15 November 2004

How Rapidly Does Science Leak Out? *

By

James D. Adams, Rensselaer Polytechnic Institute and NBER
J. Roger Clemmons, Institute for Child Health Policy, University of Florida
Paula E. Stephan, Georgia State University

November 2004

Comments Welcome

Keywords, Science, R&D, Technology, Citations, and Diffusion
JEL Codes: L3, O3

Corresponding Author: James D. Adams, Department of Economics, Rensselaer Polytechnic Institute, 3504 Russell Sage Laboratory, Troy, NY 12180-3590. Telephone: 1-518-276-2523, fax: 1-518-276-2235, E-mail: adamsj@rpi.edu.

* The Andrew W. Mellon Foundation provided generous support for this research. We thank Nancy Bayers and Henry Small of the Institute for Scientific Information (ISI) for suggestions and clarifications concerning the data. We also thank Adam Jaffe for nonlinear regression programs that we adapted for use in this paper. Ken Simons provided helpful comments on an early draft of this paper, presented at the International Industrial Organization Society meetings, Boston, April 4-5, 2003.

Abstract

This paper measures the speed of diffusion of scientific research by field and sector of the U.S. economy. The data set derives from the Institute for Scientific Information (ISI). The ISI data consist of about 2.4 million scientific papers having at least one author in the top 110 universities, and of 240 thousand papers having at least one author in the top 200 U.S. R&D firms. This evidence covers the majority of scientific research in the U.S. during the period 1981-1999.

We find that the modal lag in citations between U.S. universities, or the lag at which citations peak, occurs at 2.85 years for the baseline field (chemistry), 4.25 years for the slowest diffusing field (computer science), and 1.75 years for the fastest diffusing field (physics). The equally weighted average of the modal lags across all fields is about 3.5 years. Modal lags for citations between U.S. firms and U.S. firms citing U.S. universities are the same, about 3.5 years. However, these similarities in speed of diffusion conceal substantial differences in the rapidity with which science leaks out across industries and fields, and the modal lags differ by a year and more in these dimensions of the data.

Using the same methodology as the above, researchers have found modal lags of five years based on patent data. Taken together, the findings on scientific papers and patents suggest that the speed of diffusion of science is 30 percent faster than the speed of diffusion of technology. These results seem consistent with the idea that Open Science promotes more rapid dissemination of information than Proprietary Technology.

I. Introduction

In this paper we set about measuring the speed of diffusion of science in U.S. universities and firms. By so doing we hope to establish stylized facts about the diffusion process in academic and industrial science and to compare the diffusion of science with that of patented technology. We rely on citation lags between citing and cited scientific papers for this purpose. Our method for estimating citation functions based on scientific papers, including diffusion lags, closely resembles the methodology for estimating citation functions based on patents in Jaffe and Trajtenberg (1996, 1999). We rely on a common methodology by intention. This convergence of approaches makes measurements in science and technology more comparable, even though we do not claim that the two diffusion processes are necessarily the same.

To the extent that industrial innovation depends on recent results from science (Mansfield, 1991), an increase in the speed of diffusion of scientific research pushes technology-in-use closer to the best-practice technology. For this reason alone the speed of diffusion of science could be important, much as the speed of acceptance of industrial innovations is important: because it increases productive efficiency and real income. Griliches (1957) examines the adoption of hybrid corn by farmers in U.S. states and crop reporting districts, as well as the dynamics of the supply of hybrid corn varieties that were adapted to particular regions. His findings suggest that lags in the adoption of hybrid corn shorten, as expected future profits from adoption increase. Likewise Mansfield (1963) shows that adoption of the diesel locomotive by U.S. railroads was triggered by increasing advantages of diesel over steam, as conditioned by factors specific to the railroads.

In the same way we believe that the speed of adoption of new scientific approaches is also determined by gains to adopters. Biotechnology is a case in point. Drug companies were eager to embrace the technology, because it seemed to avoid mass testing of chemical compounds (Henderson and Cockburn, 1996). And yet, shortfalls in human resources in molecular biology and cell science hindered adoption. This bottleneck stimulated entry of biotechnology firms, whose founders were practitioners of these disciplines (Audretsch and Stephan, 1996; Zucker, Darby, and Brewer, 1998). The outcome of this impromptu allocation of scientific resources was a two-tiered structure in which newer firms possessed the science and older firms possessed testing facilities, marketing and distribution. It was this structure that helped to spread biotechnology throughout pharmaceuticals (Ruttan, Ch.10, 2001).

A related literature considers the link between intellectual property rights, diffusion, and economic performance. Mansfield (1985) reports survey-based evidence on the speed with which knowledge of a company's development efforts leaks out to competitors[1]. His findings suggest that knowledge of these efforts reaches rival firms within 12 to 18 months. Other results are that once developed, knowledge of new products and processes leaks out in 12 more months. However, imitation costs imply that actual imitation follows an opportunity to imitate with an additional lag. Mansfield, Schwartz and Wagner (1981) find that imitation costs are about two-thirds of innovation costs, and that patenting increased imitation costs. Both findings suggested substantial delays in imitation of industrial technology[2].

Weak incentives hinder invention and adoption in planned economies and in public sectors more generally. In his discussion of the Soviet bonus system, Berliner (1976) suggests that financial gains rewards from innovation were largely missing and that this contributed to the Soviet Union's falling behind the West. Dearden, Ickes, and Samuelson (1990), and Hart, Vishny, and Shleifer (1997) make this point in theoretical studies of the limits to public sector innovation. Even in the private sector, imperfect rewards to inventors could limit invention by industrial firms. Scherer (1984, chapter 9) finds that firms' innovative output rises at a decreasing rate with firm size and suggests that incentive problems are the cause. Thus a substantial literature relates lags in technology diffusion to costs of adoption and weak incentives to adopt.

Incentives are harder to observe and to quantify in science than technology, because they are more often non-pecuniary. However, the priority system in science, as Robert Merton has shown, encourages individuals to share knowledge quickly, since it is sharing that establishes property rights in science (Stephan, 2004). For this reason it is possible that the priority publication system accelerates the diffusion of science compared with technology, especially given the fact that the private gains from technology are proprietary and rely on secrecy (Levin, Klevorick, Nelson, and Winter, 1987; Cohen, Nelson, and Walsh, 2002; Furman and Stern, 2004).

The findings in this paper suggest that the speed of diffusion of science may exceed that of industrial technology. This speed advantage of science is independent of a paper's origin in universities or in R&D

[1] This paper and its title are partly inspired by Mansfield (1985).
[2] Adams (2005) reviews the survey-based literature of industrial R&D laboratories and firms.

firms. Specifically, we find that among universities the modal, or most frequent, citation lag is about 2.9 years in the baseline field of chemistry, but that this lag varies from 1.8 years in physics to 4.3 years in computer science and averages about 3.4 years. Lags in citation between fields are slightly longer by 0.4 years and average about 3.8 years. Lags on self-citations by universities reach an even shorter average of 2 years. It is striking that the diffusion lags from universities to firms and from firms to firms are the same as above, about 3.5 years for non-self citations and 1.8 years for firms citing their own research. All these results apply to science. In comparison, patented technology within the U.S. displays a modal lag that ranges from 4.6 to 5.3 years within the United States and still longer between countries (Jaffe and Trajtenberg, 1996, 1999). Since the modal lag is a robust statistic for ranking diffusion given the functional form that we employ, we argue that diffusion of science takes place about 30 percent faster than diffusion of industrial technology.

In additional results we explore variation in the speed of diffusion by industry as well as field. The range in this parameter is substantial: on average the slowest industry in a field takes about two years longer to diffuse than the fastest industry. In the more detailed results, electrical equipment and communications stand out as industries to which science diffuses rapidly, while diffusion proceeds more slowly to metals and machinery. These results indicate unexplained differences in the speed of diffusion by industry that deserve further study.

The rest of the paper is divided into six parts. Section II discusses the article citation function, which is based on the patent citation methodology of Jaffe and Trajtenberg (1996, 1999). Our use for the citation function in this paper is in specifying the speed of diffusion of new scientific research. As noted, by using a similar methodology to patent citation functions, we are able to compare the speed of diffusion of science and technology. Section III discusses the database of scientific papers and citations that underlies our findings. The data derive from the Institute for Scientific Information (ISI). They consist of about 2.7 million scientific papers and roughly 20.2 million citations to these papers over the period 1981-1999. The papers are written in the top 110 U.S. universities and in the top 200 U.S. R&D firms, and the universities contribute roughly 90 percent of total papers and citations. Section IV presents estimates of the speed of diffusion of science within the U.S. university system, while section V provides estimates of the speed of diffusion of science from universities to firms and between firms. A special feature of this section is its inclusion of results that allow the speed of diffusion to vary by industry as well as field of science. In section VI we draw comparisons of the speed of diffusion of science with that of patented technology as based on the literature of patent citation functions. Section VII concludes the paper.

II. The Citation Function

In estimating the speed of diffusion between citing and cited academic papers, we rely on the citation function. Jaffe and Trajtenberg (1996, 1999) develop and use the citation function to quantify the direction as well as speed of diffusion of new technologies, where the latter is based on lags in patent citation between citing and cited patents. In this paper we apply the citation function to scientific papers and the study of diffusion based on lags in science citations. Since the citation function is the centerpiece of our empirical analysis we provide a brief discussion of it in this section.

Before launching into its discussion, though, we would like to comment on asymmetries as well as symmetries between science citations and patent citations. Both refer to prior literature, but their motivations for doing so are obscure. Both could measure influence of earlier ideas and both could define the problem being addressed, which in patent citations limits commercial application of the invention to improvements on prior art. Science citations are more likely to refute findings, and more likely to be honorific or a strategy to achieve publication. 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. Overall, there are important differences as well as similarities in science and patent citations, although their bearing on diffusion is not clear.

We estimate the citation function on cells that are defined by grouped characteristics of citing and cited papers. Each cell includes a citation probability that is composed of citations, papers citing and papers cited. This probability is

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

In (1) p_{iTjt} is the probability that a group i paper published in year T cites a group j paper published in year t, T>t. c_{iTjt} is the number of citations, or the number of pairs of papers that are linked by citations. n_{iT} is the number of group i papers in year T that could (but might not) cite group j papers in year t. For this reason n_{iT} is the number of potentially citing papers and n_{jt} is the number of group j papers in year t that could be cited, and so are potentially cited papers. The product $n_{iT} \times n_{jt}$ is the number of pairs of papers that could be but mostly are not linked by citation.

It is useful to explain what is meant by "group" in this analysis. In the case of universities citing universities, which we present first, we use group i to mean the citing field of science, while group j is the cited field. Thus citing and cited fields and years define four-dimensional cells in these particular data.

Below we discuss the dimensions of the cells and we report descriptive statistics that are based on them. For now we indicate how the citation function offers a parametric representation of the probability of citation from group i in year T to group j in year t. First, we allow for intercept terms that correspond to each of the cell dimensions, although we do not discuss these terms in this paper, given its focus on the speed of diffusion[3]. In addition we allow for diffusion and decay processes as the lag increases between citing and cited papers. The citation function for universities citing universities is

(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. But note that the probability parameters are meaningful only when compared with a baseline value. In this article, the α_{ij} parameters are normalized by the value for chemistry, whose value is set equal to 1.0. Likewise, the α_T and α_t parameters are normalized by the earliest citing and cited periods, whose values are set equal to 1.0.

The β_1 parameter is a baseline parameter for the rate of decay in citation to chemistry, while β_{1j} is a vector of decay parameters relative to the baseline. For this reason β_{1j} for chemistry is fixed at 1.0. Finally, β_2 governs the overall diffusion as captured by citation. Since β_2 positions the overall rate of

[3] Adams, Clemmons and Stephan (2004a,b) discuss the intercept terms of the citation function at length. These parameters can be thought of as long-run linkages between science fields, industries, and other groups.

citation, the parameter is not identified by field independently of the α_{ij} vector. As noted before, we limit reported regression coefficients to β_1, β_{1j}, and β_2 for brevity and because of this paper's focus on diffusion. The error term is u_{iTjt} and the equation is estimated by nonlinear least squares.

In addition to the citation function for universities citing universities, we estimate citation functions for firms citing universities, and for firms citing firms. The following passage explains how these versions of the citation function differ from (2).

For the case of firms citing universities we include a vector of parameters α_I that captures citing industries and another α_j that captures citing fields. Again because of this paper's emphasis on diffusion, we do not report these results[4]. As before the vector of parameters must be normalized by one of the industries, since β_2 absorbs the overall citation rate. For this purpose we choose petrochemicals, whose parameter is fixed at 1.0. In summary, the citation function for firms citing universities is:

(3)
$$p_{ITjt} = \alpha_I \alpha_j \alpha_T \alpha_t \exp\left[-\beta_1 \beta_{1j} (T-t)\right]\left\{1-\exp\left[-\beta_2 (T-t)\right]\right\} + u_{iTjt}$$

In the case of firms citing firms we include three intercept vectors. Two stand for citing and cited industries and are α_I and α_i. A third vector, standing for citing field, is α_j. Thus, the citation function for this case is:

(4)
$$p_{iTjt} = \alpha_I \alpha_I \alpha_j \alpha_T \alpha_t \exp\left[-\beta_1 \beta_{1j} (T-t)\right]\left\{1-\exp\left[-\beta_2 (T-t)\right]\right\} + u_{iTjt}$$

In some of our analysis we allow for greater freedom in the form of diffusion. This takes the form of allowing the decay parameters to vary by citing industry as well as field so that β_{1j} becomes β_{1ij}. As before, in both (3) and (4) we limit reporting of the parameters to β_1, β_{1j}, and β_2.

We turn now to the interpretation of the speed of diffusion. The point of this discussion is to argue that the modal or most frequent lag in citation is a robust measure of speed given the functional form in (2),

[4] For an analysis of the intercept terms that allows for a full set of interactions between industry and field of science for firms citing universities and firms citing firms, see Adams, Clemmons and Stephan (2004a).

(3), and (4). Recall that these functional forms are motivated by comparability with the citation function in the patent literature. The modal lag for science field j, or the lag at which the citation probability peaks, is

(5) $$L_{Modal} = 1/\beta_1\beta_{1j}\,.$$

To prove this, take the derivative of the citation function, set it equal to zero, and solve for L.

The cumulative citation probability at L=∞ is found by integrating (2)-(4). The cumulative probability is given by:

(6) $$C(\infty) = \int_0^\infty \alpha \exp(-\beta_1\beta_{1j}L)\,[1-\exp(-\beta_2 L)]\,dL = \frac{\alpha\beta_2}{\beta_1\beta_{1j}(\beta_1\beta_{1j}+\beta_2)}\,.$$

In (6) we have combined the intercept terms in a single term α for convenience.

Next we compute the average lag in citation by multiplying each lag by the probability of citation and integrating out to infinity:

(7) $$M(\infty) = \int_0^\infty \alpha\, L\exp(-\beta_1\beta_{1j}L)\,[1-\exp(-\beta_2 L)]\,dL = \frac{\alpha\left[(\beta_1\beta_{1j}+\beta_2)^2 - (\beta_1\beta_{1j})^2\right]}{(\beta_1\beta_{1j})^2\,(\beta_1\beta_{1j}+\beta_2)^2}$$

The result to the right of the second equality sign can be proved by applying integration by parts to the middle expression. However, this average lag is based on a cumulative citation probability that does not sum to unity. To obtain a proper mean lag, we normalize or divide (7) by the cumulative citation probability (6). After some manipulations the result is

(8) $$\overline{L} = \frac{1}{\beta_1\beta_{1j}} + \frac{1}{\beta_1\beta_{1j}+\beta_2} \approx \frac{2}{\beta_1\beta_{1j}} = 2 \bullet L_{Modal}$$

The approximation in the third expression is based on $\beta_2 / \beta_1\beta_{1j} \approx 0$, which is true for the estimates reported below. The point of discussing the modal and mean lags consisting of (5) and (8) is to show that these lags rank the fields by speed of diffusion in the same order[5]. It follows that the modal lag is a robust way to compare the speed of diffusion, given the functional forms that we use in this paper.

[5] We can also show that the median lag, which divides the cumulative probability of citation into 50 percent before and 50 percent after, is approximately $L_{median} \approx -\ln(1/2)/\beta_1\beta_{1j} = 0.6931/\beta_1\beta_{1j} = 0.6931 \bullet L_{Modal}$. This approximation is again based on $\beta_2/\beta_1\beta_{1j} \approx 0$. So the modal lag, the median lag, and the mean lag all rank the speed of diffusion in the same order for the functional form of (2)-(4).

III. Database

The underlying data consist of 2.4 million scientific papers written in the top 110 U.S. universities during 1981-1999 and of 18.8 million citations to those papers by other papers written in the 110. Also included are 230 thousand scientific papers written by the top 200 U.S. R&D firms over the same period, as well as 640 thousand citations to these papers by other papers written in firms. In addition, cross-citations link the academic and industrial sectors together. The 200 firms make about one million citations to scientific papers by the 110 universities. The 110 universities and 200 firms account for the majority of academic and industrial research conducted in the U.S. The source of all these data is ISI, the Institute for Scientific Information, in Philadelphia, Pennsylvania.

The papers appear in 7137 scientific journals. Each journal is assigned to a unique science field, along with the papers published in them. The alternative to this journal assignment method is to assign papers according to sciences of "origin", as given by author's departments. But that approach is ruled out by incomplete information on academic departments[6].

A. Distribution of Papers and Citations

Table 1 describes the distribution of university and firm papers and citations by field of science. The first two columns contain the data for universities. About 61 percent of the university papers and 73 percent of citations originate in agriculture, biology, and medicine, so that the life sciences predominate over university science. The fields of chemistry, engineering, and physics rank second and account for 24 percent of papers and 16 percent of citations. The remaining six fields of astronomy, computer science, earth sciences, economics and business, mathematics and statistics, and psychology, which are half the total, account for just 15 of university papers and 11 percent of citations.

The last two columns of the table describe the distribution of firm papers and citations. Nine of the 12 science fields are represented. The remaining three—astronomy, economics and business, and psychology—account for one percent of industrial papers, and since they are trace elements, are dropped from the table. In the industrial distribution, agriculture, biology, and medicine account for 32 percent of

[6] As an experiment we sought to assign all the papers of Harvard University to one of the 12 main science fields in our data using address information. About one-third of the papers could not be assigned to a field using information on authors' Harvard addresses. Given this failure rate we abandoned the effort.

papers and 41 percent of citations instead of 61 and 72 percent, as in universities. Chemistry, engineering and physics account for 59 percent of industrial papers and 54 percent of industrial citations rather than 24 and 16 percent, as in academia. Thus, the relative importance of life science and natural science and engineering in industrial science are exactly the reverse of what they are in academic science. Another key feature of the industrial distribution is the much greater importance of computer science, which accounts for five percent of industrial papers and three percent of industrial citations compared with one percent or less in universities. While these differences are not surprising, they are useful to take note of. For example, they indicate that industrial citations made are much more likely to originate in physical science than are the citations made by university papers.

In this section we complement the analytical discussion of section II with a description of citation probabilities by citation lag. Our presentation is complicated by the fact that we are interested in comparing diffusion between universities, firms and universities, and between firms, although the comparison provides useful information on the diffusion process.

To undertake an analysis of diffusion we aggregate the papers and citations into cells classified by citing and cited group and year, as in Section II. For each cell we calculate the numbers of citations, potentially cited papers, and potentially citing papers. In the case of firms, the citing and cited groups that define the cells entail more dimensions, because of the important role played by the firm's industry.

Consider first universities citing other universities. In this case the data are classified by citing field and year, and cited field and year[7]. We remove citations from a school to itself in order to separate institutional self-citations from citations to other schools, since the diffusion process is very different for the two kinds of reference. The number of cells in the case of 'other' citations, where the citing and cited universities are different, is 36,834[8]. There are 21,801 cells composed of self-citations to and from the same university[9].

[7] For citations within the same university-field, where the majority of citations take place, we also keep track of citations and papers cited and citing of the top 20%, the middle 40%, and the lowest 40% of universities in a field. Thus, within a field the cells are six-dimensional, consisting of citing field, rank, and year; and cited field, rank, and year. However, taking rank of university-field into account has little bearing on the diffusion process, since diffusion does not change when rank is taken out of the analysis.

[8] Cross-field citations do not occur in some years. The number of within-field cells (allowing for rank) is 9×12 for each citing and cited year combination. Likewise the number of cross-field combinations is

When firms cite universities the cells are classified by citing industry, field and year, and by cited field and year. The number of citing industry-field-year, cited field-year cells is 30,604 in this environment. Finally consider firms citing firms. The cells are classified by citing and cited industry, field and year. There are 34,246 cells in the case of non-self citations, where citing and cited firms are different. The number of cells where firms self-cite their own research is 10,687.

Table 2 describes the cells for the three sets of data excluding self-citations. The first three columns present mean citations, mean potentially cited papers, and mean potentially cited papers for the three sectoral interactions that we have discussed. The final column on the right presents means of the citation probabilities as computed in equation (1).

Panel A presents means by field and year for universities citing universities. Mean 'other' citations and citing papers vary more than cited papers, reflecting the scale of citing academic fields. The mean number of cited papers varies less because of the rarity of cross-field citations. In general citations are more numerous in this, the largest of the three sectors. The probability of citation is on the order of 10^{-4} and it is noticeably larger in small disciplines such as astronomy and economics.

Panel B reports means by field and year, and industry group and year in the case of firms citing universities. Because astronomy, economics, and psychology are rarely practiced in industrial research, these fields are dropped from the panel. Since industrial papers are one-tenth as numerous as university papers, the number of citations and papers citing are less than in panel A[10]. The predominance of industrial chemistry, engineering, and physics shows up in the larger means of citing papers in these areas relative to biology and medicine and compared with panel A. The means by industry point out the frequency of citations and publications in pharmaceuticals and biotechnology and their scarcity for example, in metals and machinery. Notice that mean citation probabilities from firms to universities are on the order of 10^{-5} and thus are one-tenth as frequent as university-university citations. This calculation takes account of the

---

11×12. The *potential* number of citing and cited year combinations is (19×18)/2. Thus the potential number of cells is (9×12 +11×12) × (19×18)/2=41,040. But 4,206 of the cross-field cells do not exist.

[9] The definition needs to be kept in mind. True self-citations, in which the same investigators reference their own past research, are likely to diffuse even more rapidly than institutional self-citations, which is the type that we record here.

[10] The mean number of potential university papers that are cited by firms in agriculture, biology, and medicine exceeds the numbers cited by universities. This is because firms are latecomers to citation compare with universities, when article counts in these fields are larger.

smaller scale of industrial papers and suggests that citations between sectors are really less common than they are within sectors.

Panel C displays means by citing fields and industries for the case of firms citing other firms. The number of dimensions exceeds that of previous panels and, given the rarity of industrial papers, this contributes to the low citation counts and papers cited in panel C compared to panels A and B[11]. However the mean science citation probabilities between industrial firms are on the order of 10^{-4}, confirming the higher frequency of science citations within sectors as opposed to between sectors, which we have already seen in comparing the citation frequencies in panels A and B.

B. Diffusion Lags by Citing and Cited Sector

Figure 1 graphs empirical citation curves between universities, from firms to universities, and between firms. The citation curves show the mean probability of citation by the lag in years between citing and cited papers. The curves peak in the second year, indicating that diffusion is rapid, though fitted citations diffuse somewhat more slowly (see section IV). The irregular shape of the curve for firms citing each other at longer lags is due to small sample sizes in the firm-firm data.

The apparent narrowing of the three curves as the citation lag increases results in part from differences in scale between the curves. For this reason figure 2 normalizes each curve by the mean citation probability at a lag of one, which brings out differences in shape independently of scale. The normalized citation curve for universities is higher to begin with but converges on the firm citation curves. The differences are very slight in the normalized curves. From a visual perspective the diffusion of science proceeds at a very similar rate across citing and cited sectors of the U.S. economy.

C. Influence of Distance and Know-How on Diffusion

The next three figures examine the role of distance and of university "know-how", where this is represented by quality rankings of university science fields. Figure 3 graphs mean citation lags by the distance between citing and cited scientific institutions[12]. For this graph we take a close-in perspective.

---

[11] In panel B the cells are classified by citing industry, field and year; and by cited field and year. In panel C industry, field, and year on both citing and cited sides classify the cells.
[12] Since these are mean citation lags they are roughly twice as large as modal lags. See equations (5) and (8) on this point.

We examine the relationship between mean citation lag and distance in intervals of 50 miles as far as 500 miles between citing and cited universities and firms. A tendency is evident for the citation lag to increase with distance in this figure, especially in the case of firms citing universities. While it is not easy to be sure of the cause of the upward slope, it seems plausible that local partners are more likely to engage in research collaborations, which would speed up the local diffusion process. However the increase in the mean lag is small, about 0.3-0.5 years over 500 miles, less within universities, more outside, and most of this occurring within 150 miles or so..

Figure 4 re-examines the distance relationship from a wide-angle perspective. The mean citation lag is if anything declining. In the case of firms citing universities the lag reaches a peak at 1500 miles and then declines out to 3000 miles. We suspect that this "inverted U" picks up regional advantages of the east and west coasts.. In any event, it is hard to argue on the basis of figures 3 and 4 that science leaks out more slowly with distance, except for the slightly more rapid diffusion within a radius of 150 miles.

Figure 5 calculates normalized university citation curves relative to a lag of one for each curve. The graph reports separate citation curves for citing fields grouped in the top 20 percent, in the middle 40 percent, and in the bottom 40 percent of each field. The idea is to informally test whether more highly ranked universities and fields learn more rapidly about outside research. The figure suggests that top 20 percent citations occur perhaps slightly sooner than those made by lesser departments. But the difference is small and the speed of diffusion is about the same regardless of rank.

IV. Regression Findings for Universities

We turn now to quantitative results on diffusion which are derived from fitting the citation function (2) to the data for universities citing universities. These results allow us to distinguish speed of diffusion by science field separately from other aspects of the data. Tables 3 and 4 contain the results for the nonlinear estimation problem described by equation (2).

Table 3 reports findings on the speed of diffusion of science between universities. The first column of the table reports decay and diffusion parameters while the second reports estimated modal lags by field of science. At the bottom of the first column are decay and diffusion parameters β_1 and β_2, which represent the starting point of the analysis. The decay parameter is β_1 =0.351, indicating a modal lag of 2.85 years for the baseline field, chemistry ($1/\beta_1$ =1/0.353=2.85). The diffusion parameter is β_2 =0.000072. Taken together with β_1, this indicates a peak probability of citation of $\beta_2 / \beta_1 \approx 0.00020$.

The field decay rates are point estimates of the modal citation lags by field. Using equation (50 , in which the modal lag equals $1/\beta_1\beta_{1i}$, we find that the shortest lag is 1.75 years in physics, while the longest is 4.25 years, which occurs in computer science. The equally weighted mean of the modal lags across all fields is 3.37 years. One part of these differences in lags represents variation in the speed of acceptance and publication across fields, while another part reflects differences in propensities to collaborate, since joint work speeds up diffusion. Of course, fields with shorter modal lags also exhibit more rapid decay in scientific influence given the formulation of the citation function.

To see what role publication lags could be playing in the diffusion of science, we turn next to lags in institutional self-citations, where universities cite themselves. One would expect that publication delays play a less important role in this sample.

The third and fourth columns of table 3 report parameter estimates as well as estimates of the modal lags. Modal lags for self-citations occur about 1.5-2.0 years sooner than citations between universities. In chemistry the modal lag is 1.65 years and lags for the other fields are correspondingly shorter, yielding an equally weighted average modal lags of 1.95 years. This is 73 percent faster than the modal lag for citations between universities (3.37/1.95=1.73). Put another way, publication lags could account for as much as three-quarters of the lags in the diffusion of science. The shorter lags on self-citation are evidence of advanced notification: authors need not wait for publication to cite their own previous papers, as is usually required before papers by others can be cited. This circumvention of journal delays may also explain why modal lags on self-citation differ so little across fields. This narrowing of the range in modal lags can be easily seen by comparing columns four and two of table 3.

Table 4 revisits the citation function for universities citing other universities. On this occasion the analysis differentiates the speed of diffusion within fields from the speed of diffusion between fields. Decay rates between follow decay rates within, and are indicated by italics. Of course we expect cross-field diffusion to proceed more slowly because of barriers to the assimilation of technologically distant information. This hypothesis is usually accepted, though in three cases it is narrowly rejected and a field cites other fields more rapidly than itself. The three cases include the slowly diffusing fields of computer science, earth sciences, and psychology, whose publication lags could slow down within diffusion compared to between. Still, the hypothesis is on average confirmed, though the lags on cross-field citation are only slightly longer. The modal lag is 3.4 years for same-field citation but is 3.9 years for other-field citation, about half a year's difference. Notice that the within-field parameters in column one are usually the same up to the third decimal place as the total (within- and between-field) parameters in table 3, column one. This shows just how much within-field citation dominates the findings of table 3.

V. Regression Findings for Industry

A. Diffusion from University to Industry

We turn next to a study of the diffusion of science into industry. The results are based on citations of the top 200 R&D firms to the top 110 universities, on citations of the firms to each other, and on citations of the firms to themselves. We begin with the case of firms citing universities, in which citation function (3) is fitted to the data. Tables 5, 6, and figure 6 contain the findings. As before and for brevity's sake, the tables concentrate on exponential terms of the citation function.

Table 5 reports main decay and diffusion parameters, along with the estimated modal lags by field. The baseline modal lag for chemistry is slightly shorter than in table 3, column one, but the (equally weighted) mean citation lag is 3.4 years, about the same as in table 3. This confirms the visual intuition of figures 1 and 2, that there is little difference in the rate with which research diffuses within the university system and outside the system. In addition, the ranking of fields in tables 3 and 5 is roughly the same: physics remains the most rapidly diffusing field, and computer science continues to be among the slowest.

Table 6 explores the diffusion of science in greater depth. In this table we allow the rate of decay to vary by citing industry as well as field, which in turn changes the estimated modal lag (see (5) above). The analysis considers six major fields that are widely distributed across industries as well as evenly represented in citations from firms to universities as well as citations between firms. The six are biology, chemistry, computer science, engineering, medicine, and physics. Since the decay parameters vary across industries as well as field, there are 66 possible parameters representing the six fields and 11 industry groups. Making allowance for the fact that some sciences are negligible in several industries, there are 59 parameters left to be estimated. Given the large number of parameters, though, a way of summarizing the results is needed. Table 6 accomplishes this by reporting the fastest and slowest diffusion estimates within each of the six fields. In this way the table records industry variation by field. The range is substantial: in biology the fastest industry has a modal lag of 1.93 years and the slowest a modal lag of 4.56 years. Even physics with its short publication lags varies from 1.72 years to 2.56 years. On average the slowest industry in a field takes 1.7 years longer to diffuse than the fastest industry. In the more detailed results, not fully reported in the table, electrical equipment and communications stand out as industries to which university science diffuses rapidly, while academic research reaches metals and machinery more slowly.

Figure 6 looks directly at industry variation in the speed of diffusion of academic science. Figure 6 is a high-low graph where horizontal bars mark out medians of the modal lags. The medians are connected by a solid line to permit easy comparisons across industries. This way of presenting the results removes the influence of very fast or very slow fields in each industry, and suggests that certain industries are slower to absorb new university research (metals, machinery, miscellaneous), while others (electrical equipment, communications, and instruments) are faster. The average of the medians in the slower group yields a lag of 3.18 years, while the average of the medians in the faster group is 2.28 years.

B. Diffusion Within Industry

The empirical work concludes with a study of the diffusion of science between firms. Table 7 contains a set of basic results. The table confirms the earlier findings of tables 3 and 5 that science appears to leak out as rapidly in industry as in academia. The equally weighted modal lag for firms citing each other in column two is 3.5 years, which is nearly the same as the 3.4 years observed for universities citing each other in table 3, column two. That estimate in turn is virtually the same as the lag of 3.4 years observed for firms citing universities in table 5, column two. Of course, none of these estimates can *directly* capture delays in the permission to publish sensitive findings. And yet the almost identical lags across the academic and industrial sectors suggest that publication delays are quantitatively not very important in industrial science. If they were, firms would cite an older literature on average than the literature cited by universities, because an additional lag would afflict firm scientific publications but not academic publications. What the results do seem to suggest is that any delays and secrecy typically apply further down the chain of industrial innovation from pre-technology science.

Columns three and four of table 7 contain the results for self-citation by firms. The average lag in self-citation is 1.84 years, which is slightly less than the lag of 1.95 years for universities shown in column four of table 3. As in table 3, self-citations occur about a year and a half more rapidly than citations to other institutions. The lag structure is not very different from that prevailing in universities.

Table 8 explores the diffusion of science (between firms) by industry as well as field. As in table 6 we allow the decay parameter (β_1) to vary by citing industry as well as field. The analysis considers the six fields as table 6: biology, chemistry, computer science, engineering, medicine, and physics. Making allowances for the fact that several of the sciences are negligible in the citing and cited industries, 50 parameters are left to be estimated out of a possible 66. Given the large number of parameters we report the fastest and slowest diffusion estimates for each of the six fields, which displays industry variation by field in firm-firm diffusion. The range is large: in medicine the fastest industry has a modal lag of 2.40 years while the slowest has a modal lag of 6.11 years. Physics exhibits the smallest range of variation, from 1.79 years to 3.32 years, but this is still substantial. On average the slowest industry in a field takes 2.2 years longer to diffuse than the fastest industry. In the more detailed results drugs, electrical equipment and communications stand out as industries to which outside industrial science diffuses rapidly, while research results reach computers, metals and miscellaneous more slowly.

Figure 7 is a high-low graph of the variation in modal lags for citations between firms. Like figure 6 medians of the modal lags by industry are illustrated by horizontal bars. The medians as before are connected by a solid line that permits ready comparisons across industries. This suggests that certain industries are unusually slower to absorb new university research (metals, computers, and miscellaneous), while others (drugs, electrical equipment, and communications, and instruments) are unusually fast. The average of the medians in the slower group amounts to 3.97 years, while the average of the medians in the faster group is 2.53 years, about one and a half year's difference.

Taken together tables 3, 5, and 7 suggest that diffusion of scientific research on average is both rapid and similar across the academic and industrial sectors of the U.S. economy. However, as tables 6, 8 and figures 6, 7 show the close similarities in average speed of diffusion conceal substantial differences in the speed of diffusion of science across industry.

VI. Comparative Diffusion of Science and Technology

This section is concerned with two questions. First, how does the diffusion of scientific research compare with the diffusion of patented technology? And second, what is the significance for economic welfare of any of the findings on diffusion?

We begin by comparing our results to results from other papers that apply the citation function approach, but to patent citations. Jaffe and Trajtenberg (1996) estimate a citation function for U.S. patents citing patents by U.S. universities and the federal government with the goal of understanding the arrival of knowledge created in the public sector in the United States. Using a lag between patents generated by the difference between citing and cited years that the patents were granted, the paper finds that the modal lag is 4.69 years.

In a study of international patent citations Jaffe and Trajtenberg (1999) find that the modal lag for U.S. patents citing other U.S patents is 5.3 years in the baseline field of drug and medical patents, but that this lag varies slightly, from a low of 4.6 years to a high of 5.3 years, depending on the technology field. Thus the average modal lag is about five years in patent citations.

Popp (2002) studies the diffusion of energy-saving technologies as part of a project aimed at understanding the role of energy prices and of energy-related knowledge on the creation of new energy-saving technologies. His citation function uses the *application* year of citing patents and the granting year of cited patents, on the grounds that the application year captures the true date of the citing invention and the grant year on the cited invention captures the date at which the information becomes public. This approach necessarily shortens the citation lag, and indeed the modal lag is 2.8 years. Since on average patents require two years from application to grant, this result is about the same as the five year lag in Jaffe and Trajtenberg (1996, 1999). Notice that grant years for patents are analogous to publication years for scientific papers. The five year modal lag in technology is comparable to a 3.5 year modal lag in science. On this basis the rate of diffusion in science can be judged to be 30 percent faster. Part of this difference between the rate at which science and technology leak out is due to the fact the lag between application date and grant date is two years in the case of patents, whereas Ellison (2002) finds that publication lags in large areas of science are one year or less. But another part is due an intrinsically faster rate of diffusion in science.

Branstetter (2004) estimates the citation function on U.S. patents citing *scientific papers* of California's research universities. Citing year is the grant date while cited year is the publication date. This concept of citation lag compares closely with that used in Jaffe and Trajtenberg (1996, 1999), except that the lags run from invention to science rather than from invention to invention. The modal lag between patents and scientific papers is 8.33 years, and this is the most frequent time it takes for science to move between researchers in science, and then from industrial science to industrial invention[13].

One might speculate as to whether the modal lag of 3.5 years within science *plus* the modal lag of five years within technology combines to produce a lag of 8.33 years from papers to patents, but more work would be needed to verify this claim. What seems clear from all these results, though, is that citation lags in science are shorter than citation lags in technology. This could well be an advantage of the Open Science system, since shorter lags diminish the gap between technology-in-use and best-practice technology, and raise the path of output over time (Nelson and Phelps, 1966).

VII. Conclusion

In this paper we have demonstrated the rapid diffusion of science between scientific researchers, not only between universities, but also between firms, and between firms and universities. The modal lag in citation, a fairly robust indicator of the speed of diffusion, is 3.5 years, compared with 5.0 years based on patent data using the same formulation. Diffusion seems to be about 30 per cent faster between scientists than between inventors. It is important to see that this evidence provides a lower bound on the speed of diffusion between science and technology. As noted, that requires an additional diffusion of science in industry to industrial inventors as Branstetter's (2004) results imply. Our findings offer tentative support for the hypothesis that Open Science contributes to more rapid diffusion than Proprietary Technology. Curiously, diffusion of scientific research does not appear to be much delayed in industry. For those allowed to do science, incentives to publish appear to be strong regardless of sector.

[13] Adams (1990) uncovers a mean lag of 20 years for the *peak* effect of stocks of scientific papers on productivity growth. Since the mean lag for the citation function is twice the modal lag (see (8)), a modal lag of 8.33 years corresponds to a mean of 16.667 years, not very different from 20 years.

Certain scientific fields stand out for the rapidity with which their research disperses. Most rapid of all is physics, whose modal lag is 1.75 years. Certain others, such as computer science and economics, diffuse slowly. Some of the differences are due to publication lags, as the results on self-citation demonstrate. Still others are due to differences in collaboration practices, which are strong in rapidly diffusing fields such as physics (Adams, Black, Clemmons and Stephan, 2004).

Looking ahead we would like in principle to go back to the roots of the subject, as set forth in the pioneering work of Griliches (1957) and Mansfield (1963), and find out why diffusion varies across fields of science and by industry.. Part of the answer lies in field differences in the cost and time lags involved in the review process, which impact the industry differences in figures 6 and 7 as well. But another part lies purely within industries, in which some industry form part of the frontier while others lie well behind. Yet another question is what these differences imply for the efficiency of the invention production function. But these investigations must await another occasion.

References

Adams, James D., "Fundamental Stocks of Knowledge and Productivity Growth," **Journal of Political Economy** 98 (August 1990): 673-702.

______________, "Industrial R&D Laboratories: Windows on Black Boxes?" **Journal of Technology Transfer** 30 (January 2005): 129-137.

______________, Grant C. Black, J. Roger Clemmons, and Paula E. Stephan, "Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999", forthcoming in **Research Policy**.

______________, J. Roger Clemmons, and Paula E. Stephan, "Science and Industry: Tracing Basic Research Through Manufacturing and Trade," Working Paper, Rensselaer Polytechnic Institute, July 2004a.

____, "Standing on Academic Shoulders: Measuring Scientific Influence in Universities", NBER Working Paper #10875, October 2004b.

Audretsch, David, and Paula E. Stephan, "Company-Scientist Locational Links: the Case of Biotechnology," **American Economic Review** 86 (June 1996): 641-652.

Berliner, Joseph, **The Innovation Decision in Soviet Industry**, Cambridge, MA: MIT Press, 1976.

Branstetter, Lee, "Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations", Working Paper, Columbia University Business School: March 2004.

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.

Dearden, James, Barry W. Ickes, and Larry Samuelson, "To Innovate or Not To Innovate: Incentives and Innovation in Hierarchies," **American Economic Review** 80 (December 1990): 1105-1124.

Ellison, Glenn, "Evolving Standards for Academic Publishing: A q-r Theory," **Journal of Political Economy** 110 (October 2002): 994-1034.

Furman, Jeffrey L., and Scott Stern, “Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research,” Working Paper, Boston University and Northwestern University, March 2004.

Griliches, Zvi, “Hybrid Corn: An Exploration in the Economics of Technical Change,” **Econometrica** 25 (October 1957): 501-522.

Hart, Oliver, Andrei Shleifer, and Robert Vishny, “The Proper Scope of Government,” **Quarterly Journal Of Economics** 112 (November 1997): 1127-1162.

Henderson, Rebecca, and Iain Cockburn, “Scale, Scope, and Spillovers: the Determinants of Research Productivity in Drug Discovery,” **RAND Journal of Economics** 27 (Spring 1996): 32-59.

Jaffe, Adam B., and Manuel Trajtenberg, “Flows of Knowledge from Universities and Federal Laboratories,” **Proceedings of the National Academy of Sciences** 93 (November 1996): 12671-12677.

__________________________________, “International Knowledge Flows: Evidence from Patent Citations,” **Economics of Innovation and New Technology** 8 (1999): 105-136.

Levin, Richard C., Alvin K. Klevorick, Richard R. Nelson, and Sidney G. Winter, “Appropriating the Returns to Industrial Research and Development,” **Brookings Papers in Economic Activity** 1987 (Issue 3, 1987): 783-820.

Mansfield, Edwin, “Intra-firm Rates of Diffusion of an Innovation,” **The Review of Economics and Statistics** 45 (November 1963): 348-359.

_______________, “How Rapidly Does New Industrial Technology Leak Out?” **The Journal of Industrial Economics** 34 (December 1985): 217-213.

_______________, “Academic Research and Industrial Innovation,” **Research Policy** 20 (1991): 1-12.

_______________, Mark Schwartz, and Samuel Wagner, “Imitation Costs and Patents: An Empirical Study,” **The Economic Journal** 91 (December 1981): 907-918.

Nelson, Richard R., and Edmund S. Phelps, “Investment in Humans, Technological Diffusion, and Economic Growth”, **American Economic Review** 56 (March 1966): 69-75.

Popp, David, “Induced Innovation and Energy Prices”, **American Economic Review** 92 (March 2002): 160-180.

Ruttan, Vernon W., **Technology, Growth, and Development**, New York, Oxford University Press, 2001.

Scherer, F. Michael, **Innovation and Growth: Schumpeterian Perspectives**, Cambridge, MA: MIT Press, 1984.

Stephan, Paula E., "Robert K. Merton's Perspective on Priority and the Provision of the Public Good, Knowledge," **Scientometrics** 60 (2004): 81-87.

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

Table 1
Papers and Citations by 12 Main Fields of Science
The Top 110 U.S. Universities and the Top 200 U.S. R&D Firms, 1981-1999

Field	The Top 110 Universities		The Top 200 R&D Firms	
	Papers [a]	Citations from Top 110 Schools [b]	Papers [c]	Citations from Top 200 R&D Firms [d]
Agriculture	189,740 (7.8%)	730,777 (3.9%)	6,025 (2.5%)	6,850 (1.1%)
Astronomy	35,795 (1.5%)	371,982 (2.0%)	— [e]	— [e]
Biology	639,195 (26.3%)	8,339,862 (44.4%)	44,082 (18.5%)	187,617 (29.2%)
Chemistry	195,437 (8.0%)	1,371,491 (7.3%)	39,346 (16.5%)	106,943 (16.7%)
Computer Science	28,184 (1.2%)	76,424 (0.4%)	12,367 (5.2%)	17,916 (2.8%)
Earth Sciences	73,126 (3.0%)	566,280 (3.0%)	3,616 (1.5%)	5,011 (0.8%)
Economics	43,892 (1.8%)	161,813 (0.9%)	— [e]	— [e]
Engineering	170,569 (7.0%)	467,955 (2.5%)	50,203 (21.1%)	53,541 (8.3%)
Mathematics and Statistics	61,061 (2.5%)	187,484 (1.0%)	2,665 (1.1%)	3,471 (0.5%)
Medicine	659,000 (27.1%)	4,563,261 (24.3%)	26,739 (11.2%)	68,507 (10.7%)
Physics	217,026 (8.9%)	1,219,080 (6.5%)	50,346 (21.1%)	187,770 (29.2%)
Psychology	116,976 (4.8%)	727,673 (3.9%)	— [e]	— [e]

Source: Institute for Scientific Information and Computer Horizons, Inc. [a] Sum of university articles is 2,430,001. [b] Sum of university-university citations is 18,784,082. [c] Sum of firm articles is 235,389. [d] Sum of firm-firm citations is 637,626. [e] Astronomy, economics and psychology contribute less than one percent of firm papers and so are dropped from this table. Papers are assigned to journals according to the Journal-Field Assignment Method discussed in the text, in which journals are assigned to unique fields, and papers follow the field assignment of journals. Percentages are percents of column totals.

Table 2
Mean Citations, Papers Cited, Papers Citing, and Citation Probabilities
The Top 110 Universities and the Top 200 U.S. R&D Firms

Classification	Citations	Potential Papers Cited	Potential Papers Citing	Mean Citation Probability
Panel A. Citations between the Top 110 Universities				
Citing Fields				
Agriculture	487	14,609	13,537	0.19×10^{-4}
Astronomy	538	11,916	3,729	7.17×10^{-4}
Biology	4,158	16,840	51,541	0.21×10^{-4}
Chemistry	660	14,095	14,486	0.40×10^{-4}
Computer Science	55	12,262	2,462	1.37×10^{-4}
Economics and Business	207	14,317	3,720	2.28×10^{-4}
Engineering	230	12,543	13,369	0.13×10^{-4}
Earth Sciences	435	12,310	6,375	1.51×10^{-4}
Mathematics and Statistics	103	12,400	4,669	0.56×10^{-4}
Medicine	3,614	19,355	54,570	0.14×10^{-4}
Physics	1,333	13,731	19,270	0.51×10^{-4}
Psychology	605	14,025	9,403	0.81×10^{-4}
Panel B. Citations from the Top 200 R&D Firms to the Top 110 Universities				
Citing Fields				
Agriculture	15	25,723	112	0.99×10^{-5}
Biology	120	26,348	579	0.87×10^{-5}
Chemistry	33	17,689	301	1.21×10^{-5}
Computer Science	13	4,564	180	5.81×10^{-5}
Engineering	16	10,690	333	0.83×10^{-5}
Earth Sciences	20	14,926	65	11.06×10^{-5}
Mathematics and Statistics	6	3,030	32	7.54×10^{-5}
Medicine	106	30,467	556	0.85×10^{-5}
Physics	35	12,727	367	0.98×10^{-5}
Citing Industries				
Petrochemicals (13, 28 except 283, 29-30)	37	20,037	323	1.20×10^{-5}
Pharmaceuticals & Biotechnology (283)	182	22,816	994	1.49×10^{-5}
Primary and Fabricated Metals (33,34)	7	18,781	37	1.60×10^{-5}
Machinery, Except Computers (35, except 357)	6	10,009	44	1.63×10^{-5}
Computers (357)	14	13,990	124	1.95×10^{-5}
Electrical Equipment (36)	22	12,539	319	5.17×10^{-5}
Transportation Equipment (37)	21	14,964	214	5.12×10^{-5}
Instruments (38)	20	22,717	119	2.88×10^{-5}
Communications Services (48)	28	11,097	265	7.39×10^{-5}
Computer Software & Services (737)	25	13,298	268	4.64×10^{-5}
All Other	10	23,119	41	1.49×10^{-5}

Table 2
Mean Citations, Papers Cited, Papers Citing, and Citation Probabilities
The Top 110 Universities and the Top 200 U.S. R&D Firms

Classification	Citations	Potential Papers Cited	Potential Papers Citing	Mean Citation Probability
Panel C. Citations between the Top 200 R&D Firms				
Citing Fields				
Agriculture	2	606	134	0.85×10^{-4}
Biology	20	568	1172	0.81×10^{-4}
Chemistry	5	415	346	0.63×10^{-4}
Computer Science	3	232	215	1.40×10^{-4}
Engineering	3	429	466	0.25×10^{-4}
Earth Sciences	3	89	103	3.52×10^{-4}
Mathematics and Statistics	2	52	37	12.15×10^{-4}
Medicine	13	654	867	0.46×10^{-4}
Physics	7	398	410	0.59×10^{-4}
Citing Industries				
Petrochemicals (13, 28 except 283, 29-30)	5	464	400	0.59×10^{-4}
Pharmaceuticals & Biotechnology (283)	22	511	1,405	0.33×10^{-4}
Primary and Fabricated Metals (33,34)	2	642	44	0.85×10^{-4}
Machinery, Except Computers (35, except 357)	2	637	37	1.40×10^{-4}
Computers (357)	4	421	151	0.81×10^{-4}
Electrical Equipment (36)	5	341	478	0.88×10^{-4}
Transportation Equipment (37)	4	408	357	0.45×10^{-4}
Instruments (38)	3	604	126	0.63×10^{-4}
Communications Services (48)	6	363	469	0.64×10^{-4}
Computer Software & Services (737)	5	348	374	0.60×10^{-4}
All Other	2	881	38	0.91×10^{-4}

Notes: The table entries are means over all cells and not sums. The number of cells that enter into the calculations for Panel A is 36,834. Citing and cited fields and citing and cited years classify these university-university cells. For Panel B, which consists of firm-university cells, this number is 30,604. In this case citing industry, citing and cited fields, and cited fields and years classify the cells. The number of cells in Panel C is 34,246. In the firm-firm cells the classifying variables are citing and cited industries, fields, and years.

Table 3
Diffusion Estimates: Citations Between the Top 110 U.S. Universities

Parameter, Science Field	**Citations to Other Universities [a]**		**Self-Citations to the Same University [b]**	
	Estimate (St. Error)	**Modal Lag [c]**	**Estimate (St. Error)**	**Modal Lag [c]**
Decay Parameters (β_{1i})				
Agriculture	0.783 (0.022)	3.64	0.781 (0.012)	2.11
Astronomy	1.046 (0.012)	2.72	0.867 (0.005)	1.90
Biology	1.090 (0.016)	2.61	0.894 (0.010)	1.84
Chemistry	1.000 (--)	2.85	1.00 (--)	1.65
Computer Science	0.671 (0.011)	4.25	0.866 (0.011)	1.90
Economics and Business	0.678 (0.009)	4.20	0.732 (0.011)	2.25
Engineering	0.744 (0.028)	3.83	0.907 (0.017)	1.82
Earth Sciences	0.854 (0.011)	3.34	0.743 (0.005)	2.22
Mathematics and Statistics	0.708 (0.014)	4.02	0.834 (0.011)	1.98
Medicine	0.922 (0.019)	3.09	0.814 (0.018)	2.02
Physics	1.632 (0.022)	1.75	1.211 (0.012)	1.36
Psychology	0.693 (0.010)	4.11	0.698 (0.008)	2.36
Baseline Decay Parameter (β_1)*	0.351 (0.004)	—	0.607 (0.003)	—
Diffusion Parameter (β_2)*	1.08×10^{-4} (3.23×10^{-6})	—	4.95×10^{-4} (5.86×10^{-6})	—
Average Modal Lag	—	3.37	—	1.95

Notes: The equation includes intercept terms for citing and cited fields, cited years and citing intervals, which are significant. [a] The number of observations is 36,834. The adjusted R^2=0.938. The estimated standard error of the regression (root mean squared error) is 0.0010. [b] The number of observations is 21,801. The adjusted R^2=0.952. The estimated standard error of the regression (root mean squared error) is 0.0006. [c] The modal lag equals the reciprocal of $\beta_1\beta_{1i}$. See (5) in the text.

Table 4
Diffusion Estimates: Citations Between the Top 110 U.S. Universities, Within and Between Sciences

Parameter, Citing Science	Cited Group	Citations to Other Universities [a] Estimate (St. Error)	Modal Lag [b]
Decay Parameters (β_{1i})			
Agriculture	Same Field	0.786 (0.023)	3.62
"	*All Others*	*0.679 (0.107)*	*4.20*
Astronomy	Same Field	1.044 (0.012)	2.73
"	*All Others*	*0.891 (0.119)*	*3.20*
Biology	Same Field	1.090 (0.016)	2.61
"	*All Others*	*0.958 (0.097)*	*2.97*
Chemistry	Same Field	1.000 (--)	2.85
"	*All Others*	*0.908 (0.171)*	*3.14*
Computer Science	Same Field	0.671 (0.011)	4.25
"	*All Others*	*0.751 (0.203)*	*3.79* [R]
Economics and Business	Same Field	0.678 (0.009)	4.20
"	*All Others*	*0.423 (0.071)*	*6.74*
Engineering	Same Field	0.749 (0.029)	3.80
"	*All Others*	*0.577 (0.109)*	*4.94*
Earth Sciences	Same Field	0.853 (0.010)	3.34
"	*All Others*	*0.940 (0.125)*	*3.03* [R]
Mathematics and Statistics	Same Field	0.707 (0.014)	4.03
"	*All Others*	*0.685 (0.301)*	*4.16*
Medicine	Same Field	0.921 (0.019)	3.09
"	*All Others*	*0.911 (0.070)*	*3.13*
Physics	Same Field	1.631 (0.022)	1.75
"	*All Others*	*1.135 (0.216)*	*2.51*
Psychology	Same Field	0.692 (0.010)	4.12
"	*All Others*	*0.701 (0.156)*	*4.06* [R]
Baseline Decay Parameter (β_1)*		0.351 (0.004)	—
Diffusion Parameter (β_2)*		1.08×10^{-4} (3.25×10^{-6})	—
Average Modal Lag, Same Field	—	—	3.37
Average Modal Lag, *All Others*	—	—	*3.82*

Notes: The equation includes intercept terms for citing and cited fields, cited years and citing intervals, which are significant. [a] The number of observations is 36,834. The adjusted R^2=0.938. The estimated standard error of the regression (root mean squared error) is 0.0010. [R] Citation lag to other fields is shorter than citation lag within a field and represents a reversal in the relative speed of within field diffusion. [b] The modal lag equals the reciprocal of $\beta_1\beta_{1i}$. See (5) of the text.

Table 5
Diffusion Estimates: Citations from the Top 200 U.S. R&D Firms
To the Top 110 U.S. Universities

Parameter, Science Field	**Citations to Universities** [a]	
	Estimate (St. Error)	**Modal Lag** [b]
Decay Parameters (β_{1i})		
Agriculture	0.765 (0.056)	3.35
Astronomy	0.934 (0.024)	2.75
Biology	0.997 (0.036)	2.57
Chemistry	1.000 (--)	2.56
Computer Science	0.622 (0.020)	4.12
Economics and Business	0.672 (0.023)	3.82
Engineering	0.655 (0.040)	3.91
Earth Sciences	0.819 (0.023)	3.13
Mathematics and Statistics	0.553 (0.027)	4.64
Medicine	0.886 (0.043)	2.89
Physics	1.242 (0.045)	2.06
Psychology	0.532 (0.031)	4.82
Baseline Decay Parameter (β_1)*	0.390 (0.010)	—
Diffusion Parameter (β_2)*	0.69×10^{-4} (3.75×10^{-6})	—
Average Modal Lag	—	3.39

Notes: The equation includes intercept terms for citing and cited fields, cited years and citing intervals, which are significant. [a] The number of observations is 30,604. The adjusted R^2=0.711. The estimated standard error of the regression (root mean squared error) is 0.0009. [b] The modal lag equals the reciprocal of $\beta_1\beta_{1i}$. See equation (5) in the text.

Table 6
Diffusion Estimates by Field and Industry:
Citations from the Top 200 U.S. R&D Firms to the Top 110 U.S. Universities

Decay Parameter, Science Field	Citations to Universities [a]			
	Industry of Slowest Diffusion [b]		Industry of Fastest Diffusion [b]	
	β_{1ij} Estimate (St. Error)	Modal Lag [c]	β_{1ij} Estimate (St. Error)	Modal Lag [c]
Biology	0.480 (0.078)	4.56	1.131 (0.042)	1.93
Chemistry	0.609 (0.040)	3.59	1.028 (0.030)	2.13
Computer Science	0.385 (0.013)	5.68	0.697 (0.019)	3.14
Engineering	0.473 (0.021)	4.63	0.714 (0.057)	3.06
Medicine	0.648 (0.069)	3.38	1.074 (0.051)	2.04
Physics	0.852 (0.056)	2.56	1.270 (0.029)	1.72
Average Modal Lag	____	4.07	____	2.34

Notes: The equation includes intercept terms for citing and cited fields, cited years, and citing intervals, which are significant. [a] The number of observations is 22,855. The adjusted R^2=0.800. The estimated standard error of the regression (root mean squared error) is 0.000274. [b] The identity of the industry for which diffusion is the slowest or the fastest of course varies across the six science fields. [c] The modal lag equals the reciprocal of $\beta_1\beta_{1ij}$, where $\beta_1 = 0.457$ is the estimate of the rate of decay within petrochemicals in the field of chemistry, and subscripts ij stand for citing industry and field. See equation (5) in the text.

Table 7
Diffusion Estimates: Citations Within and Between the Top 200 U.S. R&D Firms

Parameter, Science Field	Citations to Other Firms [a]		Self-Citations to the Same Firm [b]	
	Estimate (St. Error)	Modal Lag [c]	Estimate (St. Error)	Modal Lag [c]
Decay Parameters (β_{1i})				
Agriculture	0.736 (0.043)	4.21	0.932 (0.113)	1.63
Biology	1.050 (0.018)	2.95	0.948 (0.038)	1.60
Chemistry	1.000 (--)	3.10	1.000 --	1.52
Computer Science	0.800 (0.020)	3.87	0.700 (0.118)	2.17
Engineering	1.019 (0.030)	3.83	0.888 (0.089)	1.71
Earth Sciences	0.742 (0.025)	4.17	0.526 (0.016)	2.89
Mathematics and Statistics [d]	0.904 (0.024)	3.42	--	--
Medicine	0.820 (0.027)	3.78	1.024 (0.040)	1.48
Physics	1.371 (0.023)	2.26	0.884 (0.032)	1.72
Baseline Decay Parameter (β_1)*	0.323 (0.005)	—	0.658 (0.019)	—
Diffusion Parameter (β_2)*	0.93×10^{-4} (5.89×10^{-6})	—	13.46×10^{-4} (2.28×10^{-4})	—
Average Modal Lag	—	3.51	—	1.84

Notes: The equation includes intercept terms for citing and cited fields, cited years and citing intervals, which are significant. [a] The number of cells is 34,246. The adjusted R^2=0.580. The estimated standard error of the regression (root mean squared error) is 0.0010. [b] The number of observations is 10,687. The adjusted R^2=0.805. The estimated standard error of the regression (root mean squared error) is 0.0124. [c] The modal lag equals the reciprocal of $\beta_1\beta_{1i}$. See equation (5) in the text. [d] Self-citations by industrial firms in mathematics and statistics are rare enough that we include them in a remainder diffusion term. That term is insignificant so it is not reported here.

Table 8
Diffusion Estimates by Field and Industry:
Citations Between the Top 200 U.S. R&D Firms

Decay Parameter, Science Field	**Citations to Other Firms [a]**			
	Industry of Slowest Diffusion [b]		**Industry of Fastest Diffusion [b]**	
	β_{1ij} Estimate (St. Error)	**Modal Lag [c]**	**β_{1ij} Estimate (St. Error)**	**Modal Lag [c]**
Biology	0.610 (0.023)	3.87	1.052 (0.046)	2.24
Chemistry	0.491 (0.028)	4.80	1.013 (0.050)	2.33
Computer Science	0.488 (0.029)	4.83	0.684 (0.031)	3.45
Engineering	0.543 (0.098)	4.34	1.119 (0.058)	2.11
Medicine	0.386 (0.029)	6.11	0.957 (0.064)	2.46
Physics	0.711 (0.035)	3.32	1.318 (0.057)	1.79
Average Modal Lag	____	4.55	____	2.40

Notes: The equation includes intercept terms for citing and cited fields, cited years, and citing intervals, which are significant. [a] The number of observations is 31,770. The adjusted R^2=0.545. The estimated standard error of the regression (root mean squared error) is 0.00103. [b] The identity of the industry for which diffusion is the slowest or the fastest of course varies across the six science fields. [c] The modal lag equals the reciprocal of $\beta_1 \beta_{1ij}$, where $\beta_1 = 0.424$ is the estimate of the rate of decay within petrochemicals in the field of chemistry, and subscripts and ij stands for citing industry and field. See equation (5) in the text.

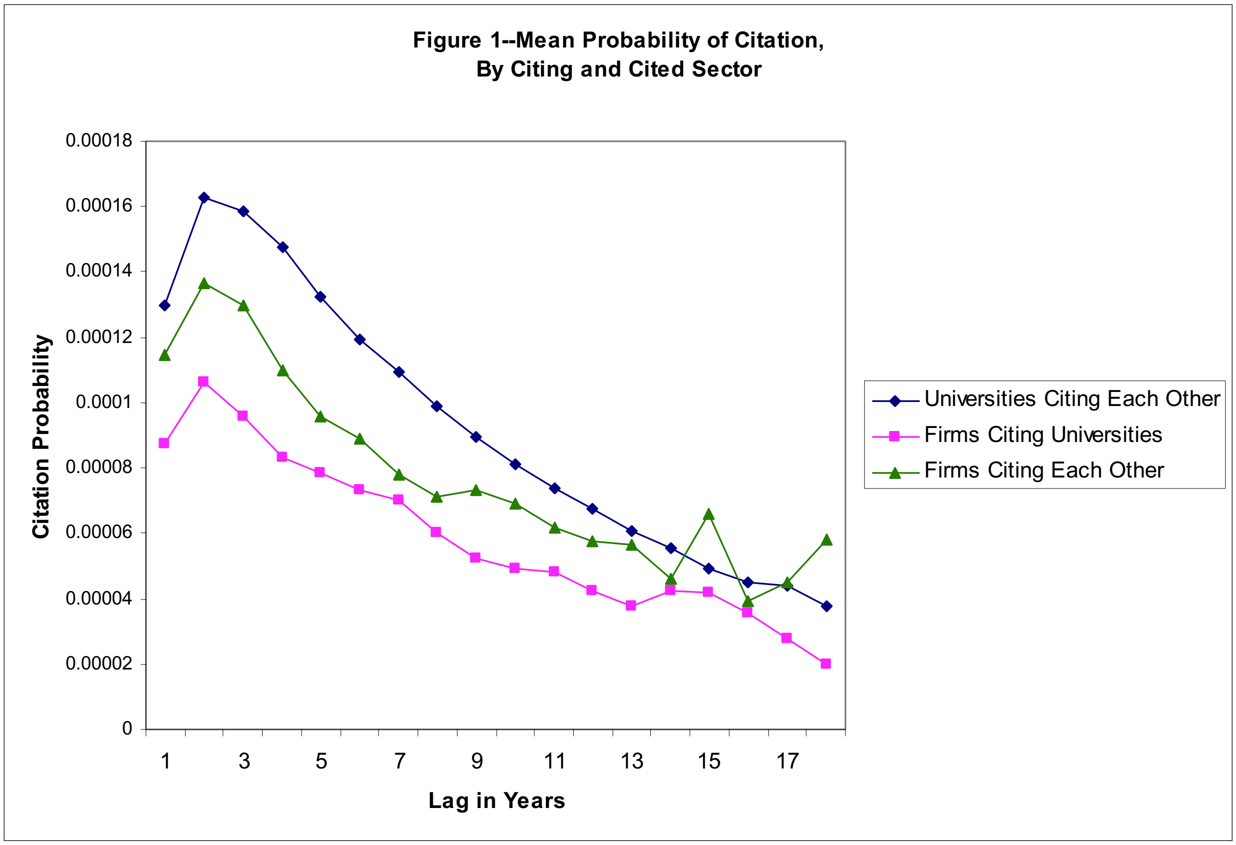
Figure 1--Mean Probability of Citation,
By Citing and Cited Sector
0.00018
0.00016
0.00014
0.00012
0.0001
0.00008
0.00006
0.00004
0.00002
0
Citation Probability
1
3
5
7
9
11
13
15
17
Lag in Years
Universities Citing Each Other
Firms Citing Universities
Firms Citing Each Other

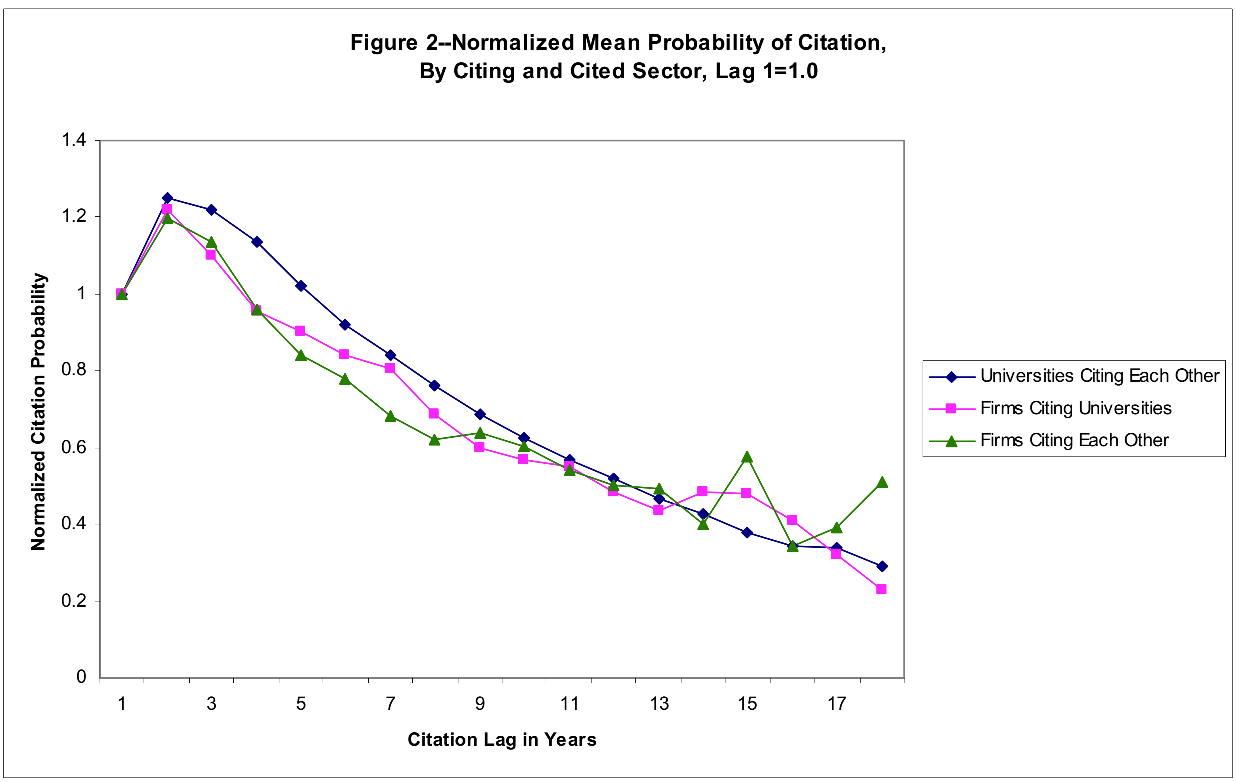
Figure 2--Normalized Mean Probability of Citation,
By Citing and Cited Sector, Lag 1=1.0
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Normalized Citation Probability
1
3
5
7
9
11
13
15
17
Citation Lag in Years
Universities Citing Each Other
Firms Citing Universities
Firms Citing Each Other

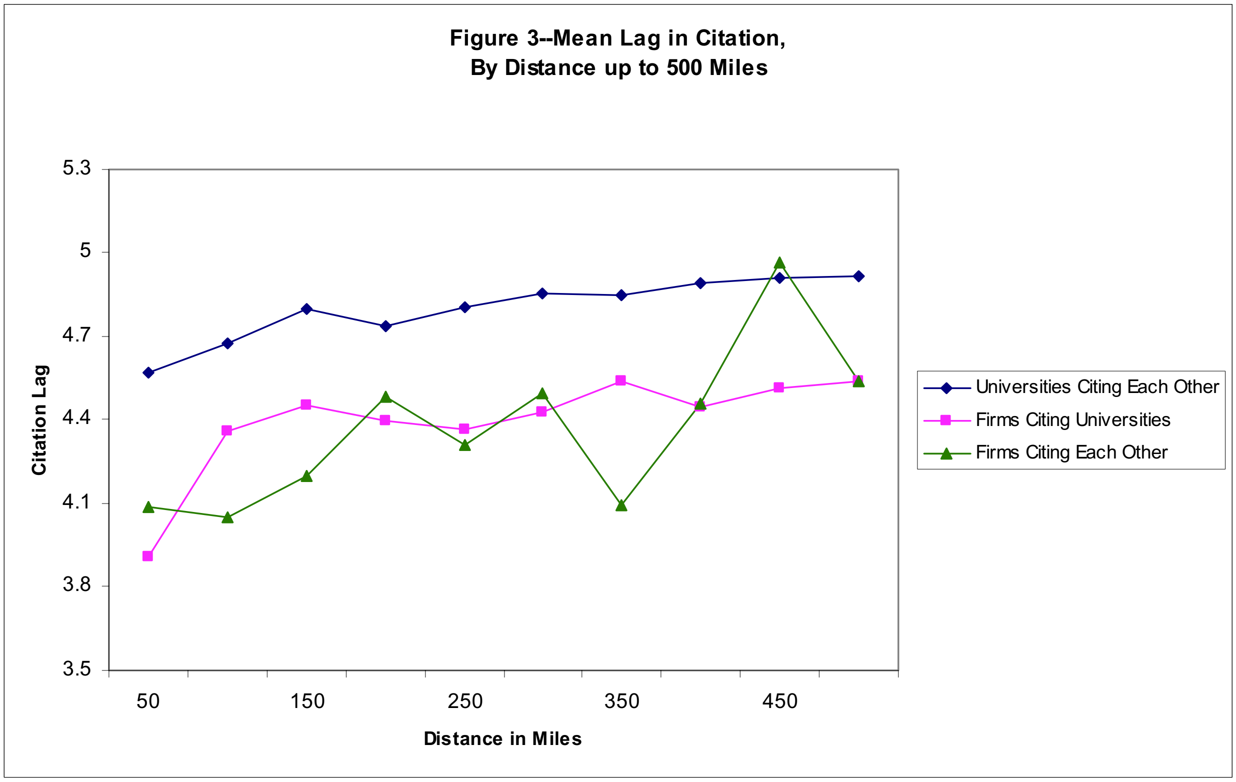
Figure 3--Mean Lag in Citation,
By Distance up to 500 Miles
5.3
5
4.7
4.4
4.1
3.8
3.5
Citation Lag
50
150
250
350
450
Distance in Miles
Universities Citing Each Other
Firms Citing Universities
Firms Citing Each Other

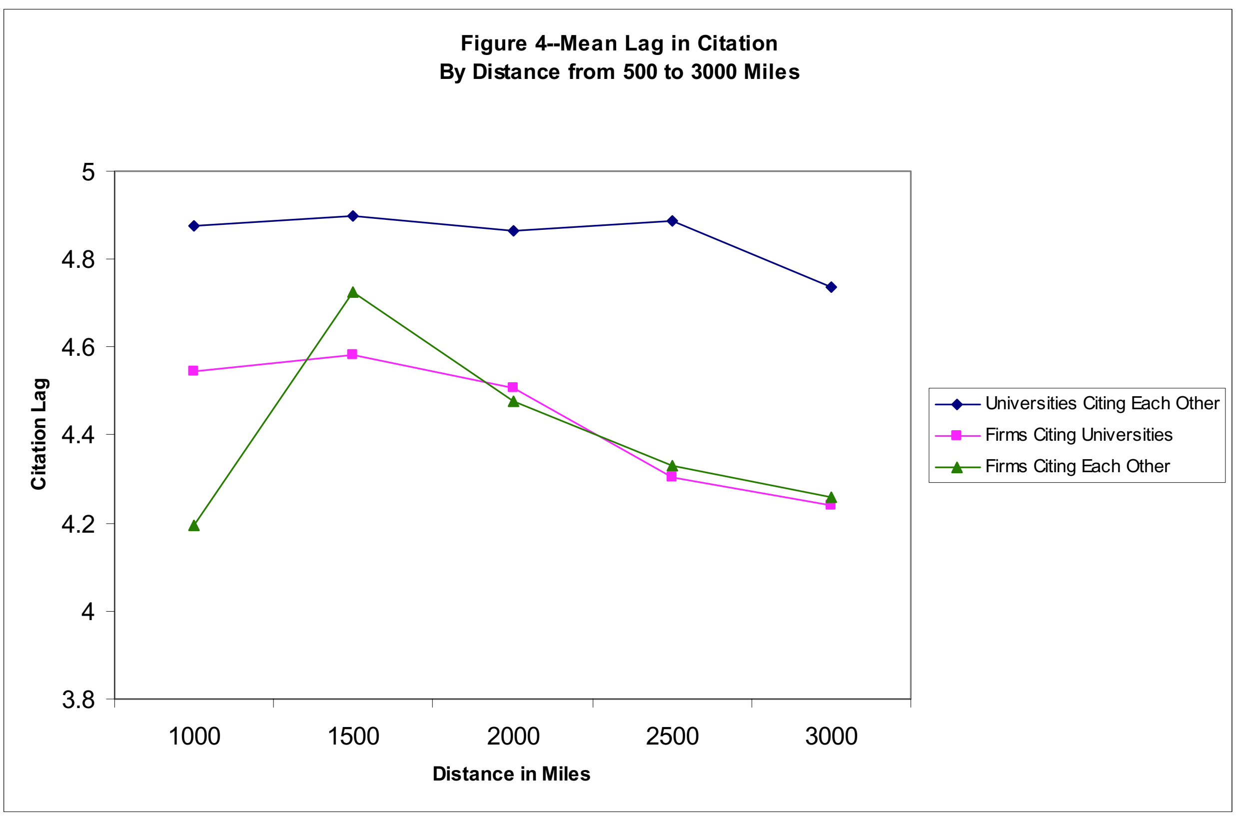
Figure 4--Mean Lag in Citation
By Distance from 500 to 3000 Miles
5
4.8
4.6
4.4
4.2
4
3.8
Citation Lag
1000
1500
2000
2500
3000
Distance in Miles
Universities Citing Each Other
Firms Citing Universities
Firms Citing Each Other

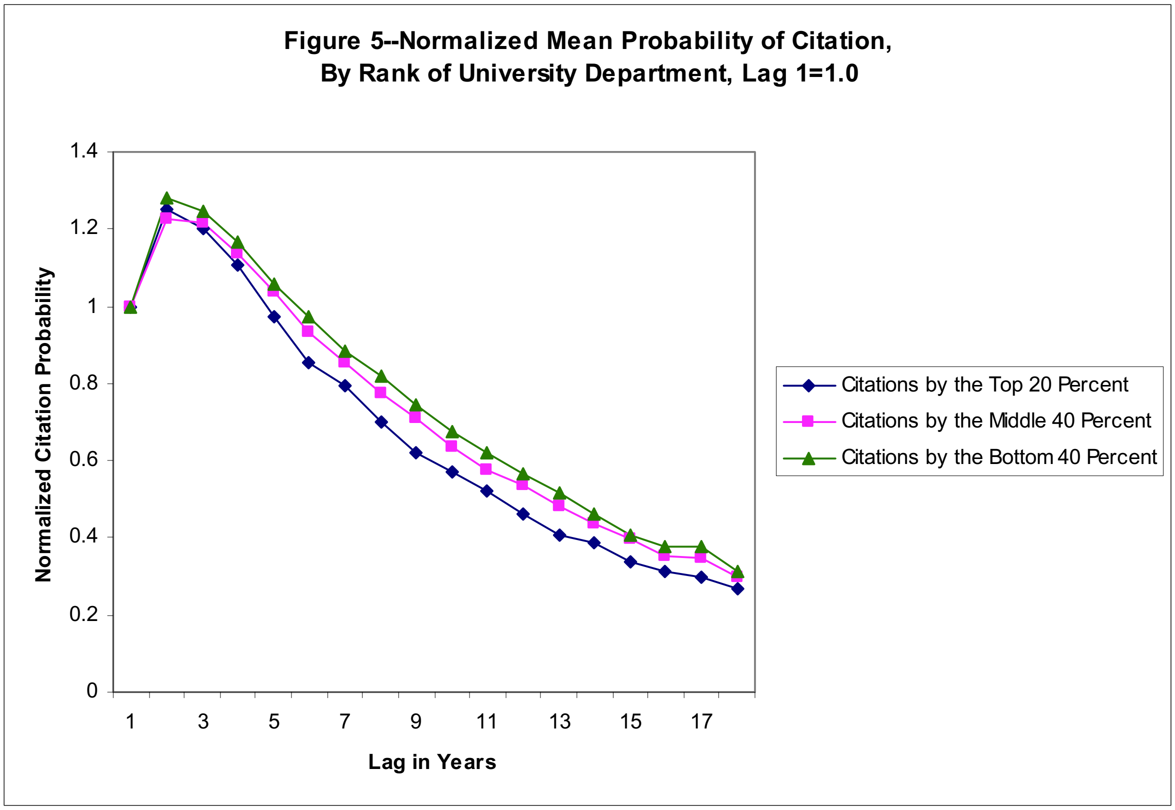

Figure 5--Normalized Mean Probability of Citation,
By Rank of University Department, Lag 1=1.0
Normalized Citation Probability
Lag in Years
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1
3
5
7
9
11
13
15
17
Citations by the Top 20 Percent
Citations by the Middle 40 Percent
Citations by the Bottom 40 Percent

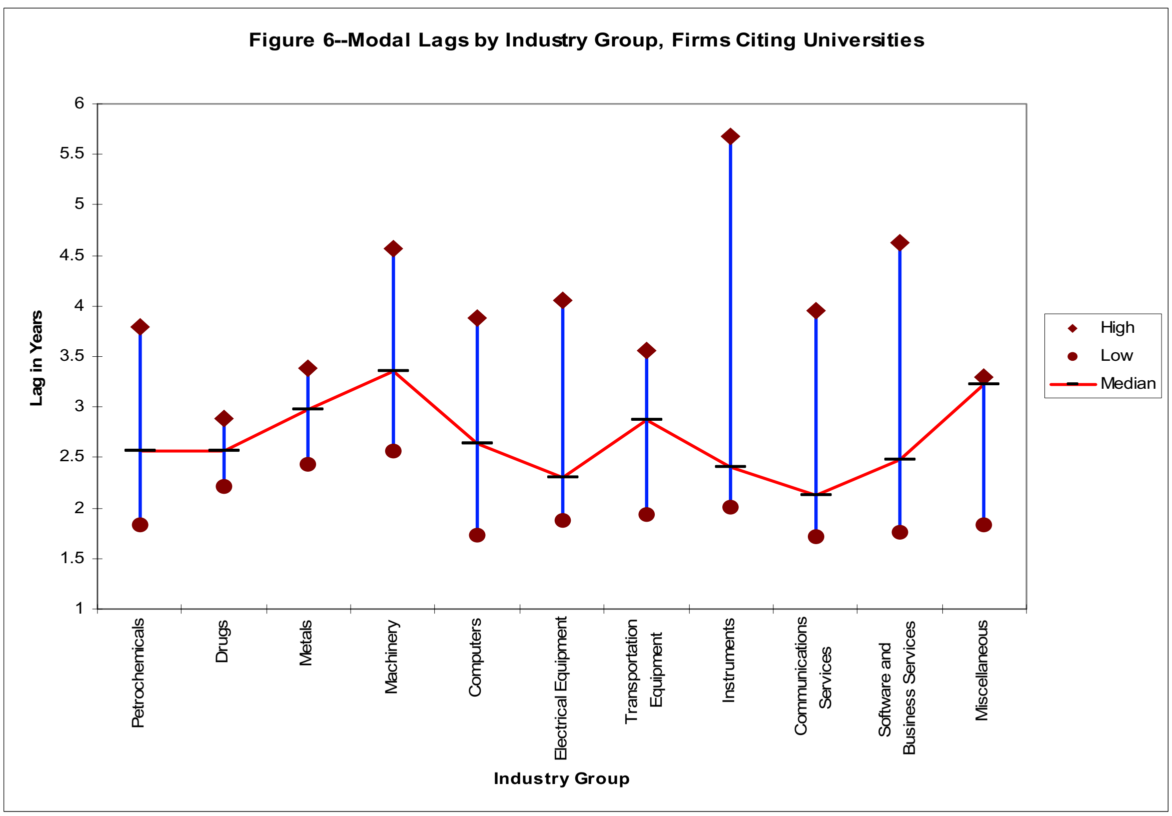

Figure 6--Modal Lags by Industry Group, Firms Citing Universities
Lag in Years
Industry Group
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
Petrochemicals
Drugs
Metals
Machinery
Computers
Electrical Equipment
Transportation Equipment
Instruments
Communications Services
Software and Business Services
Miscellaneous
High
Low
Median

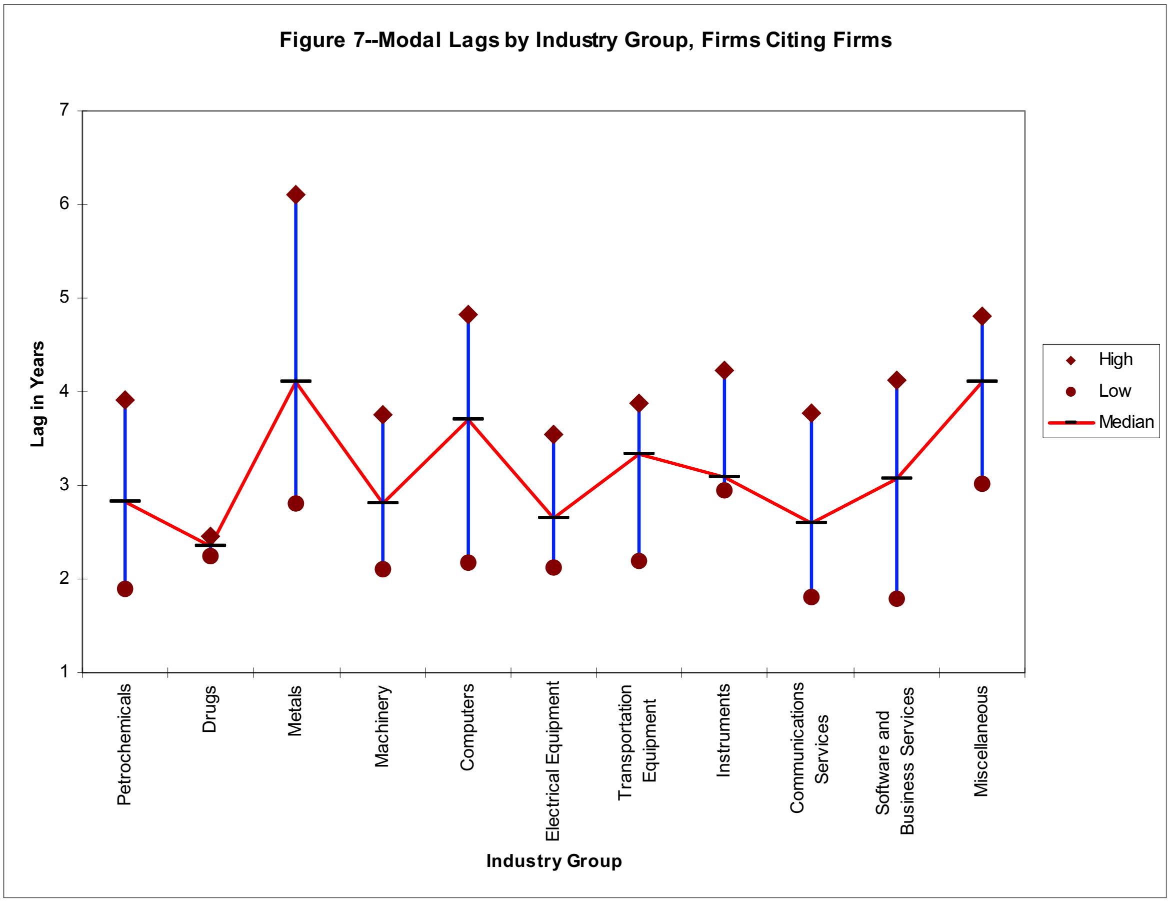

Figure 7--Modal Lags by Industry Group, Firms Citing Firms
Lag in Years
7
6
5
4
3
2
1
High
Low
Median
Petrochemicals
Drugs
Metals
Machinery
Computers
Electrical Equipment
Transportation Equipment
Instruments
Communications Services
Software and Business Services
Miscellaneous
Industry Group