NBER WORKING PAPER SERIES

THE ORIGINS OF INDUSTRIAL SCIENTIFIC DISCOVERIES

James D. Adams
J. Roger Clemmons

Working Paper 13823
http://www.nber.org/papers/w13823

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2008

The Andrew W. Mellon Foundation has generously supported this research. We thank Nancy Bayers and Henry Small of Thomson Scientific for advice concerning the data. We also thank Ernie Berndt, Nick Bloom, Wes Cohen, Wayne Gray, Mike Scherer, Scott Stern, and John Van Reenan for helpful comments. Presentations at the NBER Productivity Meetings, the NBER productivity seminar, Rensselaer Polytechnic Institute, and the University at Albany have substantially improved this paper. Any errors are our responsibility. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

The Origins of Industrial Scientific Discoveries
James D. Adams and J. Roger Clemmons
NBER Working Paper No. 13823
February 2008
JEL No. D24,L33,O31,O33

ABSTRACT

This paper estimates science production functions for R&D-performing firms in the United States using scientific papers as the measure of output, by analogy with patents. The underlying evidence covers 200 top U.S. R&D firms during 1981-1999 as well as 110 top U.S. universities. We find that industrial science builds on past scientific research inside and outside the firm, with most of the returns to scale in production deriving from outside knowledge. In turn, the largest outside contribution derives from universities rather than firms; this is especially true when papers are weighted by citations received, a measure of their importance. Consistent with the role assigned to knowledge spillovers in growth theory, the importance of outside knowledge, especially that of universities, increases from the firm to the industry level. The findings survive the inclusion of fixed effects, interactions among the effects, variations in sample and specification, and efforts to control for endogeneity.

James D. Adams
Department of Economics
Rensselaer Polytechnic Institute
3406 Russell Sage Laboratory
Troy, NY 12180-3590
and NBER
adamsj@rpi.edu

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

"Empirically [knowledge capital] is too broad a concept: it aspires to and contains too much. We can, however, focus on the contribution of identified investments in advancing the state of knowledge in a particular (or related) area(s). The contribution of 'science' in general to a particular industry is probably not measurable since there is no way of knowing how much 'science' is actually used in one industry versus another."

—Zvi Griliches, "Issues in Addressing the Contribution of R&D to Productivity," **Bell Journal of Economics** 10 (1979)

I. Introduction

Against the odds, and despite the verdict of a master, this paper sets out to show that the contribution of science to firms and industries is to some degree measurable. In this article our main interest is in the early stages of innovation and in the role that science inside and outside the firm plays in scientific discoveries. We assume that science is productive and contributes to expected profits. This is a persuasive reason as to why firms pursue science and we adopt it as our working hypothesis[1].

The analysis builds on a simple precept from the economics of search, that industrial science replenishes technological opportunities in firms and industries (Evenson and Kislev, 1976)[2]. Proceeding on this basis, we examine how opportunities are replenished and how the firm's science resources interact with stocks of knowledge to produce industrial science. Thus our concern is with industrial science, not with patents, products, and stock market values[3].

We would first like to be clear about what we do not claim. We do not claim that only industrial science replenishes technological opportunity, just that it is a factor. Nor do we assert that only academic science matters, for clearly we assume that industrial science is an essential piece of the puzzle. Instead we merely claim that the return to R&D rises in part because of science regardless of sector.

[1] But see Stern (2004), who argues that letting researchers do science also yields a wage savings.
[2] Adams (1990) draws on the insights of Evenson and Kislev (1975, 1976).
[3] Adams and Clemmons (2007) consider the role of science in patent production functions.

To answer this question we pursue a flexible approach. We allow the firm's past scientific research, its stock of basic research, and the past scientific research of universities as well as other firms to influence the rate of scientific discovery of a firm. Knowledge is allowed to flow from any science and from any university or firm to any particular firm and field[4]. In the case of universities we provide for collaboration as well as citation channels of influence[5].

We are motivated by a desire to understand the sources that firms draw on while doing their own science. Thus we are interested in the "origins" question, of the comparative contributions of universities and other firms, as well as the firm itself, to its scientific discoveries.

The data give us the freedom to explore these dimensions. They are based on scientific papers, citations, and collaborations during 1981-1999 collected by Thomson Scientific in Philadelphia, Pennsylvania. The papers are written by scientists in the top 200 U.S. R&D firms and the top 110 U.S. universities. These institutions account for most of the scientific research carried out in the U.S. during this period.

The data include 230 thousand papers written by the top 200 firms and 2.4 million papers written by the top 110 universities. They report roughly one million citations of top 200 firms to papers of the top 110 universities, about 40 thousand collaborations between firms and universities, and over 600 thousand citations to firms. From these data we extract a panel consisting of science outputs and inputs at the three-dimensional level of firms, fields, and years.

[4] Adams and Clemmons (2008, forthcoming) describe the data and characterize dimensions of the knowledge flows in terms of firms and industries, universities, and science fields.
[5] Firms write some of their papers with collaborating institutions which are almost entirely universities.

This paper draws on studies of knowledge production functions in microeconomics (Griliches, 1979, 1992) and in growth theory (Romer, 1990; Jones, 1995; Aghion and Howitt, 1998). We explore a particular implication of the growth side of this literature, that knowledge is more important at the industry level than firm level.

Given our emphasis on science inside and outside the firm, we draw on studies of the limits of the firm in R&D, especially Cohen and Levinthal (1989), Mowery (1995), and Adams (2006). Our results concur with findings from the management literature on open innovation (Chesbrough, 2003) and innovation communities (Von Hippel, 2005). Consistent with this, we find that reliance on outside knowledge has increased over time.

The historical record (Nelson, 1962; Hoddeson, 1980, 1981; Hounshell and Smith, 1988) provides examples of firms for which science is essential to strategy. The empirical literature on knowledge flows (Jaffe, 1989; Trajtenberg, 1990; Jaffe and Trajtenberg, 1999; Harhoff, Narin, Scherer, and Vopel, 1999; Cohen, Nelson, and Walsh, 2002) has proven essential to our research. In this same vein, Narin, Hamilton and Olivastro (1997) and Branstetter and Ogura (2005) find that patent citations to science have increased over time, consistent with similar findings of our own.

We assume that sampling behavior from knowledge stocks inside and outside the firm is a necessary part of knowledge flows. Thus work on search, technical progress, and growth has influenced us, especially microeconomic studies by Evenson and Kislev (1976), Nelson (1982), and Klette and Griliches (1998), but also studies of growth by Kortum (1997), Aghion and Howitt (1998), and Klette and Kortum (2004). Consistent with our results, these writings stress the role of knowledge in counteracting diminishing returns from search.

Less directly, our findings can be viewed as consistent with the literature on growth and convergence, since they suggest that productivity in research as well as goods and services is deliberately created and is probably easier for followers that are inside the technology frontier. Examples are Barro and Sala-i-Martin (2004, Ch. 8, 12), Howitt (2000), and Griffith, Redding, and Van Reenan (2004), where productivity growth compared with leaders depends on R&D and technology transfer[6].

This paper seeks to make several contributions. To our knowledge we estimate the first science production functions at the firm level. Second, we bring several knowledge flows to bear on the problem of explaining scientific output in firms and we confront these measures both with each other and with a simple alternative, the firm's basic research stock (the new measures add considerably). Third, instead of assuming the importance of knowledge flows from outside science, we test these against flows of science inside the firm (outside flows do matter). Fourth, we compare university science flows with those from other firms (university flows are more important). Fifth, rather than assume that only citations matter, we explore the alternative of collaboration-based university flows. Consistent with Adams, Black, Clemmons, and Stephan (2005), collaboration also contributes. Finally, we confirm the importance of knowledge flows after controlling for a range of individual effects and specifications, and we uncover evidence that knowledge flows are more important at higher levels of aggregation, especially university knowledge flows, as some growth theories suggest.

[6] Geography and technology can restrict knowledge flows. These frictions lie outside the scope of this paper. For agriculture, Evenson and Kislev (1975) discuss limits on knowledge flows due to geography. Keller (2002) and Peri (2005) find that distance, country, and language constrain knowledge flows. Adams and Jaffe (1996), Adams (2002), and Adams, Clemmons and Stephan (forthcoming) discuss limits imposed by field and technology. Frictions imposed by industry are less clear. Scherer (1982a, b) and Klevorick, Levin, Nelson, and Winter (1995) suggest that inter-industry flows of technology are large. Adams and Clemmons (2008, forthcoming) find weak effects of industry in impeding knowledge flows from science.

The rest of the paper contains five sections. Section II presents the knowledge production function for industrial science. Besides firm R&D, this includes knowledge flows from universities as well as knowledge flows from the firm's past research and that of other firms. In this section we define the variables. Section III discusses the data. Section IV presents time series graphs of knowledge flows from universities and firms by industry group and science field. Section V reports our findings on industrial scientific discoveries. Section VI concludes.

II. Science Production Function

We assume that production of new scientific ideas in firms depends partly on search effort. We measure effort by citations and collaborations. These require time spent perusing scientific literature, as well as time and resources spent in joint research, all of it rendered productive by scientific training. This input of labor is reflected in our data in counts of citations to scientific papers and collaborations on jointly written papers. Underlying the extent of search are stocks of advanced human capital, as well as an equilibration of the marginal benefits and costs from search.

Then, in what follows, we examine the amount of knowledge that arrives at the firm. This turns out to depend on citation or collaboration counts divided by the number of papers potentially cited or collaborated. This gives us a rate of search. To simplify the discussion we shall refer to citation rates or collaboration rates simply as sampling rates when the context is clear.

The sampling rate is multiplied by the knowledge stock from which it draws, to indicate the knowledge flow from a given institution. This is summed to arrive at the total knowledge flow. Some knowledge flows are spillovers, in the sense that they represent outside knowledge that is nonrivalrous and non-excludable (Romer, 1990). But some of the flows could represent technology transfer that is excludable. We therefore refrain from calling knowledge flows spillovers.

For the purpose of scientific discovery we identify stocks of knowledge in firms with their basic research stocks, and we replace stocks of knowledge in universities with stocks of R&D. We prefer R&D stocks as proxies for knowledge because they stand for research effort in the long run, regardless of whether it is patented, published, or even observed. In addition, R&D is associated with a larger number of products, which can be thought of as the firm's know-how or stock of knowledge (Klette and Kortum, 2004).

A. Analysis at the Firm and Science Field Level

In most of the empirical work we employ a three-dimensional panel in which a knowledge production function for science i in firm j at time t takes center stage. The panel treats scientific output as heterogeneous within the firm, so that knowledge arrives at different rates at different branches of the firm's knowledge production function for science. For example, the firm can have a biomedical branch, a chemistry branch, an engineering branch, and so forth, and the knowledge inputs that enter into each branch differ from other branches. Our approach takes this into account.

We postulate the following production function for scientific discoveries:

$$(1) \qquad n_{ijt} = A\exp(Z_D'\delta)R_{j,t-1}^{\eta_R}\prod_{v=1}^{V} S_{ijv,t-1}^{\eta_{sv}}\exp\left(u_{ijt}\right)$$

In (1) n_{ijt} is the number of papers (or citation weighted papers), A is productivity not elsewhere accounted for, and Z_D' is a vector of firm, field, and time dummies with coefficients δ. Thus total factor productivity in scientific research is $A_{ijt} \equiv A\exp(Z_D'\delta)$. Note that firm dummies control for hard-to-observe effects of firm size, efficiency, and diversification, field dummies control for differences among fields, and time dummies control for time effects[7]. In addition, the firm's stock of basic research $R_{j,t-1}$ controls for effects of the firm's knowledge that are not measured by knowledge flows[8]. $S_{ijv,t-1}$ includes lagged knowledge flows from past research in universities and other firms. These are flows of outside knowledge. $S_{ijv,t-1}$ also includes flows from the firm's past science research—this is the flow of inside knowledge. Notice that the lagged flows specifically pertain to science i and firm j at time t. Also, u_{ijt} is the error term and the η_i exponents are output elasticities. The composition of the error term is given by

$$u_{ijt} = v_i + v_j + v_t + e_{ijt} \tag{2}$$

In (2), v_i, v_j, v_t are variance-components for field, firm, and time, whose effects are absorbed by the dummy variables Z'_D. e_{ijt} is the "innovation" in the error term.

Taking logarithms of (1) and substituting (2) into the result we reach:

$$\ln(n_{ijt}) = \beta + Z'_D\delta + \eta_R \ln(R_{j,t-1}) + \sum_{v=1}^{V} \eta_v \ln(S_{ijv,t-1}) + v_i + v_j + v_t + e_{ijt} \tag{3}$$

in which β equals the logarithm of A. For identification e_{ijt} must be orthogonal to the right-hand side variables. In addition to using lagged variables as instruments we address this by including all available lags as instruments using panel GMM.

In (3) the firm's stock of basic research is

$$R_{jt-1} = b_{t-1}\sum_{\tau=1}^{5} r_{jt-1-\tau}(1-d)^{\tau} \tag{4}$$

[7] The dummies also absorb knowledge flows not controlled by researchers in a firm and field. These become endogenous at higher levels of aggregation. See Romer (1990), Griliches (1992), and Jones (1995).

[8] We also used stocks of basic plus applied research in place of basic research throughout, including the knowledge flows, but this made little difference to our results.

Here b_{t-1} is the share of basic research in the primary industry of firm *j* from NSF (various years), $r_{jt-1-\tau}$ is the lagged flow of firm R&D over the previous five years from Standard and Poor's Compustat, and $d = 0.15$ is the depreciation rate assumed for R&D[9].

There is an important issue concerning the knowledge flows $S_{ijv,t-1}$. Since these draw on several sciences they are really two-level production functions. Barring fixed proportions, concavity is needed to uniquely determine inputs from each science. But for tractability we replace the nonlinear functions with linear approximations[10].

The knowledge flows use sampling rates c_{ij} / n_j as weights that measure the intensity with which researchers search various knowledge stocks. The numerator c_{ij} is a count of citations or collaborations from group i to group j. This is divided by n_j, the number of scientific papers in group j that could have been referenced. For a group of papers n_i the rate c_{ij} / n_j is more meaningful than probability $c_{ij} / n_i n_j$. The probability is appropriate for a single paper because it captures the average proportion of knowledge flowing from j to i (Adams, Clemmons, and Stephan, forthcoming). But if n_i papers cite n_j papers in *j* then the probability $c_{ij} / n_i n_j$ reduces to the sampling rate c_{ij} / n_j. We now consider the knowledge flows in detail.

We begin with flows of outside knowledge into firms. The citation flow from universities takes the form:

---

[9] Firm R&D flows are expressed in millions of 1992 dollars. We chose a five year measure of the R&D stock because of the short length of many firm R&D histories. And since we do not have firms' basic research, we adjust the total stock of R&D by the industry ratio of basic research to total R&D.

[10] Since we include six sciences and four knowledge flow variables, estimating a nonlinear regression with a two-level production function for the R&D variables is not promising. The linear knowledge flows are simple and almost surely, are highly correlated with the ideal two-level production functions.

$$S_{ijt}^{CU} = \sum_{F=1}^{M} \sum_{k=1}^{N} \sum_{\tau=1}^{t-1} (\mathrm{CU}_{ijt}^{Fk\tau} / n_{Fk\tau})\mathrm{R}_{Fk\tau} \tag{5}$$

Here CU refers to citations to university science. Subscripts stand for the citing (receiving) side of knowledge flows, and superscripts stand for the cited (sending) side. Subscripts and superscripts consistently follow the ordering: field of science, institution (university or firm), and time. Thus, in the case of (5) subscript ijt refers to citing field i in firm j at time t and superscript $Fk\tau$ refers to cited field F, university k, and prior year τ. The citation rate in parentheses is field-specific because university R&D stocks are field-specific. $CU_{ijt}^{Fk\tau}$ is the number of citations from papers in a field, firm, and year to papers in a university, field, and year. It is inherently a six-dimensional object. The number of papers in university k in field F at time τ is $n_{kF\tau}$. Thus the sampling rate equals $CU_{ijt}^{Fk\tau} / n_{Fk\tau}$. The university citation flow is just the sum of the sampling rates times cited R&D stocks over cited sciences F, universities k, and years τ. This explains the triple summation in (5), which reduces the knowledge flow to the three-dimensional level of citing field, firm, and year, representing in a simple way all that has been learned. In (5) the definition of the stock of university R&D $R_{kF\tau}$ is

$$R_{kF\tau} = \sum_{z=1}^{8} r_{kF\tau-z}(1-d)^{z} \tag{6}$$

It is the sum of depreciated R&D flows $r_{kF\tau-z}$ over the previous eight years, where the depreciation rate is $d = 0.15$. R&D f lows are taken from the NSF CASPAR database. University R&D stocks have three advantages over firm R&D: they are consistently available over a longer period, they exist by field, and they are expenditures on science[11].

[11] Field and university R&D stocks are expressed in millions of 1992 dollars.

One disadvantage is that they could contain considerable respondent error. To reduce simultaneity bias we lag (5) in production function (3).

The collaboration knowledge flow from universities is simpler than (5). This is because joint research takes place within field—the field of the journal where it is published. Collaboration occurs in the same year—the year of publication[12]. Since collaboration occurs within field in the same year, the knowledge flow is a single sum over collaboration rates times collaborated university R&D stocks:

$$(7) \qquad S^{JU}{}_{ijt} = \sum_{k=1}^{N} (\mathrm{JU}_{ijt}^{ikt} / n_{ikt}) R_{ikt-1} \,,$$

The term JU_{ijt}^{ikt} stands for counts of joint research in field i between firm j and university k. The flow of knowledge is the sum of the sampling rate in parentheses times the R&D stocks defined in (6). Since it is within-field (7) is restricted to field i. Again we lag (7) in production function (3).

Next consider knowledge flows within a firm. The firm's basic research stock, which enters this calculation at various points in time, depends on total R&D expenditures in Compustat, since stocks of firm basic research do not exist, either by field of science or in total. Because of this, we estimate the firm's stock of basic research across fields using b_{t-1}—the ratio of basic research to total R&D in the firm's primary industry—times $R_{i\tau}$, which is the stock of total R&D[13].

The knowledge flow from the firm's past research is the product of the firm's basic research stock at various times by citation-sampling rates to its earlier papers:

[12] Tracing collaboration over time is beyond the research frontier at this time.
[13] The firm R&D stock is the deflated stock of R&D in millions of 1992 $ over the previous five years, depreciated at 15 percent per year. We have these stocks going back as far as 1977 depending on the year that the firm is first listed in Compustat. The industry ratios of basic research to total R&D are taken from National Science Foundation (various years).

$$S_{ijt}^{CS} = \sum_{F=1}^{M} \sum_{\tau=1}^{t-1} (CS_{ijt}^{Fj\tau} / \sum_{F=1}^{M} n_{Fj\tau})\, b_{t\text{-}1}\, \mathrm{R}_{j\tau}$$
$$(8) \qquad = \sum_{\tau=1}^{t-1} (\sum_{F=1}^{M} CS_{ijt}^{Fj\tau} / \sum_{F=1}^{M} n_{Fj\tau})\, b_{t\text{-}1}\, \mathrm{R}_{j\tau}$$

The term CS represents citations to papers written by the same firm. Following our notation $CS_{ijt}^{Fj\tau}$ citations are made by a firm's papers in field i, firm j and time t to the firm's previous papers $n_{Fj\tau}$ in field F in year τ. Since firm j stays the same, (8) is a double sum over fields and years cited. In the second line we use the citing field's *weighted average* citation rate. This is the number of citations from field i in year t to all fields in year τ, divided by the number of papers in year τ across all fields. Unlike (5) and (7) we use a weighted average because for firms, we lack R&D by field of science.

The citation knowledge flow from other firms is similar to (8) except that cited firms differ from citing. This requires a third sum over cited firms:

$$S_{ijt}^{CI} = \sum_{F=1}^{M} \sum_{\substack{k=1\\k\neq i}}^{N} \sum_{\tau=1}^{t-1} (\mathrm{CI}_{ijt}^{Fk\tau} / \sum_{F=1}^{M} n_{Fk\tau})\, b_{t-1}\, \mathrm{R}_{k\tau}$$
$$(9) \qquad = \sum_{\substack{k=1\\k\neq i}}^{N} \sum_{\tau=1}^{t-1} (\sum_{F=1}^{M} \mathrm{CI}_{ijt}^{Fk\tau} / \sum_{F=1}^{M} n_{Fk\tau})\, b_{t-1}\, \mathrm{R}_{k\tau}$$

We use CI for citations to other firms. Scientific papers written in field i, firm j, and time t make $CI_{ijt}^{Fk\tau}$ citations to papers $n_{Fk\tau}$ in field F, other firms k, and year τ. As in (8) the ratio in parentheses on line two is the weighted average citation rate. We lag (8) and (9) by one year in (3) to reduce issues of endogeneity.

B. Aggregation to Higher Levels

Besides the firm and field level we explore the production of science at the firm and industry and field levels. Firm level analysis aggregates over fields within a firm. Industry and field level analysis aggregates over firms, but leaves field intact.

A firm level analysis checks for factors that are external to science fields but internal to the firm. Such "quasi-external" effects might differ by source of knowledge. Another reason for a firm level analysis is that we are better able to apply panel GMM at this level, because a longer time series of data exists. One drawback, though, is that errors in variables and in aggregation could weaken the fit of the equation.

The knowledge production function for firm j is, to a first approximation,

$$\ln(n_{jt}) = \beta + Z_D'\delta + \eta_R \ln(R_{jt-1}) + \sum_{v=1}^{V} \eta_v \ln(S_{jvt-1}) + v_j + v_t + e_{jt} \tag{10}$$

The notation resembles (3) except that knowledge flows are aggregated over fields.

We also aggregate the data across firms in the same industry to undertake analysis of the industry. The idea is to uncover knowledge effects that are external to firms but internal to fields. The logic derives from Romer (1990), Griliches (1992), Jones (1995), and Aghion and Howitt (1998): returns to knowledge could increase in passing from the firm to the industry level, despite duplicative and business-stealing R&D.

Knowledge production in field i and industry I is

$$\ln(n_{iIt}) = \beta + Z_D'\delta + \eta_R \ln(R_{It-1}) + \sum_{v=1}^{V} \eta_v \ln(S_{iIvt-1}) + v_i + v_I + v_t + e_{iIt} \tag{11}$$

The notation is similar to (3) but the interpretation differs, because knowledge flows include external effects that are held constant at the firm level[14].

III. Database

The underlying data consist of 230 thousand papers of the top 200 U.S. R&D firms and 2.43 million papers of the top 110 U.S. universities that were published during 1981-1999. The data source is Thomson-Scientific in Philadelphia, Pennsylvania. The papers

[14] See Griliches (1992) for a proof. His discussion of aggregation from firms to industries uses common prices for inputs and common production functions.

appear in 7,137 scientific journals. Each journal (and the papers in it) is assigned to a single science field. The main alternative to this method is the assignment of papers according to authors' fields. But this information is all too often incomplete[15].

Top 200 firms make one million citations to papers of top 110 universities and 600 thousand citations to Top 200 papers including their own. We have seen that outside knowledge flows involve citations to universities and other firms, and are "non-self" citations. Conversely self-citations are evidence of inside knowledge flows from a firm's past research to its current research. Because firms very rarely collaborate in science, in this paper collaboration consists of joint research between firms and universities[16]. Indeed, university-firm collaborations cover 20 percent of firms' scientific papers.

In (5) and (7)-(9) we showed how to exploit citation and collaboration rates to construct knowledge flows needed for production functions (3), (10), and (11). In building the flows, we use data on R&D in universities by field from the NSF CASPAR database; data on total R&D in firms from Compustat; and data on the ratio of basic to total R&D by industry from National Science Foundation (various years).

To undertake the empirical analysis we construct a panel of firms, fields, and years. Firms are the top 200. Fields are biology, chemistry, computer science, engineering, medicine, and physics: these cover 95 percent of all scientific papers in industry. Although the data go back to 1981, we want to allow for histories of citation and collaboration, so the time period that we use is 1988 to 1999. Fifteen firms drop out

---

[15] We attempted to assign all papers of Harvard University to science fields using address information. About a third of the papers could not be assigned so we abandoned this method.
[16] Firm papers are 1/10 as many as university papers so collaborations between firms would be 1/10 as many as university and firm collaborations. This assumes that collaboration propensities are the same. But firm-firm collaborations are less common than this.

because R&D histories are missing. In addition data on fields do not exist in each year. After missing values are removed the panel consists of up to six sciences, 185 firms with adequate R&D histories, and up to 12 years of data over the period 1988-1999.

Dependent variables are scientific papers as well as papers weighted by citations during the first five years since, and including, publication. Independent variables include field, firm, and time fixed effects. Also included are logarithms of the firm's basic research stock and the various knowledge flows defined in Section II. After missing values are removed, the panel consists of 4,268 observations on papers and 2,495 observations on citation-weighted papers[17].

Tables 1 and 2 provide descriptive statistics. Table 1 shows means and standard deviations of the principal variables. The mean number of papers is 28.4. Papers are fractionally assigned, they are continuous, and they are not censored. And so we use OLS to estimate production functions for papers rather than a count model (Hausman, Hall and Griliches (1984))[18]. The mean of fractional citation-weighted papers is 71.9, but a fourth of the observations are left-censored. We use Tobit analysis for this variable[19].

The rest of Table 1 presents means and standard deviations of basic research and the knowledge flow variables (in millions of 1992 dollars), as well as their sampling rates. The mean of the firm's basic research stock—see (6)—is 145.5 million, reflecting the amount of R&D in Top 200 firms.

---

[17] Since citations received cover the first five years after publication (including publication year) and since the data end in 1999, citation-weighted papers end in 1995, leading to the drop in the observations.
[18] In the underlying calculations an institution gets half a paper or citation-weighted paper if it collaborates with one other institution, a third of a paper if it collaborates with two others, and so on. This procedure avoids multiple counting of papers across collaborating institutions.
[19] At the industry and field level none of the citation-weighted papers are zero so we use OLS at this level of analysis.

The mean university-firm citation rate is 0.006. The mean knowledge flow (5) is 41.0 million. The mean university-firm collaboration rate is 0.010; the mean knowledge flow (7) is 2.9 million. The collaboration flow is smaller than the citation flow because collaboration occurs much less frequently.

Self-citation and firm-firm knowledge flows conclude Table 1. The mean self-citation rate is 0.014, while the mean self-citation flow (8) is 41.8 million. The mean firm-firm citation rate is 0.005, slightly less than the firm-university rate. But the mean knowledge flow among firms (9) is 24.5 million, less than the university-firm flow, again because firm-firm citation occurs less frequently.

Table 2 presents *total* knowledge flows in columns and receiving fields and industry groups in rows. The top panel shows flows by field. As one would expect, biology and medicine account for more than half of university citation flows. Their share in collaboration is less. Biology and medicine are less common in industry hence their collaboration share is smaller. For the same reason, collaboration flows in computer science and engineering exceed their share in citation flows.

For related reasons citation flows within and between firms are distributed differently than flows from universities. Compared with universities chemistry and physics are more important in industry; biology and medicine are less important; and computer science and engineering are again more important.

Table 2's bottom panel shows total knowledge flows by industry. One observes a spike in drugs and biotechnology, with lesser spikes occur in petrochemicals, electrical equipment, and software and distinct drop-offs that occur in metals, machinery, and miscellaneous. These patterns follow the volume of science in an industry.

IV. Trends in Flows of Scientific Knowledge

We now turn to trends in knowledge flows. The area graphs of Figure 1 display *shares* in the total flow of scientific knowledge by type for each industry group. This total is the sum over the four knowledge flows used in this study. Following equations (5) and (7)-(9) they are: flows from citations to university papers; flows from collaborations on university papers; flows from citations to papers of the same firm; and flows from citations to papers of other firms. Use of shares allows us to see changes in relative contributions over time and to compare these across industries and sciences.

The six panels of Figure 1 represent: petrochemicals; drugs and biotechnology; metals, machinery, and miscellaneous; computers, communications, and software; electrical equipment and instruments; and transportation equipment. The share of knowledge from university collaborations is larger in early years of the sample and is shown by a downward pointing cusp. It results from the immediacy of collaboration compared with the gradual build-up of citation. The share of knowledge flows from university citations grows over time, as does the share of citation flows from other firms. Over time, knowledge flows within the firm steadily lose share to outside flows.

The university share (collaboration plus citation) grows fastest in drugs and biotechnology; electrical equipment and instruments; and transportation equipment. The share of other firms grows fastest in petrochemicals; drugs and biotechnology; and electrical equipment and instruments. These differences follow variations in the use of different sciences by industries as well as varying growth of R&D across sciences and sectors. But Figure 1 as a whole clearly shows that firms draw an increasing share of knowledge from external sources at the end of the 20th century. This agrees with the literature on open innovation, for example, Chesbrough (2003) and Von Hippel (2005).

Figure 2 reports shares of knowledge flows by field. Decline in the within-firm share is clear. It is smallest in chemistry and physics, where university research stocks have grown more slowly, yet even here it occurs. The fact that chemistry and physics are dominant sciences in petrochemicals; in metals, machinery, and miscellaneous; and in computers, communications, and software (Adams and Clemmons, 2008, forthcoming) helps to explain why the within-firm share drops less in these industries.

The university share (citations plus collaborations) grows most in biology and medicine, where university R&D has grown the fastest , and the same is true of computer science. Across fields, collaboration flows again start off large, because collaboration flows unfold more rapidly than citation flows. In addition, collaboration is more important in computer science and engineering than elsewhere.

The share of knowledge flows from other firms grows fastest in biology, medicine, and computer science. But it occurs in all fields, where between-firm flows grow relative to within-firm. This implies that knowledge sharing in industry has become more important over time.

V. Findings

A. Firm and Field Level Estimates

Tables 3 and 4 are a basic set of estimates of the knowledge production function for industrial science. Recall that this is

(3) $$\ln(n_{ijt}) = \beta + Z_D'\delta + \eta_R \ln(R_{jt-1}) + \sum_{v=1}^{V} \eta_v \ln(S_{ijvt-1}) + v_i + v_j + v_t + e_{ijt}$$

The logarithm of papers or citation-weighted papers is on the left. On the right are field, firm, and year fixed effects Z_D' and the logarithm of the firm's basic research stock.

Under the summation sign are logarithms of citation and collaboration knowledge flows from universities, the self-citation flow from the firm's past research, and the citation flow from other firms. The η_i coefficients are elasticities, and the error components complete the specification.

Since knowledge flows sometimes equal zero we handle this by adding 0.001 to the flows before taking logarithms. Given this adjustment, the elasticities are really averages over zero and positive observations[20].

Table 3 reports regression findings with the logarithm of scientific papers as the dependent variable. All equations use robust, clustered standard errors where firm and science field are the clustering variable. Firm, field and year fixed effects are also included. These are jointly significant at the 0.1 percent level.

Equation 3.1 includes the firm's basic research stock, whose elasticity is 0.124 and significant. Equation 3.2 adds the self-citation knowledge flow. This cuts the basic research elasticity to 0.061, while the self-citation elasticity is 0.195 and significant.

Equation 3.3 adds the logarithm of citation flows from universities and other firms to 3.2. The basic R&D elasticity falls to 0.027 and becomes insignificant. While the self-citation elasticity falls to 0.094, it remains significant. The university citation elasticity is 0.170 while the firm citation elasticity is 0.102 and both are significant. Accounting for outside knowledge clearly detracts from the role of knowledge inside the firm.

[20] Another approach is to include a dummy indicator equal to 1 if the knowledge flow equals zero, and to interact the dummy with the logarithm of each knowledge flow variable. When the interaction is included along with the knowledge flow, the two terms together absorb the effect of a zero flow while the main term by itself captures the effect of a positive flow. This is similar to a spline or polynomial regression across zero and positive values, except that it does not force facets of the polynomial to join at corners. For this approach see Adams and Clemmons (2008, forthcoming). For a brief introduction to splines see Greene (2008), pages 111-112.

Equation 3.4 adds university collaboration flows to 3.3. The basic research, self-citation, and firm citation elasticities do not change. But the university citation elasticity drops from 0.170 to 0.126, probably because collaboration substitutes for citation. All knowledge flow elasticities are positive and significant, but the largest effect comes from universities and other firms. Their combined elasticity is 0.343 whereas self-citation contributes 0.084. Estimated returns to scale are given by the sum of the elasticities, or 0.427, suggesting diminishing returns[21]. And yet papers are not weighted by quality or importance in Table 3.

Table 4 gets at this by using citation-weighted scientific papers as the dependent variable. To capture *industrial* relevance of a scientific paper the citations are forward citations from other firms over the first five years of a paper's existence[22]. The time period is 1988-1995. The number of observations is 2,495 of which 586 are left-censored at zero. We adopt random effects Tobit as the estimation method because the fixed effects estimator is biased and inconsistent. This is because fixed effects cannot be factored out of the Tobit likelihood, leading to an incidental parameters problem[23].

Elasticities in Table 4 tend to exceed those in Table 3. This is partly due to differences in statistical method. In Table 3 the OLS expected marginal effect is the elasticity. But in Table 4 it is the elasticity times the probability that the dependent variable is not censored (Greene, 2008, pages 872-873).

But the difference in marginal effects is not the main reason for the difference in elasticities. In 4.1 the elasticity of the firm's basic research stock is 0.562 but in 3.1 it is

---

[21] Differentiate equation (3) with respect to all variables X subject to the restriction that $dX / X = c$.

[22] Notice that citations received by the firm's papers from other firms in the future are completely separate from the citations that the firm makes to papers in the past. This is an important point, because it says that citation-weighted papers are not subject to a hidden dependency.

[23] See for example, Cameron and Trivedi (2005), pages 800-801.

0.124. The difference in elasticities is 0.438. The probability that citation-weighted papers are not censored is 0.765 ((2495-586)/2495) so the expected marginal effect in 4.1 is 0.765×0.562=0.430. The remaining difference is 0.306 (0.430-0.124), which accounts for most of the raw difference, 0.438. The jump in the expected marginal effect is real; some of this may be due to the use of random effects rather fixed effects.

Equation 4.2 adds self-citation knowledge flows to 4.1. This cuts the basic research elasticity from 0.562 to 0.474, while the self-citation elasticity is 0.180. Both are significant. Equation 4.3 adds university and firm citation spillovers to 4.2. The university and firm elasticities are 0.336 and 0.205. Including them reduces the basic research and self-citation elasticities, though these remain significant.

Equation 4.4 adds university collaboration. Its elasticity is 0.114 and it is significant. Collaboration draws some effect away from the other variables, especially university citation, though all elasticities remain positive and significant. As before, the sum of the elasticities is an estimate of the returns to scale in production. This equals 1.04, indicating constant returns and suggesting that part of the payoff to basic research partly consists of higher quality science. But internal returns to scale are 0.43, indicating diminishing returns when outside knowledge is held constant.

Table 5 reports formal tests of equations 3.4 and 4.4. Equality of the university and firm citation elasticities is rejected (line one) in favor of the alternative, that the university elasticity is greater. The hypothesis that the university citation elasticity equals the collaboration elasticity is also rejected (line two). The combined university elasticity usually exceeds the firm citation elasticity (line three). At times (line four), the combined university elasticity exceeds the total firm elasticity.

B. Robustness Checks

Table 6 explores sensitivity of the elasticities to changes in sample and specification. The method resembles that of Donohue and Levitt (2001). We use 3.4 as the baseline for Panel A, which deals with papers; and 4.4 as the baseline for Panel B, which deals with citation-weighted papers. For comparison the estimates from 3.4 and 4.4 form the top lines of each panel. Columns are elasticities—standard errors are in parentheses. Rows summarize the different experiments.

Equation 6.1 allows for interactions between firm and field, while 6.2 allows for interactions between firm and year. The elasticities fall slightly, yet all remain positive and significant. In 6.3-6.5 science-intensive industries (drugs and biotechnology; software and communications; or both) are dropped from the sample. Omitting these could diminish the elasticities, but we find little effect. At the other extreme omitting metals, machinery, and miscellaneous in 6.6 could increase the elasticities, but again there is little effect. The estimates are stable across industries.

Equation 6.7 tests the specification of the variables. Recall that the university citation and collaboration flows (5) and (7) use field-specific R&D stocks for universities, while the self-citation and firm-firm citation flows (8) and (9) use basic research stocks. Data do not exist on firm basic research by field. In 6.7 we treat R&D stocks for universities exactly like those for firms. But we find that this makes little difference. The estimates in 6.7 are similar to the original estimates (Panel A, top). Apparently the decline in errors in R&D cancels the specification error caused by summing over fields.

The final line (6.8) replaces the knowledge flows with naïve mean sampling rates. The results are significant, indicating that search effort plays a role in knowledge transmission. While this is consistent with the search hypothesis, observe the drop in R-squared from 0.74 on the top line to 0.69 in equation 6.8. This implies that R&D stocks play a role in addition to search effort.

Panel B applies similar checks to citation-weighted papers with similar results[24]. Dropping science-intensive sectors or dropping sectors that use little science (6.9-6.12) matters little. Using the sum of university R&D over fields (6.13) has little effect on the estimates. Using mean sampling rates in 6.14 yields significant elasticities, though the drop in the log likelihood suggests that R&D stocks are important in addition to search.

C. Firm Level Estimates

We now expand the investigation to higher levels of aggregation. Recall that the firm level equation is (10). Table 7 reports OLS estimates of (10) for papers in 7.1, and Tobit estimates for citation-weighted papers in 7.2.

All elasticities are positive and significant. The sum of the elasticities declines to 0.193 in 7.1, compared with 0.427 in the corresponding firm and field equation, 3.4. But in 7.2 the reverse is true: the sum of the elasticities rises to 0.925 compared with 0.782 in 4.4. The firm's basic research is significant in 7.1 while it is insignificant in 7.2. The instability of basic research suggests that collinearity is present, probably with the self-citation knowledge flow (8). And yet the self-citation elasticity is stable. Because of this we emphasize self-citation and other knowledge flows in the remaining discussion.

While it is hard to compare the firm level estimates with those at the firm and field level, the relative contribution of university knowledge does increase at the firm level. To see this note that in 7.1 the *share* of university citation and collaboration in all

---

[24] Since firm fixed effects produce inconsistent and biased estimates in Tobit analysis it is not worthwhile to create interactions with field and year effects in Panel B, as was done in Panel A, in equations 6.1-6.2.

knowledge flows is $(0.078+0.046)/(0.078+0.046+0.026+0.043)=0.642$. But at the firm and field level, in 3.4, the share of university citation and collaboration is $(0.126+0.096)/(0.126+0.096+0.084+0.088)=0.546$.

In 7.2 this share is $(0.441+0.111)/(0.441+0.111+0.175+0.198)=0.597$. But it is $(0.302+0.114)/(0.302+0.114+0.167+0.199)=0.532$ at the firm and field level in 4.4. We conclude that university knowledge flows are more important at the firm level.

In Table 8 we undertake dynamic panel estimation of the production function using panel GMM[25]. We include the logarithm of lagged scientific papers $\ln(n_{jt-1})$ as a factor of production representing the persistence of discovery. The model is

$$(12) \qquad \ln(n_{jt})=\beta+Z_D'\delta+\eta_n\ln(n_{j,t-1})+\eta_R\ln(R_{j,t-1})+\sum_{v=1}^{V}\eta_v\ln(S_{jv,t-1})+v_j+v_t+e_{jt}$$

The error term e_{jt} is serially uncorrelated. Differencing (12) to eliminate the fixed effect v_j introduces first order serial correlation in the transformed error $e_{jt}-e_{jt-1}$, as well as correlation with lagged papers $\ln(n_{j,t-1})$[26]. Therefore, instruments for lagged papers are provided by lags of $t-2$ years and earlier. Instruments for differenced knowledge and R&D variables in year t are levels of these variables dated $t-1$ and earlier. More instruments become available moving forward in time (Arellano and Bond 1991). Moment conditions for products of the instruments with the errors are then combined in a quadratic criterion and estimated using Generalized Method of Moments (GMM). Because the data are differenced this method is known as Difference GMM.

[25] It is not clear how one would do GMM estimation of citation-weighted papers, for which 11 percent of the observations are left censored at zero. Hence we omit GMM estimates for this variable.

[26] Blundell and Bond (1998, 2000) make a different assumption. The equation contains lagged papers as a result of the correction for serially correlated errors in the *original* equation. The correction for this introduces lagged dependent and exogenous variables. But in this article, lagged papers belong in the original equation and serial correlation is absent until differencing to remove fixed effects introduces it, as in Arellano and Bond (1991).

One weakness of Difference GMM is that lagged instruments in levels may exhibit low correlation with the differenced variables and create a problem of weak instruments. Blundell and Bond (1998) show that, given certain initial conditions, additional moments become available. The additional instruments are differences of the knowledge and R&D variables in the levels equations. The additional level conditions are combined with the differenced conditions to form a system, so this is known as System GMM. We estimate (12) using Difference and System GMM[27].

Table 8 contains the results. Equations 8.1 and 8.2 use Difference GMM and assume that the knowledge variables are exogenous or predetermined. The results are similar to 7.1 except for the inclusion of lagged papers $\ln(n_{j,t-1})$, which are insignificant. The System GMM equation 8.3 fits better, consistent with Blundell and Bond (1998, 2000). Lagged papers are significant, suggesting that past discovery does contribute to future discovery. In general the knowledge elasticities in 8.3 rise compared with 8.1 and 8.2. The Wald chi-Square statistics confirm the joint significance of the variables. To sum up, the firm level results validate the importance of knowledge in scientific discovery. As before, much of the knowledge that drives discovery comes from outside the firm[28].

---

[27] We use Stata 10.0's xtabond and xtdpdsys commands for Difference and System GMM respectively.

[28] We explored firm level estimates using spillovers in the style of Jaffe (1986). Here the equation is

(A) $\ln(n_{jt}) = \beta + Z'_D\delta + \gamma_R \ln(R_{jt-1}) + \gamma_U X^U_{jt-1} + \gamma_I X^I_{jt-1} + \nu_j + \nu_t + e_{jt}$. The new terms consist of the two *lagged* variables X^U_{jt-1} and X^I_{jt-1}, which are "spillovers" from universities and industry. The spillovers are defined as $X^K_{jt} = \sum_{\substack{i=1 \\ i\neq j}}^{N} P^K_{ijt} \ln(R^K_{it-1})\ K = U, I$. Here P^K_{ijt} is the un-centered correlation between vectors of lagged shares of papers by field of science in firms i, j. Estimates of (A) perform poorly. Spillovers are insignificant in three out of four cases and the elasticities are sensitive to the inclusion of zero spillovers.

C. Industry and Field Level Estimates

Table 9 presents estimates of the industry, field, and year level production function (11), where aggregation takes place over firms in the same industry. The regressions report robust, clustered standard errors where clustering is by industry and field.

Equations 9.1 and 9.2 report results for papers and citation-weighted papers. The estimation method is OLS for both equations since at this level no observations are left censored. In 9.1 the sum of the knowledge elasticities—the estimate of the returns to scale—rises to 0.675, compared with the firm and field level estimate of 0.394 in 3.4. In addition, the share of university citations and collaborations rises to 0.791 in 9.1 (0.534/0.675) compared with 0.564 in 3.4 (0.222/0.394). This is consistent with theories which predict that spillover effects rise with aggregation (Romer, 1990; Griliches, 1992; Jones, 1995; and Aghion and Howitt, 1998). This is especially true of university knowledge, which may be less excludable than firm knowledge.

Results for 9.2 are similar. In the industry and field equation 9.2 the sum of the elasticities is 0.919. This is the expected marginal effect since no censoring occurs. But in the firm and field equation 4.4 the sum of the knowledge flow elasticities is 0.782. When multiplied by the probability not censored of 0.835 to yield the expected marginal effect (Greene (2008), pages 872-873), the sum is 0.653. The knowledge elasticities increase with aggregation. University flows assume a more prominent role, suggesting larger externalities from university knowledge.

D. Marginal Products of Knowledge

So far we have discussed elasticities of knowledge. These are percentage changes in papers or citation-weighted papers per one percent change in knowledge. Here we would like to compute marginal products of the knowledge variables. Since knowledge is expressed in millions, marginal products are expressed in papers per million.

They are calculated as follows. Elasticities are defined as

(13) $$\eta_i = \partial \ln(Papers) / \partial \ln(S_i) = (\partial\, Papers / \partial\, S_i) \times (S_i \,/\, Papers) \,.$$

Here S_i is the knowledge flow from source i. Inverting (13) and evaluating at the mean the marginal product is:

(14) $$\partial\, Papers / \partial S_i = \eta_i \times (\overline{Papers} \,/\, \overline{S_i})$$

where the bar indicates a mean. Table 10 reports marginal products computed using (14).

Consider first the firm and field level (Panel A). The marginal product of university collaboration is the largest, marginal products of university and firm citations are less; and the self-citation marginal product is the smallest. This pattern would reflect an equilibrium, if collaboration is more costly than citation (because collaboration commits sizable resources), and if self-citation is less costly than external citation (because the latter requires more search). At the firm level (Panel B) and the industry and field level (Panel C), collaboration marginal products continue to be the largest and self-citation marginal products the smallest. Marginal products of the firm citation spillover vary relative to those of universities, suggesting rough equality. Also, industry and field marginal products usually exceed firm and field level marginal products. This suggests that the social product of research exceeds the private product.

VI. Conclusion

The evidence in this paper implies that industrial science progresses by drawing on past scientific research inside and outside the firm, with most of the influence deriving from outside research (Tables 3, 4, 7, and 9). Moreover, as time goes by, the share of knowledge flows from outside the firm increases (Figures 1 and 2). In turn, the contribution of outside scientific knowledge that is contributed by universities exceeds that of other firms. This is particularly true when citation-weighted papers are the dependent variable so that quality of scientific output is included. These findings survive the inclusion of fixed effects for firms, fields, and years as well as variations in sample and specification.

At the firm level we use OLS and Tobit with lagged instruments for knowledge, and for papers we use panel GMM to introduce additional lagged instruments. Both lend support to the firm and field estimates but with a few differences, especially the GMM equations (Table 8). The GMM results include lagged papers of the firm, with the goal of capturing productive effects of previous discoveries, for which we find some, albeit mixed support. As before knowledge remains positive and significant. There is evidence that the university contribution rises relative to that of firms as aggregation proceeds from the firm and field level to the firm level.

At the level of the industry and field we find that the productive role of knowledge again rises. These results are consistent with spillover effects of knowledge at higher levels of aggregation (Romer, 1990; Griliches, 1992; Jones, 1995; Aghion and Howitt, 1998). At this level the role of university knowledge increases markedly, suggesting that large spillovers from university research take place.

The next step in this process of detection is to try to understand the role of science in invention. Beyond this the role of science in commercialization of new products would be interesting as well as challenging to explore.

References

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

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

_______________, “Learning, Internal Research, and Spillovers,” **Economics of Innovation and New Technology** 15 (January 2006): 5-36.

_______________, and J. Roger Clemmons, “Science and Industry: Tracing the Flow of Basic Research through Manufacturing and Trade,” **Economics of Innovation and New Technology**, 2008, forthcoming.

___________________________________, “Invention and Discovery in Science-Based Firms,” Working Paper, Rensselaer Polytechnic Institute, Troy, NY: November 2007.

_______________, 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.

_______________, J. Roger Clemmons, and Paula E. Stephan, “Standing on Academic Shoulders: Measuring Scientific Influence in Universities,” **Les Annals d’Economie et de Statistique**, forthcoming.

______________, Grant C. Black, J. Roger Clemmons, and Paula E. Stephan, “Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999,” **Research Policy** 34 (2005): 259-285.

Aghion, Philippe, and Peter Howitt, **Endogenous Growth Theory**, 1st Edition, Cambridge, Massachusetts: MIT Press, 1998.

Arellano, Manuel, and Stephen Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," **Review of Economic Studies** 58 (1991): 277-297.

Barro, Robert J., and Xavier Sala-i-Martin, **Economic Growth**, 2nd Edition, Cambridge, Massachusetts: MIT Press, 2004.

Blundell, Richard, and Stephen Bond, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," **Journal of Econometrics** 87 (1998): 115-143.

______________________________, "GMM Estimation with Persistent Panel Data: An Application to Production Functions," **Econometric Reviews** 19 (2000): 321-340.

Branstetter, Lee, and Yoshiaki Ogura, "Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations," NBER Working Paper 11561, August 2005.

Cameron, A. Colin, and Pravin K. Trivedi, **Microeconometrics: Methods and Applications**, New York: Cambridge University Press, 2005.

Chesbrough, Henry W., **Open Innovation**, Boston, Massachusetts: Harvard Business School Press, 2003.

Cohen, Wesley M., and Daniel A. Levinthal, "Innovation and Learning: The Two Faces of R&D," **Economic Journal** 99 (September 1989): 569-596.

_______________, 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.

Donohue III, John J., and Levitt, Steven D., "The Impact of Legalized Abortion on Crime," **Quarterly Journal of Economics** CXVI (May 2001): 379-420.

Evenson, Robert E., and Yoav Kislev**, Agricultural Research and Productivity**, New Haven, Connecticut: Yale University Press, 1975.

________________________________, "A Stochastic Model of Applied Research," **Journal of Political Economy** 84 (April 1976): 265-282.

Greene, William H., **Econometric Analysis**, 6th edition, Pearson-Prentice-Hall: Upper Saddle River, New Jersey, 2008.

Griffith, Rachel, Stephen Redding, and John Van Reenan, "Mapping the Two Faces of R&D: Productivity Growth in a Sample of OECD Industries," **Review of Economics and Statistics** 87 (November 2004): 883-895.

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 1991): 29-47.

Harhoff, Dietmar, Frances Narin, F. M. Scherer, and Katrin Vopel, "Citation Frequency and the Value of Patented Inventions," **Review of Economics and Statistics** 81 (August 1999): 511-515.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," **Econometrica** (July 1984): 909-938.

Hoddeson, Lillian H., "The Entry of the Quantum Theory of Solids into the Bell Laboratories, 1925-1940: A Case Study of the Industrial Application of Fundamental Science," **Minerva** 18 (1980): 422-447.

__________________, "The Discovery of the Point Contact Transistor," **Historical Studies in the Physical Sciences** 12 (1981): 41-76.

Hounshell, David A., and John Kenley Smith, **Science and Corporate Strategy: DuPont R&D, 1902-1980**, Cambridge, UK: Cambridge University Press, 1988.

Howitt, Peter, "Endogenous Growth and Cross-Country Income Differences," **American Economic Review** 90 (September 2000): 829-846.

Jaffe, Adam B., "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," **American Economic Review** 76 (December 1986): 984-1001.

_____________, "Real Effects of Academic Research," **American Economic Review** 79 (December 1989): 957-970.

_____________, 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.

Keller, Wolfgang, "Geographic Localization of International Technology Diffusion," **American Economic Review** 92 (March 2002): 120-142.

Klette, Tor Jakob, and Zvi Griliches, "Empirical Patterns of Firm Growth and R&D Investment: A Quality Ladder Model Interpretation," **Economic Journal** 110 (April 1998): 363-387.

_______________, and Samuel Kortum, "Innovating Firms and Aggregate Innovation," **Journal of Political Economy** 112 (October 2004): 986-1018.

Klevorick, Alvin K., Richard C. Levin, Richard R. Nelson, and Sidney G. Winter, "On the Sources and Significance of Interindustry Differences in Technological Opportunities," **Research Policy** 24 (1995): 185-205.

Kortum, Samuel, "Research, Patenting, and Technological Change," **Econometrica** 65 (November 1997): 1389-1419.

Mowery, David C., "The Boundaries of the U.S. Firm in R&D," Naomi Lamoreaux and Daniel Raff, editors, **Coordination and Information: Historical Perspectives on the Organization of Enterprise**, Chicago: University of Chicago Press for NBER, 1995.

Narin, Francis, Kimberly S. Hamilton, and Dominic Olivastro, "The Increasing Linkage Between U.S. Technology and Public Science," **Research Policy** 26 (1997): 317-330.

National Science Foundation, **CASPAR** (Computer-Aided Science Policy Analysis and Research) **Database**, Arlington, Virginia, www.nsf.gov/statistics

National Science Foundation, **Funds for Performance of Basic Research, Applied Research, and Development, by Industry and Source of Funds**, Arlington, Virginia, www.nsf.gov/statistics/iris, various years.

Nelson, Richard R., “The Link between Science and Invention: The Case of the Transistor,” in Richard R. Nelson, editor, **The Rate and Direction of Inventive Activity**, Princeton, New Jersey: Princeton University Press for the National Bureau of Economic Research, 1962.

________________, “The Role of Knowledge in R&D Efficiency,” **Quarterly Journal of Economics** 97 (August 1982): 453-370.

Peri, Giovanni, “Determinants of Knowledge Flows and Their Effect on Innovation,” **Review of Economics and Statistics** 87 (May 2005): 308-322.

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 (1982a): 227-245.

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

Stern, Scott, “Do Scientists Pay to be Scientists?” **Management Science** 50 (June 2004): 835-853.

Trajtenberg, Manuel, “A Penny for Your Quotes: Patent Citations and the Value of Innovation,” **RAND Journal of Economics** 20 (Spring 1990): 172-187.

Von Hippel, Eric, **Democratizing Innovation**, Boston: MIT Press, 2005.

Table 1
Principal Variables for a Study of Scientific Discovery in Firms:
Simple Statistics from a Panel of Firms, Fields, and Years

Variable	Mean	Standard Deviation
Papers	28.4	60.0
Citation-Weighted Papers	71.9	241.4
Basic Research Stock	145.5	242.4
Science Citations, Firms to Universities		
Probability of Citation [a]	0.006	0.008
Citation Knowledge Flow, Universities [b]	41.0	101.6
Science Collaborations, Firms and Universities		
Probability of Collaboration [a]	0.010	0.024
Collaboration Knowledge Flow, Universities [b]	2.9	6.5
Science Citations, Firms to Themselves		
Probability of Citation [a]	0.014	0.025
Citation Knowledge Flow, Same Firm [b]	41.8	141.5
Science Citations, Firms to Other Firms		
Probability of Citation [a]	0.005	0.011
Citation Knowledge Flow, Other Firms [b]	24.5	58.7

Source: Thomson Scientific and authors' calculations. [a] Mean probability is the average of the citation or collaboration rate for each cell, denoted $\bar{x}_i = \left(\overline{c_{ij} / n_j}\right)$. [b] Knowledge Flow is the citation or collaboration rate times the stock of R&D (in millions of 1992 dollars) by firm, field, and year, with a lag of one year. See the text for additional discussion.

Table 2
Total Knowledge Flows by Type, Field, and Industry

Field or Industry	Citation Knowledge Flow, Universities	Collaboration Knowledge Flow, Universities	Citation Knowledge Flow, Same Firm	Citation Knowledge Flow, Other Firms
Fields of Science				
Biology	71,284	2,025	51,897	30,727
Chemistry	18,820	810	43,420	25,136
Computer Science	8,153	1,759	4,734	3,043
Engineering	21,657	4,201	13,545	11,571
Medicine	43,429	2,903	27,217	12,123
Physics	21,423	1,570	43,358	25,992
Total Across Fields	184,766	13,267	184,171	108,591
Industries				
Petrochemicals	18,893	1,226	27,518	15,886
Drugs and Biotechnology	102,836	4,018	76,011	43,800
Metals	1,446	184	806	1,149
Machinery	748	178	333	928
Computers	5,025	793	1,999	5,708
Electrical Equipment	11,805	1,718	4,809	10,893
Transportation Equipment	9,344	1,442	7,547	6,131
Instruments	7,057	625	11,956	4,656
Communications	11,075	1,161	8,089	10,654
Software and Business Services	14,980	1,744	42,314	7,546
Misc. Agriculture & Manufacturing	1,556	177	2,789	1,242
Total Across Industries	184,766	13,267	184,171	108,591

Notes: Except for rounding error totals equal sums across fields and industries. Total knowledge flow is the sum over citing and cited (collaborating and collaborated) cells of the citation/collaboration rate, times the cited R&D stock (in millions of 1992 dollars). The citation/collaboration rate c_{ij} / n_j is the number of citations from citing industry and industrial science field i to the cited academic science field j divided by the number of potentially cited papers.

Table 3
Firm and Field Level Science Production Functions, Papers
(Robust, Clustered Standard Errors in Parentheses)

Variable or Statistic (Parameter in Italics)	Dependent Variable: Log (Papers)			
	3.1	3.2	3.3	3.4
Time, Field, & Firm Dummies Included	Yes	Yes	Yes	Yes
Test for Joint Significance of Dummies	206.8^{+++}	37.7^{+++}	87.4^{+++}	156.1^{+++}
Log (Basic Research Stock)	0.124** (0.027)	0.061* (0.024)	0.027 (0.024)	0.033 (0.023)
Log (Citation Knowledge Flow, Universities) ($\beta_{univ,cit}$)			0.170** (0.015)	0.126** (0.013)
Log (Collaboration Knowledge Flow, Universities) ($\beta_{univ,coll}$)				0.096** (0.008)
Log (Citation Knowledge Flow, Same Firm) ($\beta_{firm,selfcit}$)		0.195** (0.016)	0.094** (0.008)	0.084** (0.007)
Log (Citation Spillover, Other Firms) ($\beta_{firm,cit}$)			0.102** (0.008)	0.088** (0.007)
Root Mean Squared Error (σ)	1.15	0.96	0.83	0.81
R^2	0.47	0.63	0.72	0.74

Notes: Number of observations is N=4,340. Time period is 1988-1999. Method is OLS. Data are a panel of numbers of papers produced in a given firm, science, and year. ** Variable is significantly different from zero at the one percent level. * Variable is significantly different from zero at the five percent level. $^{+++}$ F-statistic is significant at the 0.1 percent level.

Table 4
Firm and Field Level Science Production Functions,
Citation-Weighted Papers
(Standard Errors in Parentheses)

Variable or Statistic (Parameter in Italics)	Dependent Variable: Log (Citation-Weighted Papers)			
	4.1	4.2	4.3	4.4
Time Dummies Included	Yes	Yes	Yes	Yes
Test for Joint Significance of Dummies	13.3	12.4	31.7^{+++}	33.5^{+++}
Log (Basic Research Stock)	0.562**	0.474**	0.282**	0.258**
	(0.075)	(0.071)	(0.065)	(0.064)
Log (Citation Knowledge Flow, Universities) ($\beta_{univ,cit}$)			0.336**	0.302**
			(0.039)	(0.040)
Log (Collaboration Knowledge Flow, Universities) ($\beta_{univ,coll}$)				0.114**
				(0.029)
Log (Citation Knowledge Flow, Same Firm) ($\beta_{firm,selfcit}$)		0.180**	0.167**	0.167**
		(0.026)	(0.027)	(0.027)
Log (Citation Knowledge Flow, Other Firms) ($\beta_{firm,cit}$)			0.205**	0.199**
			(0.028)	(0.028)
σ_u	4.20	3.53	2.21	2.09
σ_e	2.65	2.71	2.83	2.84
Log Likelihood	-6,489.8	-6,465.5	-6,397.7	-6,390.2

Notes: Number of observations is N_0=2,719. Number of left censored observations is N_1=448. Time period is 1988-1995. Method is Random Effects Tobit. Data are a panel of citation-weighted numbers of papers produced in a given firm, science field, and year. Citation weights are citations received by papers in their first five years. Papers published after 1995 have incomplete five-year histories and are omitted. σ_u is the square root of the variance component for firm and science field. σ_e is the square root of the variance component for firm, science field, and time. ** Variable is different from zero at the one percent level. * Variable is different from zero at the five percent level. $^{+++}$ F-statistic is significant at the 0.1 percent level.

Table 5
Wald Tests of Equality of the Coefficients,
Firm and Field Level Science Production Functions

Description of Null Hypothesis	Restriction (Parameter in Italics)	F-Statistics for Papers	F-Statistics for Citation-Weighted Papers
		Eq. 3.4	Eq. 4.4
1. Equality of university and firm citation knowledge flow elasticities	$\beta_{univ,cit} = \beta_{firm,cit}$	8.0^{++}	4.3^{+}
2. Equality of university citation and collaboration knowledge flow elasticities	$\beta_{univ,cit} = \beta_{univ,coll}$	4.1^{+}	12.0^{++}
3. Equality of university citation plus collaboration knowledge flow elasticities, with firm citation knowledge flow elasticity	$\beta_{univ,cit} + \beta_{univ,coll} = \beta_{firm,cit}$	67.3^{++}	16.4^{++}
4. Equality of university citation plus collaboration knowledge flow elasticities with self-citation and firm knowledge flow elasticities	$\beta_{univ,cit} + \beta_{univ,coll} = \beta_{firm,cit} + \beta_{firm,selfcit}$	8.8^{++}	0.7

Notes: $^{++}$ Equality restriction is rejected at the one percent level. $^{+}$ Equality restriction is rejected at the five percent level.

Table 6
Sensitivity of the Knowledge Flow Coefficients
To Alternative Specifications
(Standard Errors in Parentheses)

Specification	Log (Citation Knowledge Flow, Universities)	Log (Collaboration Knowledge Flow, Universities)	Log (Citation Knowledge Flow, Same Firm)	Log (Citation Knowledge Flow, Other Firms)	R^2
Panel A. Dependent Variable: Log (Papers), Method: OLS [a]					
Baseline Specification (Table 3, Eq. 3.4)	0.126** (0.013)	0.096** (0.008)	0.084** (0.007)	0.088** (0.007)	0.74
6.1 Allow Interactions of Field and Firm Dummies	0.037** (0.010)	0.027** (0.007)	0.020** (0.006)	0.018** (0.007)	0.89
6.2 Allow Interactions of Year and Firm Dummies	0.151** (0.020)	0.106** (0.012)	0.092** (0.009)	0.099** (0.010)	0.83
6.3 Drop Drug and Biotechnology Firms	0.099** (0.012)	0.070** (0.008)	0.071** (0.007)	0.074** (0.008)	0.72
6.4 Drop Software and Communications Firms	0.131** (0.012)	0.094** (0.009)	0.082** (0.007)	0.085** (0.007)	0.74
6.5 Drop Drug and Biotechnology, Software and Communications Firms	0.101** (0.011)	0.068** (0.008)	0.068** (0.007)	0.069** (0.007)	0.70
6.6 Drop Metals, Machinery, and Misc. Agriculture, Manufacturing Firms	0.130** (0.014)	0.102** (0.009)	0.087** (0.008)	0.092** (0.008)	0.74
6.7 Use Averaged University Citation and Collaboration Knowledge Flows	0.137** (0.014)	0.096** (0.008)	0.084** (0.007)	0.087** (0.007)	0.74
6.8 Use Naïve Citation and Collaboration Probabilities instead of knowledge flows	0.249** (0.034)	0.171** (0.023)	0.379** (0.035)	0.306** (0.037)	0.69

Table 6
Sensitivity of the Knowledge Flow Coefficients
To Alternative Specifications
(Standard Errors in Parentheses)

Specification	Log (Citation Knowledge Flow, Universities)	Log (Collaboration Knowledge Flow, Universities)	Log (Citation Knowledge Flow, Same Firm)	Log (Citation Knowledge Flow, Other Firms)	Log Likelihood
Panel B. Dependent Variable: Log (Citation-Weighted Papers), Method: Random Effects Tobit [b]					
Baseline Specification (Table 4, Eq. 4.4)	0.302** (0.040)	0.114* (0.029)	0.167** (0.027)	0.199** (0.028)	-6,390.2
6.9 Drop Drug and Biotechnology Firms	0.269** (0.044)	0.115** (0.033)	0.162** (0.030)	0.202** (0.032)	-5,332.5
6.10 Drop Software and Communications Firms	0.338** (0.042)	0.091** (0.030)	0.155** (0.028)	0.195** (0.029)	-5,929.3
6.11 Drop Drug and Biotechnology, Software and Communications Firms	0.304** (0.047)	0.087** (0.034)	0.149** (0.032)	0.196** (0.034)	-4,872.1
6.12 Drop Metals, Machinery, and Misc. Agriculture, Manufacturing Firms	0.288** (0.040)	0.119** (0.030)	0.138** (0.027)	0.194** (0.028)	-5,738.8
6.13 Use Averaged University Citation and Collaboration knowledge flows	0.348** (0.041)	0.138** (0.029)	0.162** (0.026)	0.196** (0.028)	-6,380.2
6.14 Use Naïve Citation and Collaboration Probabilities instead of knowledge flows	0.353** (0.098)	0.124 (0.069)	0.340** (0.071)	0.438** (0.087)	-6,450.1

Notes: [a] Robust, clustered standard errors appear in parentheses. All equations include fixed effects for firm, field, and year. [b] Equations include random effects or variance components for firm and field (σ_u); and for firm, field, and year observations (σ_e).

Table 7
Firm Level Science Production Functions
(Standard Errors in Parentheses)

Variable or Statistic (Parameter in Italics)	Dependent Variable: Log (Papers) [a] 7.1	Dependent Variable: Log (Citation-Weighted Papers) [b] 7.2
Statistical Method	Fixed Effects OLS	Random Effects Tobit
Time Period	1988-1999	1988-1995
Dummies Included	Firm, Year	Year
Test for Joint Significance of Dummies	282.5+++	19.6++
Log (Basic Research Stock)	0.100**	0.087
	(0.026)	(0.071)
Log (Citation Knowledge Flow, Universities) ($\beta_{univ,cit}$)	0.078**	0.441**
	(0.019)	(0.060)
Log (Collaboration Knowledge Flow, Universities ($\beta_{univ,coll}$)	0.046**	0.111**
	(0.012)	(0.043)
Log (Citation Knowledge Flow, Same Firm) ($\beta_{firm,selfcit}$)	0.026**	0.175**
	(0.009)	(0.033)
Log (Citation Knowledge Flow, Other Firms) ($\beta_{firm,cit}$)	0.043**	0.198**
	(0.012)	(0.044)
Number of Observations	1,486	928
Left Censored Observations	0	101
Root Mean Squared Error (σ)	0.493	--
R^2	0.93	--
σ_u	--	1.32
σ_e	--	2.27
Log Likelihood	--	-2,058.8

Notes: [a] OLS specification uses robust, clustered standard errors where the clustering variable is the firm. ** Variable is significantly different from zero at the one percent level. * Variable is significantly different from zero at the five percent level. +++ Statistic is significant at the 0.1 percent level. ++ Statistic is significant at the one percent level.

Table 8
Dynamic Panel Data Estimates, Firm Level Science Production Functions
(Standard Errors in Parentheses)

Variable or Statistic (Parameter in Italics)	Dependent Variable: Log (Papers)		
	8.1	8.2	8.3
Statistical Method	Difference GMM	Difference GMM [a]	System GMM [b]
Assumption on Knowledge Flows	Exogenous	Predetermined	Predetermined
Time Period	1988-1999	1988-1999	1988-1999
Dummies Included	Year	Year	Year
Log (Lagged Scientific Papers)	0.127	0.050	0.245**
	(0.085)	(0.119)	(0.044)
Log (Basic Research Stock)	0.045	0.198	0.212**
	(0.040)	(0.217)	(0.076)
Log (Citation Knowledge Flow,	0.091**	0.101**	0.126**
Universities) ($\beta_{univ,cit}$)	(0.022)	(0.031)	(0.018)
Log (Collaboration Knowledge Flow,	0.051**	0.026	0.044**
Universities) ($\beta_{univ,coll}$)	(0.009)	(0.028)	(0.015)
Log (Citation Knowledge Flow , Same	0.043**	0.043**	0.052**
Firm) ($\beta_{firm,selfcit}$)	(0.011)	(0.016)	(0.013)
Log (Citation Knowledge Flow, Other	0.053**	0.041*	0.063**
Firms) ($\beta_{firm,cit}$)	(0.013)	(0.019)	(0.016)
Number of Observations	1,210	1,210	1,210
Number of Firms	151	151	151
Wald χ^2 (degrees of freedom= 15)	264.4^{+++}	440.7^{+++}	929.6^{+++}
Test for autocorrelation of order 1	-3.88^{+++}	-3.01^{++}	-4.89^{+++}
Test for autocorrelation of order 2	1.26	0.75	1.78

Notes: [a] Instruments in the differenced equation include lag 2 and greater on papers, and lags 1 to 4 on knowledge flows. [b] Instruments in the differenced equation are as above. Instruments in the level equation are differences in papers and the knowledge flows. ** Variable is significantly different from zero at the one percent level. * Variable is significantly different from zero at the five percent level. $^{+++}$ Statistic is significant at the 0.1 percent level. $^{++}$ Statistic is significant at the one percent level.

Table 9
Industry, Field, and Year Level Science Production Functions
(Robust, Clustered Standard Errors in Parentheses)

Variable or Statistic (Parameter in Italics)	Dependent Variable: Log (Papers) [a]	Log (Citation-Weighted Papers) [b]
	9.1	9.2
Statistical Method	OLS	OLS
Time Period	1988-1999	1988-1995
Dummies Included	Industry, Field, Year	Year
Test for Joint Significance of Dummies	11.3[+++]	17.4[++]
Log (Basic Research Stock)	0.262	-0.215
	(0.157)	(0.201)
Log (Citation Knowledge Flow, Universities)	0.374**	0.290**
($\beta_{univ,cit}$)	(0.063)	(0.116)
Log (Collaboration Knowledge Flow, Universities)	0.160**	0.300*
($\beta_{univ,coll}$)	(0.040)	(0.117)
Log (Citation Knowledge Flow, Same Firm)	0.054*	0.147**
($\beta_{firm,selfcit}$)	(0.026)	(0.023)
Log (Citation Knowledge Flow, Other Firms)	0.087*	0.182*
($\beta_{firm,cit}$)	(0.035)	(0.084)
Number of Observations	728	485
Root Mean Squared Error (σ)	0.672	1.416
R^2	0.85	0.75

Notes: Data are a panel of industries, firms, and years. [a] Time period is 1988-1999. Standard Errors are robust and clustered by industry-field. [b] Time period is 1988-1995. ** Variable is different from zero at the one percent level. * Variable is different from zero at the five percent level. [+++] Statistic is significant at the 0.1 percent level. [++] Statistic is significant at the one percent level.

Table 10
Marginal Products of the Knowledge Flows

Level of the Analysis And Statistic	Citation Knowledge Flow, Universities	Collaboration Knowledge Flow, Universities	Citation Knowledge Flow, Same Firm	Citation Knowledge Flow, Other Firms
Panel A. Firm and Field				
Table 3, Eq. 3.4: Papers (Mean=28.4)				
Mean (Mill. $)	41.0	2.9	41.8	24.5
Elasticity	0.126	0.096	0.084	0.088
Marginal Product (Papers/Mill. $)	0.087	0.940	0.057	0.102
Table 4, Eq. 4.4: Cite-Weighted Papers (Mean=71.9)				
Mean (Mill. $)	35.6	2.7	42.0	22.2
Elasticity [a]	0.302	0.114	0.167	0.199
Marginal Product (Papers/Mill. $)	0.509	2.535	0.239	0.538
Panel B. Firm				
Table 7, Eq. 7.1: Papers (Mean=86.1)				
Mean (Mill. $)	124.2	8.9	123.9	73.0
Elasticity	0.078	0.046	0.026	0.043
Marginal Product (Papers/Mill. $)	0.054	0.445	0.018	0.051
Table 7, Eq. 7.2: Cite-Weighted Papers (Mean=218.9)				
Mean (Mill. $)	106.0	8.1	124.9	65.3
Elasticity [b]	0.441	0.111	0.175	0.198
Marginal Product (Papers/Mill. $)	0.911	2.673	0.273	0.591
Panel C. Industry and Field				
Table 9, Eq. 9.1: Papers (Mean=177.0)				
Mean (Mill. $)	106.0	18.3	254.4	150.0
Elasticity	0.374	0.160	0.054	0.087
Marginal Product (Papers/Mill. $)	0.625	1.548	0.038	0.103
Table 9, Eq. 9.2: Cite-Weighted Papers (Mean=420.9)				
Mean (Mill. $)	203.8	15.6	239.9	125.7
Elasticity	0.290	0.300	0.147	0.182
Marginal Product (Papers/Mill. $)	0.599	8.094	0.258	0.609

Notes: [a] Elasticity is the Tobit estimate shown in equation 4.4, Table 4, multiplied by 0.835, the probability that the data are not censored. [b] Elasticity is the estimate shown in equation 7.2, Table 7, multiplied by 0.891, the probability that the data are not censored.

Figure 1a--Shares of Knowledge Flows
Petrochemicals

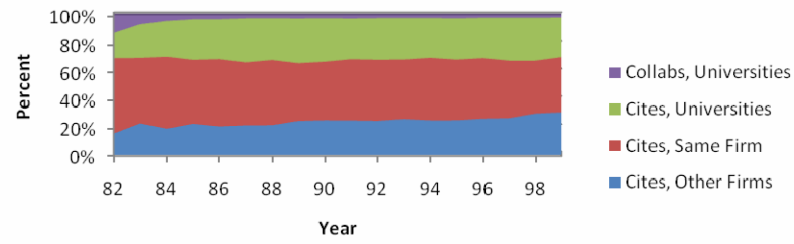

Figure 1b--Shares of Knowledge Flows
Drugs and Biotechnology

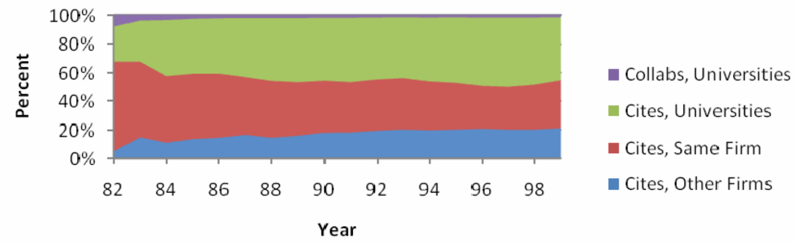

Figure 1c--Shares of Knowledge Flows
Metals, Machinery, Miscellaneous

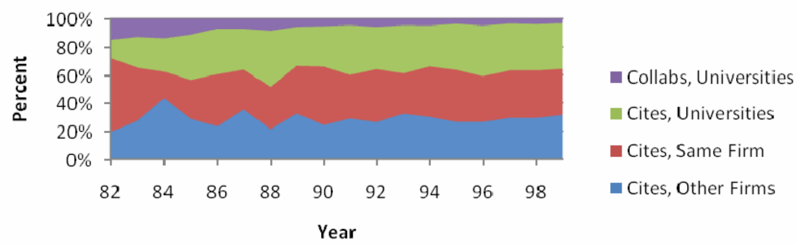

Figure 1d--Shares of Knowledge Flows
Computers, Communications and Software

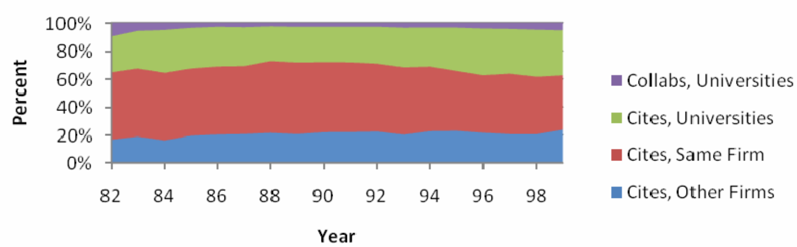

Figure 1e--Shares of Knowledge Flows
Electrical Equipment and Instruments

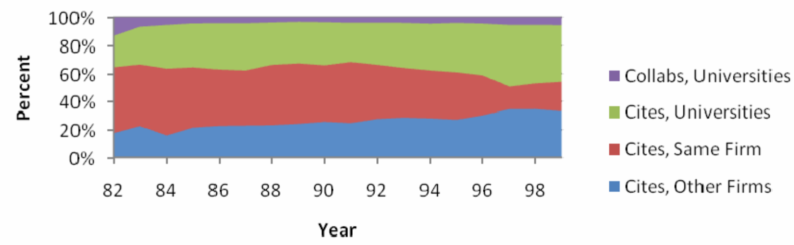

Figure 1f--Shares of Knowledge Flows
Transportation Equipment

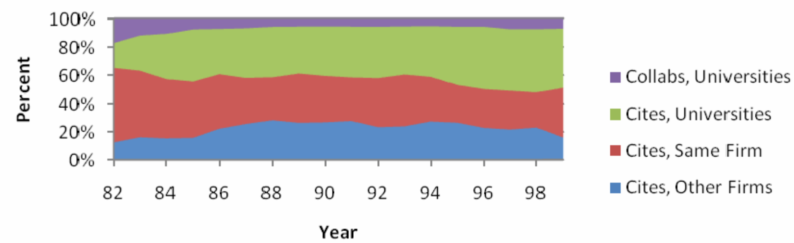

Figure 2a--Shares of Knowledge Flows
Biology

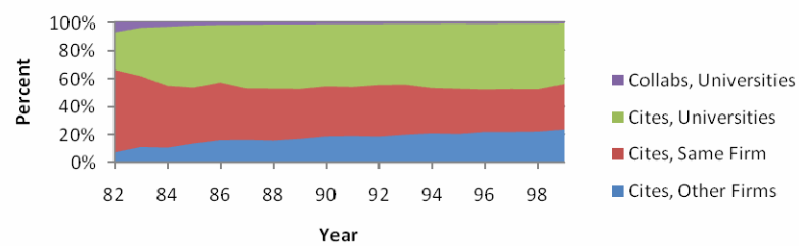

Figure 2b--Shares of Knowledge Flows
Chemistry

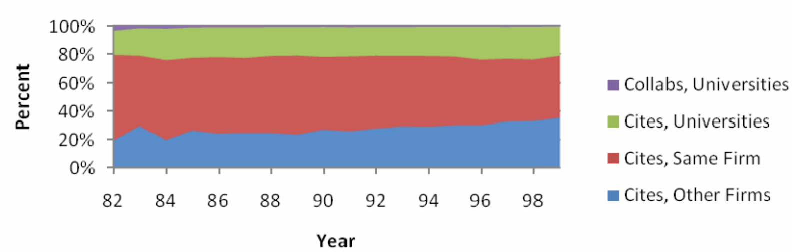

Figure 2c--Shares of Knowledge Flows
Computer Science

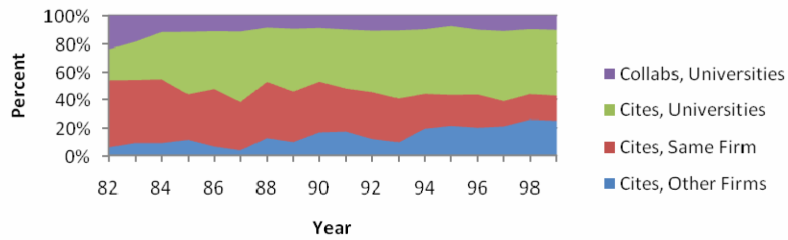

Figure 2d--Shares of Knowledge Flows
Engineering

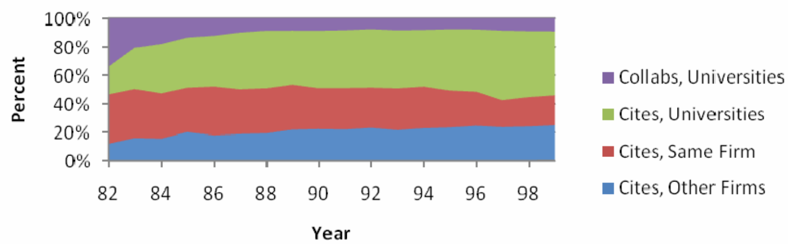

Figure 2e--Shares of Knowledge Flows
Medicine

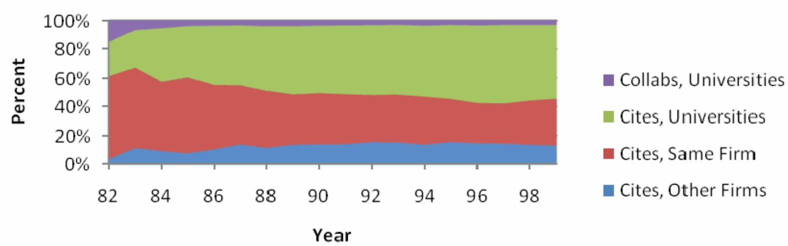

Figure 2f--Shares of Knowledge Flows
Physics