Exploring the Link Between Academic Science and Industrial Innovation[*]

Lee Branstetter

Discussion Paper No. 29

Lee Branstetter

Associate Professor
Columbia Business School
813 Uris Hall
New York, NY 10027
and NBER

This Version: October, 2003

Discussion Paper Series
APEC Study Center
Columbia Business School

May 2004

[*] **Acknowledgements**: I wish to thank Jim Adams, Iain Cockburn, Robert Feenstra, Rebecca Henderson, Marvin Lieberman, David Mowery, Ariel Pakes, and participants at the 2000 NBER Summer Institute (Innovation Policy and the Economy workshop), the 2001 NBER Conference on Data and Policy Issues of the Scientific Workforce, the UCLA Anderson Graduate School of Management Policy Seminar, and the 2003 NBER-CRIW-CREST Zvi Griliches Memorial Conference for useful comments and suggestions. I am especially grateful to Adam Jaffe and Joshua Lerner for detailed feedback on an earlier draft. I also wish to thank a number of academic scientists and industrial R&D managers for providing me with their insights into the process by which knowledge flows from academia to industry. I am indebted to Masami Imai, Hiau-Looi Kee, and Kaoru Nabeshima for excellent research assistance. I would like to thank Tony Breitzman and Francis Narin of CHI-Research, Adam Jaffe, and Marie and Jerry Thursby for their help in obtaining the data used in this study. This project was funded by grants from the University of California Industry-University Cooperative Research Program, the NBER Project on Industrial Technology and Productivity, the Japan Foundation Center for Global Partnership, the National Science Foundation, and the Institute for Governmental Affairs at UC-Davis.

I. Introduction

This paper is motivated by Figure 1. The figure counts the number of citations made by U.S. patents to published scientific papers. As the reader can see, these counts have gone up dramatically over the last few years. In fact, these counts have grown much more quickly than the aggregate number of patents, the number of scientific publications, or the level of public sector R&D.[1]

Figure 1 is interesting because it may imply that the nature of the relationship between "academic science" and "industrial innovation" has changed. In some sense, inventors may be finding that academic science is more useful in the process of creating new technology than it used to be. This is significant, because in the United States and elsewhere, the large postwar expansion in public funding for scientific research has been predicated on the belief that investments in "basic science" would stimulate economic growth in the long run. If the positive impact of these investments in basic science on industrial innovation has been increasing in recent years, then this would have interesting implications for U.S. science policy and, potentially, for the prospects of continued technology-driven economic growth.

To better understand what Figure 1 implies, I begin with an in-depth study of the citing patents, themselves. What kinds of patents cite academic science? How do citing patents differ from patents that do not cite academic research? How have these patterns changed over time? What kinds of academic articles get cited? How has the distribution of citations across patent assignees changed over time? The paper directly addresses these questions and takes a first step towards measuring the impact of these knowledge flows on research productivity. Finally, using data from a sample of California research universities, the paper produces estimates of the innovative output of these sample campuses, controlling for public R&D expenditure across campuses and disciplines. The implications of these results are discussed in the conclusion.

[1] I am certainly not the first person to document this increase. These trends have been analyzed by Narin, Hamilton, and Olivastro (1997) and are clearly evident in the most recent edition of the National Science Board's *Science and Engineering Indicators*. See also Hicks et. al. (2001).

II. Patent Citations of Academic Papers as Indicators of Knowledge Spillovers

Prior Research on University-Industry Research Interaction

A number of studies have examined aspects of the impact of academic science on industrial innovation. A full review of even the recent literature is beyond the scope of this paper, so I will only mention a few studies from the streams of research on which my paper directly builds.[2] One such stream has used case studies or surveys in an attempt to assess both the magnitude of this impact and the channels through which it flows.[3] Mansfield (1995) directly interviewed industrial research directors to obtain their assessments of the impact of academic research on industrial R&D. Cohen et. al. (1994) have continued in this tradition, surveying a large cross-section of firms on the impact of academic science on their own research productivity and the means by which these knowledge flows are mediated.[4] A second stream of research has undertaken quantitative studies of knowledge spillovers from academic research. Jaffe (1989) and Adams (1990) were early contributors to this literature. More recently, Jaffe et. al. (1993, 1996, 1998) have used data on university patents and citations to these patents to quantify knowledge spillovers from academic science.[5] A related stream of research has undertaken quantitative analysis of university-industry research collaboration. Contributors include Zucker et. al. (1998) and Cockburn and Henderson (1998, 2000). A number of papers in this literature have studied "start-up" activity related to academic science or academic scientists, such as Zucker et. al. (1998) or Audretsch and Stephan (1996). Finally, several recent studies have examined university licensing of university generated inventions, such as Barnes et al. (1998), Mowery et. al. (1998), Shane (2000, 2001), and Thursby and Thursby (2002).

Patent Citations to Academic Papers

[2] For a more comprehensive review of the more recent literature, see Agrawal (2001). The literature on this topic by both economists and noneconomists goes back several decades. See, for instance, Marquis and Allen (1966), Price (1965), Lieberman (1978), Schmookler (1966), and Sherwin and Isenson (1967).
[3] A more historical perspective is provided by Rosenberg and Nelson (1994). See also Mowery (1981).
[4] Many others have contributed to this "case study/survey" literature, including Faulkner and Senker (1995) and Gambardella (1995).
[5] Barnes, Mowery, and Ziedonis (1998) and Mowery, Nelson, Sampat, and Ziedonis (1998) have undertaken a similar study for a smaller number of universities. See also Agrawal and Cockburn (2003).

This paper will use patent citations to academic papers to measure "knowledge spillovers" between academic science and industrial R&D. It is not the first research project to use such data – Francis Narin and his collaborators have pioneered the use of these data in large-sample "bibliometric" analysis.[6] As indicators of knowledge spillovers from academia to the private sector, these data have a number of advantages. The academic promotion system creates strong incentives for academic scientists to publish all research results of scientific merit. As a consequence, the top-ranked research universities generate thousands of academic papers each year. The available data show that patent citations to these papers have been growing very rapidly over the last several years, far outstripping the growth rates in patenting, publications, or R&D spending. Figure 1, which shows national growth in such citations, is taken directly from a recent edition of the NSF's *S&E Indicators*.[7]

In response to the Bayh-Dole Act and other public policy measures, universities have increased the extent to which they patent the research of university-affiliated scientists. They have also increased the extent to which they license these patented technologies to private firms. Nevertheless, it is clear to observers that only a *tiny fraction* of the typical research university's research output is ever patented, and only a fraction of this set of patents is ever licensed. Agrawal and Henderson (2002) strongly emphasize this point in their recent study, and point to additional corroborating evidence. Given this, patent citations to academic papers may provide a much broader window through which to observe knowledge spillovers from academia to the private sector than the available alternatives.

I will investigate patent citations to academic science using two distinct empirical approaches. I start with a random sample of 30,000 utility patents, approximately 4,500 of which make at least one citation to "science." Using these data, I can examine how citing patents differ from others. I will also present a number of tests of the geographic localization of knowledge spillovers. Finally, I can assess whether patents that make

[6] See Narin et. al. (1997) and Hicks et. al. (2001) for recent examples of this work. Other recent studies employing patent citations to papers include Fleming and Sorenson (2001), Sorenson and Fleming (2001), and Lim (2001).
[7] Figure 1 is constructed from data placed on the NSF web site which tracks the increase in patent citations to scientific and technical articles, of which those authored by scientists at research universities constitute the largest part.

references to scientific publications are "better" than patents that do not. This would be consistent with, but not proof of, the view that knowledge spillovers from "academic science" are actually *beneficial* to the citing inventors.

In tracking patent citations to papers, one would like to control for the R&D inputs associated with the generation of these papers. I do this by using information on the universe of patent citations to the papers generated by two university systems -- the University of California and Stanford -- over the 1987-1996 period. In the context of this data set, I have constructed measures of academic output for the specific university systems mentioned above. These measures will include annual counts of papers, counts weighted by the degree to which these papers are cited in subsequent academic papers, and counts weighted by the degree to which these papers are cited in subsequently granted patents, broken down by campus and field of technology. These output measures can then be regressed on measures of institutional attributes, including levels of R&D input. One can think of this as estimating an academic "production function," where one of the output measures provides a quantitative indication of the extent to which scientific research in field *i* at institution *j* impacts the subsequent activity of commercial inventors.

III. Evidence from the Random Sample

I begin by presenting evidence drawn from a random sample of nearly 30,000 U.S. utility patents.[8] Given the sample size, I can be reasonably confident that sample trends will be reflective of trends in the underlying population. Because this data source may be unfamiliar to some readers, it is worthwhile to spend some time simply reviewing basic trends and tendencies in the raw data.

Graphing Trends in Patent Citations to Academic Science

Figure 2 examines the distribution of citations across scientific fields, where the designation of field reflects the nature of the science being cited by the patent, rather than

[8] Comprehensive data on the citations made by these patents to nonpatent documents were provided by CHI Research, under a contract which prevents them from being made publicly available to all researchers. Regression analysis of these patents required that the data from CHI Research be linked to data in the NBER Patent Citation database. For a small number of the CHI patents, no link could be found, so the data base on which regression analysis is based has slightly more than 29,800 observations.

the technological class of the patent itself. It is immediately obvious that the most frequently cited fields are "biomedical research" and "clinical medicine." While there are also substantial numbers of citations to other sciences, these two fields, which we will collectively refer to as "biomedical science," constitute the majority of citations when one aggregates across time.

Figure 3 gives the reader a sense of how changes in the number of patents citing science over time have largely been driven by changes in the number of patents making citations to "biomedical science." The date here is the date of application of the citing patent, and the graphs demonstrate that there has been a sharp acceleration in the number of citing patents in recent years, with most of this increase driven by an increase in the number of patents citing biomedical science. Figure 4 extends this analysis by counting not the number of citing patents, but rather the number of citations to science made by those patents, and graphing that together with the total number of citations made by patents in classes typically associated with drugs, medicines, and other applications of "biotechnology." Figure 4 strongly suggests that much of the increase in the increase in aggregate patent citations to science in Figure 1 is driven by citation activity in these patent classes.

Figure 5 illustrates the breakdown in citations to scientific papers by category of cited institution. As one can see, scientific papers generated by universities are the most frequently cited kind of scientific publication. However, there are a substantial number of citations made to the academic publications of private firms. This underscores the reality that many large corporate R&D departments, particularly in the chemicals and pharmaceuticals industries, encourage their staff to publish in the scientific literature.[9] A much smaller fraction of total citations go to papers generated by nonprofit, non-university research organizations, including government agencies, national laboratories, and private research foundations.

Figure 6 illustrates the extent to which the science pool upon which U.S. patents draw is a global one. As can be seen, approximately 41% of the science citations made by U.S. patents are made to institutions located outside the United States. In a sense, this

---

[9] For studies on why firms might allow R&D personnel to publish, see Cockburn and Henderson (1998) and Stern (1999).

is not surprising, given that approximately half of the patents granted by the U.S. Patent and Trademark Office over the last 20 years have been made to foreign inventors. Although this graph does not do so, it is possible to look at science citations made by U.S. inventors. These tend to be made to predominately U.S. institutions, and foreign inventors show a similar tendency to disproportionately cite science generated in their own countries.[10]

It is also of interest to plot the histogram for lags between the publication of an academic article application date of the patent that cites it. This is graphed out in Figure 7. As is immediately obvious, the modal lag between paper publication and grant application is three years, suggesting that spillovers from academic science to industrial research are quite rapid. Note that the left tail of this distribution actually includes a small number of negative lags. In some cases, this appears to represent coding error in terms of the publication date of the paper, but in others, it seems that (forthcoming) papers are being cited by patent applicants even prior to their publication in the literature. The distribution is clearly skewed to the right – papers continue to receive citations long after publication – and the picture suggested by this graph is strongly reminiscent of the double-exponential functional form used by Jaffe and Trajtenberg (1996, 2002) in their pathbreaking work on the modeling of patent citations to previous patents.

Finally, a geographic perspective on patent citation activity is provided in Figure 8. This figure shows the density of patent citations to academic science over the entire sample period, mapped to geographic space. Citing patents are assigned to U.S. counties based on the address of the first inventor, and the vertical dimension in this figure gives the number of patents making citations to science. Clearly, citation activity is localized in the East and West Coasts, and reflects the "bicoastal" distribution of both industrial and academic research activity.

Practitioner Perspectives from Field Interviews

One of the useful features of the citations data used in this section is that it is quite easy to identify highly cited scholars and "intensively citing" firms – that is, firms whose patents frequently cite scientific research conducted by various kinds of institutions,

---

[10] This pattern of disproportionately "intranational" citation was noted by Narin (1995).

including universities. To date, I have contacted 7 academics from various scientific disciplines and campuses, most of them "highly cited" by patents. I have also interviewed 7 corporate executives from "intensively citing" firms who were closely connected to corporate patenting and R&D efforts as well as 2 independent patent lawyers involved in intellectual property issues.

One issue of particular interest in these interviews was the extent to which the recent surge in citations to academic papers represents a real increase in the incidence of knowledge spillovers versus simply a change in citations practices. *Every interviewee involved in bioscience agreed with the view that the "knowledge spillovers" from academic research to corporate R&D have grown increasingly strong over recent years, and most believed that the increase in citations, in part, reflected that*. These increased spillovers stem in part from a fundamental change in the way pharmaceutical and biomedical products firms conduct research. Modern drug discovery techniques are closely based on relatively recent biological science. The compression of the product development cycle in these industries has led to a situation in which major new discoveries in academic science touch off a product development race in which leading firms immediately start applied research programs based on these new scientific discoveries. Pharmaceutical firms no longer patent only narrowly defined chemical compounds but also gene sequences, CDNA products, and research methodologies – and these more complicated inventions generally draw heavily on multiple sources of academic science.

Several interviewees also pointed out that, in the biotechnology sector, the boundary between "academic" and "commercial" research has blurred as academic scientists have founded independent firms, served on the scientific advisory boards of biotech firms, or entered into long-term consulting relationships with firms without ever relinquishing their academic positions. The presence of hundreds of such individuals with a foot in both worlds constitutes a kind of human bridge through which important new scientific knowledge quickly diffuses to corporations. Since individual academics are now competitors and collaborators in commercial product development, it is not surprising that citations to their scientific work are increasing.

Econometric Results from the Random Sample

Having described some features of the raw data, and having related some of the impressions of practitioners in the field, I now seek to obtain econometric estimates of the *conditional* impact of various attributes of citing patents on the probability of citation of science, holding others constant. The nature of the data suggests several alternative approaches. One approach is to run logit regressions in which the dependent variable is a binary variable equal to 1 if the patent in question made at least one reference to "science." Independent variables of interest include dummy variables for the (application) year of the patent cohort, the technology category of the patent, the category of organization to which the patent is assigned, and a crude measure of geographic proximity between the region in which the (first) inventor of the patent is located and the region(s) in which academic science is produced.

Collapsing the number of patent citations made to science to a binary variable may "throw away" useful information. An obvious alternative approach is to take as the dependent variable the number of citations made to science, generating a count variable that varies from zero to 103 in my sample. The appropriate statistical model is one which takes into account the count nature of the dependent variable. I have used the negative binomial model, running regressions of the number of academic citations on the same set of independent variables listed in the previous paragraph.

A third potential approach is to focus only on those observations for which I observe at least one citation to academic science. The paper presents results obtained from a truncated negative binomial model. While this approach ignores the differences between citing and nonciting patents, it generates some benefits in that I can conduct more nuanced comparisons of citing patents to one another. In particular, I can include in these regressions a continuous measure of geographic distance between the location of the patent inventor and the location where the cited science was generated – a measure which does not exist for nonciting patents.

With all of these statistical approaches, there are also different ways in which I can define "academic science" and thus measure patent citations to it. The most comprehensive such measure is to include in our counts all nonpatent citations which appear to be to "scientific documents," including conference proceedings and technical manuals. A narrower measure would be to count all references to articles in journals tracked by the Institute for Scientific Information database. This database tracks many (if not most) of the peer-reviewed journals across all major scientific disciplines, more precisely corresponding to the output of “academic science.” A still narrower measure would count only references to *university*-authored papers in tracked journals. This distinction is useful because, in some scientific disciplines, large corporate R&D labs and public science agencies generate a substantial contribution to “academic science,” publishing in the same journals as their university-affiliated peers.

Table 1 presents results based on a logit specification. The first five rows present the coefficients on dummy variables equal to one if the patent assignee falls into one of the five listed categories: university, non-profit R&D organization (many of these are research hospitals), U.S. government agency (i.e., NASA), foreign (foreign firms, individuals, and government agencies are all placed in this category), and “other” (the largest fraction of which are U.S. individuals). The reference category here is private firms. It is immediately clear that universities, nonprofit R&D organizations, and U.S. government agencies are all more likely to cite academic research than are firms. This differential gets generally more pronounced as one restricts the definition of what constitutes academic science. That being said, the vast majority of citing patents are generated by firms. Though not shown here, regressions run using only firm data do not reveal patterns of citation qualitatively different from those shown in this and the next several tables.

The next set of dummy variables corresponds to the technology class of the citing patent. Using a taxonomy developed by Adam Jaffe and Manuel Trajtenberg, I have aggregated the primary patent classes of the U.S. Patent and Trademark Office patent classification system into six groups – chemicals, communications/computers, drugs/medical, electronics/electrical machinery (not directly computer related), mechanical devices, and a catch-all “other” category which constitutes the reference group in these regressions.[11] Patents in the drugs/medical category stand out as being disproportionately likely to cite. This differential effect gets dramatically stronger as I

[11] I thank Adam Jaffe for providing this taxonomy in electronic form. Note that there are several hundred primary patent classes.

narrow the definition of academic science across columns. Depending on the definition of the dependent variable, the chemicals category ranks second in terms of likelihood of citing.

The next variable attempts to gauge the extent to which invention building on academic research is geographically localized. “Science center” is a dummy variable equal to 1 if the patent inventor is located in one of the top 100 U.S. counties in terms of generation of scientific publications. Not surprisingly, this variable is positive and statistically significant at conventional levels. Patents generated in regions with high levels of proximate scientific research are disproportionately likely to cite science. However, it is hard to view this as strong evidence of geographic localization of spillovers from academic research. I will return to this point in a moment.

All regression specifications are run with patent application year cohort effects. While the coefficients are not shown in Tables 1-3, the results from Table 1, column 1 are graphed out in Figure 9, along with the 95% confidence bounds. What is evident from this graph is a pronounced rise in the tendency of patents to cite over time, controlling for the increase in university patenting and changes in the distribution of patents over classes with different tendencies to cite science. However, the most pronounced rise in this tendency came during the early 1980s. Unfortunately, due to data limitations in the random sample, I am unable to examine in detail the increase in citations reflected in the cohorts of patents granted after 1996.

Table 2 presents a similar set of regression results using the full count of citations as the dependent variable and a negative binomial regression approach. For the most part, the qualitative picture of citation patterns is similar to that of the logit regressions. The interpretation of the coefficients has changed slightly, in that the reported coefficients indicate the percent change in citations associated with a unit change in the reported variable. Therefore, column 3 indicates that drug/medical patents generate nearly 400% more citations than the reference group. Though not shown here, I also experimented with so-called “zero-inflated” negative binomial (ZINB) regressions, along the lines of Lambert (1992). This technique allows the econometrician to allow the propensity to cite at all to be determined by a different statistical process than that which determines how much a patent cites, conditional on it citing at least once. These experiments were motivated by the observation that the majority of patents in the data set make no citations at all, and there may plausibly be some unobserved latent variable (proximity to academic science in the technology space) which determines citation. However, experiments with ZINB regressions yielded results that were not qualitatively different from the results reported here. Furthermore, specifications tests did *not* indicate that a ZINB specification was preferred. Figure 10 graphs out the application year cohort effects for the regression shown in column 1.

Table 3 presents results that focus only on observations where at least one citation to academic science is observed. The appropriate statistical model is a truncated negative binomial model. A brief technical treatment of the estimator is given in the appendix. These regressions were motivated by the desire to compare citing patents to one another along dimensions for which there are no corresponding measures for nonciting patents. Among these are the average linear distance between the patent inventor and the cited science source. Note that this variable is only available in cases where the cited science sources could be geographically located, which means that some citing patents are not included in this regression. In addition, only measured distance between the patent inventor and cited science sources *located in the United States* was used in calculating average linear distance.[12] Care must therefore be taken in comparing these regression results to those of the earlier tables. The coefficient on the distance variable "Ave_Dist" is positive but statistically indistinguishable from zero. This suggests that, while citing patents are disproportionately likely to arise in regions with high levels of academic research, they cite research from a range of sources, not all of which are geographically proximate. This finding, in turn, suggests that care needs to be taken in thinking about the policy implications of the results indicative of geographic localization of academic knowledge spillovers.

Given the somewhat ambiguous findings regarding geographic localization, I present one more test, which explicitly takes into account the skewed distribution of innovative activity across U.S. regions. Following Jaffe et. al. (1993), I match each of the citing patents with a nonciting "control" patent issued on the same date in the same

[12] This stems from a limitation in the data set initially provided by CHI Research, which provided information on the geographic address of the cited institution only when that institution was located inside the United States.

patent class as the citing patent. Let p_c be the probability that a citation comes from the same county as that in which the cited "science source" is located. Let p_0 be the corresponding probability for a randomly drawn control patent. I test for "geographic localization of knowledge spillovers" using the following test statistic:

$$t = \frac{\hat{p}_c - \hat{p}_0}{\sqrt{[\hat{p}_c(1-\hat{p}_c) + \hat{p}_0(1-\hat{p}_o)]/n}} \qquad (1)$$

where the two terms in the numerator are the sample proportion estimates of p_c and p_0. The null hypothesis that $p_c = p_0$ is easily rejected at conventional levels.[13]

Other than the results on geographic localization, the general qualitative patterns of the earlier tables are consistent with the results of Table 3. One interesting difference is that the coefficient on the foreign assignee dummy is positive and significant. Evidently, foreign patents are less likely to cite, but, conditional on citing at least once, they cite relatively heavily.

Does Citation of Academic Science Make Inventions Better?

The discussion of trends in the citations data above is of limited interest unless the knowledge spillovers indicated by these citations are actually enhancing the research productivity of the firms and other organizations that receive them. Are innovators learning from academic science in such a way that they are able to produce more inventions than they otherwise could or better inventions than they otherwise could? Alternatively, does the information generated by academic science allow them to invent in areas in which they could not work without the pre-existing foundation of academic science on which to build?

It is very difficult to establish the technological *dependence* of a particular invention on a cited scientific article without engaging in an in-depth study of the invention and extensive interviews with its inventors. However, I can seek to measure whether or not patented inventions that cite UC or Stanford academic science are systematically "better" than patents that do not. The micro literature on patents has suggested several measures of patent "quality" – quantitative features of the patent

---

[13] This test was conducted using both the "state" and the "county" as the regional unit of analysis. The t-statistic of the difference in ratios was 11.27 for state-level comparisons, 11.03 for county-level comparisons.

document – that have been demonstrated to be positively correlated with the *ex-post* commercial and technological importance of the patent. Three such measures include counts of *ex-post* (or "forward") citations, counts of claims contained in the patent document, and a measure of "generality" proposed by Henderson, Jaffe, and Trajtenberg (1998). This latter measure is a quantitative index of the diversity of technological fields across which *ex-post* citations occur. An invention whose citations come from multiple technological fields can be thought of as having a more "general" impact than an invention whose citations come from a single technological field. The formal definition of the index is

$$Generality_i = 1 - \sum_{k=1}^{N_i} \left(\frac{Nciting_{ik}}{Nciting_i} \right)^2 \qquad (2)$$

where the numerator in the expression measures the number of citations to patent *i* coming from patent class *k*, while the denominator measures the total number of citations to patent *i* across all classes.

Table 4 presents the results of regressions in which these three measures of quality are the dependent variable, a dummy variable indicating patents which cite academic research is the chief independent variable of interest, and I use as controls measures of the patent cohort (application year) and technological field. The results in Table 4 suggest that patents citing academic research are significantly better according to all three indices of patent quality.

However, in this context, it is very difficult to interpret this result in a *causal* way. Are patents that cite academic research "better" because they cite, or do they tend to cite academic research from UC and Stanford more frequently because they are "better"? At this level of aggregation, it is difficult to determine which interpretation is correct.

IV. Evidence from an Academic Production Function

Basic Empirical Approach

Since the early 1970s, there has been a large shift in federal government research resources away from the physical sciences and toward the life sciences. Might this shift in the distribution of research resources account, in part, for the large measured impact of "biotech-related" sciences on industrial innovation?[14] In this section, I seek to address this problem by estimating what amounts to an "academic production function." Taking as my sample the University of California's campuses and research institutes and Stanford University, I will measure the "science output" of these university systems across campuses, scientific fields, and time. This science output, measured using a number of alternative indices, will be regressed on, among other things, total R&D spending, the number of enrolled graduate students, the number of postdoctoral researchers, and measures of average faculty salary across institutions and years. For concreteness, the estimating equation can be written *as*

$$Q_{ijt} = \alpha_i + \alpha_j + \alpha_t + \beta_1 R_{ijt} + \beta_2 G_{ijt} + \beta_3 P_{ijt} + \beta_4 Sal_{it} + \varepsilon_{ijt} \qquad (3)$$

where Q is some measure of science output for institution *i* in scientific field *j* in year *t*.

While a number of alternative output measures could be constructed and used, I will consider three. First, I will include a simple count of papers published by campus *i*, in field *j*, in year *t*. Second, I will weight these counts by the subsequent citation of these papers in the scientific literature. Third, I will weight the paper counts by the subsequent citation of these papers in patents. This last measure of scientific output will be of particular interest, as I can view it as a measure of the impact of a "campus-field" on subsequent private sector invention.

On the right hand side of the equation, the α's are estimated campus, scientific field, and year fixed effects. *R* is a measure of R&D input, the largest part of which is publicly funded. It makes sense to think of academic output as being the result of a *stock* of current and appropriately depreciated past R&D investment. Therefore, results are presented using an R&D stock rather than a flow. *G* and *P* measure the numbers of enrolled graduate students and employed postdoctoral researchers, respectively.[15] *Sal* is a measure of the average salary of scientific faculty across all departments. Since I do not have the breakdown by scientific field, it is indexed *i,t*.

Preliminary results from regressions of various version of (7) are presented in Table 5. Table 6 presents the regression coefficients on the campus and field dummy variables for column 3 of Table 5. Table 7 presents the results of a simple ANOVA that

[14] I wish to thank Ariel Pakes for pointing out the importance of controlling for inputs.
[15] The use of simple lags of *G* and *P* yielded results qualitatively similar to those shown in this draft.

gives the reader a sense of the fraction of the total variance in outcomes that can be ascribed to campus and field effects. The lessons of these results can be summarized as follows.

First, field effects account for a rather large share of the total variance in outcomes. Examination of the production function results indicates that the biotech-related fields of science are "productive," even accounting for their higher levels of funding. Second, campus effects also account for a nontrivial share of the total variance in outcomes, though this is a much smaller share than field effects. The campuses that are generally regarded as more "prestigious" are also the ones that seem to be more "productive." Third, increases in research funding yield increases in research output, but the measured impact is far less than one-for-one. To some extent, this may reflect the inadequacy of the measures of research input as well as long lags in the academic "production" process.

VII. Conclusions

Based on the results presented in this paper, I would like to conclude by making five observations. First, relative to other indicators of knowledge flow from academia to the private sector, citations to academic papers are relatively numerous, rich, and available across campuses and scientific disciplines. Quite simply, there is a great deal of information to be mined from this source, and the existing literature has only begun this process. In making this statement, I do not mean to imply that study of, for instance, licensing of university-generated patents is unimportant or uninteresting – far from it. I merely wish to emphasize that patent citations to science are worthy of academic scrutiny.

Second, this paper presents evidence that the propensity of successive cohorts to cite academic science has been growing over time, controlling for changes in the distribution of patents across fields over time and controlling for the rise in university patenting. One interpretation of this is that industrial research is steadily getting "closer" to academic science. Of course, this is not the only plausible interpretation of my finding, and further research is needed to confirm just what this increasing propensity to cite science actually means.[16]

Third, I have presented ample evidence that citations to academic science are highly concentrated in a small number of technologies. In particular, "biotech" related technologies and sciences play an extremely strong role in explaining overall trends in citations. In the light of recent research on the interaction between public science and private innovation in the pharmaceutical industry, this is hardly surprising.

Fourth, analysis at the patent level suggests that the incidence of citation of academic science is positively associated with measures of invention "quality." While this evidence is consistent with the idea that knowledge spillovers from academia make private inventions better, for reasons discussed at length in the paper, this does not constitute proof that the chain of causality runs from citation to invention quality. Certainly, further analysis is needed at the assignee (firm) level.

Finally, preliminary evidence based on an academic production function suggests that the relatively high impact of bioscience research on commercial invention does not seem to be driven purely by the relatively generous funding that university researchers in these fields have received.

There are a number of directions in which the research contained herein could be extended. For instance, to understand better the forces driving the aggregate increase in patent citations to academic papers, it would be useful to control for changes in the level and distribution across fields of potentially cited papers, as well as persistent differences across fields of science in the degree to which papers tend to be cited by patents. In their studies of patent citations to previous patents, Jaffe and Trajtenberg (1996, 2002) have taken such an approach. Branstetter (2003) has begun an investigation along these lines, using a "citation function" approach inspired by the work of Jaffe and Trajtenberg.

[16] Branstetter (2003) examines this issue in greater detail.

Technical Appendix: The Truncated Negative Binomial Estimator

Sketch Derivation of the Estimation Technique

A complete derivation of this model is given in Cameron and Trivedi (1998). This brief description draws heavily on Cameron and Trivedi (1998) and uses their notation.

The mean and variance of the Poisson distribution truncated at zero are

$$E[y_i \mid y_i > 0] = \frac{\mu_i}{1 - e^{-\mu_i}} \tag{4}$$

and

$$\begin{aligned} V[y_i \mid y_i > 0] &= E[y_i \mid y_i > 0][1 - \Pr[y = 0]E[y_i \mid y_i > 0] \\ &= \frac{\mu_i}{1 - e^{-\mu_i}} \left[1 - \frac{\mu_i e^{-\mu_i}}{1 - e^{-\mu_i}} \right] \end{aligned} \tag{5}$$

A more general negative binomial distribution truncated at zero, which is what is used in the paper, would have the following first two moments:

$$E[y_i \mid y_i > 0] = \frac{\mu_i}{1 - (1 + \alpha\mu_i)^{-\frac{1}{\alpha}}} \tag{6}$$

and

$$V[y_i \mid y_i > 0] = \frac{\mu_i}{1 - (1 + \alpha\mu_i)^{-\frac{1}{\alpha}}} \times \left[1 - (1 + \alpha\mu_i)^{-\frac{1}{\alpha}} \frac{\mu_i}{1 - (1 + \alpha\mu_i)^{-\frac{1}{\alpha}}} \right] \tag{7}$$

Note that the truncated Poisson, unlike the standard Poisson model, does *not* have equal first and second moments. As pointed out by Cameron and Trivedi (1998), misspecification of the distribution implies that the first conditional truncated moment, which depends on the correct probability of zero value, will also be misspecified, resulting in inconsistent estimates of our parameters if the parent distribution is incorrectly specified.

The left-truncated Poisson model can be estimated by maximum likelihood methods. Let the log likelihood estimation be based on n independent observations, such that

$$L(\beta) = \sum_{i=1}^{n} \left[y_i \ln(\mu_i) - \mu_i - \ln\left(1 - \exp(-\mu_i) \sum_{j=0}^{r-1} \mu_i^j / j! \right) - \ln(y_i !) \right] \tag{8}$$

where the MLE of β is the solution of

$$\sum_{i=1}^{n} [y_i - \mu_i - \delta_i] \mu_i^{-1} \frac{\partial \mu_i}{\partial \beta} = 0 \tag{9}$$

where

$$\delta_i = \frac{\mu_i h(r, \mu_i)}{[1 - H(r-1, \mu_i]} \tag{10}$$

Bibliography

Adams, J., 1990, "Fundamental Stocks of Knowledge and Productivity Growth," *Journal of Political Economy* 98: 673-702.

Adams, J. and Z. Griliches, 1996, "Research Productivity in a System of Universities," NBER working paper no. 5833.

Agrawal, A., 2001, "Research on University-to-Industry Knowledge Transfer: Framework of Existing Literature and Future Questions," *International Journal of Management Reviews*.

Agrawal, A. and I. Cockburn, 2003, "The Anchor Tenant Hypothesis: Exploring the Role of Large, Local, R&D-Intensive Firms in Regional Innovation Systems," forthcoming in the *International Journal of Industrial Organization*.

Agrawal, A. and R. Henderson, 2002, "Putting Patents in Context: Exploring Knowledge Transfer from MIT," *Management Science*, Vol. 48, No. 1.

Audretsch, D. and P. Stephan, 1996, "Company-Scientist Locational Links: The Case of Biotechnology," *American Economic Review*, Vol. 86, No. 3.

Barnes, M., D. Mowery, A. Ziedonis, 1998, "The Geographic Reach of Market and Nonmarket Channels of Technology Transfer: Comparing Citations and Licenses of University Patents," working paper.

Bush, V., 1945, *Science – the Endless Frontier: A Report to the President on a Program for Postwar Scientific Research*, Washington, D.C.: U.S. Government Printing Office.

Branstetter, L., 2003, "Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations," unpublished working paper.

Cameron, A. C. and P. Trivedi, 1998, *The Regression Analysis of Count Data*, Econometric Society Monograph No. 30, Cambridge: Cambridge University Press.

Cockburn, I. and R. Henderson, 2000, "Publicly Funded Science and the Productivity of the Pharmaceutical Industry," paper prepared for the NBER Conference on Science and Public Policy.

Cockburn, I. and R. Henderson, 1998, "The Organization of Research in Drug Discovery,"*Journal of Industrial Economics*, Vol XLVI, No. 2.

Cohen, W., R. Florida, L. Randazzese, and J. Walsh, 1998, "Industry and the Academy: Uneasy Partners in the Cause of Technological Advance," in R. Noll, ed., *Challenges to the Research University*. Washington, D.C.: Brookings Institution

Evenson, R. and Y. Kislev, 1976, "A Stochastic Model of Applied Research," *Journal of Political Economy* 84 (2): 265-282.

Faulkner, W. and J. Senker, 1995, *Knowledge Frontiers: Public Sector Research and Industrial Innovation in Biotechnology, Engineering Ceramics, and Parallel Computing*, Oxford: Clarendon Press.

Fleming, L. and O. Sorenson, 2001, "Science as a Map in Technological Search," working paper.

Gambardella, A, 1995, *Science and Innovation: The U.S. Pharmaceutical Industry during the 1980s*, Cambridge: Cambridge University Press.

Henderson, R., A. B. Jaffe, and M. Trajtenberg, 1998, "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988," *Review of Economics and Statistics*, 119-127.

Hicks, D., T. Breitzman, D. Olivastro, K. Hamilton, 2001, "The Changing Composition of Innovative Activity in the US – A Portrait Based on Patent Analysis," *Research Policy* 30, pp. 681-703.

Jaffe, A., 1989, "The Real Effects of Academic Research," *American Economic Review*, 79 (5), pp. 957-70

Jaffe, A., M. Trajtenberg, and R. Henderson, 1993, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, Vol. CVIII, No. 3.

Jaffe, A. and M. Trajtenberg, 1996, "Flows of Knowledge from Universities and Federal Labs: Modeling the Flow of Patent Citations over Time and across Institutional and Geographic Boundaries," NBER working paper no. 5712.

Jaffe, A. and M. Trajtenberg, 2002, *Patents, Citations, and Innovations: A Window on the Knowledge Economy*, MIT Press.

Jaffe, A., M. Fogarty, and B. Banks, (1998), "Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation," *Journal of Industrial Economics*, Vol. XLVI, No. 2.

Jensen, R. and M. Thursby, 1999, "Proofs and Prototypes for Sale: The Licensing of University Inventions," *American Economic Review*.

Kortum, S. and J. Lerner, 1998, "Stronger Protection or Technological Revolution: Which is Behind the Recent Surge in Patenting?" Carnegie-Rochester Conference Series on Public Policy, 48, pp. 247-304.

Lambert, D., 1992, "Zero-inflated Poisson Regression, with an Application to Defects in Manufacturing," *Technometrics* 34: 1-14.

Lim, K., 2001, "The Relationship Between Research and Innovation in the Semiconductor and Pharmaceutical Industries," working paper.

Mansfield, E., 1995, "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing," *The Review of Economics and Statistics* 77: 55-65.

Marquis, D. G. and T. Allen, 1966, "Communications Patterns in Applied Technology," *American Psychologist*, 21, pp. 1052-1060.

Mowery, D., 1981, The Emergence and Growth of Industrial Research in American Manufacturing, 1899-1945, Ph.D. dissertation, Stanford University.

Mowery, D., R. Nelson, B. Sampat, and A. Ziedonis, 1998, "The Effects of the Bayh -Dole Act on U.S. University Research and Technology Transfer: An Analysis of Data from Columbia University, the University of California, and Stanford University," working paper

Narin, F., K. Hamilton, and D. Olivastro, 1997, "The Increasing Linkage Between U.S. Technology and Public Science," *Research Policy* 197: 101-121.

Narin, F., 1995, "Linking Biomedical Research to Outcomes – The Role of Bibliometrics and Patent Analysis," CHI Working Paper.

Office of Technology Transfer, University of California, 1997, Annual Report: University of California Technology Transfer Program. Oakland, CA: University of California.

Rosenbloom, R. and W. Spencer, 1996, *Engines of Innovation: U.S. Industrial Research at the End of an Era*, Boston: Harvard Business School Press.

Price, D., 1965, "Is Technology Historically Independent of Science? A Study in Statistical Historiography," *Technology and Culture*, 6, pp. 553-568.

Rosenberg, N. and R. Nelson, 1994, "American Universities and Technical Advance in Industry," *Research Policy*, 23, pp. 323-348.

Schmookler, J., 1966, *Invention and Economic Growth*, Cambridge, MA: Harvard University Press.

Shane, S., 2000, “Prior Knowledge and the Discovery of Entrepreneurial Opportunities,” *Organization Science*, 11, pp. 448-469.

Shane, S., 2001, “Technological Opportunities and New Firm Creation,” *Management Science*, 47, pp. 205-220.

Sherwin, C. W. and R. S. Isenson, 1967, “Project Hindsight,” *Science*, 156, pp. 1571-1577.

Sorenson, O. and L. Fleming, 2001, “Science and the Diffusion of Knowledge,” working paper.

Stephan, P., 1996, “The Economics of Science,” *Journal of Economic Literature* 34: 1199-1235.

Stern, S., 1999, “Do Scientists Pay to Be Scientists?” NBER Working Paper No. 7410.

Thursby, J. and M. Thursby, 2002, “Who is Selling the Ivory Tower? Sources of Growth in University Licensing,” *Management Science*, 48, pp. 90-104.

Zucker, L., M. Darby, and M. Brewer, 1998, “Intellectual Capital and the Birth of U.S. Biotechnology Enterprises,” *American Economic Review*, 88: 290-306.

Figure 1 Patent Citations to Academic Research

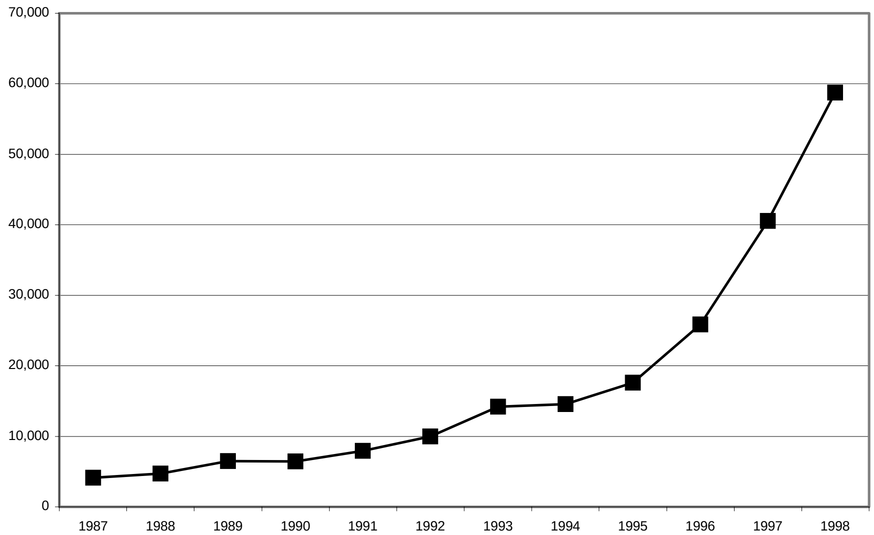

Figure 2 Distribution of Citations Across Science Categories

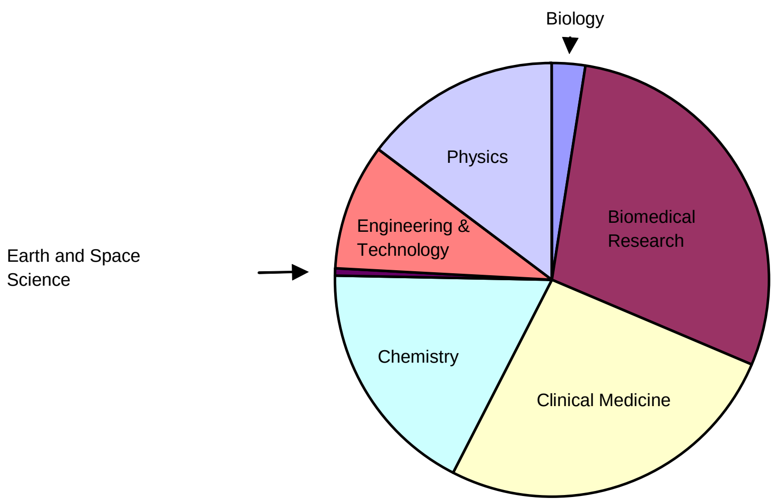

Figure 3 Increase in Citations over Time

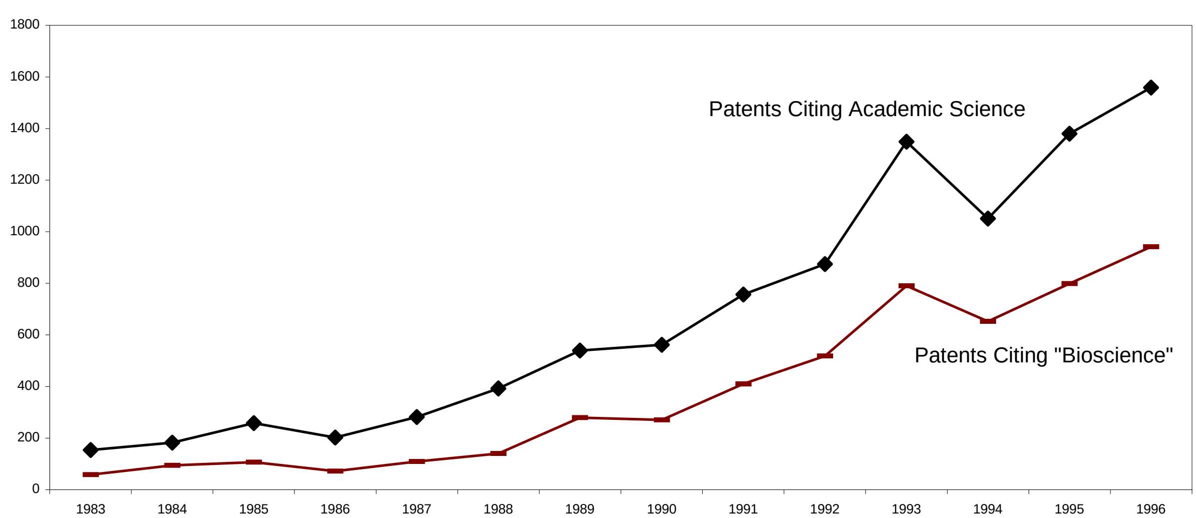

Figure 4 Total Citations and Citations by Drug/Medicine Patents

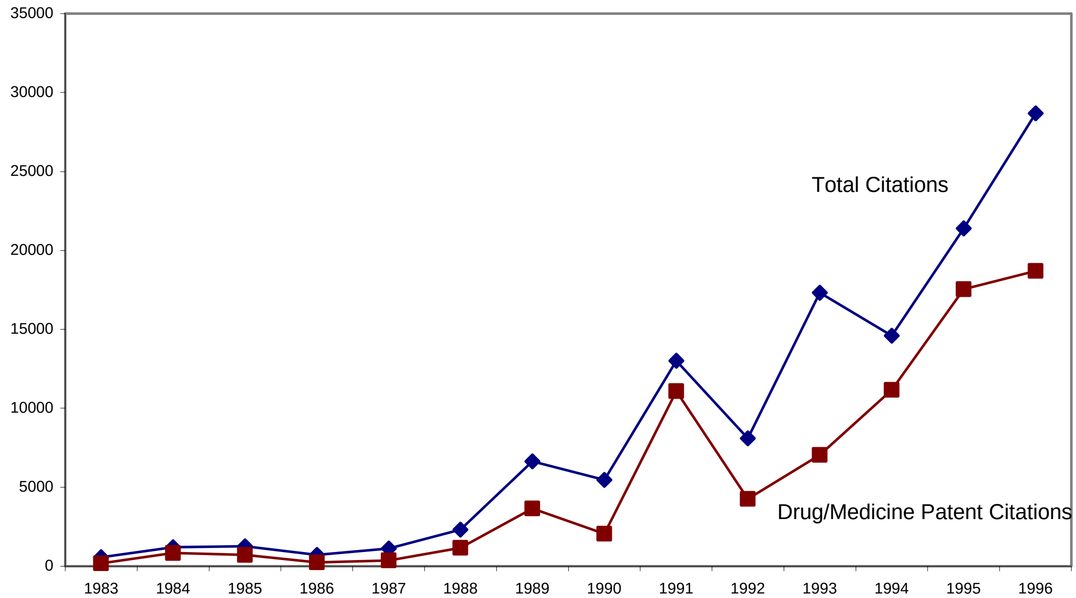

Figure 5 Academic Citations by Cited Institution Category

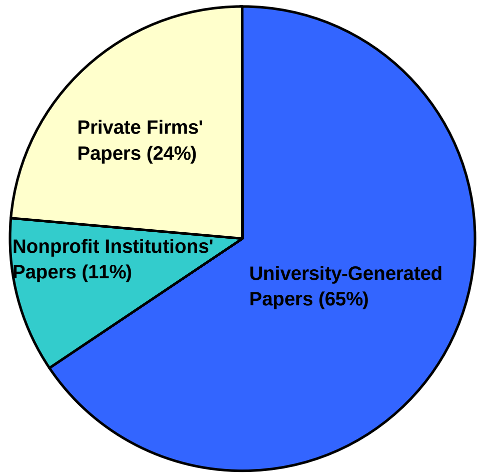

Figure 6 U.S. versus Foreign Institutions

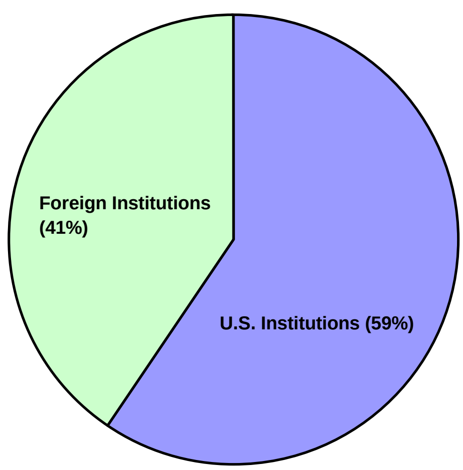

Figure 7 Lags between paper publication and patent application

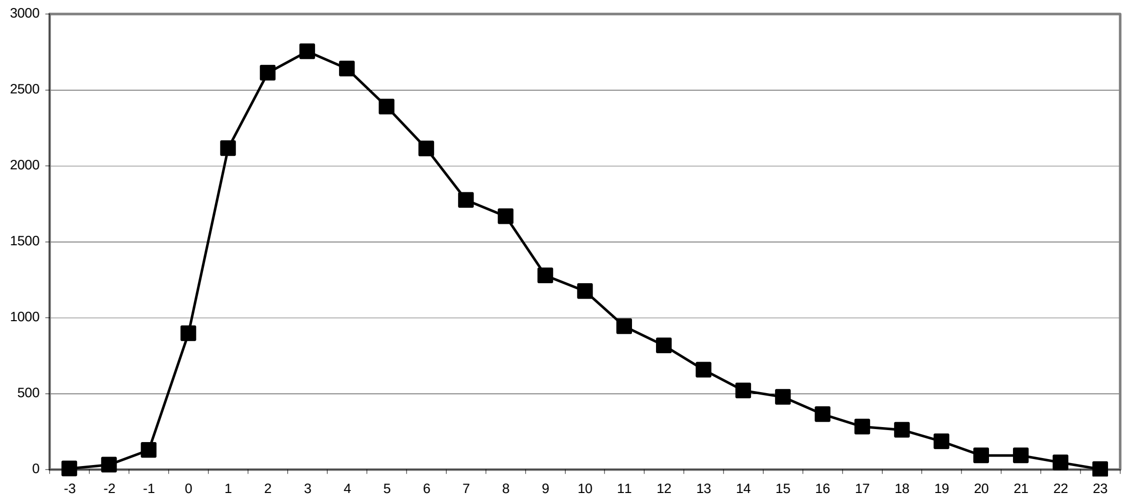

Figure 8 Geographic Location of Citing Patents

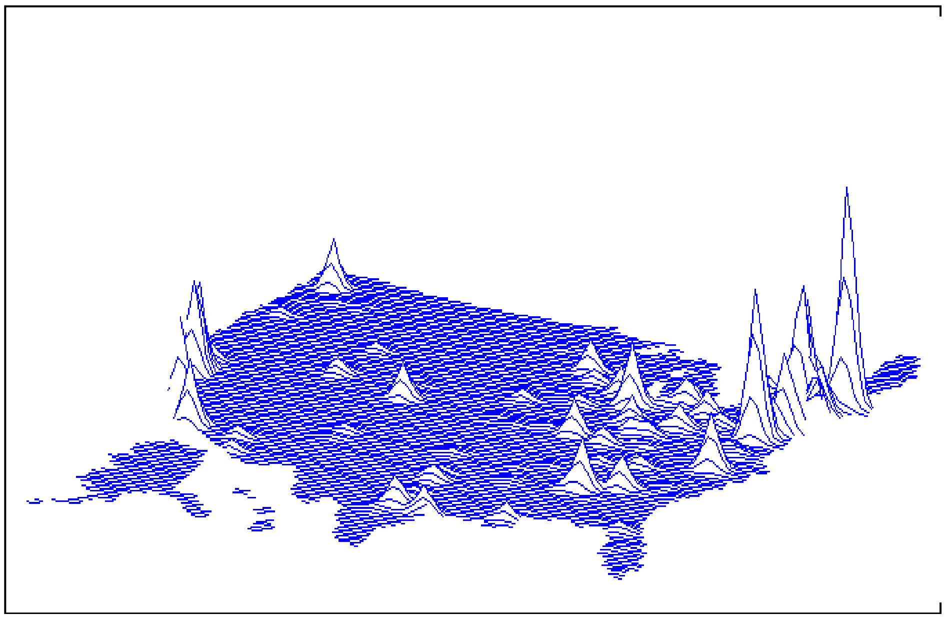

Table 1 Logit Regressions on the Determinants of Academic Citation

Variable Category	*Variable*	*All citations*	*ISI journals only*	*University-affiliated authors*
Type of Assignee	**Universities**	1.698 (.117)	1.73 (.115)	2.27 (.123)
	Nonprofit R&D organization	1.329 (.247)	1.23 (.251)	1.53 (.274)
	U.S. government agency	.584 (.126)	.818 (.141)	.640 (.204)
	Foreign assignee	-.158 (.043)	-.161 (.055)	-.371 (.084)
	Other	-1.04 (.070)	-1.07 (.094)	-.831 (.124)
Technology Class	**Chemicals**	1.63 (.069)	2.02 (.104)	2.14 (.172)
	Communications/ Computers	1.63 (.073)	1.58 (.113)	1.57 (.187)
	Drugs/Medical	2.36 (.075)	2.94 (.107)	3.40 (.171)
	Electronics	1.42 (.071)	1.64 (.107)	1.59 (.179)
	Mechanical devices	.028 (.085)	.001 (.137)	.081 (.228)
Science Center		.364 (.044)	.421 (.055)	.503 (.075)
Application Cohort Effects		Yes	Yes	Yes
Obs		29,876	29,876	29,843
Log-Likelihood		-10,882.552	-7,302.910	-3,928.787

Figure 9 Application Year Cohort Effects, Logit Regression

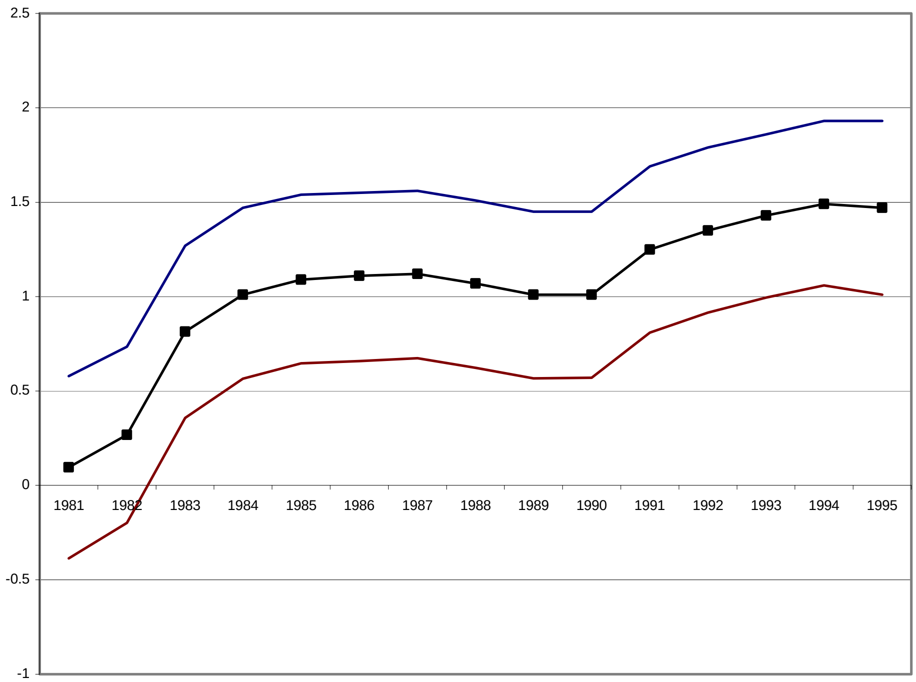

Figure 10 Application Year Cohort Effects, Negative Binomial Regressions

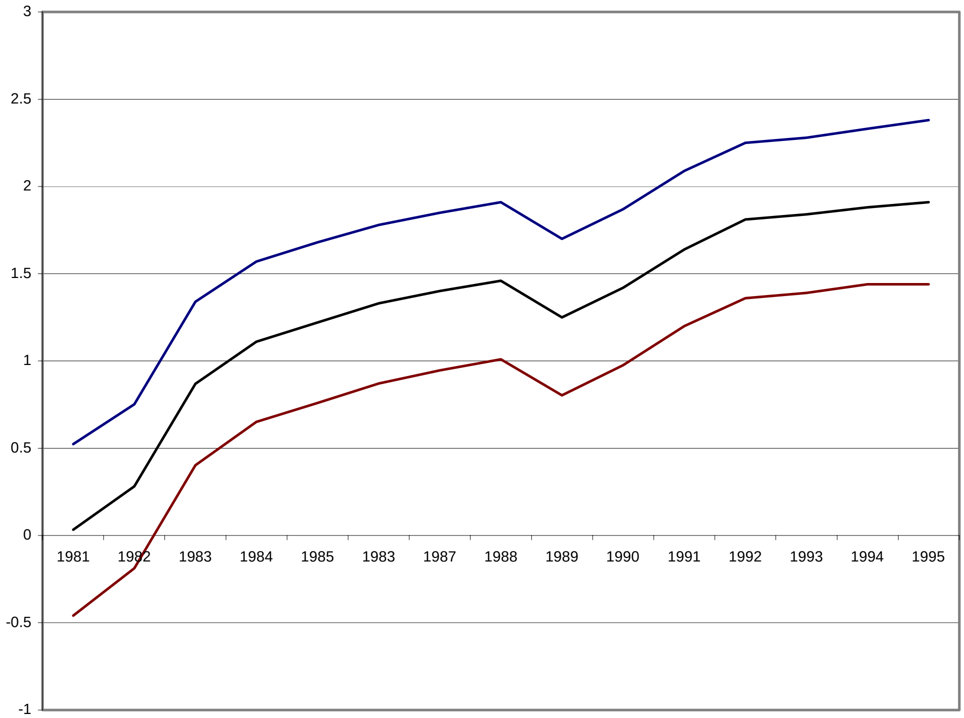

Table 2 Negative Binomial Regressions on the Determinants of Academic Citation

Variable Category	*Variable*	*All citations*	*ISI journals only*	*University-affiliated authors*
Type of Assignee	**Universities**	1.60 (.146)	1.90 (.173)	2.51 (.225)
	Nonprofit R&D organization	1.09 (.319)	1.42 (.377)	1.56 (.495)
	U.S. government agency	.623 (.165)	.758 (.206)	.618 (.290)
	Foreign assignee	-.527 (.048)	-.474 (.063)	-.579 (.095)
	Other	-1.089 (.065)	-1.06 (.088)	-.685 (.125)
Technology Class	**Chemicals**	1.94 (.066)	2.30 (.092)	2.45 (.140)
	Communications/ Computers	1.65 (.074)	1.42 (.105)	1.34 (.161)
	Drugs/Medical	2.90 (.078)	3.46 (.104)	3.97 (.151)
	Electronics	1.42 (.069)	1.61 (.097)	1.41 (.150)
	Mechanical devices	.020 (.075)	-.041 (.113)	-.403 (.194)
Science Center		.436 (.050)	.492 (.064)	.649 (.091)
Application Cohort Effects		Yes	Yes	Yes
Obs		29,876	29,876	29,876
Log-Likelihood		-20,695.575	-12,510.80	-6,641.674

Table 3 Truncated Negative Binomial Regressions on Citing Patents

Variable Category	*Variable*	*All citations*	*ISI journals only*	*University-affiliated authors*
Type of Assignee	**Universities**	.653 (.095)	.837 (.121)	1.11 (.173)
	Nonprofit R&D organization	.592 (.234)	.919 (.300)	.774 (.442)
	U.S. government agency	-.066 (.161)	-.122 (.209)	.353 (.393)
	Foreign assignee	.584 (.366)	.960 (.478)	1.41 (.752)
	Other	.176 (.137)	.150 (.177)	.185 (.260)
Technology Class	**Chemicals**	.471 (.167)	.782 (.217)	1.07 (.358)
	Communications/ Computers	.024 (.180)	-.155 (.236)	.047 (.390)
	Drugs/Medical	.711 (.167)	1.21 (.217)	1.71 (.354)
	Electronics	.018 (.176)	.274 (.228)	.024 (.376)
	Mechanical devices	-.123 (.222)	-.142 (.292)	-.871 (.507)
Ave_Dist		.249 (.204)	.412 (.274)	.492 (.474)
Citations to Patents		.016 (.003)	.009 (.003)	.001 (.005)
Application Cohort Effects		Yes	Yes	Yes
Log-Likelihood		-3,240.90	-2,573.40	-1,645.36

Table 4 Results on Quality Differentials

Variable	***Coefficient***	***Std. Err.***	***Implied Difference***
Claims	2.25	.220	15.1%
Forward Citations	.581	.101	5.5%
Generality	.029	.008	7.5%

Table 1 reports the regression coefficient on a dummy variable identifying patents that cite scientific research. Regressions control for technological field and application year effects. The measures of patent quality in the table are used as the dependent variable in the regression, as in Henderson, Jaffe, and Trajtenberg, 1998.

Table 5 Results of an Academic Production Function

Variables	Total publications	Publications weighted by academic citations	Publications weighted by patent citations
Log(total R&D stock)	.174 (.022)	.136 (.023)	.049 (.024)
Log(post-docs)	.148 (.021)	.158 (.026)	.061 (.033)
Log(grad enrollment)	.120 (.034)	.126 (.042)	.217 (.057)
Log (real avg salary)	.433 (.144)	-.991 (.182)	.432 (.827)
Constant	0.76 (.589)	8.41 (.757)	-3.77 (3.12)
Field Effects	Yes	Yes	Yes
Institution Effects	Yes	Yes	Yes

Results of column three include year fixed effects.

Table 6 Campus and Field Dummy Variables

Variable	Coefficient	Standard Error
Stanford	.595	.313
Berkeley	.506	.222
Davis	.577	.144
Irvine	.169	.137
UCLA	.257	.181
UCSD	.368	.174
UCSF	.396	.261
UCSB	.501	.136
Riverside	.294	.134
Biology	.303	.152
Biomedical Research	3.95	.168
Chemistry	1.40	.138
Geoscience	.112	.108
Engineering/Technology	1.18	.145
Physics	1.35	.114
Psychology	.033	.108

The reference campus is UC-Santa Cruz. The reference field is “mathematics/statistics.” The coefficients in this table are obtained from the regression of Table 3, column 3.

Table 7 ANOVA of the Variance in Outcomes

Root MSE=.8499 **R-squared=.7165**

Source	**Partial SS**	**df**	**MS**	**F**	**Prob>F**
Model	1241.17	16	77.57	107.39	0.00
Field	1019.60	7	145.66	201.65	0.00
Institution	86.05	9	9.56	13.24	0.00
Residual	491.19	680	.722		
Total	1732.36	696	2.489		