

NBER WORKING PAPER SERIES

STANDING ON ACADEMIC SHOULDERS:  
MEASURING SCIENTIFIC INFLUENCE IN UNIVERSITIES

James D. Adams  
J. Roger Clemmons  
Paula E. Stephan

Working Paper 10875  
<http://www.nber.org/papers/w10875>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2004

Paper prepared for the International Conference on R&D, Education, and Productivity in Memory of Zvi Griliches, Paris, France; August 25-27, 2003. The Andrew W. Mellon Foundation provided generous support for this research. Nancy Bayers and Henry Small of the Institute for Scientific Information (ISI) offered numerous clarifications concerning the data, while Adam Jaffe provided nonlinear regression programs that we adapted for use in the estimation procedures. We thank three referees for their comments and suggestions. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

© 2004 by James D. Adams, J. Roger Clemmons, and Paula E. Stephan. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Standing on Academic Shoulders: Measuring Scientific Influence in Universities

James D. Adams, J. Roger Clemmons, and Paula E. Stephan

NBER Working Paper No. 10875

October 2004

JEL No. L30, O30

### **ABSTRACT**

This article measures scientific influence by means of citations to academic papers. The data source is the Institute for Scientific Information (ISI); the scientific institutions included are the top 110 U.S. research universities; the 12 main fields that classify the data cover nearly all of science; and the time period is 1981-1999. Altogether the database includes 2.4 million papers and 18.8 million citations. Thus the evidence underlying our findings accounts for much of the basic research conducted in the United States during the last quarter of the 20th century. This research in turn contributes a significant part of knowledge production in the U.S. during the same period.

The citation measure used is the citation probability, which equals actual citations divided by potential citations, and captures average utilization of cited literature by individual citing articles. The mean citation probability within fields is on the order of  $10^{-5}$ . Cross-field citation probabilities are one-tenth to one-hundredth as large, or  $10^{-6}$  to  $10^{-7}$ . Citations between pairs of citing and cited fields are significant in less than one-fourth of the possible cases. It follows that citations are largely bounded by field, with corresponding implications for the limits of scientific influence.

Cross-field citation probabilities appear to be symmetric for mutually citing fields. Scientific influence is asymmetric within fields, and occurs primarily from top institutions to those less highly ranked. Still, there is significant reverse influence on higher-ranked schools. We also find that top institutions are more often cited by peer institutions than lower-ranked institutions are cited by their peers. Overall the results suggest that knowledge spillovers in basic science research are important, but are circumscribed by field and by intrinsic relevance. Perhaps the most important implication of the results are the limits that they seem to impose on the returns to scale in the knowledge production function for basic research, namely the proportion of available knowledge that spills over from one scientist to another.

James D. Adams  
Department of Economics  
3504 Russell Sage Laboratory  
Rensselaer Polytechnic Institute  
110 8<sup>th</sup> Street  
Troy, NY 12180-3590  
and NBER  
adamsj@rpi.edu

J. Roger Clemmons  
Institute for Child Health Policy  
University of Florida  
PO Box 100147  
Gainesville, FL 32610-0147  
jrc@ichp.ufl.edu

Paula E. Stephan  
Department of Economics  
and Andrew Young School of  
Policy Studies  
Georgia State University  
Atlanta, GA 30303-3083  
pstephan@gsu.edu

# I. Introduction

This paper is part of a larger project that has occupied much of our time in recent years<sup>1</sup>. In the early going the project's main goal is to describe basic research interactions among the top 110 U.S. universities, among the top 200 U.S. R&D firms, and between the universities and firms. Owing to space constraints, this paper concentrates on measurement of scientific influence among universities. The time period is 1981-1999; the scope of our investigation includes all of science. The data cover 2.4 million scientific papers by universities and 18.8 million citations to these papers<sup>2</sup>. To date our research emphasizes pre-technology science rather than patented technology, though the two overlap. Branstetter (2003), Darby and Zucker (2003), Jensen and Thursby (1999), and other papers show that this overlap is important to technical progress. This paper's study of scientific influence within academia, however, lays the groundwork for understanding its influence on industry, since fertile areas of science can give rise to industrial science, in part by means of faculty consultants and entrepreneurs (Audretsch and Stephan, 1996; Zucker, Darby and Brewer, 1998).

This paper describes scientific influence, which in principle refers to the productive role that earlier work plays in later work. Our goal is to trace scientific influence across institutions, sciences, and time. Space limitations confine our measurements not just to universities, but to a single indicator, citations to scientific papers from other papers. But publications are the most common channel of knowledge flow in a recent survey of industrial research (Cohen, Nelson, and Walsh, 2002), let alone universities. The findings in this paper are a first step towards allowing for many channels of influence, especially collaboration and mobility of graduate students and faculty. We assume that papers represent new knowledge, though the amount of this knowledge varies. We also assume that citations to papers on the whole indicate intellectual influence rather than pre-publication strategy or honorific referencing.

The citation measures used in this paper owes a great deal to bibliometrics. Examples of this literature are De Solla Price (1965, 1986), Garfield (1972), Narin and Hamilton (1996), Narin, Hamilton, and Olivastro (1997), and Van Raan (1990), among others.

---

<sup>1</sup> During the planning phases of the project we had the indisputable advantage as well as undeniable privilege of working with Zvi Griliches. We deeply and profoundly regret that he did not live to see this work near completion.

<sup>2</sup> The full data set consisting of all papers and citations of the universities and firms amounts to 2.7 million papers and more than 20 million citations.

The following are key questions raised in the course of our research. What determines the level of influence within different fields? How does influence within fields compare with influence across fields? How common are the “between” components compared to “within” components? Are cross-field influences symmetric or not? What role does quality of scientific institutions play in scientific influence? Could scientific influence increase with quality in such a way as to reinforce scientific excellence?<sup>3</sup>

This paper draws on a literature which argues that the role of outside knowledge in productivity growth is important but limited. Its common thread is that knowledge flows are constrained by limited relevance, so that increasing returns from knowledge are correspondingly limited. The various studies cover agriculture (Evenson and Kislev, 1975), manufacturing (Scherer, 1982a, b), basic research (Adams, 1990), applied industrial research (Adams and Jaffe, 1996), as well as surveys of the externalities from R&D and their limits (Griliches, 1979, 1992). Our approach to measurement of these limits has been shaped by Jaffe (1986), Trajtenberg (1990), and especially Jaffe and Trajtenberg (1999).

Likewise the findings in this paper imply that limits apply to the relevance of scientific ideas. And since science could replenish research opportunities in industry (and conversely), this result bears on economic growth. In the evolution of growth theory, scale effects in the knowledge production function have been steadily curtailed over time. Thus, opportunities for growth have been viewed as deriving from the growth of R&D rather than its level, supported by a contribution from spillovers that is sufficient to avoid diminishing returns to research. However, growth models (Romer (1990), and Jones (1995, 2002)) tend to assume that knowledge flow without friction throughout countries and even the world<sup>4</sup>. And yet our findings suggest that the frictions are substantial, so that influence is limited by the narrow applicability of most discoveries.

The measure of scientific influence used in this paper is the citation probability, which divides actual citations by potential citations. This is the average number of citations per potentially cited paper, divided by the number of potentially citing papers. This measure captures the rarity of citation by individual science papers by normalizing actual citations by potential citations.

---

<sup>3</sup> This is the Matthew Effect. For more on this topic see Merton (1973) and Zuckerman (1977).

<sup>4</sup> Marshall (1920, p. 220) observes that “Many of those economies in the use of specialized skill and machinery which are commonly regarded as within the reach of very large factories, do not depend on the size of individual factories. Some depend on the aggregate volume of production of the kind in the neighborhood; while others again, especially those connected with the growth of knowledge and the progress of the arts, depend chiefly on the aggregate volume of production in the whole civilized world.”

Main findings of the paper are as follows. First, the within-field citation probability is on the order of  $10^{-5}$  and is 10 to 100 times the between-field citation probability. The comparative rarity of cross-field citations holds even when cross-field interactions are required to be statistically significant.

Second, the number of significant cross-field interactions is one-fourth of the potential number. Since the data include 12 science fields, there are  $11 \times 12 = 132$  cross-field interactions. Thirty-two of these differ significantly from zero. This suggests that knowledge flows are bounded by scientific distance, just as knowledge flows in industry are bounded by technological distance (Adams and Jaffe, 1996).

Third, the findings support the hypothesis of symmetry of cross-citation between pairs of mutually citing fields. The rate at which biology cites medicine is nearly the same as the rate at which medicine cites biology. However, this finding is conditional on two-way citation. Asymmetries exist, but take a more subtle form of applied fields citing underlying basic fields more than the reverse.

Fourth, the modal or most frequent lag in science citations based on publication years, a measure of the speed of diffusion, is slightly more than three years. This compares with a modal lag of more than five years in patent citations based on analogous grant years (Jaffe and Trajtenberg, 1999).

Fifth, the rank of scientific institutions increases scientific influence. Sixth, citation interactions with peer institutions increase with rank. University-fields in the top 20% cite one another at a significantly higher rate more than university-fields in the middle 40% or bottom 40% cite one another. This suggests that knowledge flows from peer programs increase with rank.

The rest of this paper consists of five sections. Section II motivates the paper by reexamining the knowledge production function in light of our findings. Section III discusses the meaning of science citations, presents the citation probability, and describes the science citation function. Section IV describes the data, specifies the science fields, and discusses measurement of the quality of scientific institutions. Section V reports econometric estimates of the science citation function. Section VI concludes.

## **II. Motivation**

We motivate the empirical findings by exploring a knowledge production function in which searching the literature increases absorption of knowledge into current research. The research process accumulates heterogeneous stocks of knowledge. Accumulation of one stock depends on several knowledge stocks, as the findings suggest, and on the allocation of labor to invention and searching the literature. The stocks

consist of basic scientific knowledge. Using this construct we illustrate the constrained output-maximizing allocation of scientific labor. We then show how this allocation might replicate the empirical results below.

Let  $\dot{A}_i$  be the time derivative of science  $i$ , let  $L_i$  be “inventive” labor engaged in discovery, and let  $K_{ij}$  be “absorptive” labor in  $i$  engaged in searching stock  $A_j$ . Then the knowledge production function is

$$(1) \quad \dot{A}_i = \delta L_i^\alpha \sum_{j=1}^N (K_{ij} A_j)^{\beta_{ij}} \quad 1 > \alpha, \beta_{ij} > 0, \quad i = 1 \dots N$$

Returns to scale are constant or diminishing so  $\alpha + \sum_{j=1}^N \beta_{ij} \leq 1$ . Inventive labor and some absorptive labor are required for accumulation. If  $\beta_{ii}$  exceeds  $\beta_{ij}$  within-field effects exceed cross-field effects.

Let the wage of researchers be  $W_R$  and let total research cost be  $C$ . If we take the objective of scientists to be maximization of output subject to a cost constraint, then the Lagrangian is

$$H = \delta L_i^\alpha \sum_{j=1}^N (K_{ij} A_j)^{\beta_{ij}} + \lambda \left[ C - W_R \left( L_i + \sum_{j=1}^N K_{ij} \right) \right]$$

where  $\lambda$  is the Lagrangian multiplier. Adopting this as the objective avoids the assumption of profit maximization, which is not clearly relevant to university research. Maximization of  $H$  yields the following first order conditions for absorptive labor in the same and other fields:

$$(2) \quad \begin{aligned} \delta \beta_{ii} L_i^\alpha K_{ii}^{\beta_{ii}-1} A_i^{\beta_{ii}} &= \lambda W_R \\ \delta \beta_{ij} L_i^\alpha K_{ij}^{\beta_{ij}-1} A_j^{\beta_{ij}} &\leq \lambda W_R \end{aligned} ,$$

In the interest of brevity we omit the first order condition for  $L_i$  and we assume that the second order conditions are satisfied. The top equation determines absorption of science in the same field while the second inequality determines absorption of other fields. In the case of generally small  $\beta_{ii}$  and  $\beta_{ij}$ , absorptive labor is small in the face of limited research budgets. Limiting values of  $\beta_{ij}$  imply that the second member of (2) is a strict inequality. In that case science  $j$  drops out of the production function for  $i$ , as we frequently find. If, as also appears to hold empirically, technological distance generates an

extreme drop off in the cross-term  $\beta_{ij}$  relative to  $\beta_{ii}$ , absorptive labor in the same science  $K_{ii}$  exceeds  $K_{ij}$  committed to other sciences. To see this, let the second member of (2) be an equation, compute the ratio of the two equations, and solve for  $K_{ii}$ . After some manipulations the result is

$$(3) \quad K_{ii} = \left( \frac{\beta_{ii} A_i^{\beta_{ii}}}{\beta_{ij} A_j^{\beta_{ij}}} \right)^{\frac{1}{1-\beta_{ii}}} K_{ij}^{\frac{1-\beta_{ij}}{1-\beta_{ii}}}$$

If  $A_i \approx A_j$ , and if, as already postulated,  $\beta_{ii} > \beta_{ij}$ , then it follows that  $K_{ii} > K_{ij}$ . To put this point another way, unless  $A_j$  greatly exceeds  $A_i$ , absorptive labor is concentrated within the same field.

In the next section we argue that science citations proxy for absorptive labor, even if with error, by representing the outcome of that labor. If this is so the implications of this section are reflected in a small propensity to cite, dominance of within-field citations, and absence of many between-field citation possibilities.

### III. Citation Analysis

#### A. The Meaning of Science Citations

Science citations refer to prior literature, but their motivation is obscure. Citations could measure influence of earlier ideas, they could limit the problem being addressed, they could seek to refute findings, or they could be a strategy to raise the odds of acceptance. Of these motives the first two are most likely to represent scientific influence. Given negative or strategic citations, though, we have to regard citations as measuring scientific influence with error<sup>5</sup>.

Science citations are typically controlled by authors. Referees and editors can suggest references, but their inclusion requires author's assent, suggesting some knowledge of the references. In contrast patent citations are often suggested by examiners and attorneys and do not imply the same acquaintance as science citations do. This is an advantage of science citations, but it should not be overstated. The quality of science citations varies with the people that make them. Again the citations could be strategic because

---

<sup>5</sup> See Banks, Fogarty, and Jaffe (1996) for an analysis that uses a set of NASA patents, as well as expert opinion on the patents, to test the validity of patent citations, answered in the affirmative, as an indicator of the importance of patents.

authors choose references with no monetary punishment, unlike patent citations. Citations have gotten easier to make because of search engines, though whether this raises or lowers quality is unclear.

Suppose that science citations reflect credible investments of time spent searching the literature. What would be the earmarks of such investments? For starters, the number of citations would set marginal benefit of another citation equal to its marginal cost, as above. This suggests that citations would span larger fractions of smaller disciplines, since similar marginal benefit and cost relationships across disciplines would lower the proportion of larger literatures that is cited. Furthermore, literatures requiring larger investments of time per cited paper would result in lower citations, size of literature held constant.

## B. Citation Measures

The citation probability used by Jaffe and Trajtenberg (1999) is one way to take account of the size of citing and cited populations. The citation probability is

$$(4) \quad P_{iTjt} = \frac{c_{iTjt}}{n_{iT}n_{jt}}$$

where  $i$  and  $j$  are citing and cited groups, defined below, and  $T$  and  $t$  are citing and cited years,  $T > t$ . The term  $c_{iTjt}$  in the numerator is the citation count from group  $i$  in year  $T$  to group  $j$  in year  $t$ . Citations are the number of linked papers in  $(i, T)$  and  $(j, t)$ . The product of  $n_{iT}$  and  $n_{jt}$  in the denominator are the numbers of potentially citing and potentially cited papers in  $(i, T)$  and  $(j, t)$  that could be linked. Thus (4) is bounded between 0 and 1 and has a probability interpretation. If not one paper in  $(i, T)$  cites not one paper in  $(j, t)$  then (4) equals zero. If every paper in  $(i, T)$  cites every paper in  $(j, t)$  then (4) equals 1.0. In fact (4) is closer to zero than 1.0, reflecting limits on citation that operate on individual scientific papers.

Of course the citation probability is one of several measures that could be examined. It is a measure of average (not total) scientific influence by group  $j$  on papers in group  $i$ . The logic is this. Suppose that a paper in  $iT$  (with superscript  $k$ ) references a fraction of the literature in  $jt$ , as indicated by citations per paper  $c_{iT}^k / n_{jt}$ . This resembles a utilization rate. The *average* utilization rate of the literature in  $jt$  by papers in  $iT$  is then the sum over  $k$  of  $c_{iT}^k / n_{jt}$  divided by  $n_{iT}$ , or (4). For a group of citing papers it makes more sense to use the group utilization rate, which is  $c_{i\bar{t}jt} / n_{jt}$  or (4) multiplied by the number of

citing papers  $n_{iT}$ . But an analysis of alternative metrics is beyond the scope of this article, which deals with measuring scientific influence on individual papers.

## C. Citation Function

The citation probability (4) is defined on cells that are specified by citing and cited groups  $i$  and  $j$  and by years. The groups are first of all fields. But if the citing and cited fields are the same, then we make a further distinction. In that case we supplement field with rankings of scientific institutions into high, medium, or low “rank-stratification” classes. The use of rank-stratification classes in the citation function is discussed below, while section IV.C explains the empirical procedure that defines the classes. That procedure relies heavily on National Research Council peer rankings of the quality of graduate programs, which are statistically independent of the bibliometric evidence that we explore below.

To fit the citation probability to the data we adapt a procedure used in Jaffe and Trajtenberg (1999) to explain patterns of patent citation. We incorporate the following features of the data. As already noted, we incorporate differences in own and cross-field citation propensities. Second, we allow for top-down asymmetries in citation which favor leading scientific institution. Third, we take account of the early peaking of citation, followed by a long decline, as the lag in citation increases. In this section we model these effects using nonlinear regression<sup>6</sup>. The baseline citation function is:

$$(5) \quad p_{iTjt} = \alpha_{ij} \alpha_T \alpha_t \exp \left[ -\beta_1 \beta_{1j} (T - t) \right] \left\{ 1 - \exp \left[ -\beta_2 (T - t) \right] \right\} + u_{iTjt}$$

In (5)  $\alpha_{ij}$  is the average probability that field  $i$  cites field  $j$ ,  $\alpha_T$  is the average probability that a citation is made in period  $T$ , and  $\alpha_t$  is the average probability that a citation is received in period  $t$ . In the data  $i$  and  $j$  represent the 12 fields of science in table 1 below. Notice that the parameters are defined relative to a baseline value<sup>7</sup>. We normalize  $\alpha_{ij}$  by the parameter for chemistry citing itself, whose value accordingly equals 1.0. Likewise  $\alpha_T$  and  $\alpha_t$  are parameters that are normalized by the earliest periods citing and cited, whose values equal 1.0. The estimation procedure does not converge when we specify a full set of

<sup>6</sup> We thank Adam Jaffe for nonlinear regression programs that we modified and used in this paper.

<sup>7</sup> Use of a baseline value avoids indeterminacy in the multiplicative specification (4). This indeterminacy is the analogue to the dummy variable trap in additive regression models.

citing and cited years, so we aggregate citing years into periods: 1981-1985, 1986-1990, 1991-1995, and 1996-1999. Thus  $T$  and  $\alpha_T$  refer to citing intervals, while  $t$  and  $\alpha_t$  refer to single years cited.

The parameter  $\beta_1$  stands for the decay rate in citations by the baseline field of chemistry. The  $\beta_{1j}$  terms are thus decay parameters relative to chemistry. The parameter  $\beta_2$  governs overall diffusion. Since  $\beta_2$  positions the overall rate of citation, this parameter is not identified by field independently of the  $\alpha_{ij}$  vector. Finally, the error term is  $u_{iTjt}$ . Estimates of the citation function (5) are shown in table 4 below. The exponential form of (5) handles the familiar hump-shaped pattern of citations over different lags, as in figure 1 below.

Recall that in addition to (5) we consider a more elaborate specification of the citation function. This version allows for rank-stratification class effects within sciences. To allow for such effects we replace the within-field parameter  $\alpha_{ii}$  with this 3×3 matrix of citation possibilities:

$$(6) \quad \alpha_{(i)(i)} = \begin{pmatrix} \alpha_{i,11} & \alpha_{i,12} & \alpha_{i,13} \\ \alpha_{i,21} & \alpha_{i,22} & \alpha_{i,23} \\ \alpha_{i,31} & \alpha_{i,32} & \alpha_{i,33} \end{pmatrix}$$

Equation (6) takes rank-stratification class effects into account within fields, where most citations occur<sup>8</sup>. The leading subscript of (6) refers to field while trailing subscripts refer to the top 20%, the middle 40%, and the bottom 40% of institutions. Again the parameters are identified up to a baseline value, which in this case is the probability that top 20% institutions in chemistry cite each other. Hence  $\alpha_{4,11}=1$ .

The first row of (6) consists of probabilities that top 20% schools cite themselves, middle 40% institutions, and bottom 40% institutions. The second row capture probabilities that the middle 40% cites the top 20%, themselves, and the bottom 40%. And in the third row are probabilities that the bottom 40% cites the top 20%, the middle 40%, and themselves. We report estimates of (6) in table 5.

We group the data on citations, potentially citing papers, and potentially cited papers into cells for the purpose of estimation. The cells are as follows. For each field  $i$  and field  $j$  combination there are 171

---

<sup>8</sup> Citations between fields may be too uncommon to allow estimates of rank-stratification effects.

citing and cited year combinations, given a citation lag of at least one year<sup>9</sup>. For each such combination we define nine cells within fields, as in (6). Between-field cells are fields potentially citing 11 other fields<sup>10</sup>. Given 12 fields there are  $9 \times 12 = 108$  within-field cells, and up to  $11 \times 12 = 132$  between-field cells, or up to 240 cells, for each of the 171 citing and cited year combinations. The potential number of cells is  $240 \times 171 = 41,040$ . But 4,206 of these are missing. This happens because citing and cited year pairs are missing if cross-field citations are rare. Thus the number of cells is  $41,040 - 4,206 = 36,834$ .

## IV. Description of the Database

### A. Scientific Papers and Citations Data

The data set consists of 2.4 million publications of the top 110 U.S. universities and 18.8 million citations to those papers during 1981-1999. The top 110 universities that are included in the data account for about three-fourths of the academic R&D conducted in the U.S. The schools are identified in the appendix to Adams, Black, Clemmons, and Stephan (2004).

The source of the data is the Institute for Scientific Information (ISI) in Philadelphia, Pennsylvania. The set of publications includes articles, reviews, notes, and proceedings, or ISI's standard set of communications, in 12 main fields of science that span scientific research. The fields are agriculture, astronomy, biology, chemistry, computer science, earth sciences, economics and business; engineering, mathematics and statistics; medicine, physics, and psychology.

The papers appear in 7,137 scientific journals. The set expands to include new journals through the end of the sample period. The journal set is also world-wide in scope. It is important to note that journals are assigned to a single dominant field<sup>11</sup>. This assignment is reasonably accurate for the vast majority of specialized journals, in part because of the breadth of the 12 fields. But the method produces serious assignment errors for about one percent or 70 journals that fall into the Multidisciplinary category. ISI

---

<sup>9</sup> Papers in 1999 can cite papers from 1998 through 1981, forming 18 combinations. Papers in 1998 can cite papers from 1997 through 1981, forming 17 combinations. This continues through 1982, when only 1981 papers can be cited. The sum of the series is  $S = 18 + 17 + 16 + \dots + 1$ , or  $S = (19 \times 18) \div 2 = 171$ .

<sup>10</sup> In table 4 we ignore rank-stratification classes and average over all nine probabilities that correspond to all possible interactions between the different rank-stratification classes for each science.

<sup>11</sup> In the case of 5,507 journals that are currently published, we follow the assignment of journals to fields by ISI and rely on ISI's experience to provide an accurate assignment. In the case of 1,630 journals that are formerly published we rely on field assignments of CHI (Computer Horizons Inc.). The argument is the same, that experience of an established firm in bibliometrics is likely to be more accurate than idiosyncratic assignments by ourselves.

treats this category as part of biology, which accounts for the largest number of its papers. General Multidisciplinary journals include **Nature**, **Science**, **Proceedings of the National Academy of Sciences USA**, and **Philosophical Transactions of the Royal Society**. To classify all articles in these journals in biology is a mistake, though solutions to the problem are scarce. Moreover, quite a few Multidisciplinary journals are linked to biology, so that this error applies to less than one percent of the journals<sup>12</sup>.

The main alternative is to assign papers according to academic departments of the authors. But current bibliometric information rules this out, and in any event this would not generate multiple field assignments<sup>13</sup>. This is unlike patents, where multiple class assignments of individual patents are common. To carry out multiple assignments clear criteria would be needed and these would have to be agreed on by a Scientific Papers Office, much like the Patent Office. Neither of these conditions is currently satisfied.

Another problem is that science citation data do not and cannot include publication date but not first submission dates at the start of perhaps several submissions. The use of publication dates creates an upward bias in the citation lag. The reason is that the lag between cited publication dates and citing submission dates captures the lag when the paper is written. The extra “frictional” lag between citing submission and publication dates overstates the true lag. This point suggests that differences in diffusion across science are biased by the use of citing publication dates.

While voluminous, the scientific papers and citations data are only a window on scientific research. They are truncated on the left and right in time. As a result we lack most of the citations to papers from the late 1990s, since the citations are not yet made. We know nothing about the papers that influenced research in the early 1980s since citations to these papers are excluded. The data are also limited by sector and country. They are limited to papers that have at least one author in the top 110 U.S. universities. Citations made by U.S. researchers to foreign institutions are excluded, as are citations received by U.S. researchers from foreign institutions. For this reason many of the rich interactions of the international scientific enterprise are left out of our analysis. But the *system* of U.S. university citations and papers is still a great improvement over what we have had.

---

<sup>12</sup> Examples include **Bioinformatics**, **Biomaterials**, **Biometrics**, **Biometrika**, **Journal of Mathematical Biology**, **Journal of Theoretical Biology** and many others.

<sup>13</sup> As an experiment we tried to assign all Harvard papers to one of the 12 main science fields using address information. A third of the papers could not be assigned to a field using information on authors' Harvard addresses. This led us to abandon the effort, though more could be done on this in the future were journals to codify fields of authors and classify papers by field of author rather than field of journal.

## B. Field Dimensions of Citation

Table 1 looks at field dimensions of the data. The 12 main science fields are shown on the left. The second column reports total papers, percent of all papers, total citations received, and percent of all citations received. The third column reveals the composition of main fields in terms of sub-fields. Besides variation in size and complexity, the table brings out differences in citation practices. Biology ranks second in publications and first in citations received. Computer science ranks last in both, pointing out how differences in the size of citing populations and propensities to cite can affect citation.

For the field combinations that are significant in the regressions, table 2 reports means of citations, potential papers citing, and potential papers cited. Table 2 excludes self-citation by a university to itself, as do all later tables. Such “institutional” self-citation reflects influence of a university’s past research rather than knowledge spillovers. Out of similar concerns we exclude same-university citations between fields. Even these precautions do not eliminate hidden self-citations due to collaborations with other universities, but they are the best that we can do with the present information.

The top entry for each science displays within-field citations, potential papers citing, and potential papers cited. This is followed by similar statistics on cross-field interactions. For example, biology interacts with agriculture, chemistry, earth sciences, medicine, and psychology. By and large the cross-field interactions in table 2 seem reasonable. For example, biology and medicine are significantly linked through cross-citation and chemistry and physics are similarly linked. While this information confirms expectations as to the structure of the sciences, it also shows selectivity of cross-field interactions. Each of the 12 fields can interact with any of the others, yielding a total of  $12 \times 11 = 132$  possible interactions, and yet only 32 or about one-fourth are significant.

Not surprisingly, citations within fields are more common than between-field citations. This is partly because of our field definitions, which sweep up sub-fields into large aggregates. But it represents as well greater scientific influence within disciplines, as explained in sections II and IIIA. This discussion neglects differences in the size of cross- and within field interactions. Comparing the highest cross-field citation count to the within field count shows that astronomy, mathematics and statistics, and physics are almost autonomous from the rest of science. But the life science fields, agriculture, biology, medicine, and psychology display strong cross-field dependencies.

Table 3 reports mean citation probabilities. We report means, standard deviations, maxima, and minima to show variations in the probabilities. On average within-field probabilities are 10 to 100 times greater than cross-field probabilities. In small fields such as astronomy, computer science, earth sciences, and economics, within field probabilities exceed the corresponding probabilities in large fields like biology, chemistry, and medicine. Following sections II and IIIA, our explanation for this is that research tends to span a larger fraction of the smaller fields, holding constant the difficulty of searching the literature.

Figure 1 explores lags in citation. The figure graphs the mean citation probability by citation lag. Since the data cover 19 years, the maximum lag is 18 years. Four curves are represented in the figure. Higher curves represent actual and fitted citation probabilities within fields by citation lag. As in other applications, both curves are hump-shaped and positively skewed. The fitted curve declines at a faster rate because it controls for the higher propensity to cite in recent years, which dominates longer lags. The lower curves show the result of including actual and fitted citation probabilities between fields.

### **C. Citations and the Quality of Scientific Institutions**

As noted, we distinguish high, medium, and low rank-stratification classes in part of the analysis. The expanded regressions require nine parameters for different combinations of citing and cited rank-stratification class. Four classes would imply 16 parameters; five classes 25 parameters, and so on. Allowing for three classes is a compromise that limits the parameters that are handled in the estimation procedure while still allowing estimates of rank-stratification effects.

We classify schools into rank-stratification classes as follows. First, we use the 1993 National Research Council (NRC) department chair rankings of graduate programs (National Research Council, 1995) to assign quality ranks to schools. NRC collects rankings in these 10 broad sciences: astronomy, biology, chemistry, computer science, earth sciences, economics and business; engineering, mathematics and statistics; physics, and psychology. NRC does not rank agriculture and medicine. As a very imperfect substitute we use federal R&D of the top 110 schools in 1998 to rank agriculture and medicine. The strength of the 1993 NRC rankings is their emphasis on quality of program rather than quantity of funding. Use of federal R&D to rank agriculture and medicine weakens the link between quality and citations. And yet peer rankings in agriculture and medicine are lacking.

The number of ranked graduate programs varies with the size of disciplines. The scale of engineering, biology, and medicine leads us to break out 75 schools in these fields with the rest of the schools treated as a residual institution. In the case of eight fields—agriculture, chemistry, computer science, earth sciences, economics and business; mathematics and statistics; physics, and psychology—we consider 50 separate schools, with the remainder again treated as a residual<sup>14</sup>. In the case of astronomy, where there are few ranked programs we break out 25 schools and treat the rest as a residual. The size of residual “institutions” treated in this way is the same as an average ranked program.

We classify the rank of a school as high if it falls in the top 20% of ranked schools in a field, as medium if it falls in the middle 40% and as low if it falls in the bottom 40% (including the remainder)<sup>15</sup>. In this way we construct the rank-stratification classes used in the study. Next we calculate the number of citations, the number of potentially citing papers, and the number of potentially cited papers for every citing and cited field, every rank-stratification class combination, and every citing and cited pair of years.

Figure 2 sums up the evidence on citation interactions between rank-stratification classes, averaged across the 12 sciences. Clearly the citation probability from the top 20% to the middle 40% lies below the probability from the middle 40% to the top 20%. To see this, compare the middle bar in the leftmost group to the first bar in the middle group. The same holds true of other groups. Scientific influence is top-down rather than bottom-up, though less highly ranked schools do influence those higher ranked. In addition the figure shows that the probability of citation among top 20% schools exceeds the probability among the middle 40%. And the probability of citation among the middle 40% again exceeds the probability among the bottom 40%. By this measure, knowledge flows among peer schools increase with rank. To an extent this counters the leveling effect of the top-down asymmetry of knowledge flows.

Figures 3 and 4 illustrate field differences in the rank-stratification effects. Figure 3 contains findings for engineering, where these effects are small. Thus in figure 3 the citation probability varies by a factor of 1.75:1 across classes and this citation gradient is smaller than the average for the 10 fields that are ranked by quality of graduate program.

---

<sup>14</sup> There are 48 formally recognized schools of agriculture. All other research in agriculture derives from researchers scattered through related disciplines.

<sup>15</sup> The percentage ranking implies that there are 15 institutions in the top 20% of engineering, biology, and medicine. There are 10 schools in the top 20% of agriculture, chemistry, computer science, earth sciences, economics and business; mathematics and statistics; physics, and psychology. The top 20% of astronomy consists of five schools. In this way scale conditions size of the rank-stratification classes.

Results for economics and business are shown in figure 4. Rank-stratification effects are large: the citation probability varies by 4.25:1 across classes, and the gradient is the steepest of any field. The difference in figures 3 and 4 is likely due to greater inequality of program quality in economics than in engineering<sup>16</sup>. The crossing of the citation curves in figure 4 also suggests some separation of research in the bottom 40% of economics from the rest of the field. This is because the bottom 40% cites itself at a higher rate than it is cited by higher-ranked departments in this field.

## V. Findings

### A. Citation Function Estimates

We turn next to econometric results from fitting the citation function to the data. Table 4 reports estimates of the baseline function (5). We begin by discussing the within- and between-field intercept parameters. Recall that the parameters are normalized by chemistry. All the within-field parameters differ significantly from zero. There is considerable variation in the within-field rate of citation, from a low 0.234 in engineering to a high 13.346 in astronomy, and this is a 50 to 1 range in citation probabilities. These differences are highly significant compared with the baseline value of 1.0.

Even the leading cross-field parameters are considerably smaller than the within field parameters. Indeed, citation parameters within fields are typically 10 to 100 times larger than the cross-field parameters, though the cross-field parameters only include those that are near or above the 5% level of significance<sup>17</sup>.

Fields vary in the extent to which they cite other fields. Judging by the ratio of the leading cross-field parameter to the within-field parameter, the following fields—agriculture, biology, engineering, and medicine—are strongly dependent on other sciences. This shows up in the closeness of agriculture and biology, biology and medicine, engineering and computer science, and medicine and biology, where the cross-field parameter is 1/5 or more the size of the within-field parameter. Some of the cross-field effects may reflect the difficulty of drawing distinctions between fields, but some are real. In contrast, mathematics and statistics shows no significant dependence on other disciplines.

---

<sup>16</sup> The much higher citation probability in economics and business is due in part to the smaller size of the economics and business literature. This issue is discussed in sections II and III.

<sup>17</sup> The results reported in the table are about the same whether interactions that are insignificantly different from zero are included or not in the estimation procedure and thus whether the data cells on which such estimates are based are or are not included. This implies that insignificant cells add very little information.

Turning to year effects, we find that these are U-shaped and reach a minimum in 1991. This pattern controls for citing period effects, which drop slightly during the late 1980s and increase thereafter. The cited year parameters seem to measure vintage effects, although the most recent papers have not had the same opportunity to be cited as earlier papers. The upward drift in citation over time is shown by the rising parameters over more recent citing intervals.

This discussion of table 4 concludes with the decay and diffusion parameters. Differentiation of the citation function (5) shows that the reciprocal of the baseline diffusion parameter  $\beta_1$  (chemistry) times the diffusion parameter for each science  $\beta_{1i}$  yields the modal lag, or the lag at which citations peak:

$$(7) \quad L_{Modal,i} = 1 / \beta_1 \beta_{1i}.$$

To prove this, take the derivative of the citation function, set it equal to zero, and solve for  $L_{Modal,i}$ . The modal lag is a measure of the speed of diffusion. Table 4 shows that this lag ranges from 1.75 years in physics to 4.2 years in computer science. Average modal lags between citing and cited publication years are about two years less than lags for patented technology (Jaffe and Trajtenberg, 1999) based on citing and cited grant years. Since publication and grant years are to a degree analogues, one interpretation of this difference could be that Open Science institutions accelerate the spread of knowledge through society compared with proprietary technology<sup>18</sup>.

The peak citation probability is approximately equal to

$$(8) \quad P_i^{\max} = \beta_2 / \beta_1 \beta_{1i}$$

For chemistry (where  $\beta_{1i} = 1$ ) the peak citation probability is approximately  $2.0 \times 10^{-4}$ . This is roughly 100 times larger than the peak citation probability for patent citations, which is about  $1.5 \times 10^{-6}$ , indicating the greater volume of science citations compared with patent citations<sup>19</sup>.

Table 5 reports estimates of the rank-stratification effects within sciences that follow the expanded citation function (6). Reporting in table 5 is limited to rank-stratification parameters, since the other estimates are closely similar to those of table 4.

---

<sup>18</sup> See David (1998) for an account of the creation of Open Science institutions as a result of competing patronage arrangements, and of the importance of these institutions for the rise of science in Europe.

<sup>19</sup> This comparison draws on Appendix B of Jaffe and Trajtenberg (1999), which reports a baseline value for  $\beta_1$  of 0.190, and a value for  $\beta_2$  of  $0.289 \times 10^{-6}$ .

Results are as follows. First, the diagonal parameters generally confirm the view that institutions in the top 20% are more likely to cite other top 20% institutions than middle 40% institutions are to cite each other; and these are more likely to each other than bottom 40% institutions are. For example, the diagonal parameters in computer science for the top 20%, middle 40%, and bottom 40% are 1.490, 1.176, and 0.712. The same pattern holds everywhere but agriculture. Citation probabilities increase with rank of institutions, suggesting that knowledge spillovers among peer institutions rise with quality.

A second feature is the tendency for less prestigious schools to cite those more prestigious at a higher rate than the reverse. The implication is that scientific influence is primarily from the top-down. Chemistry provides an example. The parameter for the top 20% citing the middle 40% is 0.700, while the parameter for the middle 40% citing the top 20% is 0.924. The difference turns out to be highly significant. The same pattern holds in other comparisons within chemistry and most other disciplines and implies that scientific influence increases with quality of program. The main exception again is agriculture.

## **B. Conditional Symmetry Tests**

The regression findings of tables 4 and 5 suggest opportunities to test for symmetry of the citation parameters across fields and rank-stratification classes. Table 6 summarizes the results. Note that these are pair-wise tests of symmetry of the parameters, which we shall call conditional symmetry tests, because they assume statistically significant two-way citation between pairs of fields or classes.

The first line of table 6 reports tests of equality of the cross-field citation parameters in table 4. For example, does the rate at which medicine cites biology differ from the rate at which biology cites medicine? The first line provides a summary of answers to this question, if citation takes place in both directions. The answer is that cross-citation effects are symmetric. The null hypothesis of equality of the parameters is usually accepted. One exception is economics, which cites mathematics and statistics more than the reverse at the 1% level of significance. A second case is physics, which cites astronomy at a higher rate than the reverse at the 2% level of significance. Two other cross-citation parameters are unbalanced and are missed by the above evaluation: agriculture cites earth sciences and astronomy cites biology, but neither is cited in return. Besides this, the conditional symmetry tests miss asymmetries in scientific influence that occur in the main body of scientific papers. Many papers use older mathematics and statistics but feedback effects from these fields to modern mathematics and statistics are less common.

Our method ignores such hidden asymmetries, which require encoding of article content that is beyond current frontiers of bibliometrics.

The second line of table 6 tests for differences in the probability of citation by rank-stratification class and refers to table 5. Equality and symmetry are rejected in most cases, with agriculture the main exception. Findings for computer science illustrate the result. The top 20% institutions in computer science cite each other more than the middle or bottom 40% institutions do, and the differences are significant at the 1% level. The third line tests for asymmetries in top-down versus bottom-up citation. Symmetry in the cross-effects is rejected at the 1% level in most cases, with agriculture and medicine the exceptions<sup>20</sup>. Thus asymmetry of the rank-stratification effects in table 6 is accepted most of the time.

## VI. Conclusion

This paper has described scientific influence among U.S. universities measured by means of citations to scientific papers. One finding is that scientific influence mostly occurs within fields. Another is that cross-field interactions are selective: significant cross-field interactions represent less than a quarter of potential cases. Collectively this suggests that academic knowledge flows are bounded by scientific distance, much as industrial knowledge flows seem to be hemmed in by technological distance (Adams and Jaffe, 1996).

We also find that two-way interactions between fields are typically symmetric, so that field A cites field B as often as B cites A. Still, we are convinced that hidden asymmetries are present in field-to-field interactions. This is because applied fields cite other fields more often than more basic fields cite, and because of deeply buried content in applied science papers that uses and interprets basic science materials, while the reverse is not true.

In addition, knowledge seems to diffuse rapidly, and may turn out to diffuse more rapidly within science than technology does within industry. Scientific paper citations are also more abundant than patent citations judged by citation frequency. Our evidence confirms that rank of field in a university is correlated with scientific influence. Tests of the null hypothesis that higher ranked university-fields are more often cited than those lower-ranked are accepted in 30 out of 36 cases. We test whether interactions with peer

---

<sup>20</sup> The reappearance of agriculture and medicine on the list of exceptions calls for an explanation. In part the pattern recurs because we rank programs according to quantity of federal R&D rather than quality of graduate program. But also the two fields may show greater equality than most other sciences. This tendency seems to hold for engineering, where rank-stratification classes follow the NRC rankings.

institutions increase with rank. The answer is again yes, in 30 out of 36 cases. This implies that surrounding programs reinforce research in a given program more as rank increases.

The work presented here is but one ingredient in a full-fledged knowledge production for the academic sector. The production process would explain papers and patents of universities and it would include as explanatory variables knowledge spillovers as well as current and past contributions from a university's own research<sup>21</sup>. Of course, pursuit of this agenda requires information on different channels of interaction among universities and fields, as well as the spillback of a university-field's own past research, besides its current research support.

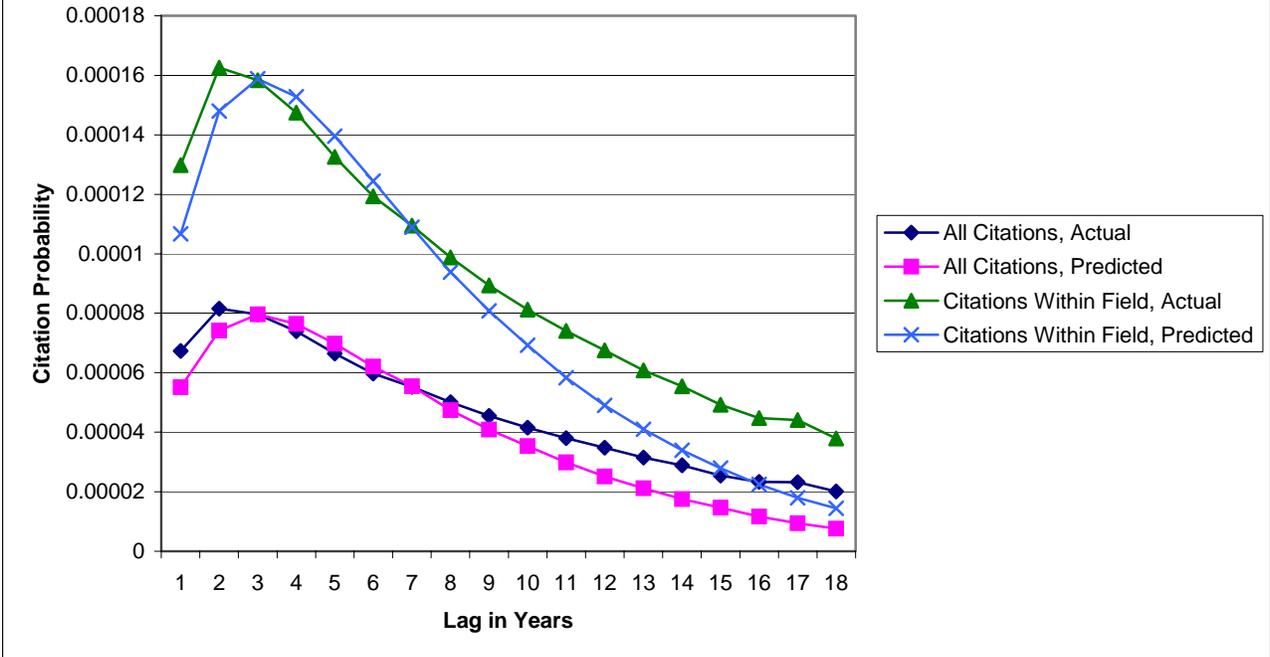
Still further ahead our goal is to extend this methodology to consider effects on firm papers and patents of multi-dimensional spillovers from universities to firms and from firms to firms, in addition to the role of firm's own research efforts in the determination of innovative success<sup>22</sup>. The resulting edifice of the knowledge production function in industry is itself an ingredient, though a crucial one, in the economics of growth and technical change.

---

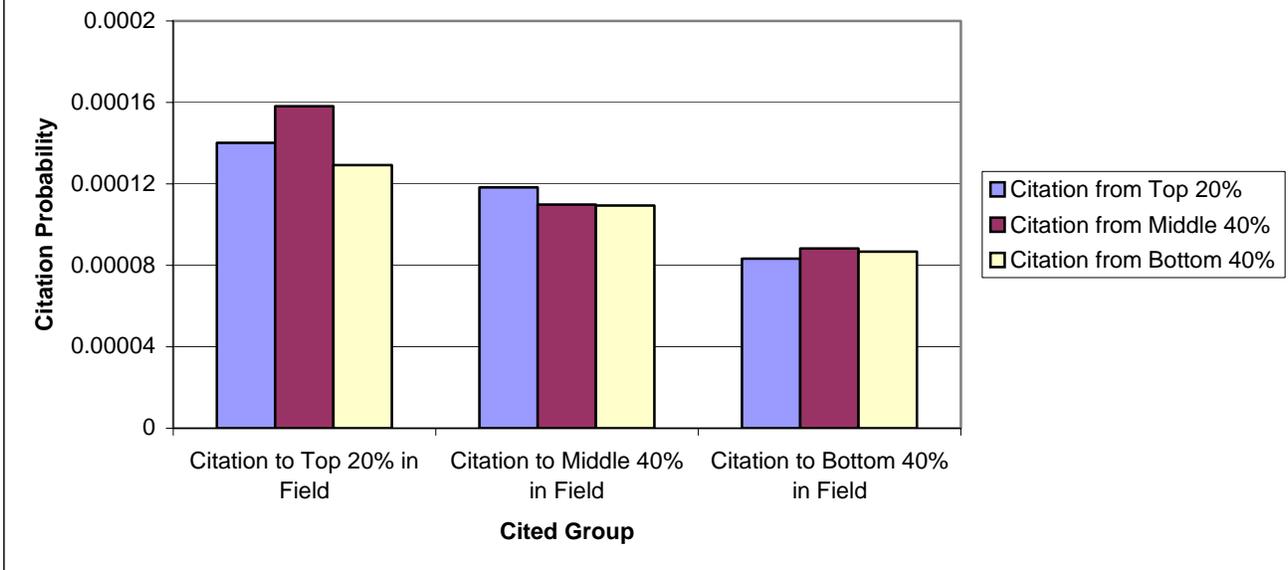
<sup>21</sup> Adams and Griliches (1998) studied production of academic research for samples of university-fields, but without knowledge spillovers. Their findings suggest that production obeyed constant returns at the aggregate level, but decreasing returns at the individual level. This may follow from knowledge externalities, or another factor operating more strongly at the individual level, such as errors in variables.

<sup>22</sup> Popp (2002) represents an approach to this question.

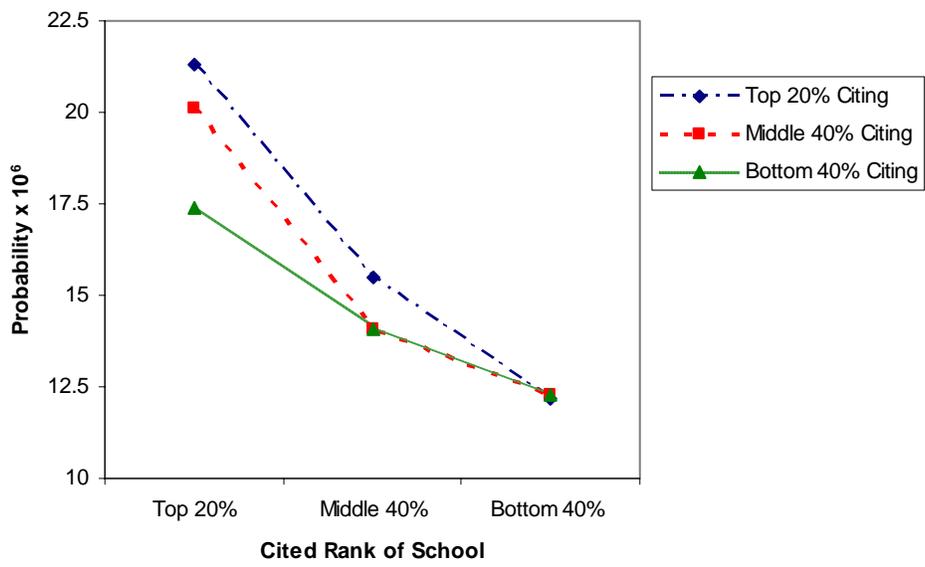
**Figure 1--Mean Citation Probabilities by Citation Lag**



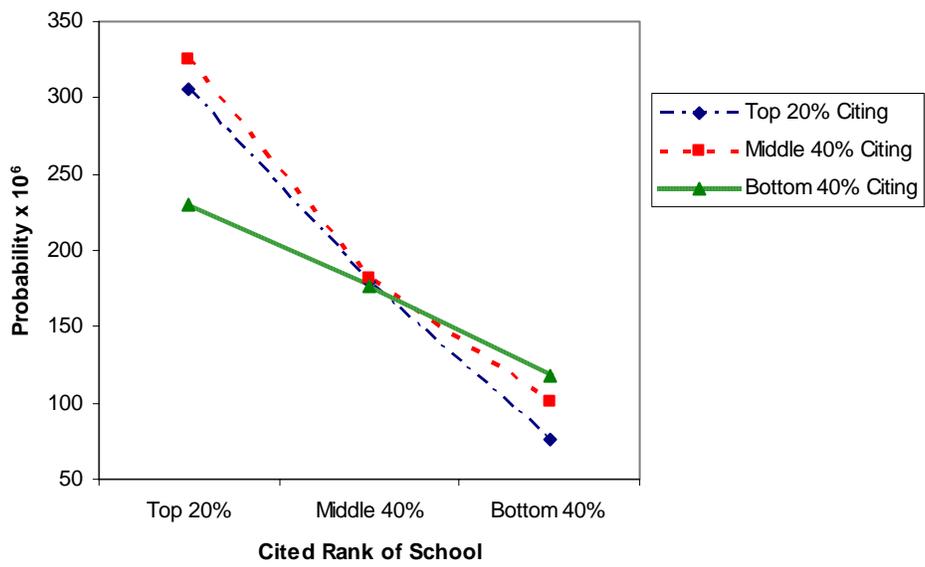
**Figure 2--Mean Citation Probabilities Within Fields Between Schools of Different Rank**



**Figure 3--Probability of Citation by Citing and Cited Rank Of School, Engineering**



**Figure 4--Probability of Citation by Citing and Cited Rank Of School, Economics and Business**



**Table 1**  
**Definition, Size, and Composition of 12 Main Science Fields**  
**Practiced in the Top 110 U.S. Universities, 1981-1999**

<b>Main Science Field</b>	<b>Total Papers</b> (% of Total Papers) [Total Citations Received] {% of Total Citations Received}	<b>Sub-Field Composition of Main Science Field</b>
Agriculture	189,740 (7.8%) [730,777] {3.9%}	General agriculture and agronomy; aquatic sciences; animal sciences; plant sciences; agricultural chemistry; entomology and pest control; food science and nutrition; veterinary medicine and animal health
Astronomy	35,795 (1.5%) [371,982] {2.0%}	Astronomy and astrophysics
Biology	639,195 (26.3%) [8,339,862] {44.4%}	General biological sciences; biochemistry and biophysics; cell and developmental biology; ecology and environment; molecular biology and genetics; biotechnology and applied microbiology; microbiology; experimental biology; immunology; neurosciences and behavior; pharmacology and toxicology; physiology; oncogenesis and cancer research
Chemistry	195,437 (8.0%) [1,371,491] {7.3%}	General chemistry; analytical chemistry; inorganic and nuclear chemistry; organic chemistry and polymer science; physical chemistry and chemical physics; spectroscopy, instrumentation, and analytical science
Computer Science	28,184 (1.2%) [76,424] {0.4%}	Computer science and engineering; information technology and communications systems
Earth Sciences	73,126 (3.0%) [566,280] {3.0%}	Atmospheric sciences; geology and other earth sciences; geological, petroleum, and mining engineering; oceanography
Economics and Business	43,892 (1.8%) [161,813] {0.9%}	Economics; accounting; decision and information sciences; finance, insurance, and real estate; management; marketing
Engineering	170,569 (7.0%) [467,955] {2.5%}	Aeronautical engineering; biomedical engineering; chemical engineering; civil engineering; electrical and electronics engineering; engineering mathematics; environmental engineering and energy; industrial engineering, materials science; mechanical engineering; metallurgy; nuclear engineering

**Table 1**  
**Definition, Size, and Composition of 12 Main Science Fields**  
**Practiced in the Top 110 U.S. Universities, 1981-1999**

Main Science Field	Total Papers (% of Total Papers) [Total Citations Received] {% of Total Citations Received}	Sub-Field Composition of Main Science Field
Mathematics and Statistics	61,061 (2.5%) [187,484] {1.0%}	Mathematics; biostatistics and statistics
Medicine	659,000 (27.1%) [4,563,261] {24.3%}	General and internal medicine; anesthesia and intensive care; cardiovascular and hematology research; cardiovascular and respiratory systems; clinical immunology and infectious disease; clinical psychology and psychiatry; dentistry and oral surgery; dermatology; endocrinology, metabolism, and nutrition; environmental medicine and public health; gastroenterology and hepatology; health care sciences and services; hematology; medical research, diagnosis, and treatment; medical research, general topics; medical research, organs and systems; neurology; oncology; ophthalmology; orthopedics, rehabilitation, and sports medicine; otolaryngology; pediatrics; radiology, nuclear medicine, and imaging; reproductive medicine; research, laboratory medicine, and medical technology; rheumatology; surgery; urology and nephrology
Physics	217,026 (8.9%) [1,219,080] {6.5%}	General physics; applied physics, condensed matter, and materials science; optics and acoustics
Psychology	116,976 (4.8%) [727,673] {3.9%}	Psychology and psychiatry

Notes: Citations received derive from top 110 universities during the period 1981-1999. They are not a census of citations received from all scientific institutions or all papers in the future. Citations can be from any field to any field among the sciences listed in the table. The total number of papers is 2,430,001. The total number of citations received is 18,784,082.

**Table 2**  
**Mean Citations and Papers by Citing and Cited Field of Science,**  
**The Top 110 U.S. Universities, 1981-1999**

Citing Field	Cited Field	Citations	Potential Papers Citing	Potential Papers Cited
Agriculture	Agriculture	2,543	11,326	10,671
“	Biology	1,843	“	34,411
“	Earth Sciences	106	“	3,979
Astronomy	Astronomy	3,218	2,879	2,127
“	Biology	212	“	34,411
“	Earth Sciences	123	“	3,979
“	Physics	118	“	11,747
Biology	Biology	40,349	44,135	34,411
“	Agriculture	905	“	10,671
“	Chemistry	625	“	10,035
“	Earth Sciences	351	“	3,979
“	Medicine	6,454	“	36,725
“	Psychology	530	“	7,007
Chemistry	Chemistry	4,989	12,166	10,035
“	Biology	1,101	“	34,411
“	Physics	492	“	11,747
Computer Science	Computer Science	326	2,031	1,410
“	Mathematics & Statistics	28	“	3,572
“	Engineering	81	“	8,205
Earth Sciences	Earth Sciences	3,324	5,312	3,979
“	Astronomy	113	“	2,127
“	Biology	668	“	34,411
Economics & Business	Economics & Business	1,315	2,966	2,632
“	Mathematics & Statistics	153	“	3,572
“	Psychology	54	“	7,007
Engineering	Engineering	1,501	11,434	8,205
“	Computer Science	133	“	1,410
“	Mathematics & Statistics	99	“	3,572
“	Physics	328	“	11,747
Mathematics & Statistics	Mathematics & Statistics	828	3,975	3,572
“	Computer Science	19	“	1,410
“	Economics & Business	32	“	2,632
“	Engineering	48	“	8,205
Medicine	Medicine	26,714	45,734	36,725
“	Biology	9,764	“	34,411
“	Psychology	737	“	7,007

**Table 2**  
**Mean Citations and Papers by Citing and Cited Field of Science,**  
**The Top 110 U.S. Universities, 1981-1999**

Citing Field	Cited Field	Citations	Potential Papers Citing	Potential Papers Cited
Physics	Physics	13,561	16,272	11,747
“	Astronomy	174	“	2,127
“	Chemistry	315	“	10,035
“	Engineering	226	“	8,205
Psychology	Psychology	4,399	7,804	7,007
“	Biology	555	“	34,411
“	Economics & Business	50	“	2,632
“	Medicine	789	“	36,725

Notes: Entries are means over as many as 171 citing and cited year pairs for each citing and cited field combination, where the lags range from one to eighteen years. The statistics are based on 36,834 cells that report number of citations and numbers of potentially citing and cited papers, classified by citing and cited groups and years. Self-citations within a field and citations between fields in the same university are excluded from this analysis.

**Table 3**  
**Moments of the Citation Probabilities, By Citing and Cited Field**  
**Papers and Citations of the Top 110 U.S. Universities, 1981-1999**

<b>Citing Field</b>	<b>Cited Field</b>	<b>Mean</b>	<b>S.D.</b>	<b>Min</b>	<b>Max</b>
<b>(All Entries in Units of 10<sup>-6</sup> a )</b>					
Agriculture	Agriculture	20.7	6.6	8.6	38.7
“	Biology	4.6	1.2	2.3	7.1
“	Earth Sciences	2.3	0.6	0.8	4.0
Astronomy	Astronomy	526.0	226.3	108.2	1081.0
“	Biology	2.2	1.7	0.3	11.2
“	Earth Sciences	11.8	7.1	2.1	40.9
“	Physics	3.6	2.2	0.6	14.7
Biology	Biology	25.7	11.9	5.0	44.6
“	Agriculture	1.9	0.7	0.7	3.3
“	Chemistry	1.4	0.5	0.5	2.5
“	Earth Sciences	1.9	0.7	0.6	3.4
“	Medicine	3.9	1.7	0.9	7.0
“	Psychology	1.7	0.6	0.7	3.7
Chemistry	Chemistry	40.9	16.5	11.1	84.1
“	Biology	2.5	1.0	0.8	5.1
“	Physics	3.4	1.2	1.1	5.8
Computer Science	Computer Science	125.5	52.6	41.0	302.3
“	Mathematics & Statistics	4.0	2.1	0.3	11.5
“	Engineering	4.9	2.3	0.7	12.9
Earth Sciences	Earth Sciences	159.1	56.9	67.8	416.6
“	Astronomy	10.3	5.8	2.3	42.1
“	Biology	3.5	1.5	1.4	10.4
Economics & Business	Economics & Business	165.1	47.4	80.2	262.4
“	Mathematics & Statistics	15.1	7.6	1.6	39.0
“	Psychology	2.7	1.2	0.2	6.0
Engineering	Engineering	15.9	5.0	6.1	27.3
“	Computer Science	8.6	3.0	2.0	17.7
“	Mathematics & Statistics	2.5	0.7	0.7	4.5
“	Physics	2.4	0.8	0.8	4.8
Mathematics & Statistics	Mathematics & Statistics	58.2	13.6	31.3	86.8
“	Computer Science	3.6	1.7	0.4	10.2
“	Economics & Business	3.3	1.9	0.2	11.0
“	Engineering	1.5	0.6	0.3	3.9
Medicine	Medicine	15.5	6.1	4.1	25.8
“	Biology	5.9	2.5	1.2	10.5
“	Psychology	2.3	0.7	0.9	4.3

**Table 3**  
**Moments of the Citation Probabilities, By Citing and Cited Field**  
**Papers and Citations of the Top 110 U.S. Universities, 1981-1999**

Citing Field	Cited Field	Mean	S.D.	Min	Max
(All Entries in Units of $10^{-6}$ <sup>a</sup> )					
Physics	Physics	60.2	49.0	10.0	238.0
“	Astronomy	5.0	3.2	0.5	14.1
“	Chemistry	2.0	0.9	0.5	4.5
“	Engineering	1.6	0.7	0.4	3.5
Psychology	Psychology	79.6	23.9	30.7	145.5
“	Biology	2.0	0.6	0.8	3.5
“	Economics & Business	2.5	1.8	0.1	11.6
“	Medicine	2.7	0.9	1.0	4.9

Notes. The entries are means over as many as 171 citing and cited year pairs for each of the citing and cited field combinations. The calculations are based on 36,834 citing and cited field, rank-stratification class, and year observations. See the text for more discussion. <sup>a</sup> The statement that all entries are in units of  $10^{-6}$  means that 20.7 is  $20.7 \times 10^{-6}$ , 6.6 is  $6.6 \times 10^{-6}$ , and likewise for all the other entries in the table.

**Table 4**  
**Baseline Citation Function, with Cross-Field Effects**  
**The Top 110 U.S. Universities, 1981-1999**

Variable or Statistic		Regression Parameter	Asymptotic Standard Error	Asymptotic t-Statistic, $H_0=0$	Asymptotic t-Statistic, $H_0=1$
<b>Field Intercepts (<math>\alpha_i</math> s)</b>					
Citing Field	Cited Field				
Agriculture	Agriculture	0.334	0.020	16.7	-33.3
“	Biology	0.073	0.012	6.1	-77.3
“	Earth Sciences	0.036	0.019	1.9	-50.7
Astronomy	Astronomy	13.346	0.337	39.6	36.6
“	Biology	0.057	0.024	2.4	-39.3
“	Earth Sciences	0.265	0.042	6.3	-17.5
“	Physics	0.086	0.031	2.8	-29.5
Biology	Biology	0.702	0.023	30.5	-13.0
“	Agriculture	0.048	0.017	2.8	-56.0
“	Chemistry	0.035	0.017	2.1	-56.8
“	Earth Sciences	0.048	0.021	2.3	-45.3
“	Medicine	0.102	0.013	7.8	-69.1
“	Psychology	0.041	0.019	2.2	-50.5
Chemistry	Chemistry	1.000	--	--	--
“	Biology	0.058	0.016	3.6	-58.9
“	Physics	0.076	0.020	3.8	-46.2
Computer Science	Computer Science	1.616	0.056	28.9	11.0
“	Engineering	0.065	0.021	3.1	-44.5
“	Mathematics & Statistics	0.053	0.026	2.0	-36.4
Earth Sciences	Earth Sciences	2.929	0.079	37.1	24.4
“	Astronomy	0.186	0.031	6.0	-26.3
“	Biology	0.066	0.015	4.4	-62.3
Economics & Business	Economics & Business	2.358	0.066	35.7	20.6
“	Mathematics & Statistics	0.190	0.024	7.9	-40.9
“	Psychology	0.035	0.019	1.8	-50.8
Engineering	Engineering	0.234	0.018	13.0	-42.6
“	Computer Science	0.124	0.025	5.0	-35.0
“	Mathematics & Statistics	0.036	0.019	1.9	-50.7
“	Physics	0.035	0.014	2.5	-68.9
Mathematics & Statistics	Mathematics & Statistics	0.867	0.035	24.8	-3.8
“	Computer Science	0.049	0.029	1.7	-32.8
“	Economics & Business	0.047	0.025	1.9	-38.1
“	Engineering	0.021	0.019	1.1	-51.5

**Table 4**  
**Baseline Citation Function, with Cross-Field Effects**  
**The Top 110 U.S. Universities, 1981-1999**

Variable or Statistic		Regression Parameter	Asymptotic Standard Error	Asymptotic t-Statistic, $H_0=0$	Asymptotic t-Statistic, $H_0=1$
<b>Field Intercepts (<math>\alpha_i</math> s)</b>					
Citing Field	Cited Field				
Medicine	Medicine	0.324	0.014	23.1	-48.3
“	Biology	0.126	0.011	11.5	-79.5
“	Psychology	0.045	0.015	3.0	-63.7
Physics	Physics	3.414	0.096	35.6	25.1
“	Astronomy	0.239	0.059	4.1	-12.9
“	Chemistry	0.086	0.040	2.2	-22.9
“	Engineering	0.069	0.041	1.7	-22.7
Psychology	Psychology	1.137	0.034	33.4	4.0
“	Biology	0.028	0.011	2.5	-88.4
“	Economics & Business	0.034	0.020	1.7	-48.3
“	Medicine	0.037	0.010	3.7	-96.3
<b>Cited Year Effects</b>					
1981		1.000	--	--	--
1982		1.012	0.009	112.4	1.3
1983		1.031	0.010	103.1	3.1
1984		1.027	0.011	93.4	2.5
1985		0.993	0.011	90.3	-0.6
1986		0.956	0.011	86.9	-4.0
1987		0.862	0.011	78.4	-12.5
1988		0.798	0.011	72.5	-18.4
1989		0.761	0.011	69.2	-21.7
1990		0.739	0.012	61.6	-21.8
1991		0.722	0.012	60.2	-23.2
1992		0.775	0.013	59.6	-17.3
1993		0.725	0.013	55.8	-21.2
1994		0.744	0.015	49.6	-17.1
1995		0.766	0.016	47.9	-14.6
1996		0.822	0.018	45.7	-9.9
1997		0.832	0.019	43.8	-8.8
1998		1.110	0.028	39.6	3.9
<b>Citing Interval Effects</b>					
1981-1985		1.000	--	--	--
1986-1990		0.925	0.008	115.6	-9.4
1991-1995		1.070	0.015	71.3	4.7
1996-1998		1.160	0.022	52.7	7.3

**Table 4**  
**Baseline Citation Function, with Cross-Field Effects**  
**The Top 110 U.S. Universities, 1981-1999**

Variable or Statistic	Regression Parameter	Asymptotic Standard Error	Asymptotic t-Statistic, $H_0=0$	Asymptotic t-Statistic, $H_0=1$
<b>Decay Parameter (<math>\beta_1</math>)</b>	0.353	0.006	58.8	--
<b>Diffusion Parameter (<math>\beta_2</math>)</b>	$7.2 \times 10^{-5}$	$1.86 \times 10^{-6}$	387.1	--
<b>Field Decay Parameters (<math>\beta_{1i}</math> s)</b>				
Agriculture	0.778	0.029	26.8	-7.7
Astronomy	1.044	0.016	65.3	2.8
Biology	1.068	0.021	50.9	3.2
Chemistry	1.000	--	--	--
Computer Science	0.675	0.015	45.0	-21.7
Earth Sciences	0.849	0.014	60.6	-10.8
Economics & Business	0.679	0.012	56.6	-26.8
Engineering	0.738	0.037	19.9	-7.1
Mathematics & Statistics	0.716	0.018	39.8	-15.8
Medicine	0.917	0.024	38.2	-3.5
Physics	1.623	0.028	58.0	22.3
Psychology	0.691	0.013	53.2	-23.8

Notes. The number of cells, classified by citing and cited fields and years, is 36,834. The adjusted  $R^2=0.900$  and the standard error of the regression (the root mean squared error) is 0.0013. Citations from the same university are treated as self-citations and hence are excluded from the equation. Reported cross-field citation parameters are at or near the margin of significance for a test of  $H_0=0$ .

**Table 5**  
**Citation Function: Effects of Rank Stratification-Class**  
**The Top 110 Universities, 1981-1999**  
**(Asymptotic Standard Errors in Parentheses)**

Field and Citing Rank- Stratification Class	Cited Rank-Stratification Class		
	Top 20%	Middle 40%	Bottom 40%
<b>Agriculture</b>			
Top 20%	0.198 (0.017)	0.224 (0.016)	0.198 (0.017)
Middle 40%	0.244 (0.017)	0.232 (0.016)	0.203 (0.017)
Bottom 40%	0.195 (0.017)	0.184 (0.016)	0.326 (0.022)
<b>Astronomy</b>			
Top 20%	8.733 (0.263)	9.753 (0.291)	7.969 (0.239)
Middle 40%	11.254 (0.335)	8.133 (0.244)	7.902 (0.236)
Bottom 40%	9.893 (0.295)	8.852 (0.264)	7.723 (0.231)
<b>Biology</b>			
Top 20%	0.867 (0.031)	0.528 (0.021)	0.303 (0.015)
Middle 40%	0.733 (0.026)	0.451 (0.018)	0.307 (0.015)
Bottom 40%	0.531 (0.021)	0.395 (0.017)	0.275 (0.014)
<b>Chemistry</b>			
Top 20%	1.000 (--)	0.700 (0.028)	0.516 (0.023)
Middle 40%	0.924 (0.032)	0.660 (0.026)	0.519 (0.022)
Bottom 40%	0.809 (0.028)	0.623 (0.023)	0.490 (0.019)
<b>Computer Science</b>			
Top 20%	1.490 (0.059)	1.079 (0.047)	0.598 (0.035)
Middle 40%	1.636 (0.063)	1.176 (0.049)	0.739 (0.037)
Bottom 40%	1.280 (0.051)	1.066 (0.045)	0.712 (0.035)
<b>Earth Science</b>			
Top 20%	2.685 (0.086)	2.093 (0.068)	1.671 (0.056)
Middle 40%	2.556 (0.081)	1.852 (0.061)	1.608 (0.053)
Bottom 40%	2.140 (0.069)	1.801 (0.059)	1.519 (0.050)
<b>Economics and Business</b>			
Top 20%	2.693 (0.086)	1.572 (0.054)	0.678 (0.030)

**Table 5**  
**Citation Function: Effects of Rank Stratification-Class**  
**The Top 110 Universities, 1981-1999**  
**(Asymptotic Standard Errors in Parentheses)**

Field and Citing Rank-Stratification Class	Cited Rank-Stratification Class		
	Top 20%	Middle 40%	Bottom 40%
Economics and Business (Cont.)			
Middle 40%	2.853 (0.090)	1.597 (0.054)	0.897 (0.035)
Bottom 40%	2.001 (0.065)	1.528 (0.051)	1.046 (0.037)
Engineering			
Top 20%	0.211 (0.018)	0.156 (0.016)	0.122 (0.016)
Middle 40%	0.198 (0.017)	0.140 (0.015)	0.122 (0.015)
Bottom 40%	0.173 (0.017)	0.141 (0.016)	0.124 (0.016)
Mathematics and Statistics			
Top 20%	0.857 (0.040)	0.637 (0.033)	0.365 (0.025)
Middle 40%	0.910 (0.040)	0.575 (0.029)	0.382 (0.023)
Bottom 40%	0.677 (0.032)	0.543 (0.027)	0.395 (0.022)
Medicine			
Top 20%	0.248 (0.013)	0.226 (0.012)	0.180 (0.011)
Middle 40%	0.252 (0.013)	0.211 (0.011)	0.187 (0.011)
Bottom 40%	0.222 (0.013)	0.206 (0.012)	0.187 (0.012)
Physics			
Top 20%	2.184 (0.077)	2.258 (0.078)	1.572 (0.059)
Middle 40%	2.634 (0.089)	2.840 (0.094)	2.253 (0.076)
Bottom 40%	2.120 (0.073)	2.622 (0.087)	1.994 (0.068)
Psychology			
Top 20%	0.929 (0.034)	0.806 (0.031)	0.658 (0.025)
Middle 40%	0.996 (0.036)	0.755 (0.029)	0.634 (0.024)
Bottom 40%	0.880 (0.031)	0.748 (0.027)	0.610 (0.022)

Notes. The number of citing and cited group and year observations is 36,834. The adjusted  $R^2=0.938$  and the standard error of the regression (root mean squared error) is 0.0010. \* The t-statistic is reported for the null hypothesis  $H_0=0$ . \*\* The t-statistic is reported for the null hypothesis that  $H_1=1$ . Citations from the same university are treated as self-citations and hence are excluded from the equation. The regression includes all the cross-field citation parameters, cited year effects, and citing year interval effects of Table 5.

**Table 6**  
**Conditional Symmetry Tests of the Citation Function**  
**Top 110 Universities, 1981-1999**

Test	Null Hypothesis	Purpose	Summary	Exceptions
Equality of Between-Field Citation Parameters	$\alpha_{ij} = \alpha_{ji}$	Check for asymmetries in the direction of citation between fields $i$ and $j$	Equality is accepted by 13 of 15 tests at the 5% level of significance	Economics and Business cites Mathematics and Statistics more than the reverse ( $\chi^2=16.7$ , $P<0.0001$ ); Physics cites Astronomy more than the reverse ( $\chi^2=5.1$ , $P=0.0240$ )
Equality of Within-Field, Within Rank-Stratification Class Parameters	$\alpha_{i,kk} = \alpha_{i,ll}$	Check for asymmetries in citation within quality groups $k$ and $l$ within field $i$	Equality is rejected by 33 of 36 tests at the 1% level of significance. Citation increases with quality of institution in 30 of 36 tests	Top 20% of Agriculture is cited less than the bottom 40% ( $\chi^2=33.7$ , $P<0.0001$ ); middle 40% is cited less than the bottom 40% ( $\chi^2=21.5$ , $P<0.0001$ ). Top 20% of Physics is cited less than the middle 40% ( $\chi^2=139.7$ , $P<0.0001$ );
Equality of Within-Field, Between Rank-Stratification Class Parameters	$\alpha_{i,kl} = \alpha_{i,lk}$	Check for asymmetries in citation across quality groups $k$ and $l$ and within field $i$	Equality is rejected by 30 of 36 tests at the 1% level of significance. Citation is less top to bottom than it is bottom to top in 30 of 36 tests	All tests accept equality in agriculture. Equality between the top 20% and middle 40% of engineering is accepted at the 1% but not 2% levels ( $\chi^2=5.1$ , $P=0.0237$ ). Equality between the top 20% and middle 40% of medicine is accepted at the 1% but not 3% levels ( $\chi^2=4.6$ , $P=0.0317$ ). Equality between the middle 40% and bottom 40% of medicine is accepted.

Notes. All  $\chi^2$  tests are Wald Tests that evaluate the difference in the parameters from zero evaluated at the unrestricted likelihood function.

## References

- Adams, James D., "Fundamental Stocks of Knowledge and Productivity Growth," **Journal of Political Economy** 98 (August 1990): 673-702.
- \_\_\_\_\_, and Adam B. Jaffe, "Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data," **RAND Journal of Economics** 27 (Winter 1996): 700-721.
- \_\_\_\_\_, and Zvi Griliches, "Research Productivity in a System of Universities," **Annales D'Economie et de Statistique** 49/50 (1998): 127-162.
- Audretsch, David, and Paula E Stephan, "Company-Scientist Locational Links: The Case of Biotechnology," **American Economic Review** 86 (June 1996): 641-652.
- Branstetter, Lee, "Measuring the Impact of Academic Science on Industrial Innovation: The Case of California's Research Universities," Working Paper, Columbia University, New York, NY: August 2003.
- Cohen, Wesley M., Richard R. Nelson, and John P. Walsh, "Links and Impacts: The Influence of Public Research on Industrial R&D," **Management Science** 48 (January 2002): 1-23.
- Darby, Michael R., and Lynne G. Zucker, "Grilichesian Breakthroughs: Inventions of Methods of Inventing and Firm Entry in Nanotechnology," NBER Working Paper 9825, Cambridge, MA: July 2003.
- David, Paul A., "Common Agency Contracting and the Emergence of 'Open Science' Institutions," **American Economic Review, Papers and Proceedings** 88 (May 1998): 15-21.
- De Solla Price, Derek J., "Networks of Scientific Papers," **Science**, New Series, 149 (July 30, 1965): 510-515.
- \_\_\_\_\_, **Little Science, Big Science...and Beyond**, New York: Columbia University Press, 1986.
- Griliches, Zvi, "Issues in Assessing the Contribution of Research and Development to Productivity Growth," **Bell Journal of Economics** 10 (Spring 1979): 92-116.
- \_\_\_\_\_, "The Search for R&D Spillovers," **Scandinavian Journal of Economics** 94 (Supplement, 1992): S29-S47.

- Evenson, Robert E., and Yoav Kislev, **Agricultural Research and Productivity**. New Haven, Connecticut: Yale University Press, 1975.
- Garfield, Eugene, "Citation Analysis as a Tool in Journal Evaluation," **Science**, New Series 178 (November 3, 1973): 471-479.
- Jaffe, Adam B., "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," **American Economic Review** 75 (December 1986): 984-1002.
- \_\_\_\_\_, Michael S. Fogarty, and Bruce R. Banks, "Evidence from Patents and Patent Citations on The Impact of NASA and Other Federal Labs on Commercial Innovation," **Journal of Industrial Economics** 96 (1998): 183-205.
- \_\_\_\_\_, and Manuel Trajtenberg, "International Knowledge Flows: Evidence from Patent Citations," **Economics of Innovation and New Technology** 8 (1999): 105-136.
- Jensen, Richard, and Marie Thursby, "Proofs and Prototypes for Sale: The Licensing of University Inventions," **American Economic Review** 91 (March 2001): 240-259.
- Jones, Charles I., "R&D-Based Models of Economic Growth," **Journal of Political Economy** 103 (August 1995): 759-784.
- \_\_\_\_\_, "Sources of Economic Growth in a World of Ideas," **American Economic Review** 92 (March 2002): 220-239.
- Marshall, Alfred, **Principles of Economics**, 8<sup>th</sup> Edition. London, Macmillan, 1920
- Merton, Robert K., **The Sociology of Science**. Chicago, Illinois: University of Chicago Press, 1973.
- National Research Council, **Research-Doctorate Programs: Continuity and Change**, Washington, DC: National Academy Press, 1995.
- Narin, F., and Kimberly S. Hamilton, "Bibliometric Performance Measures," **Scientometrics** 36 (1996): 293-310.
- \_\_\_\_\_, Kimberly S. Hamilton, and Dominic Olivastro, "The Increasing Linkage between U.S. Technology and Public Science," **Research Policy** 26 (1997): 317-330.
- Popp, David, "Induced Innovation and Energy Prices," **American Economic Review** 92 (March 2002): 160-180.

Romer, Paul M., "Endogenous Technological Change," **Journal of Political Economy** 98 (October 1990, Part 2): S71-S102.

Scherer, F. Michael, "Interindustry Technology Flows in the United States," **Research Policy** 11 (August 1982a): 227-245.

\_\_\_\_\_, "Interindustry Technology Flows and Productivity Growth," **The Review of Economics and Statistics** 64 (November 1982b): 627-634.

Trajtenberg, Manuel, "A Penny for your Quotes: Patent Citations and the Value of Innovations," **RAND Journal of Economics** 21 (1990): 172-187.

Van Raan, A. F. J., "Fractal Dimensions of Co-Citations," **Nature** 347 (October 18, 1990): 626.

Zucker, Lynne G., Michael R. Darby, and Marilyn B. Brewer, "Intellectual Human Capital and the Birth of Biotechnology Enterprises," **American Economic Review** 88 (March 1998): 290-306.

Zuckerman, Harriet A., **Scientific Elite: Nobel Laureates in the United States**. New York: The Free Press, 1977.