period 1991-1995, which coincides with the commercial expansion of the Internet, also coincides with an unusually rapid increase in the size of scientific teams, suggesting some linkage between the two developments. Second, we uncover systematic patterns across the sciences in which larger teams appear to be driven by larger instrumentation and by the intensity of research assistance. Larger teams for example, are more common in physics, astronomy, and medicine, but are less common in "pen and pencil" fields like economics and mathematics. Third, we find that team size increases with funding in a university and field, suggesting that funding drives large projects that are intensive in the use of equipment and research assistants.

Our findings on collaboration between institutions suggest that a somewhat different set of forces determines institutional collaboration than team size. The data seem to say that collaboration with foreign universities increases more rapidly over time than team size, but that collaboration between U.S. universities increases less rapidly. We take this as evidence that the location of team members is shifting and becoming more geographically dispersed, but we lack information on the likely causal factors directing this dispersion. It seems plausible, that domestic collaboration has for a longer time been more feasible than international collaboration and that modern communications technologies have only recently made international science viable for most researchers.

The pattern of inter-institutional collaboration across fields also differs from the pattern for team size. Our view, which we cannot as yet confirm, is that inter-institutional collaboration is more likely in fields and universities where departments are smaller than average. In this way collaborators in other universities permit attainment of second-best team sizes in the face of budgetary limitations, and satisfy an excess demand for team members.

References

Adams, James D., “Comparative Localization of Academic and Industrial Spillovers,” **Journal of Economic Geography** 2 (July 2002): 253-278.

_______________, And Zvi Griliches, “Research Productivity in a System of Universities,” **Annales D’Economie et de Statistique** 49/50 (1998): 127-162.

_______________, Eric P. Chiang, and Katara Starkey, “Industry-University Cooperative Research Centers,” **Journal of Technology Transfer** 26 (January 2001): 73-86.

_______________, Eric P. Chiang, and Jeffrey L. Jensen, “The Influence of Federal Laboratory R&D on Industrial Research,” forthcoming in **The Review of Economics and Statistics**.

Arora, Ashish, and Alfonso Gambardella, “Complementarity and External Linkages: the Strategies of The Large Firms in Biotechnology,” **Journal of Industrial Economics** 38 (June 1990): 361-379.

Becker, Gary S., and Kevin M. Murphy, “The Division of Labor, Coordination Costs, and Knowledge,” **Quarterly Journal of Economics** 107 (November 1992): 1137-1160.

Boning, Brent, Casey Ichniowsky, and Kathryn Shaw, “Opportunity Counts; Teams and the Effectiveness Of Production Incentives,” Cambridge, Massachusetts: NBER Working Paper #8306, May 2001.

Clarke, Beverly L., “Multiple Authorship Trends in Scientific Papers,” **Science** 143 (February 21, 1964): 822-824.

Hicks, Diana, and J.S. Katz, “Science Policy for a Highly Collaborative Science System,” **Science and Public Policy** 23 (February 1996): 39-44.

Holmstrom, Bengt, “Moral Hazard in Teams,” **Bell Journal of Economics** 13 (1982): 324-340.

Kandel, Eugene, and Edward P. Lazear, “Peer Pressure and Partnerships,” **Journal of Political Economy** 100 (August 1992): 801-817.

Katz, J.S., “Geographical Proximity and Scientific Collaboration,” **Scientometrics** 31 (1994): 31-43.

Marshall, Alfred, **Principles of Economics**, 8th Edition, Macmillan, 1920.

Mowery, David C., “International Collaborative Ventures and the Commercialization of New

Technologies," in Nathan Rosenberg, Ralph Landau, and David C. Mowery, editors, **Technology and the Wealth of Nations**, Stanford, California, Stanford University Press, 1992.

National Research Council, **Research-Doctorate Programs in the United States: Continuity and Change**, Washington, DC, National Academy Press, 1995.

National Research Council, **Funding a Revolution: Government Support for Computing Research**, Washington, DC, National Academy Press, 1999.

Powell, Walter W., "Inter-Organizational Collaboration In the Biotechnology Industry," **Journal of Institutional and Theoretical Economics** 152 (1996): 197-215.

Price, Derek J. de Solla, **Big Science, Little Science, and Beyond**, Revised Edition, New York: Columbia University Press, 1986.

Presser, Stanley, "Collaboration and the Quality of Research," **Social Studies of Science** 10 (1980): 95-101.

Rosen, Sherwin, "Authority, Control, and the Distribution of Earnings," **Bell Journal of Economics** 13 (Autumn 1982): 311-323.

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

Stephan, Paula E., and Sharon G. Levin, "The Importance of Implicit Contracts in Collaborative Scientific Research," in Philip Mirowski and Esther-Mirjam Sent, ed., **The Economics of Science**, University of Chicago Press, forthcoming.

Wiener, Norbert, **Invention: The Care and Feeding of Ideas**, Boston, MIT Press, 1994.

Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong, "Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance," Cambridge, Massachusetts, NBER Working Paper #8499, October 2001.

Zuckerman, Harriett, and Robert K. Merton, "Age, Aging, and Age Structure in Science," in Robert K. Merton, **The Sociology of Science: Theoretical and Empirical Investigations**, Norman W. Storer, editor, Chicago: University of Chicago Press, 1973.

Appendix 1
The 110 Top Universities

Appendix Table A-1
The Top 110 U.S. Universities in the Institute for Scientific Information (ISI) Database
Ranked By 1998 Federal R&D

University Name (Rank)	1998 Federal R & D Expenditures	University Name (Rank)	1998 Federal R & D Expenditures
Johns Hopkins University (1)	752.983*	Emory University (36)	118.045
Stanford University (2)	342.426	University of Iowa (37)	115.312
University of Washington – Seattle (3)	336.748	University of California-Davis (38)	114.912
University of Michigan, All Campuses (4)	311.450	Georgia Institute of Technology, All Campuses (39)	113.643
Massachusetts Institute of Technology (5)	310.741	Baylor College of Medicine (40)	110.610
University of California-San Diego (6)	262.303	University of Florida (41)	106.510
Harvard University (7)	251.876	Vanderbilt University (42)	106.325
University of Pennsylvania (8)	247.914	Boston University (43)	104.428
University of Wisconsin-Madison (9)	240.513	University of Miami (44)	101.492
University of California-Los Angeles (10)	233.702	New York University (45)	101.426
Columbia University, All Campuses (11)	229.723	University of Utah (46)	100.722
University of Colorado, All Campuses (12)	228.342	University of Massachusetts, All Campuses (47)	100.122
University of California-San Francisco (13)	219.912	University of Texas Southwestern Med Center Dallas (48)	97.200
University of Alabama, All Campuses (14)	205.511	Indiana University, All Campuses (49)	95.840
Yale University (15)	205.046	Carnegie Mellon University (50)	95.046
University of Minnesota, All Campuses (16)	204.741	University of Virginia, All Campuses (51)	93.328
Cornell University, All Campuses (17)	204.187	Purdue University, All Campuses (52)	92.844
University of Southern California (18)	190.547	SUNY at Stony Brook, All Campuses (53)	91.531
Washington University (19)	187.173	University of Cincinnati, All Campuses (54)	90.307
Pennsylvania State University, All Campuses (20)	186.274	University of Hawaii at Manoa (55)	86.886
California Institute of Technology (21)	177.748	Georgetown University (56)	84.801
Duke University (22)	172.532	University of New Mexico, All Campuses (57)	84.365
University of North Carolina at Chapel Hill (23)	171.505	Virginia Polytechnic Institute and State University (58)	82.734
University of California-Berkeley (24)	171.135	Oregon State University (59)	82.416
University of Illinois at Urbana-Champaign (25)	168.871	Michigan State University (60)	81.146
University of Pittsburgh, All Campuses (26)	168.511	Colorado State University (61)	80.451
University of Texas at Austin (27)	165.082	Yeshiva University (62)	80.000
University of Arizona (28)	161.999	North Carolina State University at Raleigh (63)	79.533
Texas A&M University, All Campuses (29)	144.938	University of Maryland at Baltimore (64)	78.037
Case Western Reserve University (30)	132.274	SUNY at Buffalo, All Campuses (65)	76.037
University of Rochester (31)	130.773	University of Illinois at Chicago (66)	73.797
University of Maryland at College Park (32)	129.198	Oregon Health Sciences University (67)	71.054
Northwestern University (33)	127.911	University of Texas Health Science Center Houston (68)	70.446
University of Chicago (34)	125.982	Rutgers the State University of NJ, All Campuses (69)	69.829
Ohio State University, All Campuses (35)	124.177	University of Tennessee, All Campuses (70)	69.793

Appendix Table A-1
The Top 110 U.S. Universities in the Institute for Scientific Information (ISI) Database
Ranked By 1998 Federal R&D

University Name (Rank)	1998 Federal R & D Expenditures	University Name (Rank)	1998 Federal R & D Expenditures
Princeton University (71)	69.005	Louisiana State University, All Campuses (91)	67.090
University of California-Santa Barbara (72)	68.408	University of California-Irvine (92)	65.902
Woods Hole Oceanographic Institution (73)	64.765	Washington State University (93)	44.510
University of Missouri, All Campuses (74)	63.556	Brown University (94)	44.412
Tufts University (75)	61.167	Rockefeller University (95)	43.845
University of Kentucky, All Campuses (76)	60.760	Arizona State University Main (96)	41.359
University of Nebraska, All Campuses (77)	58.482	Rice University (97)	34.772
Wayne State University (78)	57.646	University of Delaware (98)	33.688
Wake Forest University (79)	56.705	CUNY, All Campuses (99)	32.412
New Mexico State University, All Campuses (80)	56.587	University of AK Fairbanks, All Campuses (100)	31.505
University of Texas Health Science Center San Antonio (81)	55.004	University of Vermont (101)	31.460
Utah State University (82)	54.903	University of California-Santa Cruz (102)	29.849
University of Georgia (83)	54.712	Syracuse University, All Campuses (103)	29.200
University of Connecticut, All Campuses (84)	53.189	Brandeis University (104)	28.098
Tulane University (85)	52.924	University of Oregon (105)	27.041
Iowa State University (86)	51.196	University of New Hampshire (106)	25.913
University of Kansas, All Campuses (87)	50.567	West Virginia University (107)	24.985
Florida State University (88)	50.451	University of California-Riverside (108)	22.988
Virginia Commonwealth University (89)	48.167	Loyola University of Chicago (109)	17.685
Dartmouth College (90)	45.053	Lehigh University (110)	13.019

Notes. Federal R&D is taken from the CASPAR database of the National Science Foundation. * Figure for Johns Hopkins University includes R&D expense for the Applied Physics Laboratory.

Appendix 2
Definitions of Science Fields

Appendix Table A-2
Crosswalk Between 12 and 18 Main CASPAR Sciences
And 88 ISI Disciplines

12 Main CASPAR Sciences	18 Main CASPAR Sciences	88 ISI Disciplines
Engineering		
“	Aerospace Engineering	Aerospace Engineering
“	Chemical Engineering	Chemical Engineering
“	Civil Engineering	Civil Engineering
“	Electrical Engineering	Electrical & Electronics Engineering
“	Mechanical Engineering	Mechanical Engineering
“	Materials Science	Materials Science & Engineering
“	“	Metallurgy
“	Industrial Engineering	Industrial Engineering
“	Other Engineering	Biomedical Engineering
“	“	Environmental Engineering & Energy
“	“	Engineering Mathematics
“	“	Nuclear Engineering
Natural Sciences		
Astronomy	Astronomy	Space Science
Chemistry	Chemistry	Chemistry
“	“	Chemistry & Analysis
“	“	Organic Chemistry/Polymer Science
“	“	Physical Chemistry/Chemical Physics
“	“	Inorganic & Nuclear Chemistry
“	“	Spectroscopy/Instrumentation/Analytical Science
Physics	Physics	Physics
“	“	Applied Physics/Condensed Matter/Materials Science
“	“	Optics & Acoustics
Earth Sciences		
Earth Sciences	Geology	Earth Sciences
“	Geology	Geological, Petroleum & Mining Engineering
“	Oceanography	Oceanography

Appendix Table A-2
Crosswalk Between 12 and 18 Main CASPAR Sciences
And 88 ISI Disciplines

12 Main CASPAR Sciences	18 Main CASPAR Sciences	88 ISI Disciplines
Mathematics and Computer Science		
Mathematics and Statistics	Mathematics and Statistics	Mathematics
"	"	Statistics
Computer Science	Computer Science	Computer Science & Engineering
"	"	Information Technology & Communications Systems
Life Sciences		
Agriculture	Agriculture	Agriculture/Agronomy
"	"	Aquatic Sciences
"	"	Animal Sciences
"	"	Plant & Animal Sciences
"	"	Agricultural Chemistry
"	"	Entomology/Pest Control
"	"	Food Science/Nutrition
"	"	Plant Sciences
"	"	Veterinary Medicine/Animal Health
Biology	Biology	Biochemistry & Biophysics
"	"	Cell & Developmental Biology
"	"	Ecology/Environment
"	"	Molecular Biology & Genetics
"	"	Biotechnology & Applied Microbiology
"	"	Microbiology
"	"	Biology
"	"	Experimental Biology
"	"	Immunology
"	"	Neurosciences & Behavior
"	"	Pharmacology & Toxicology
"	"	Physiology
"	"	Oncogenesis & Cancer Research
Medicine	Medicine	Anesthesia & Intensive Care
"	"	Cardiovascular & Hematology Research
"	"	Cardiovascular & Respiratory Systems
"	"	Dentistry/Oral Surgery & Medicine
"	"	Dermatology
"	"	Clinical Immunology & Infectious Disease

Appendix Table A-2
Crosswalk Between 12 and 18 Main CASPAR Sciences
And 88 ISI Disciplines

12 Main CASPAR Sciences	18 Main CASPAR Sciences	88 ISI Disciplines
Medicine	Medicine	Clinical Psychology & Psychiatry
"	"	Endocrinology, Metabolism & Nutrition
"	"	Environmental Medicine & Public Health
"	"	Gastroenterology and Hepatology
"	"	General & Internal Medicine
"	"	Health Care Sciences & Services
"	"	Hematology
"	"	Medical Research, Diagnosis & Treatment
"	"	Medical Research, General Topics
"	"	Medical Research, Organs & Systems
"	"	Neurology
"	"	Oncology
"	"	Ophthalmology
"	"	Orthopedics, Rehabilitation & Sports Medicine
"	"	Otolaryngology
"	"	Pediatrics
"	"	Radiology, Nuclear Medicine & Imaging
"	"	Reproductive Medicine
"	"	Research/Lab Medicine & Medical Technology
"	"	Rheumatology
"	"	Surgery
"	"	Urology & Nephrology
Social & Behavioral Science		
Economics	Economics	Economics
Psychology	Psychology	Psychiatry
"	"	Psychology

Figure 1--Number of Scientific Papers In the Top 110 Universities, 1981-1999

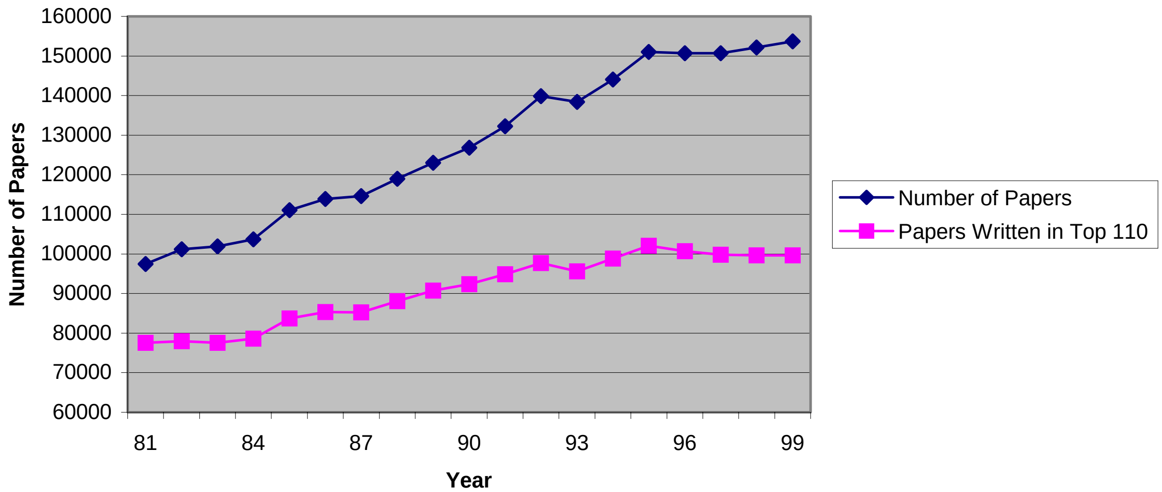


Figure 2--Papers Contributed by the Top 10 Collaborating U.S. Universities To Other Top 110 Institutions, 1981-1999

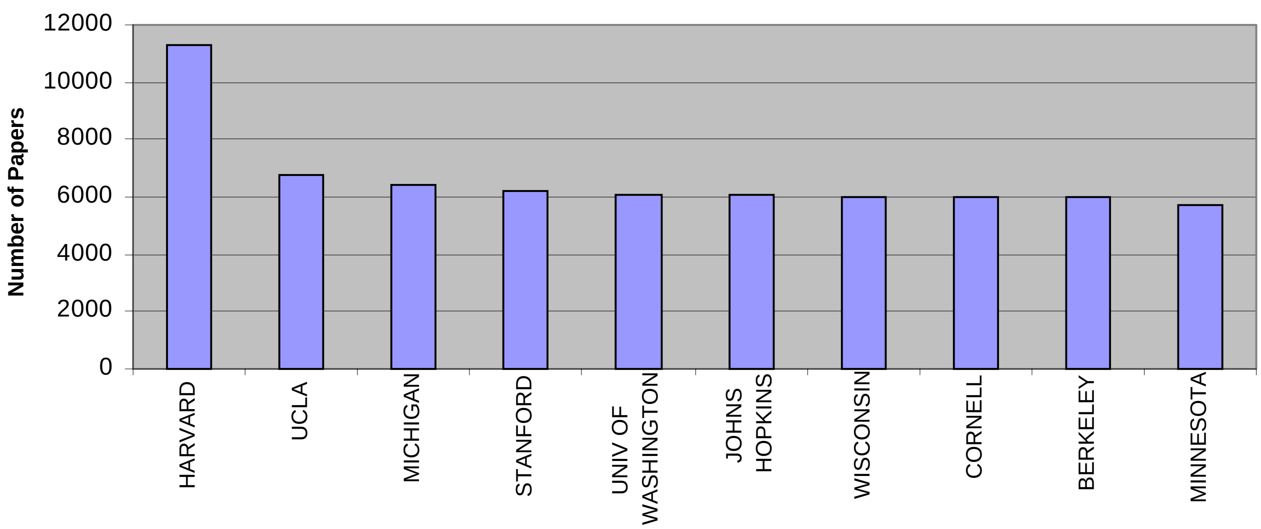


Figure 3--Papers Contributed by the Top 10 Collaborating Countries to the Top 110 U.S. Universities

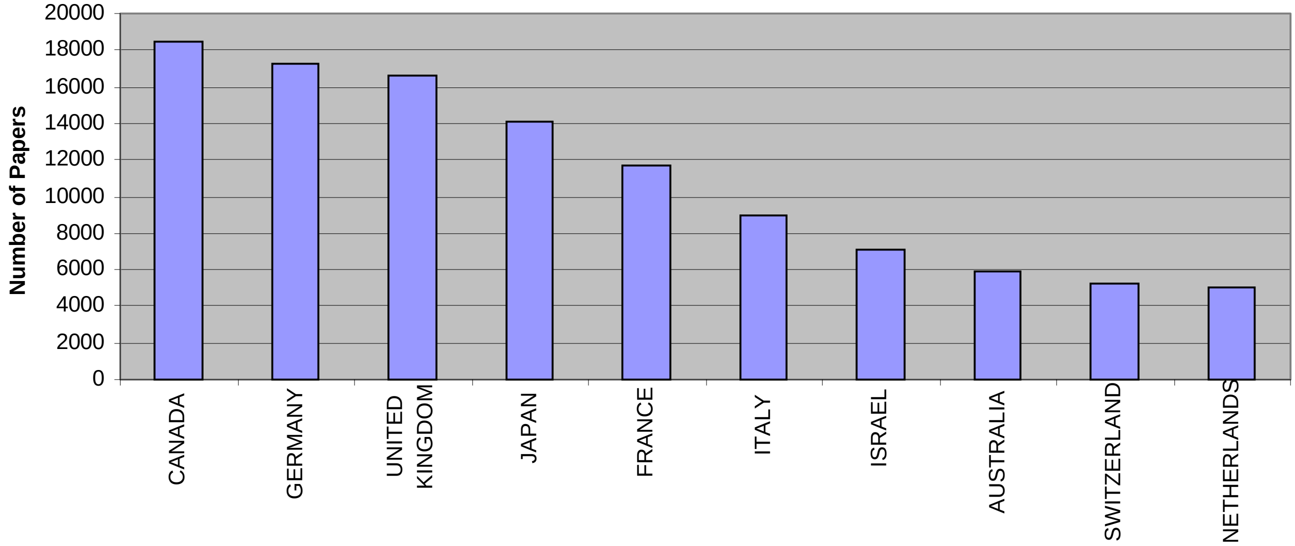


Figure 4--Mean Number of Authors per Paper, for Papers With at Least One Author In the Top 110 U.S. Universities, 1981-1999

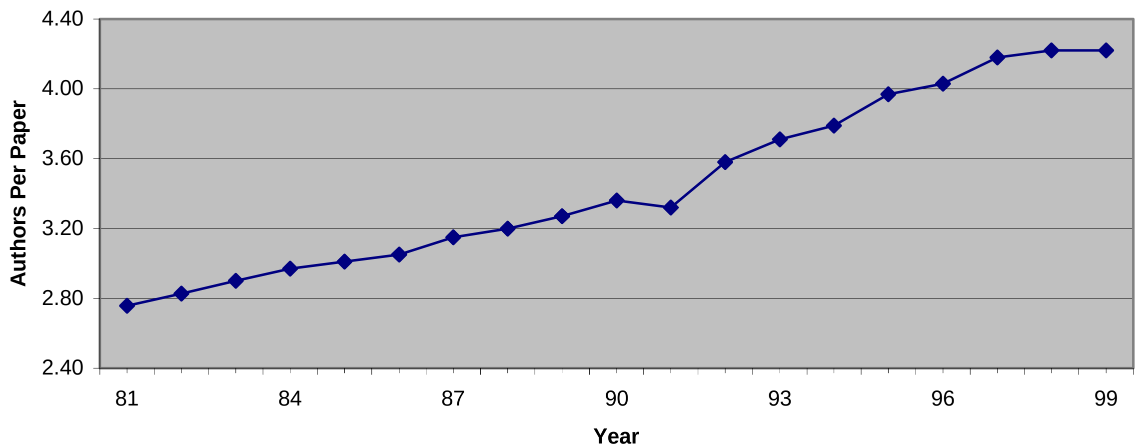


Figure 5--Mean Number of Top 110 Universities per Paper, for Papers With at Least One Author in the Top 110 Universities, 1981-1999

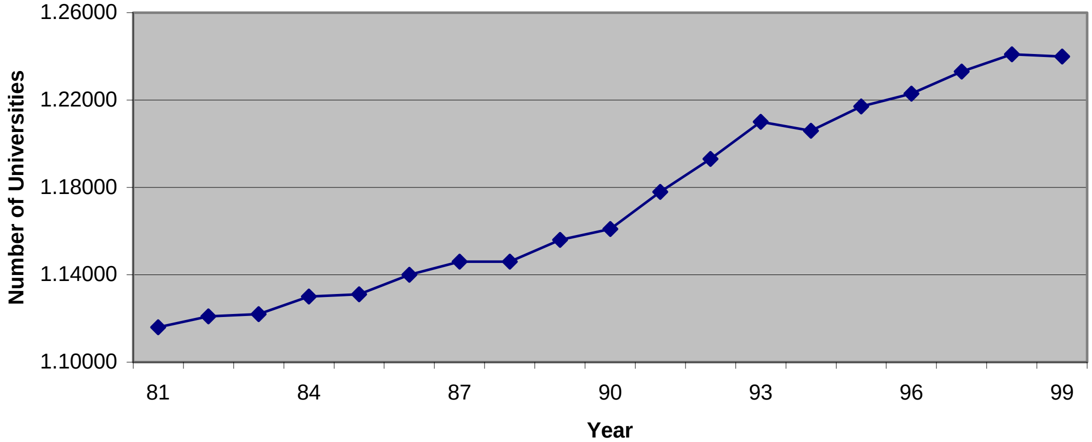


Figure 6--Percentage of Papers Written Inside and Outside the Top 110 U.S. Universities, 1981-1999

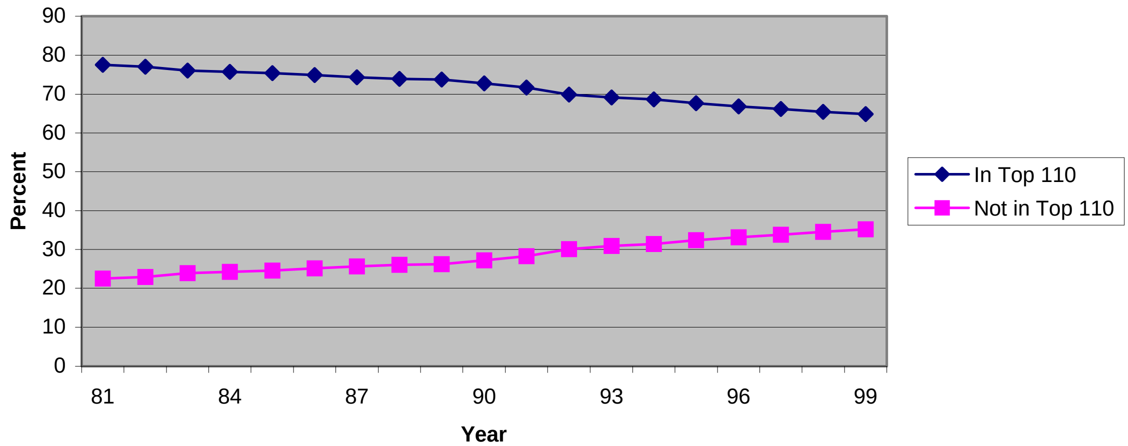


Figure 7--Percentage of Top 110 Papers Written With Non-Top 110 Institutions in the U.S., 1981-1999

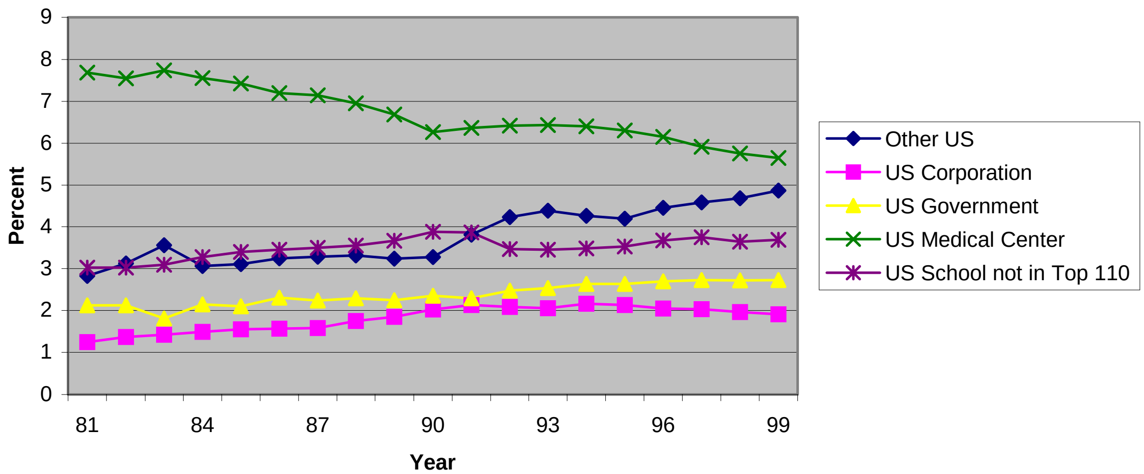


Figure 8--Percentage of Top 110 Papers Written with Other Countries, 1981-1999

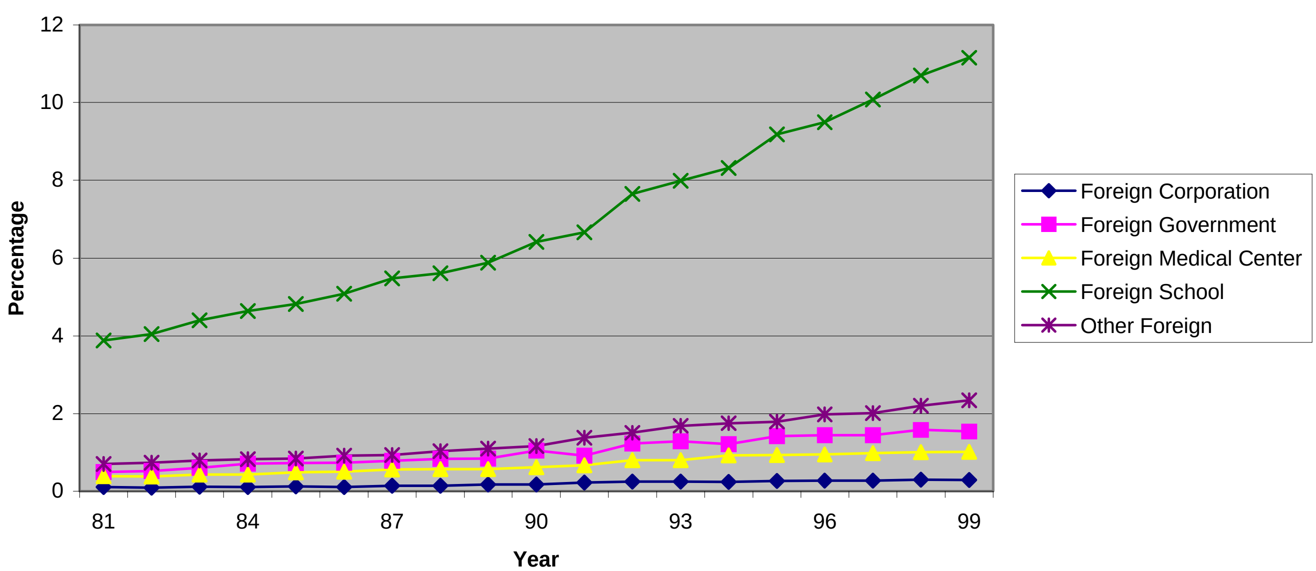


Figure 9--Mean Number of Authors Per Paper
In 12 Fields of Science, Top 110 Universities, 1981-1999

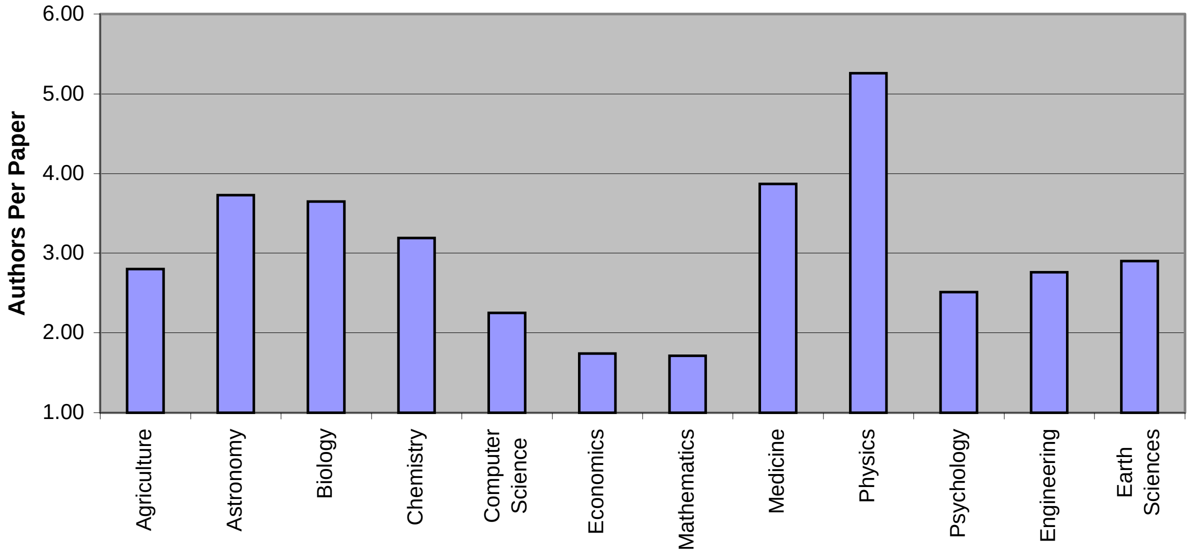


Figure 10--Mean Number of Universities Per Paper
In 12 Fields of Science, Top 110 Universities, 1981-1999

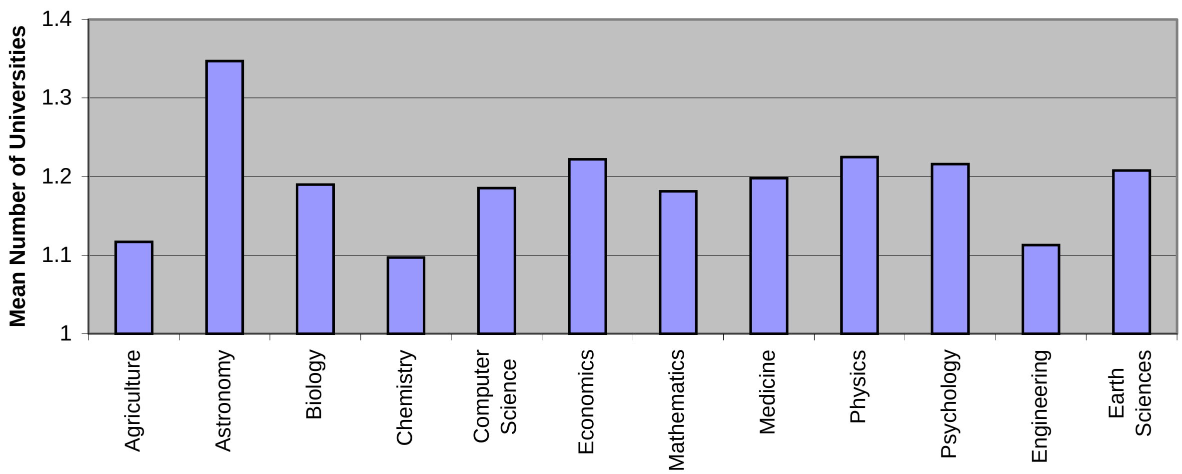
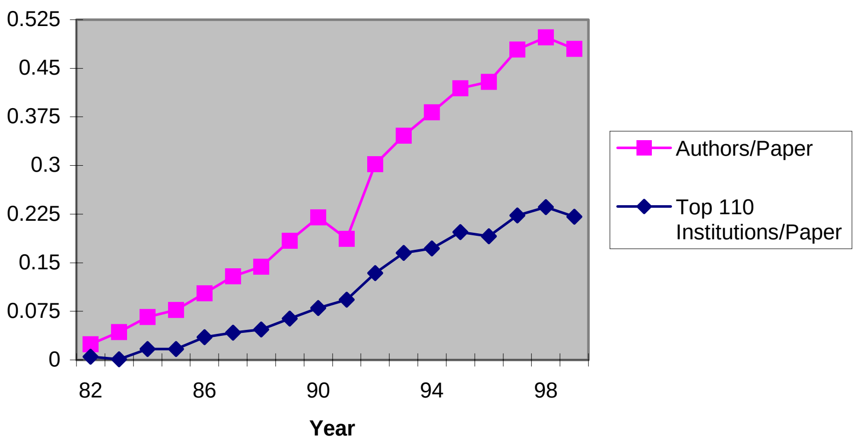
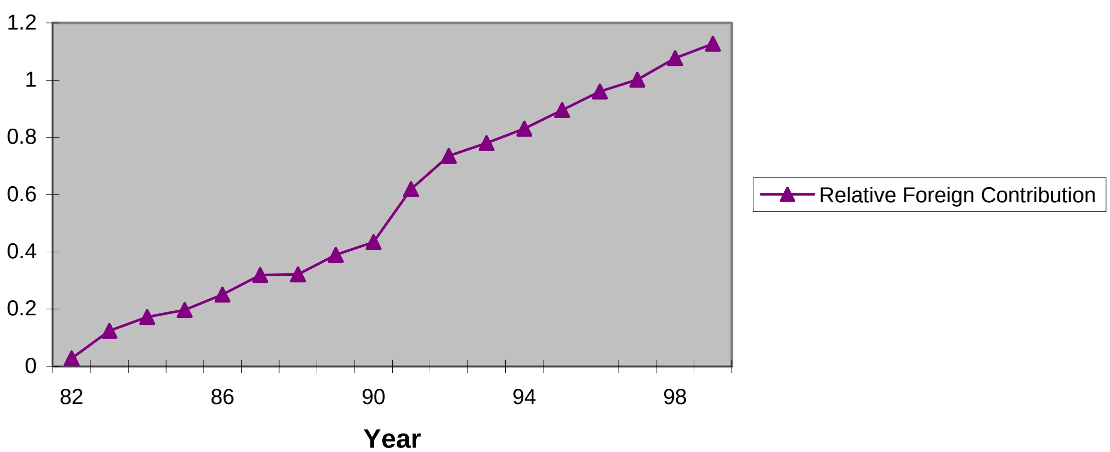


Figure 11--Time Effects in Regressions Explaining Authors and Top 110 Institutions per Paper



Source: Authors/Paper (Table 6, Equation 6.1);
Top 110 Institutions/Paper (Table 7, Equation 7.1).

Figure 12--Time Effects in Regressions Explaining Relative Foreign Contribution to Papers



Source: Relative Foreign Contribution (Table 8, Equation 8.1).

Table 1
Principal Variables in Data Covering
110 Top Universities and 12 Main Sciences

Variable	Description
A. Institute for Scientific Information (ISI) Papers Data	
Paper ID	Character ID of a scientific paper
University ID	Character ID standing for one of the 110 institutions; based on the federal government FICE code
Major Science Classification Field	Field identifier assigned to the journal in which an article was published.
Year	Year of publication
Name of University or Institution	Text data on the university or institution listed on each address; one of the top 110 universities
City	City of each address listed on a paper
State	State of each address Listed on a Paper if address is in the U.S.
Zip Code	Postal zip code if address is in the U.S.
Country	Country of each non-U.S. address listed on a paper
Number of Authors	Number of unique author names listed on each paper; authors are not linked to addresses
Top 10 overall university	Universities that were ranked most frequently in the top 10 in terms of citation impact per paper across 21 fields of science
B. National Science Foundation (NSF) CASPAR Data	
Private University	1 if primarily under private control, else 0
Log (Stock of R&D in a University and Field)	Logarithm of the eight-year stock of R&D in thousands of 1992 dollars in a university and field, depreciated at 15% per year.
Log (Stock of R&D in the same field, but in other Universities)	Logarithm of the eight-year stock of R&D in thousands of 1992 dollars in the same field, but other universities, often arrayed by distance, and depreciated at 15% per year.
C. National Research Council (NRC) Data [a]	
Number of Nobel Prizes	The number of Nobel prizes in a given university and science field from the 1993 NRC Survey
Number of Career Achievement Awards	The number of Fields Medals, McArthur awards, the National Medals of Science, National Medals of Technology, National Academy of Science Fellows in a given university and field as of the 1993 NRC Survey.
Top 10 in Field [b]	Top 10 in Ranking by field in the 1993 NRC survey, with the exception of Agriculture and Medicine, which are ranked by 1998 federally funded R&D taken from CASPAR.
D. NSF Survey of Earned Doctorates (SED) [c]	
Numbers of PhDs in University and Field	Numbers of PhDs in each year, arrayed in the case of students with definite plans, by destination (specific university, sector, or country)

Notes. [a] The data used are the microdata underlying National Research Council (1995) on the list of references. [b] Since program rankings in the NRC data are available only for PhD programs, we used federally funded R&D in 1993 in CASPAR to choose the largest 50 programs in agriculture and the largest 75 schools of medicine. [c] The data used are the microdata on students who are due to receive their PhD in a given year, which includes information on their destinations, if the student has definite plans.

Table 2
Papers Written by the Top 110 U.S. Universities
In 12 Main Fields of Science, 1981-1999

Field of Science	Number of Papers	Number of Papers Times Mean Proportion Written Within the Top 110
Agriculture	192,013	161,543
Astronomy	36,104	24,470
Biology	618,205	501,062
Chemistry	195,189	165,102
Computer Science	28,261	22,407
Earth Sciences	75,759	57,793
Economics	42,481	35,301
Engineering	184,001	149,794
Mathematics	60,875	49,892
Medicine	689,116	535,721
Physics	206,992	157,523
Psychology	96,417	78,542
Total	**2,425,413**	**1,939,150**

Table 3
Percentages of Papers Written With Other U.S. Institutions,
The Top 110 U.S. Universities By 12 Main Fields of Science, 1981-1999

Field of Science	Percent With U.S. Government	Percent With Other U.S. Universities	Percent With U.S. Corporations	Percent With U.S. Medical Centers	Percent With All Other U.S. Institutions
Agriculture	3.9	4.0	1.4	0.8	5.1
Astronomy	5.2	3.5	1.3	0.0	10.0
Biology	2.3	3.0	1.4	6.6	3.3
Chemistry	2.4	3.5	2.6	0.6	2.2
Computer Science	1.3	3.8	8.1	0.1	3.8
Earth Sciences	6.1	4.4	2.3	0.0	6.6
Economics	1.7	7.0	1.3	0.0	5.3
Engineering	3.0	3.2	5.7	0.8	4.1
Mathematics	0.8	4.6	1.2	0.4	1.6
Medicine	1.6	3.1	1.2	13.6	4.4
Physics	4.2	3.9	3.1	0.1	2.6
Psychology	0.8	9.1	0.7	6.3	5.4

Table 4
Percentage of Papers Written With Other Countries,
The Top 110 U.S. Universities, By 12 Main Fields of Science, 1981-1999

Field of Science	Percent With Foreign Governments	Percent With Foreign Universities	Percent With Other Foreign Institutions
Agriculture	0.7	5.3	2.6
Astronomy	4.1	13.9	6.9
Biology	0.9	6.9	2.3
Chemistry	0.9	9.5	1.7
Computer Science	0.5	9.4	2.4
Earth Sciences	1.5	9.8	4.3
Economics	0.3	8.0	1.2
Engineering	0.7	7.4	2.6
Mathematics	0.6	15.5	1.7
Medicine	0.4	4.7	2.3
Physics	4.8	18.0	3.7
Psychology	0.2	4.1	0.9

Table 5
Means and Standard Deviations of Principal Variables

Variable	Mean (S.D.)
Number of Authors per Paper	2.36 (1.02)
Number of Top 110 Universities per Paper	1.41 (0.51)
Share of Foreign Contribution per Paper	0.08 (0.05)
Citations per Paper in all Remaining Years	8.36 (7.60)
Citations per Paper over a Five-Year Window	3.63 (3.89)
Stock of Federally Funded R&D in a University and Field in Thousands of 1992 Dollars	45,605.41 (73,072.81)
Private University	0.35 (0.48)
Top 10 University Overall	0.16 (0.37)
Top 10 in Field	0.19 (0.39)
Number of Nobel Prizes	0.03 (0.19)
Number of Career Achievement Awards	0.17 (0.50)
Number of PhD Students Retained in a University and Field	1.44 (2.27)

Table 6
Determinants of Log (Authors per Paper)
(T-Statistics in Parentheses)

Variable or Statistic	Eq. 6.1	Eq. 6.2	Eq. 6.3	Eq.6.4	Eq.6.5
Method	OLS	OLS	OLS	OLS	OLS
Time Period	1981-1999	1981-1999	1981-1999	1981-1999	1981-1999
Fields Included in Regressions	All 12 Main Fields	10 Fields, Excepting Agriculture and Medicine			
Year Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant
Field Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant
Log (Stock of Federally Funded R & D in University and Field)	0.016 (6.5)	0.012 (3.4)	0.016 (4.2)	0.011 (3.0)	0.015 (3.9)
Private University (1=yes, 0=No)	0.046 (8.4)	0.046 (7.0)	0.040 (5.7)	0.046 (6.9)	0.039 (5.6)
Top 10 Overall University (1=yes, 0=no)	-0.031 (-3.7)	-0.021 (-2.1)	-0.023 (-2.3)	-0.027 (-2.6)	-0.028 (-2.8)
Top 10 in Field (1=yes, 0=no)	-0.001 (-0.1)	-0.007 (-0.7)	-0.005 (-0.5)	-0.015 (-1.5)	-0.013 (-1.3)
Number of Nobel Prizes				0.004 (0.2)	0.003 (0.2)
Number of Career Achievement Awards [a]				0.025 (3.5)	0.025 (3.6)
Log (Count of Faculty in 1993 NRC Survey)			-0.020 (-2.9)		-0.020 (-2.9)
Log (PhD Students Retained)		-0.001 (-1.6)	-0.001 (-1.1)	-0.001 (-1.6)	-0.001 (-1.1)
Log (PhD Students Sent Elsewhere in the US)		0.001 (0.6)	0.001 (1.0)	0.001 (0.6)	0.001 (1.0)
Root MSE	0.263	0.287	0.287	0.287	0.286
Adjusted R^2	0.76	0.75	0.75	0.75	0.75
Number of Observations	12,127	9,820	9,820	9,820	9,820

Notes. The dependent variable is the logarithm of the mean number of authors per paper in a university, field, and year observation. [a] See the text for the definition of career achievement awards.

Table 7
Determinants of Log (Top Universities per Paper)
(T-Statistics in Parentheses)

Variable or Statistic	Eq. 7.1	Eq. 7.2	Eq. 7.3	Eq.7.4	Eq.7.5
Method	OLS	OLS	OLS	OLS	OLS
Time Period	1981-1999	1981-1999	1981-1999	1981-1999	1981-1999
Fields Included in Regressions	All 12 Main Fields	All Except Agriculture and Medicine	All Except Agriculture and Medicine	All Except Agriculture and Medicine	All Except Agriculture and Medicine
Year Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant
Field Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant
Log (Stock of Federally Funded	-0.007	-0.008	-0.004	-0.008	-0.004
R & D in University and Field)	(-5.0)	(-4.1)	(-1.7)	(-4.3)	(-1.9)
Private University	0.029	0.033	0.026	0.032	0.025
(1=yes, 0=No)	(9.2)	(8.9)	(6.6)	(8.8)	(6.5)
Top 10 Overall University	-0.004	-0.006	-0.008	-0.008	-0.010
(1=yes, 0=no)	(-0.8)	(-1.1)	(-1.4)	(-1.5)	(-1.8)
Top 10 in Field	0.003	0.007	0.009	0.004	0.006
(1=yes, 0=no)	(0.7)	(1.3)	(1.7)	(0.7)	(1.0)
Number of Nobel Prizes				-0.005	-0.007
				(-0.5)	(-0.7)
Number of Career Achievement				0.013	0.013
Awards [a]				(3.2)	(3.3)
Log (Count of Faculty in 1993			-0.021		-0.021
NRC Survey)			(-5.5)		(-5.5)
Log (PhD Students Retained)		-0.002	-0.002	-0.002	-0.002
		(-4.0)	(-3.1)	(-4.1)	(-3.1)
Log (PhD Students Sent		0.001	0.002	0.001	0.002
Elsewhere in the US)		(2.4)	(3.1)	(2.3)	(3.1)
Root MSE	0.152	0.161	0.161	0.161	0.161
Adjusted R^2	0.51	0.50	0.50	0.50	0.50
Number of Observations	12,127	9,820	9,820	9,820	9,820

Notes. The dependent variable is the logarithm of the mean number of authors per paper in a university, field, and year observation. See the text for a further discussion. [a] See the text for the definition of career achievement awards.

Table 8
Determinants of the Relative Foreign Contribution
Dependent Variable: Log (Foreign/U.S. Contribution to Papers)
(t-Statistics in Parentheses)

Variable or Statistic	Eq. 8.1	Eq. 8.2	Eq. 8.3
Method		Grouped Logit	
Time Period	1981-1999	1981-1999	1981-1999
Year Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant
Field Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant
Log (Stock of Federally Funded R & D in University and Field)	-0.005 (-3.2)	-0.004 (-2.7)	-0.0005 (-0.3)
Private University (1=yes, 0=No)	0.004 (0.6)	0.032 (4.3)	0.017 (2.2)
Top 10 Overall University (1=yes, 0=no)	0.084 (9.4)	0.069 (7.7)	-0.003 (-0.2)
Top 10 in Field (1=yes, 0=no)	0.040 (5.3)	0.041 (5.5)	0.037 (3.8)
Number of Nobel Prizes			0.035 (2.5)
Number of Career Achievement Awards [a]			0.004 (0.6)
PhD Students Retained	-0.0010 (-0.8)	-0.0012 (-1.0)	-0.0004 (-0.3)
PhD Students Sent Elsewhere in the U.S.	-0.0056 (-6.6)	-0.0049 (-5.7)	-0.0031 (-3.5)
PhD Students Sent Overseas	0.0014 (0.8)	0.0016 (0.9)	0.0018 (1.0)
Log (Federal University R & D) Within 500 Miles		-0.016 (-7.8)	
Log (Federal University R & D) More Than 500 Miles Away		0.061 (3.8)	
Root MSE	0.319	0.317	0.324
Adjusted R^2	0.72	0.73	0.68
Number of Observations	11,982	11,982	9,702

Notes. The Minimum χ^2 method is a weighted, two-stage, ordinary least squares method used to explain relative proportions in grouped data. [a] See the text for the definition of career achievement awards.

Table 9
Determinants of Research Output,
Captured by Papers and Citations
(T-Statistics in Parentheses)

Variable or Statistic	Log (Papers)		Log (Citations in All Remaining Years) [1]		Log (Citations Over A 5-Year Window) [2]	
	Eq. 9.1	Eq. 9.2	Eq. 9.3	Eq. 9.4	Eq. 9.5	Eq. 9.6
Method	OLS	OLS	OLS	OLS	OLS	OLS
Time Period	1981-1999	1981-1999	1981-1999	1981-1999	1981-1995	1981-1995
Fields Included in Regressions	All 12 Main Fields	All Except Agriculture and Medicine	All 12 Main Fields	All Except Agriculture and Medicine	All 12 Main Fields	All Except Agriculture and Medicine
Year Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant
Field Dummies Included	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant	Yes, Significant
Log (Stock of Federally Funded	0.326	0.186	0.370	0.241	0.378	0.252
R & D in University and Field)	(71.6)	(34.4)	(59.8)	(29.3)	(56.5)	(28.5)
Private University	-0.169	-0.086	0.026	0.123	0.032	0.122
(1=yes, 0=No)	(-16.6)	(-8.5)	(1.9)	(8.0)	(2.2)	(7.5)
Top 10 Overall University	0.107	0.147	0.281	0.252	0.242	0.218
(1=yes, 0=no)	(7.1)	(10.0)	(13.7)	(11.4)	(11.0)	(9.3)
Top 10 in Field	0.349	0.281	0.534	0.472	0.530	0.469
(1=yes, 0=no)	(24.6)	(19.5)	(27.7)	(21.8)	(25.5)	(20.3)
Number of Nobel Prizes		-0.065		0.094		0.117
		(-2.5)		(2.4)		(2.8)
Number of Career Achievement		0.099		0.189		0.192
Awards [a]		(9.7)		(12.2)		(11.7)
Log (Count of Faculty in 1993		0.420		0.386		0.376
NRC Survey)		(42.1)		(25.6)		(23.4)
Log (PhD Students Retained)		0.017		0.019		0.017
		(12.3)		(9.0)		(7.9)
Log (Authors/Paper)	0.949	0.849	0.718	0.678	0.593	0.544
	(32.5)	(31.3)	(18.1)	(16.5)	(13.5)	(12.2)
Log (Top 110 Universities/Paper)	-0.650	-0.564	0.394	0.423	0.187	0.229
	(-22.3)	(-21.3)	(10.0)	(10.5)	(3.8)	(4.6)
Log (Foreign Share in Papers)	0.126	0.115	0.114	0.093	0.121	0.101
	(22.6)	(22.3)	(14.8)	(11.9)	(15.9)	(13.2)
Root MSE	0.483	0.415	0.653	0.625	0.626	0.591
Adjusted R^2	0.83	0.86	0.87	0.88	0.86	0.87
Number of Observations	12,127	9,820	12,058	9,753	9,555	7,739

Notes. External citations are citations by other universities to research in a university and field.
[1] Citations in all remaining years include citations from the year of publication through 1999.
[2] Citations over a 5- year window include citations from the four succeeding years. [a] See the text for the definition of career achievement awards.