Measuring Scientific Influence

By

James D. Adams, Rensselaer Polytechnic Institute and NBER
Roger Clemmons, University of Florida
And
Paula E. Stephan, Georgia State University

July 2003

Preliminary First Draft
Comments Welcome

Abstract

This paper measures scientific influence by means of citations to scientific papers. Our analysis considers the top 110 research universities over the period 1981-1999 that account for the majority of academic research in the U.S. The analysis takes into consideration 12 main fields of science that cover nearly all research in these institutions. The data set derives from the Institute for Scientific Information (ISI) and consists of 2.4 million scientific papers and 18.5 million citations to these papers by top 110 schools.

We find that citation channels of scientific influence are well defined and occur primarily within fields, and that cross-field citation channels are highly selective. Cross-field channels appear to be symmetric between pairs of fields. Scientific influence is primarily from top-ranked institutions to those less highly ranked, though there is a significant influence of lesser institutions on those that are higher-ranked. In addition we find evidence suggesting that quality of university-fields is more strongly reinforced by surrounding institutions in the case of top-ranked university-fields. Overall the results suggest that knowledge spillovers are important in academia, but are tightly circumscribed by field and intrinsic relevance. This suggests another limit on the returns to scale in the knowledge production function, in this case for basic research.

Keywords: Academic Science, R&D, Papers, Citations, and Spillovers
JEL Codes: L30, O30

Corresponding Author: James D. Adams, Department of Economics, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590. Telephone: 1-518-276-2523, fax: 1-518-276-2235. E-mail until August 10, 2003: James.D.Adams@notes.cba.ufl.edu. After August 10, 2003: E-mail to be assigned.

*Paper prepared for the International Conference in 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. We thank Nancy Bayers and Henry Small of the Institute for Scientific Information (ISI) for suggestions and clarifications concerning the data and Adam Jaffe for assistance with the estimation procedures.

I. Introduction

This paper is part of a larger project, now reaching its maturity, which has engaged much of our attention in recent years[1]. In the early going the goals of the project are to describe basic research interactions among the top 110 U.S. universities, among the top 200 U.S. R&D firms in the late 1990s, and between the universities and firms. The time period of our investigation is 1981-1999; its scope includes all of science, and the data cover almost 3 million scientific papers and more than 20 million citations to these papers. The focus of our work to date is not on patented technology, but rather on pre-technology basic science, though clearly the two can and do overlap.

The present work provides detailed evidence on interactions among the 110 universities[2]. Its purpose is to trace scientific influence across institutions and fields of science. On this occasion space limitations confine our measurements to scientific influence as captured by a single indicator, citations to scientific papers from other scientific papers. We assume that papers represent amounts of new knowledge produced, albeit variable amounts, and that, on the whole, citations to the papers indicate scientific influence. We restrict the investigation to article citations even though other paths of influence can easily be imagined, not the least of which are close collaboration and mobility of graduate students from one school to another.

We are interested in questions having to do with scientific influence of the following kind. What is the comparative influence of research in one field on research in another? How important are cross-field influences compared with influence within a field? Are cross-field influences asymmetric, so that as between a pair of fields the more cited might be regarded as the more basic, in the sense of the one more relied upon? What role does quality of graduate teaching and research play in scientific influence? Could quality of a program interact with other high quality programs to reinforce excellence in science, as if to demonstrate an institutional Matthew Effect?[3]

---

[1] During the planning phases of this project we had the indisputable advantage as well as the undeniable privilege of working with Zvi Griliches. We deeply and profoundly regret that he did not live to see this work come to fruition.

[2] See the Appendix, table A-1 for the list of universities.

[3] The Matthew Effect says that past individual success in science leads to additional support, which in turn leads to future success. Thus scientific achievement is cumulative. See Merton (1973) and Zuckerman (1977) for more on this topic, including the intergenerational aspects of scientific excellence.

This paper has been strongly influenced by a larger literature. Studies of wheat and maize and other crops by Evenson and Kislev (1975) point out limits to the usefulness of research aimed at one climatic region for productivity in other climatic regions. In this way their work suggests geographic limitations on the diffusion of research. Scherer (1982a, b) uses input-output tables based on patent citations to map interindustry flows of technology. By and large he finds that R&D used, as defined by the flow-through of product R&D from all industries to the using industry plus process R&D in the same industry, outperforms R&D from the industry of origin. The implication is that pecuniary or technological externalities of R&D proceed along well-defined channels between industries. Griliches (1979, 1992), besides providing a guide to the locus and kind of externality associated with R&D, also provides an overview of the literature of knowledge externalities and how one might go about measuring them. He argues that knowledge externalities are more powerful than pecuniary externalities because of non-rivalry in the use of knowledge. But he emphasizes that knowledge externalities associated with R&D decrease with technological distance.

Our measurements of scientific influence have been strongly affected by Jaffe (1986), which derives and implements the un-centered correlation between the distributions of two entities' R&D among patent classes as a concrete measure of the distance between these entities, as discussed by Griliches (1979, 1992). This paper likewise owes much to the citations literature, including Trajtenberg (1990) and especially Jaffe and Trajtenberg (1999).

Adams (1990) shows that the influence of stocks of scientific papers that are relevant to particular industries on productivity growth of other industries are limited by technological distance using the cosine measure of Jaffe (1986). Adams and Jaffe (1996), in a plant-level study based on Census data, finds that the influence of R&D in the rest of the firm is not only limited by technological distance between firms but also by the technological distance within firms.

The common thread that runs through these papers is that technology flows and knowledge externalities are tightly restricted by the limited relevance of the majority of knowledge generated in society, so that increasing returns are correspondingly limited. The tenor of the findings reported in this paper as regards knowledge creation in academia are similar, in that knowledge flows proceed very selectively from one institution and science to another. In this sense the knowledge production function for basic research that might be conceived of, based on the information presented here, while it would undoubtedly provide evidence of significant knowledge flows between academic entities, would also suggest that the outcome of most academic R&D is tightly circumscribed. Since academic science replenishes research opportunities in industry (and conversely), this result has a bearing on growth models, as represented for example by Romer (1990), and Jones (1995, 2002). In the evolution of these models, scale effects in the knowledge production function have been steadily curtailed over time. In this way opportunities for perpetual growth have increasingly been viewed as deriving from growth of R&D rather than its level, supported by a contribution from knowledge spillovers that is sufficient to avoid diminishing returns to research. However, models of economic growth continue to assume that knowledge flows freely throughout entire economies and even the entire world[4]. Our findings confirm that knowledge flows are extensive within the university system, but are still limited by the narrow applicability of most discoveries, as captured by citation frequency. And yet the overall picture conveyed by the results is one of a vibrant, interactive system.

The main findings are as follows. First, citation within fields occurs at rates that are an order of magnitude greater than citation between fields. This is so even if one stacks the deck in favor of those select cross-field interactions that are significantly different from zero. Second, the number of significant cross-field interactions is less than one-fourth of the potential number. Since we classify the data as falling into 12 fields, there are $11\times12=132$ potential cross-field interactions. Roughly 30 of these differ from zero at the 5% level or better. Together these observations suggest that knowledge flows are indeed bounded by scientific distance, just as there is evidence that knowledge flows are hemmed in by technological distance in industry (Adams and Jaffe, 1996).

Third, tests of the symmetry of cross-citation accept the null hypothesis in most cases. The rate at which biology cites medicine is not significantly different from the rate at which medicine cites biology. One exception is that economics cites mathematics and statistics at a higher rate than mathematics and statistics cites economics. Another is that physics cites astronomy more than the reverse. However, we

[4] Marshall (1920, p. 220) is notably circumspect in observing 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." What is noteworthy about this is that he does not say that the externality *has to depend* on the volume of production in the whole civilized world, but that it may, depending on the facts of the case.

remain convinced that asymmetries in field-to-field interactions do exist, but that they take a different form. We find that applied life sciences exhibit thicker cross-citation trails than fields that are more basic, such as mathematics and statistics. Indeed, mathematics and statistics receives citations from other fields, but fails to cite any other field at a significant rate.

Fourth, compared with patent citations, we find that knowledge diffuses more rapidly within science as evidenced by science citations, than within technology as measured by patent citations. The modal lag of scientific citations, a measure of the speed of diffusion, is slightly more than three years, compared with a modal lag of more than five years in patent citations (Jaffe and Trajtenberg, 1999).

Fifth, our evidence confirms the role of stratification in scientific research. We find that quality of a program does increase scientific influence. We carry out symmetry tests of this proposition by checking whether higher ranked university-fields are more often cited than those lower-ranked. In 30 out of 36 comparisons the answer is yes, with most of the exceptions in agriculture. Finally, we test whether interactions in citation with peer institutions increase with quality of program. In other words, do university-fields in the top 20% cite one another than university-fields in the bottom 40% cite one another? The answer again is yes, in 30 out of 36 cases. If knowledge spillovers result from citation then this suggests that reinforcement of research by surrounding peer programs increases with quality.

The rest of this paper consists of three sections. Section II explains the database and provides an extensive description of various cuts of the data that turn out to be important by means of tables and figures. Section III reports the econometric estimates and carries out tests of symmetry and equality of citation propensities within and across fields. A brief conclusion is contained in section IV.

II. Description of the Data

The data consist of 2.4 million publications of the top 110 U.S. universities from 1981-1999 and 18.7 million citations to those papers from the same period. The 110 universities account for the majority of U.S. academic R&D. The schools are documented in Table A-1 of the appendix. The data source is the Institute for Scientific Information (ISI) in Philadelphia, Pennsylvania. The data cover all articles, reviews, notes, and proceedings, or the standard set of communications, in 12 fields of science that account for most of university research. The 12 main 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 7137 scientific journals. All are assumed to belong to a single dominant field where the journal is assigned[5]. This assignment is often accurate for journals that are designed to reach a specialized audience, despite the difficulty of drawing hard and fast distinctions between fields. But the method is bound to produce errors for general journals of science, where a dominant field is assumed despite the inclusion of many other fields. The main alternative to the journal assignment method, though, is the dismal one of assigning papers according to academic "departments" of the authors, and this is ruled out by incompleteness of the information[6]. Notice that, unlike patent data, multiple field assignments are almost non-existent in the scientific papers data. In order to consistently carry out such a multiple assignment, clear and unambiguous criteria would be required, and the right to carry out the assignments would have to be vested in a Scientific Papers Office that is a precise analogue to the Patent Office.

It is critically important to see that these data, while voluminous, are still only a window on scientific research. The data are truncated on the left and right in time, they are geographically restricted to research having at least one author in a U.S. university; and industry, government, and non-profit sectors, where scientific research also takes place, are all left out. Thus for example, we lack information on citations made to papers in the late 1990s, since these citations for the most part have not yet been made. In fact we know almost nothing about papers that influenced research in the early 1980s since citations to these early papers are excluded from the database[7]. Citations made by U.S. researchers to foreign literature are excluded, and likewise citations received by U.S. researchers from foreign literature are excluded. The rich interactions of the international scientific enterprise are thus left out of our analysis. But one has to begin somewhere.

---

[5] In the case of the 5,507 journals that are currently published, we follow the assignment of journals to fields practiced by the Institute for Scientific Information, relying on ISI 's experience to provide a more accurate assignment. In the case of 1,630 journals that are formerly published we rely on the field assignments of CHI, Computer Horizons Inc. The argument is the same, that the experience of an established firm in bibliometrics is likely to be more accurate than an idiosyncratic assignment by the principal investigators.

[6] As an experiment we set out to assign all the papers of Harvard University to one of the 12 main science fields in our data using purely address information. About one-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 in the future to codify the origins of papers by field.

[7] We hope that in the future ISI will extend the database of papers and citations backward to the 1960s, in the interest of supporting bibliometric investigation of science over the long run.

Table 1 provides a look at the field dimensions of the data. Main fields are in the first column. The second column reports total papers, percent of all papers, total citations received, and percent of citations received. The third column reports composition of main fields by component sub-fields. In addition to variation in the size and complexity of the fields the table illustrates differences in citation practices. Biology ranks second in publications but dominates all other fields in citations received. Computer science ranks last in both papers and citations. It is not clear what to make of these differences. The size of citing populations differs among fields and is larger in biology than computer science, suggesting that size should be considered before drawing conclusions from the raw data.

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

$$p_{iTjt} = \frac{c_{iTjt}}{n_{iT} n_{jt}} \tag{1}$$

where i and j are citing and cited groups and T and t are citing and cited years (T>t). The term c_{iTjt} in the numerator is the number of citations from group i in year T to group j in year t. In reality the citations are the number of papers in (i, T) and (j, t) that are actually linked by citation. The terms n_{iT} and n_{jt} in the denominator of (1) are the number of potentially citing and potentially cited papers in (i, T) and (j, t) that could be linked. Thus equation (1) is bounded by 0 and 1 and has a probability interpretation. If not one paper in (i, T) cited a single paper in (j, t) then (1) would equal zero. If instead every paper in (i, T) cited every paper in (j, t) then (1) would equal 1.0. In reality (1) is much closer to zero than to one.

In the empirical work, citing and cited fields as well as citing and cited years define the citation probabilities and provide the data cells noted in (1). In addition, citations within field are distinguished according to a quality dimension that ranks institutions as high, medium, or low within a given discipline. For any i-j combination there are 171 possible citing and cited year combinations, assuming a citation lag of at least one year[8]. Of course not all of the i-j combinations are significant in the statistical sense, or even take place. Table 2 reports means of citations, potential papers citing, and potential papers cited

[8] 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 process continues for 18 years until 1982, when only 1981 papers can be cited. The sum of the series is S=18+17+16+…+1, or S=(19×18)/2=171.

across up to 171 citing and cited year combinations, but only within fields and for interactions between fields that prove to be statistically significant. In this sense table 2 looks forward to the regression findings and by this method manages the length of the table. Even with these restrictions, though, the table lists 32 interactions between fields.

Unlike table 1, table 2 and all that follow exclude self-citation and even citation from same university. Self-citation reflects the influence of a university-field's own past research rather than external scientific influence. Exclusion of same-university citations between different fields is motivated by the same concerns. Since papers are classified by journal assignment, the same authors in the same schools can write in multiple fields and cite themselves. Even these precautions do not eliminate hidden self-citation to collaborations with other universities. But it is the best that can be done with the information that we have.

For each science the top entry shows citations and papers citing and cited within field, followed by cross-field interactions in alphabetic order. For example, the table shows that agriculture has two significant cross-field interactions, with biology and earth sciences. On the contrary, biology interacts with five others: agriculture, chemistry, earth sciences, medicine, and psychology, and is in this sense a leader in cross-field interactions. By and large the significant cross-field interactions in table 2 are consistent with expectations. For example, biology and medicine, chemistry and physics are significantly linked through cross-citation. But this information, while it confirms expectations as to the structure of the sciences, also shows that interactions between fields are selective. To see this, note that each of the fields can interact with any of the remaining 11, yielding a total of $12 \times 11 = 132$ possible interactions. And yet only a fourth of these (32 of 132) are even close to significant.

Table 2 shows that mean citations within fields are an order of magnitude greater than mean citations between fields, even restricting this comparison to significant cross-field interactions. In some sense this results from the broad definitions of the fields shown in table 1. But the higher rate of interaction within fields is real in that it represents greater scientific influence within disciplines. It thus represents "localization" in scientific space. This overall description glosses over differences in the size of cross- and within field interactions. If one compares highest cross-field citation counts to the within field counts, then astronomy, mathematics and statistics, and physics seem to be autonomous from the rest of science. By the same criterion the life science fields of agriculture, biology, medicine, and psychology display strong interdependence. We revisit this topic below using mean citation probabilities and find broadly similar results.

Table 2 suggests that simple measures of citations received or even citations per paper are incomplete measures of scientific influence, because they do not consider the counterfactual of all the citations that might have taken place but did not. Put another way, very high citation counts within field, where most citations take place, are strongly correlated with the number of potentially citing papers, which in turn bring down the probability that a paper is actually cited by taking into account potential citation. This observation underscores the value of the citation probability (1) as a criterion for judging whether citation rates are high or low.

Table 3 reports means of the citation probabilities (1) for the groups shown in table 2. We provide various moments of the probabilities across the 171 citing and cited year combinations to provide a sense of the variation in the probabilities. Consistent with what has gone before, within field probabilities are an order of magnitude greater than the cross-field probabilities. The large within field probabilities in small fields such as astronomy, computer science, earth sciences, and economics and business are somewhat startling. Perhaps the skills of a given team of researchers more readily cover these fields. This suggests that time costs associated with the accumulation of knowledge *permit* a higher citation rate in smaller fields. But since the probability of citation is not especially high within mathematics and statistics, another small field, this explanation has a problem. Perhaps *certain* small fields are more homogeneous or less costly, but mathematics and statistics does not meet this additional criterion.

In some of our analysis we distinguish citing and cited cells within fields, in which schools are ranked high, medium, and low, and form "rank-stratification" classes. Since the regression analysis in below allows for a full set of interactions among rank-stratification classes, there are already nine interactions for each of the sciences. Four classes imply 16 interactions per science; five classes 25 interactions, and so on. Allowing for three classes is a compromise position. It permits us to study the role played by institutional quality without generating huge numbers of parameters for the estimation procedure.

We classify the quality of schools as follows. First we use the 1993 National Research Council (NRC) peer rankings of graduate programs (National Research Council, 1995) to rank the ten sciences (out of 12) that are included in the NRC data. The ten are astronomy, biology, chemistry, computer science, earth sciences, economics and business; engineering, mathematics and statistics; physics, and psychology. The remaining two fields, agriculture and medicine, are not ranked by NRC. As an imperfect substitute we rank institutions in agriculture and medicine by means of their 1998 federal R&D. Since a key strength of the 1993 NRC rankings is their emphasis on quality of programs rather than quantity of funding, use of federal R&D to rank agriculture and medicine could have the effect of concealing the link between quality and citations. We are not aware of peer rankings in agriculture and medicine that are comparable to the NRC rankings and we have little choice but to follow this alternative procedure in these cases.

The size of disciplines varies markedly (see table 1) and the number of ranked graduate and professional programs varies accordingly. The large scale of the biomedical enterprise leads us to break out 75 schools within biology and medicine, with the remainder of the 110 treated as a residual institution of the lowest rank. For nine other fields—agriculture, chemistry, computer science, earth sciences, economics and business; engineering, mathematics and statistics; physics, and psychology—we consider 50 separate schools, with the remainder again treated as a residual. In the case of astronomy, where there are many fewer ranked programs than the other fields, we break out 25 schools, treating the remainder as a residual. The size of residual "institutions" treated in this way is about the same as that of a single ranked university-field.

We classify a school as high quality if it falls in the top 20% of schools in a field, as medium if it falls in the next 40%, and as low quality if it falls in the bottom 40%, including the remainder. 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 field, every rank-stratification class combination, and every citing and cited year pair. As we have seen there are nine possible citing and cited rank-stratification class combinations for every field and citing and cited year combination.

Table 4 brings together mean citations, papers citing and cited, and the citation probability for citing and cited rank-stratification classes within fields of science. As before we take averages across citing and cited years. We arrange this information in a 3×3 matrix, with citing classes forming the rows and cited classes forming the columns. Thus the first and row and column for each field refers to top 20% schools citing each other, the second element in the first row to top 20% schools citing middle 40%

schools, and the third element to top 20% schools citing bottom 40% schools. Likewise the second and third rows represent the citation behavior of the middle 40% and bottom 40% schools to the top 20%, middle 40%, and bottom 40% of institutions in that field.

The top line of each entry reports mean citations, papers citing, and papers cited. Thus, the top 20% schools in chemistry make an average of 322 citations to other top 20% schools for each citing and cited year combination. Also, an average of 12,056 top 20% chemistry papers could cite; and allowing for lags in citation, an average of 9,430 papers top 20% chemistry papers could be cited. The second line of each entry reports the mean citation probability in parentheses. Thus the average citation probability with which top 20% schools in chemistry cite each other is 47.2×10^{-6}. The other entries have a similar interpretation.

While thus arranged the data can seem overwhelming, several useful observations emerge from the table. First, the probability that the top 20% cites the middle 40% and the bottom 40% is nearly always less than the probability that the middle 40% or the bottom 40% cite the top 20%. Likewise the probability that the middle 40% cite the bottom 40% is less than the probability that the bottom 40% cites the middle 40%. To see this, compare the off-diagonal elements in the upper triangle of each 3×3 matrix with their counterparts in the lower triangle. One observes that scientific influence is more top-down than it is bottom-up, although bottom-up influence is clear. The only exception to this pattern is agriculture[9].

Another important feature is that top 20% institutions usually cite each more often than the middle 40% institutions cite each other. In turn the middle 40% tend to cite each other more often than the bottom 40%, and so on. Out of 36 comparisons across the 12 sciences only six exceptions are recorded. This suggests that interactions among the top 20% institutions are typically more intense than among the middle 40%, and that interactions among the middle 40% are more intense than among the bottom 40%. This is an earmark of differences in research activity among schools. But in addition, on a knowledge

[9] Besides the fact that the method of ranking schools of agriculture is based on the size of federal R&D, which this field shares with medicine, agriculture has one other characteristic that makes it unique among the sciences in the U.S. This is the fact that all the schools of agriculture are located in state universities that were founded under the Morrill Act of 1862. The Hatch Act of 1887, founding the state agricultural experiment stations, further enhanced the standing of these schools. Huffman and Evenson (1993), Chapter 1 contains an illuminating discussion of the pressure groups that mobilized to bring about this result. research. For our purposes what this means is that there are no private universities among ranked schools of agriculture.

spillovers interpretation, it suggests that the top institutions are a greater source of support for one another than institutions of lesser rank. In raising this possibility, notice that citation practices are being held constant at the field level.

A third feature of the table that stands out is the difference among fields in top-down asymmetries of citation. We refer to the fact that schools of a lower rank cite schools of a higher rank at a higher rate than the reverse. This pattern is non-existent in agriculture, moderate in engineering and medicine, but is pronounced in computer science and economics. These differences probably have to do with the varying depth of science departments. In some sciences high quality programs are widely dispersed, such as chemistry and medicine, while in others, like economics, a few top programs dominate the field.

Figures 1 and 2 provide a graphical summary. Figure 1 includes line graphs of the average probability of citation by lag between citing and cited years, where the average is taken across all 12 fields of science. The top two lines are graphs of citations within fields. The actual data on citations within science peak at around two years and decline thereafter, reflecting a familiar diffusion-decay process. The fitted data peak later but also decline faster, reflecting the adjustment for citing and cited year effects. Peaks in the within-field citation rates occur much earlier than for patents, whose citation probability crests at something more than five years (Jaffe and Trajtenberg, 1999). The lower two lines in the figure consist of all-inclusive citations showing the average probability of citation within and between fields. Again the fitted curve peaks later but falls faster. At all lags the all-inclusive probability of citation is about ½ the within-field probability, but peaks at two to three years as before.

Figure 2 summarizes in one graph what we have learned about the role of rank-stratification class. The figure shows for example that the probability of citation from the top 20% to the middle 40% is less than the probability of citation 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 figure also shows that the probability of citation from the bottom 40% to the top 20% is greater than the converse probability, and that the probability of citation is greater from the bottom 40% to the middle 40% than the converse. Scientific influence is more often from the top-down than it is from the bottom-up, even though there is an appreciable influence of less highly ranked schools. Finally, figure 2 again shows that the probability of citation among the top 20% schools exceeds the probability among middle 40% schools. The probability of citation among the middle

40% is again higher than the probability of among the bottom 40%. By this measure knowledge flows increase with quality of university-fields, which could make already strong institutions still stronger.

III. Citation Function Analysis

A. Estimation Procedure

In this section we model the citation probability as a function of citing and cited sciences, years, and the quality of scientific institutions, and we estimate this function using papers and citations of the top 110 universities. To this end we adapt a procedure developed by Jaffe and Trajtenberg (1999) for the purpose of explaining patterns of patent citation. The application of this approach to scientific paper citations is based upon the findings just presented. There we found sizable differences in own and cross-field citation and we found evidence within each science of top-down asymmetries that give a predominant role to higher quality institutions. As usual we found that citation reaches an early peak and then declines, with a long tail off to the right as the lag between citing and cited scientific papers increases. In this section we model these effects using nonlinear regression[10]. Throughout we make the assumption that citations and the field, year, and institutional effects that drive them indicate greater scientific influence[11].

The baseline citation function used in this paper takes the form:

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

In (2) α_{ij} captures the average probability that field i cites field j, α_T is the average probability that a citation is made in period T, and α_t is the average probability that a citation is received in period t. In the case of a field citing itself, the parameter α_{ij} becomes α_{ii}. In our data i and j run from 1 to 12, representing the 12 fields of science in table 1. One important point is that the probability parameters are only defined relative to a baseline value. We chose to normalize the α_{ij} by the value for chemistry citing itself, whose transformed value is accordingly set equal to 1.0. Likewise, the various α_T and α_t parameters are

[10] We thank Adam Jaffe for providing us with nonlinear regression programs that we modified for present purposes.

[11] 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.

normalized by the earliest periods citing and cited, whose transformed values are accordingly set equal to 1.0. We experienced difficulties obtaining convergence during estimation when we specified a full set of citing and cited years. For this reason we aggregate the citing years into the four periods: 1981-1985, 1986-1990, 1991-1995, and 1996-1999. Thus T and α_T refer to the four time intervals, while t and α_t refer to single years cited.

The β_1 parameter is a parameter standing for the rate of decay in citation for the baseline field, which we again take to be chemistry, while the β_{1j} parameters are decay parameters relative to chemistry. The parameter β_2 governs overall diffusion as captured by citation. Since β_2 positions the overall rate of citation this parameter is not identified by field independently of the α_{ij} vector. Finally, the error term is u_{iTjt}. The estimates for citation function (2) are reported in table 5 below.

In addition to (2) we consider a more elaborate specification of the citation function. This version allows for differences in same-field citation parameters by quality of scientific institution. We turn to this more elaborate specification in view of the evidence in table 4. That suggests that lower-ranked institutions cite those that are higher ranked more than these higher-ranked institutions cite them. We also found confirmation that on the whole, higher-ranked institutions cite each other more than lower-ranked institutions cite each other. To allow for these effects, we replace the within-field parameter α_{ii} in (2) with the following 3×3 matrix of citation possibilities:

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

The leading subscript i refers to field, while the trailing subscripts 1, 2, and 3 refer to the top 20%, the middle 40%, and the bottom 40% of institutions. As in (2) the parameters are identified up to a baseline parameter. Here we choose to normalize the parameters by the top 20% institutions in chemistry citing each other. Thus $\alpha_{4,11}$=1 in (3), so that all the other within-field parameters, as well as the between-field parameters α_{ij}, are defined relative to $\alpha_{4,11}$.

The parameters in the first row of (3) represent the probabilities that top 20% schools cite themselves, the middle 40%, and the bottom 40% institutions, in that order. The parameters in the second row capture citations of the middle 40% to themselves and other rank-stratification classes, and those of the third row spell out citations of the bottom 40%, both to themselves and to other groups. The matrix of citation possibilities allows us to take account of quality effects within fields, where most citations occur[12]. Moreover, these quality effects are statistically significant. We report estimates of the expanded citation function, which incorporates (3), in table 6 below.

Construction of the citation probabilities is our first priority. We begin by grouping the data on citations, potentially citing papers, and potentially cited papers into cells. The definitions of the cells are as follows. We define nine cells that correspond to (3) within fields. The between field cells consist of the interactions of fields with the other 11 fields, at least where these exist[13]. Thus there are 9×12=108 within-field cells and up to 11×12=132 cells between fields, and therefore up to 240 cells for every citing and cited year combination. Since there are (19×18)/2=171 potential citing and cited year combinations, the number of cells that could exist is 240×171=41,040. But 4,206 or about 10% of the cells are missing. This occurs mostly because particular citing and cited year pairs are missing in cases of rare cross-field citations. As a result, the actual number of cells in the data set is 41,040-4,206= 36,834.

B. Regression Findings

The next step is to fit the citation probabilities to field and year dummies and the lag in citation. Table 5 reports estimates of the baseline function. We begin by discussing within- and between-field parameters in the intercept of (2). All the within parameters differ significantly from zero and are significantly larger than the leading cross-field parameters. Thus citation decreases with scientific distance. There is also considerable variation in the rate of citation, from a low of 0.234 in engineering to a high of 13.346 in astronomy, these differences are significant compared with the baseline value of 1.0 for chemistry. The findings indicate the more than 50 to 1 range in citation probabilities across disciplines.

As noted in the descriptive section, citation parameters within fields are typically an order of magnitude larger than the cross-field parameters, even though the cross-field parameters reported in table 5

[12] Citations between fields are typically not so thick on the ground as to permit rank-stratification effects for cross-field citations.
[13] Since we the ignore rank-stratification effects in table 5, in that table we average over the nine probabilities for each science that correspond to all possible interactions between the different rank-stratification classes in that science.

include only those that are near or above a cutoff representing the 5% level of significance[14]. Cross-field citation is highly selective.

Another result is that 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—may be seen as strongly dependent on other sciences. This appears in the close connection between agriculture and biology, biology and medicine, engineering and computer science, and medicine and biology. In all four cases the cross-field parameter is 1/5 or more the size of the within-field parameter. Some of the cross-field effects may in part reflect the difficulty of drawing distinctions between fields that study similar processes. In contrast, mathematics and statistics shows no significant dependence on other disciplines.

We turn now to a discussion of the cited year and citing period effects in table 5. Cited year effects 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 reflect vintage effects, though the most recent papers have not had the same opportunity to be cited as earlier papers. The upward drift in citation is shown by the rising parameters over citing intervals.

This discussion of table 5 concludes with the decay and diffusion parameters. Differentiation of the citation function shows that the inverse of the baseline diffusion parameter β_1 (chemistry) times the diffusion parameter for each of the sciences β_{1i} yields the modal lag for that science, or the lag at which citations peak:

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

To prove this result 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 5 shows that the modal lag varies from 1.75 years for the fastest field (physics), to 2.83 years for chemistry, to 4.2 years for the slowest field (computer science). All the modal lags lie appreciably under similar lags for patented technologies (Jaffe and

[14] 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 the estimates are based are or are not included. This is consistent with the fact that the insignificant cells add very little information to what we know about scientific influence across disciplines.

Trajtenberg, 1999). In fact science diffuses about two years faster than technology, suggesting that Open Science institutions do in fact accelerate the spread of knowledge through society.

Another feature of the data is the peak citation probability, which is approximately equal to

$$P_i^{\max} = \beta_2 / \beta_1 \beta_{1i} \tag{5}$$

It turns out that for the baseline field of chemistry (where β_{1i} =1) the peak citation probability is approximately 2.0×10^{-4}. This is roughly 100 times larger than the peak citation probability for patent citations, which is on the order of 1.5×10^{-6}, indicating the greater volume of science citations compared with patent citations[15].

Table 6 reports estimates of the expanded citation function, that allows for rank-stratification classes within each of the sciences. In this table we report only the rank-stratification parameters since the other estimates are quite similar to table 5. The table confirms patterns already seen in table 4 and figure 2. First, the probability within a given field that an institution in the top 20% cites another top 20% institution typically exceeds the probability that middle 40% institutions cite each other, and that probability in turn exceeds the probability that bottom 40% institutions cite each other. For example, in computer science the parameters in question, measured relative to chemistry, are 1.490, 1.176, and 0.712. The differences are significant. The same pattern holds in every field save agriculture. Citation trails thicken as quality of institution increases, suggesting increased scientific activity and knowledge spillovers that reinforce differences among institutions.

A second feature of table 6 also predominates. This is the tendency for citations to proceed more from the bottom-up than from the top-down. This pattern suggests 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 the case of the off-diagonal comparisons in chemistry and implies that scientific influence increases with quality of institution. The primary exception to this pattern is again agriculture.

---

[15] This comparison draws on Appendix B of Jaffe and Trajtenberg (1999), which reports a baseline value for β_1 of 0.190, and a value for β_2 of 0.289×10^{-6}.

C. Symmetry Tests

The regression findings of tables 5 and 6 point out opportunities for statistical tests, especially of the symmetry of the citation parameters across fields and rank-stratification classes. Table 7 summarizes the results of pair-wise tests of symmetry of the parameters.

The first line of the table reports tests of equality of the cross-field citation parameters shown in table 5. For example, does the rate at which medicine cites biology differ from the rate at which biology cites medicine? The first line provides a round up of our answers to this question for cases in which citation takes place in both directions. Cross-citation effects do not differ significantly from symmetry in the majority of cases. One exception is economics, which cites mathematics and statistics more than the reverse at the 1% level of significance. Another 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 missed by the above evaluation: agriculture cites earth sciences and astronomy cites biology, but neither is cited in return. More deeply, these tests miss "underground" asymmetries in influence, which appear in the main body of scientific papers. Many papers, for example, make of techniques in mathematics and statistics, but feedback effects to mathematics and statistics are far less common. Our method passes over such hidden asymmetries, which require a method of encoding content that is beyond current frontiers.

The second line of table 7 tests for differences in the probability of citation by rank-stratification classes and refers to table 6. Equality is rejected in the majority of cases, with agriculture the primary exception. The top 20% institutions in computer science cite each other more often than the middle or bottom 40% institutions do, and the differences are significant at the 1% level. The third line tests for the significance of asymmetries in top-down versus bottom-up citation. Asymmetry is accepted, and equality in the cross-effects is rejected at the 1% level in the majority of cases, with agriculture and medicine being the main exceptions[16]. Thus the point estimates of table 6 are significantly asymmetric in most cases.

---

[16] 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, which may mix up institutional quality in these fields. But also these two fields may be more egalitarian than most other sciences. This tendency towards greater equality also holds to an extent for engineering, where rank-stratification classes follow the usual filter provided by the NRC rankings.

IV. Conclusion

This paper has described interactions among the top 110 U.S. universities by means of citations to scientific papers. One finding is that citation and scientific influence mostly occur within fields of science. Also, cross-field interactions are highly selective: statistically significant interactions occur in less than one-fourth of the potential cases. Collectively this suggests that knowledge flows are bounded by scientific distance, just as industrial knowledge flows are hemmed in by technological distance (Adams and Jaffe, 1996).

We find besides that most two-way interactions between fields are symmetric so that most of the time field A cites field B about as often as B cites A. Even so, we are convinced that hidden asymmetries are present in field-to-field interactions. This is because of the fact that applied fields cite other fields more often than more basic, underlying fields, and because of deeply buried course content in applied courses that interprets basic science materials, but not as much in reverse.

In addition we find that knowledge diffuses more rapidly within science than technology, and that paper citations are more abundant than patent citations, as judged by citation frequency. Our evidence confirms that quality of a university-field increases its scientific influence. When we test whether higher ranked university-fields are cited more than those lower-ranked, we accept the null hypothesis in five of every six cases. We also test for whether interactions with peer institutions increase with quality. The answer is again yes in five of every six cases. This suggests that surrounding programs reinforce the research process in a given university-field to a larger extent as quality increases.

The way ahead seems clear. The work presented here is but one ingredient of a full-fledged knowledge production for the academic sector. The production process would explain papers and perhaps patents of universities and it would include as explanatory variables, knowledge spillovers as well as current and past contributions from one's own research [17]. Of course, adequate pursuit of this agenda would require for starters, information on several 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.

[17] 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.

Looking still further ahead our goal is to extend this methodology to consider the effects on firm 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 [18]. 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.

[18] Popp (2002) represents a promising approach to this question.

Figure 1--Mean Citation Probabilities by Citation Lag

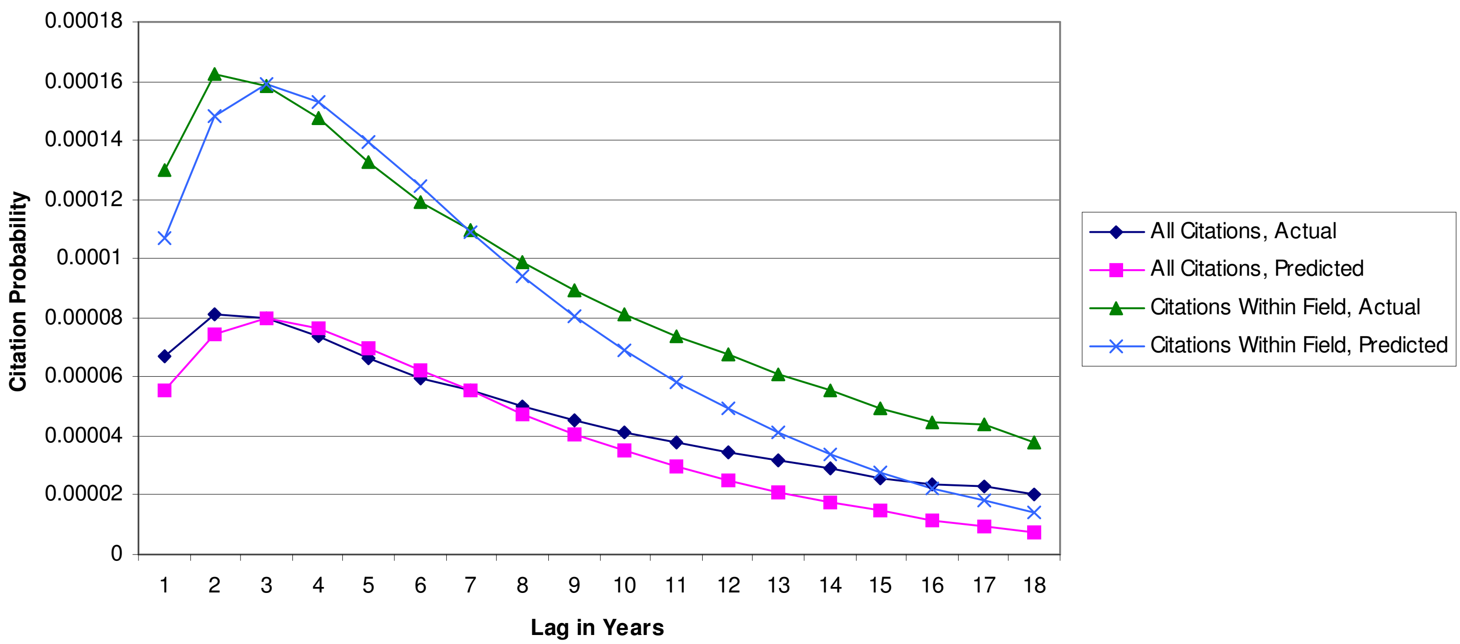

Figure 2--Mean Citation Probabilities Within Fields Between Schools of Different Rank

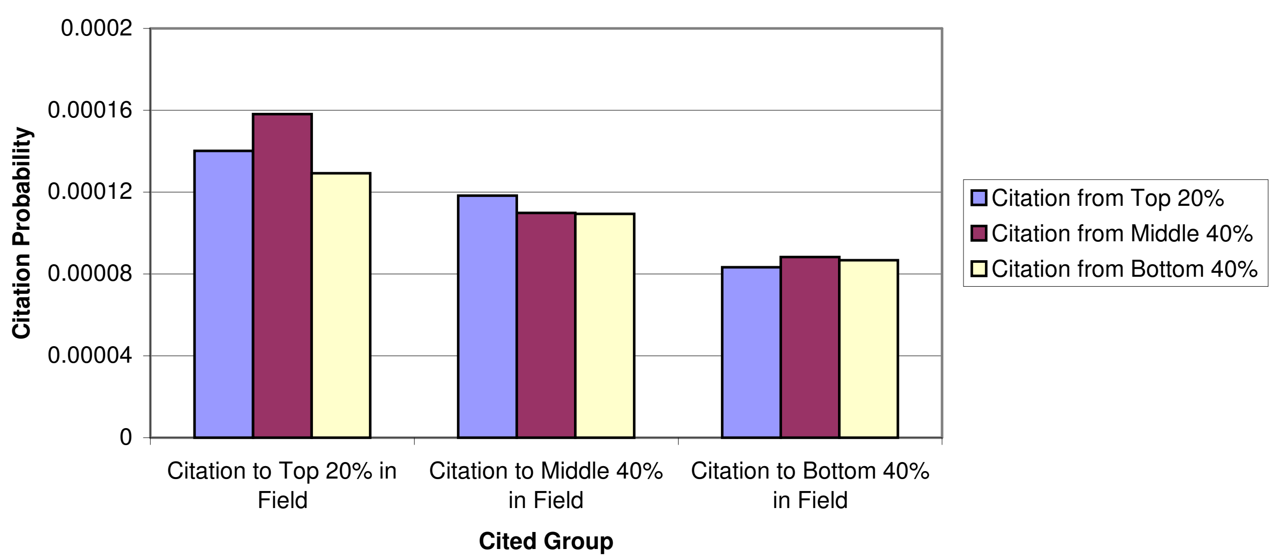

Table 1
Definition, Size, and Composition of 12 Main Science Fields
Practiced in the Top 110 U.S. Universities, 1981-1999

Main Science Field	Total Papers (Percent of Total Papers) [Total Citations Received [a]] {Percent of Total Citations Received}	Sub-Field Composition of Main Science Field
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

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 (Percent of Total Papers) [Total Citations Received [a]] {Percent of Total Citations Received}	Sub-Field Composition of Main Science Field
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
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. [a] Citations received derive from top 110 universities during the period 1981-1999. They are not a census of citations received, though citations can originate in any of the 12 main science fields listed in the table to any of the fields. Total number of papers across all 12 fields is 2,430,001. The total number of citations received is 18,784,082.

Table 2
Mean Citations and Papers Citing and Cited, by 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 Citing and Cited, by 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. The 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

Citing Field	**Cited Field**	**Mean**	**S.D.**	**Min**	**Max**
		(All Entries in Units of 10^{-6} [a])			
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

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 $10^{-6\ a}$)			
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
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 all 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 group and year observations. See the text for an additional discussion. [a] The statement that all entries are in units of 10^{-6} simply says that 20.7 is 20.7×10^{-6}, 6.6 is 6.6×10^{-6}, and similarly for all the other entries in the table.

Table 4
Mean Citations, Papers Citing and Papers Cited, and Citation Probabilities
Within Field, by Rank Stratification-Class
The Top 110 Universities, 1981-1999

Field and Citing Rank-Stratification Class	Statistics — Cited Rank-Stratification Class: **Citations/Papers Citing/ Papers Cited (Probability of Citation)** Top 20%	**Citations/Papers Citing/ Papers Cited (Probability of Citation)** Middle 40%	**Citations/Papers Citing/ Papers Cited (Probability of Citation)** Bottom 40%
Agriculture			
Top 20%	221 / 3,529/ 3,312 (18.7×10^{-6})	356 /3,529 / 4,743 (21.0×10^{-6})	174 / 3,529 / 2,616 (18.6×10^{-6})
Middle 40%	388 / 5,011/ 3,312 (23.0×10^{-6})	524 / 5,011 / 4,743 (21.7×10^{-6})	252 / 5,011 / 2,616 (18.9×10^{-6})
Bottom 40%	171 / 2,787 / 3,312 (18.2×10^{-6})	231 / 2,787 / 4,743 (17.2×10^{-6})	226 / 2,787 / 2,616 (30.5×10^{-6})
Astronomy			
Top 20%	160 / 647 / 486 (541.0×10^{-6})	259 / 647 / 692 (591.5×10^{-6})	293 / 647 / 949 (469.9×10^{-6})
Middle 40%	292 / 872 / 486 (706.1×10^{-6})	303 / 872 / 692 (496.6×10^{-6})	402 / 872 / 949 (468.6×10^{-6})
Bottom 40%	395 / 1,361 / 486 (615.8×10^{-6})	503 / 1,361 / 692 (539.6×10^{-6})	611 / 1,361 / 949 (451.3×10^{-6})
Biology			
Top 20%	5,502 / 12,056 / 9,430 (47.2×10^{-6})	4,486 / 12,056 / 12,569 (28.5×10^{-6})	2,569 / 12,056 / 12,412 (16.4×10^{-6})
Middle 40%	6,371 / 16,056 / 9,430 (40.8×10^{-6})	5,225 / 16,056 / 12,569 (24.8×10^{-6})	3,520 / 16,056 / 12,412 (16.9×10^{-6})
Bottom 40%	4,742 / 16,023 / 9,430 (30.6×10^{-6})	4,681 / 16,023 /12,569 (22.5×10^{-6})	3,524 / 16,023 / 12,412 (15.7×10^{-6})
Chemistry			
Top 20%	322 / 2,378 / 2,088 (65.7×10^{-6})	313 / 2,378 / 2,881 (45.3×10^{-6})	406 / 2,378 / 5,066 (33.4×10^{-6})
Middle 40%	438 / 3,467 / 2,088 (61.1×10^{-6})	434 / 3,467 / 2,881 (43.2×10^{-6})	599 / 3,467 / 5,066 (33.7×10^{-6})
Bottom 40%	699 / 6,320 / 2,088 (54.0×10^{-6})	746 / 6,320 / 2,881 (40.8×10^{-6})	1,031 / 6,320 / 5,066 (31.9×10^{-6})
Computer Science			
Top 20%	33 / 518 / 377 (177.1×10^{-6})	28 / 518 / 449 (129.1×10^{-6})	20 / 518 / 584 (70.9×10^{-6})
Middle 40%	43 / 615 / 377 (198.8×10^{-6})	36 / 615 / 449 (142.8×10^{-6})	30 / 615 / 584 (90.0×10^{-6})
Bottom 40%	48 / 898 / 377 (152.3×10^{-6})	48 / 898 / 449 (125.2×10^{-6})	41 / 898 / 584 (83.4×10^{-6})

Table 4
Mean Citations, Papers Citing and Papers Cited, and Citation Probabilities Within Field, by Rank Stratification-Class
The Top 110 Universities, 1981-1999

Field and Citing Rank-Stratification Class	**Statistics** — Cited Rank-Stratification Class: **Citations/Papers Citing/ Papers Cited (Probability of Citation)** Top 20%	**Citations/Papers Citing/ Papers Cited (Probability of Citation)** Middle 40%	**Citations/Papers Citing/ Papers Cited (Probability of Citation)** Bottom 40%
Earth Sciences			
Top 20%	274 / 1,211 / 1,027 (224.8×10^{-6})	266 / 1,211 / 1,260 (174.9×10^{-6})	286 / 1,211 / 1,692 (137.9×10^{-6})
Middle 40%	371 / 1,738 / 1,027 (216.9×10^{-6})	339 / 1,738 / 1,260 (156.8×10^{-6})	394 / 1,738 / 1,692 (133.3×10^{-6})
Bottom 40%	427 / 2,362 / 1,027 (181.3×10^{-6})	454 / 2,362 / 1,260 (151.7×10^{-6})	513 / 2,362 / 1,692 (125.3×10^{-6})
Economics and Business			
Top 20%	141 / 714 / 633 (306.3×10^{-6})	94 / 714 / 725 (178.9×10^{-6})	71 / 714 / 1,275 (76.4×10^{-6})
Middle 40%	168 / 801 / 633 (324.7×10^{-6})	108 / 801 / 725 (181.7×10^{-6})	106 / 801 / 1,275 (101.1×10^{-6})
Bottom 40%	215 / 1,451 / 633 (229.7×10^{-6})	188 / 1,451 / 725 (176.3×10^{-6})	224 / 1,451 / 1,275 (117.7×10^{-6})
Engineering			
Top 20%	238 / 3,824 / 2,944 (21.3×10^{-6})	182 / 3,824 / 2,978 (15.5×10^{-6})	109 / 3,824 / 2,283 (12.2×10^{-6})
Middle 40%	251 / 4,282 / 2,944 (20.1×10^{-6})	183 / 4,282 / 2,978 (14.1×10^{-6})	123 / 4,282 / 2,283 (12.3×10^{-6})
Bottom 40%	173 / 3,328 / 2,944 (17.4×10^{-6})	144 / 3,328 / 2,978 (14.1×10^{-6})	97 / 3,328 / 2,283 (12.3×10^{-6})
Mathematics and Statistics			
Top 20%	61 / 828 / 777 (94.8×10^{-6})	64 / 828 / 1,115 (69.3×10^{-6})	55 / 828 / 1,680 (39.2×10^{-6})
Middle 40%	97 / 1,244 / 777 (100.7×10^{-6})	88 / 1,244 / 1,115 (62.9×10^{-6})	87 / 1,244 / 1,680 (41.2×10^{-6})
Bottom 40%	111 / 1,903 / 777 (75.3×10^{-6})	127 / 1,903 / 1,115 (59.4×10^{-6})	137 / 1,903 / 1,680 (42.6×10^{-6})
Medicine			
Top 20%	3,374 / 15,210 / 12,183 (17.8×10^{-6})	3,659 / 15,210 / 14,464 (16.2×10^{-6})	2,010 / 15,210 / 10,079 (12.8×10^{-6})
Middle 40%	4,158 / 18,349 / 12,183 (18.2×10^{-6})	4,133 / 18,349 / 14,464 (15.2×10^{-6})	2,537 / 18,349 / 10,079 (13.4×10^{-6})
Bottom 40%	2,453 / 12,175 / 12,183 (16.2×10^{-6})	2,698 / 12,175 / 14,464 (14.9×10^{-6})	1,693 / 12,175 / 10,079 (13.5×10^{-6})

Table 4
Mean Citations, Papers Citing and Papers Cited, and Citation Probabilities Within Field, by Rank Stratification-Class
The Top 110 Universities, 1981-1999

Field and Citing Rank-Stratification Class	**Statistics** Cited Rank-Stratification Class **Citations/Papers Citing/ Papers Cited (Probability of Citation)** Top 20%	**Citations/Papers Citing/ Papers Cited (Probability of Citation)** Middle 40%	**Citations/Papers Citing/ Papers Cited (Probability of Citation)** Bottom 40%
Physics			
Top 20%	952 / 4,260 / 3,347 (64.4×10^{-6})	1,062 / 4,260 / 3,524 (60.9×10^{-6})	1,060 / 4,260 / 4,876 (40.6×10^{-6})
Middle 40%	1,315 / 4,815 / 3,347 (76.8×10^{-6})	1,505 / 4,815 / 3,524 (74.9×10^{-6})	1,686 / 4,815 / 4,876 (56.2×10^{-6})
Bottom 40%	1,604 / 7,197 / 3,347 (61.4×10^{-6})	2,110 / 7,197 / 3,524 (68.0×10^{-6})	2,266 / 7,197 / 4,876 (49.3×10^{-6})
Psychology			
Top 20%	242 / 1,624 / 1,450 (102.4×10^{-6})	253 /1,624 / 1,736 (88.9×10^{-6})	455 / 1,624 / 3,822 (72.5×10^{-6})
Middle 40%	304 / 1,880 / 1,450 (110.9×10^{-6})	276 /1,880 / 1,736 (83.7×10^{-6})	510 / 1,880 / 3,822 (70.0×10^{-6})
Bottom 40%	613 / 4,300 / 1,450 (97.7×10^{-6})	625 /4,300 / 1,736 (82.8×10^{-6})	1,121 / 4,300 / 3,822 (67.4×10^{-6})

Notes. For this table, the number of citing and cited group and year observations is $171\times12\times9=18{,}468$. As in previous tables, 171 is the number of citing and cited year combinations, 12 is the number of fields, and 9 is the number of rank-stratification class combinations within each field.

Table 5
Baseline Citation Function, with Cross-Field Effects
The Top 110 U.S. Universities, 1981-1999

Citing Field	Cited Field	Regression Parameter	Asymptotic Standard Error	Asymptotic t-Statistic, H_0=0	Asymptotic t-Statistic, H_0=1
Field Intercepts (α_i s)					
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 5
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
Field Intercepts (α_i s)					
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
Cited Year Effects					
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
Citing Interval Effects					
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 5
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**
Decay Parameter (β_1)	0.353	0.006	58.8	--
Diffusion Parameter (β_2)	7.2×10^{-5}	1.86×10^{-6}	387.1	--
Field Decay Parameters (β_{1i} s)				
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 6
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%
Agriculture			
Top 20%	0.198	0.224	0.198
	(0.017)	(0.016)	(0.017)
Middle 40%	0.244	0.232	0.203
	(0.017)	(0.016)	(0.017)
Bottom 40%	0.195	0.184	0.326
	(0.017)	(0.016)	(0.022)
Astronomy			
Top 20%	8.733	9.753	7.969
	(0.263)	(0.291)	(0.239)
Middle 40%	11.254	8.133	7.902
	(0.335)	(0.244)	(0.236)
Bottom 40%	9.893	8.852	7.723
	(0.295)	(0.264)	(0.231)
Biology			
Top 20%	0.867	0.528	0.303
	(0.031)	(0.021)	(0.015)
Middle 40%	0.733	0.451	0.307
	(0.026)	(0.018)	(0.015)
Bottom 40%	0.531	0.395	0.275
	(0.021)	(0.017)	(0.014)
Chemistry			
Top 20%	1.000	0.700	0.516
	(--)	(0.028)	(0.023)
Middle 40%	0.924	0.660	0.519
	(0.032)	(0.026)	(0.022)
Bottom 40%	0.809	0.623	0.490
	(0.028)	(0.023)	(0.019)
Computer Science			
Top 20%	1.490	1.079	0.598
	(0.059)	(0.047)	(0.035)
Middle 40%	1.636	1.176	0.739
	(0.063)	(0.049)	(0.037)
Bottom 40%	1.280	1.066	0.712
	(0.051)	(0.045)	(0.035)
Earth Science			
Top 20%	2.685	2.093	1.671
	(0.086)	(0.068)	(0.056)
Middle 40%	2.556	1.852	1.608
	(0.081)	(0.061)	(0.053)
Bottom 40%	2.140	1.801	1.519
	(0.069)	(0.059)	(0.050)

Table 6
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			
Top 20%	2.693	1.572	0.678
	(0.086)	(0.054)	(0.030)
Middle 40%	2.853	1.597	0.897
	(0.090)	(0.054)	(0.035)
Bottom 40%	2.001	1.528	1.046
	(0.065)	(0.051)	(0.037)
Engineering			
Top 20%	0.211	0.156	0.122
	(0.018)	(0.016)	(0.016)
Middle 40%	0.198	0.140	0.122
	(0.017)	(0.015)	(0.015)
Bottom 40%	0.173	0.141	0.124
	(0.017)	(0.016)	(0.016)
Mathematics and Statistics			
Top 20%	0.857	0.637	0.365
	(0.040)	(0.033)	(0.025)
Middle 40%	0.910	0.575	0.382
	(0.040)	(0.029)	(0.023)
Bottom 40%	0.677	0.543	0.395
	(0.032)	(0.027)	(0.022)
Medicine			
Top 20%	0.248	0.226	0.180
	(0.013)	(0.012)	(0.011)
Middle 40%	0.252	0.211	0.187
	(0.013)	(0.011)	(0.011)
Bottom 40%	0.222	0.206	0.187
	(0.013)	(0.012)	(0.012)
Physics			
Top 20%	2.184	2.258	1.572
	(0.077)	(0.078)	(0.059)
Middle 40%	2.634	2.840	2.253
	(0.089)	(0.094)	(0.076)
Bottom 40%	2.120	2.622	1.994
	(0.073)	(0.087)	(0.068)
Psychology			
Top 20%	0.929	0.806	0.658
	(0.034)	(0.031)	(0.025)
Middle 40%	0.996	0.755	0.634
	(0.036)	(0.029)	(0.024)
Bottom 40%	0.880	0.748	0.610
	(0.031)	(0.027)	(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 7
Symmetry Tests of the Citation Function
Of the 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 (χ^2=16.7, P<0.0001); Physics cites Astronomy more than the reverse (χ^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 and within fields	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% (χ^2=33.7, P<0.0001); middle 40% is cited less than the bottom 40% (χ^2=21.5, P<0.0001). Top 20% of Physics is cited less than the middle 40% (χ^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 and within fields	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 (χ^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 (χ^2=4.6, P=0.0317). Equality between the middle 40% and bottom 40% of medicine is accepted.

Notes. All χ^2 tests are Wald Tests that evaluate the difference in the parameters from zero for the unrestricted likelihood.

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.

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.

Huffman, Wallace E., and Robert E. Evenson, **Science for Agriculture: A Long-term Perspective**. Ames, Iowa: Iowa State University Press, 1993.

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.

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,** 8th Edition. London, Macmillan, 1920

McDonald, Arthur B., Joshua R. Klein, and David L. Wark, "Solving the Solar Neutrino Problem," **Scientific American** 288 (April 2003): 40-49.

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.

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 1982): 227-245.

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

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

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

Appendix Table A-1
The Top 110 U.S. Universities

FICE Code	University Name	1998 Federal R & D Expenditures
2077	Johns Hopkins University	752.983*
1305	Stanford University	342.426
3798	University of Washington - Seattle	336.748
X9091	University of Michigan, All Campuses	311.450
2178	Massachusetts Institute of Technology	310.741
1317	University of California-San Diego	262.303
2155	Harvard University	251.876
3378	University of Pennsylvania	247.914
3895	University of Wisconsin-Madison	240.513
1315	University of California-Los Angeles	233.702
X7963	Columbia University, All Campuses	229.723
X8717	University of Colorado, All Campuses	228.342
1319	University of California-San Francisco	219.912
X1051	University of Alabama, All Campuses	205.511
1426	Yale University	205.046
X8761	University of Minnesota, All Campuses	204.741
X8779	Cornell University, All Campuses	204.187
1328	University of Southern California	190.547
2520	Washington University	187.173
X8813	Pennsylvania State University, All Campuses	186.274
1131	California Institute of Technology	177.748
2920	Duke University	172.532
2974	University of North Carolina at Chapel Hill	171.505
1312	University of California-Berkeley	171.135
1775	University of Illinois at Urbana-Champaign	168.871
X8815	University of Pittsburgh, All Campuses	168.511
3658	University of Texas at Austin	165.082
1083	University of Arizona	161.999
X3632	Texas A&M University, All Campuses	144.938
3024	Case Western Reserve University	132.274
2894	University of Rochester	130.773
2103	University of Maryland at College Park	129.198
1739	Northwestern University	127.911
1774	University of Chicago	125.982
X8802	Ohio State University, All Campuses	124.177
1564	Emory University	118.045
1892	University of Iowa	115.312
1313	University of California-Davis	114.912
X8723	Georgia Institute of Technology, All Campuses	113.643

Appendix Table A-1
The Top 110 U.S. Universities

FICE Code	University Name	1998 Federal R & D Expenditures
4949	Baylor College of Medicine	110.610
1535	University of Florida	106.510
3535	Vanderbilt University	106.325
2130	Boston University	104.428
1536	University of Miami	101.492
2785	New York University	101.426
3675	University of Utah	100.722
X8755	University of Massachusetts, All Campuses	100.122
3660	University of Texas Southwestern Med Center Dallas	97.200
X8731	Indiana University, All Campuses	95.840
3242	Carnegie Mellon University	95.046
X3745	University of Virginia, All Campuses	93.328
X8732	Purdue University, All Campuses	92.844
9555	SUNY at Stony Brook, All Campuses	91.531
8805	University of Cincinnati, All Campuses	90.307
1610	University of Hawaii at Manoa	86.886
1445	Georgetown University	84.801
2663	University of New Mexico, All Campuses	84.365
3754	Virginia Polytechnic Institute and State University	82.734
3210	Oregon State University	82.416
2290	Michigan State University	81.146
1350	Colorado State University	80.451
2903	Yeshiva University	80.000
2972	North Carolina State University at Raleigh	79.533
2104	University of Maryland at Baltimore	78.037
X9554	SUNY at Buffalo, All Campuses	76.037
1776	University of Illinois at Chicago	73.797
4882	Oregon Health Sciences University	71.054
11618	University of Texas Health Science Center Houston	70.446
X8771	Rutgers the State University of NJ, All Campuses	69.829
X8051	University of Tennessee, All Campuses	69.793
2627	Princeton University	69.005
1320	University of California-Santa Barbara	68.408
X8745	Louisiana State University, All Campuses	67.090
1314	University of California-Irvine	65.902
2230	Woods Hole Oceanographic Institution	64.765
X2515	University of Missouri, All Campuses	63.556
2219	Tufts University	61.167
X8744	University of Kentucky, All Campuses	60.760
X8025	University of Nebraska, All Campuses	58.482
2329	Wayne State University	57.646

Appendix Table A-1
The Top 110 U.S. Universities

FICE Code	University Name	1998 Federal R & D Expenditures
2978	Wake Forest University	56.705
8773	New Mexico State University, All Campuses	56.587
3659	University of Texas Health Science Center San Antonio	55.004
3677	Utah State University	54.903
1598	University of Georgia	54.712
X8718	University of Connecticut, All Campuses	53.189
2029	Tulane University	52.924
1869	Iowa State University	51.196
X9001	University of Kansas, All Campuses	50.567
1489	Florida State University	50.451
3735	Virginia Commonwealth University	48.167
2573	Dartmouth College	45.053
3800	Washington State University	44.510
3401	Brown University	44.412
2807	Rockefeller University	43.845
1081	Arizona State University Main	41.359
3604	Rice University	34.772
1431	University of Delaware	33.688
X2686	CUNY, All Campuses	32.412
29094	University of AK Fairbanks, All Campuses	31.505
3696	University of Vermont	31.460
1321	University of California-Santa Cruz	29.849
X8789	Syracuse University, All Campuses	29.200
2133	Brandeis University	28.098
3223	University of Oregon	27.041
2589	University of New Hampshire	25.913
3827	West Virginia University	24.985
1316	University of California-Riverside	22.988
1710	Loyola University of Chicago	17.685
3289	Lehigh University	13.019

Notes. * Includes R&D expense for the Applied Physics Laboratory. The FICE code is the federal ID for U.S. universities as well as university systems.