Measuring the Link Between Academic Science and Industrial Innovation
The Case of California's Research Universities[?]

Lee Branstetter
Department of Economics
University of California
Davis, CA 95616
and NBER

June 30, 2000

Preliminary and Incomplete

Do Not Quote or Cite

I.   Introduction

The impact of academic science on industrial innovation has received a great deal of attention. A full review of even the recent literature is beyond the scope of this draft, and I will only mention a few studies from the streams of research on which this current paper directly builds. One such stream of research has used case studies or surveys in an attempt to assess both the magnitude of this impact and the channels through which it flows. Mansfield (1990) directly interviewed industrial research directors to obtain their assessments of the impact of academic research on industrial R&D, finding that that this impact is substantial across a wide-range of industries. 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. Other qualitative studies of this phenomenon include Faulkner and Senker (1995).

A second stream of research has attempted to quantitatively assess the real effects of academic research. Jaffe (1989) examined the impact of university R&D spending on the patenting of "technologically proximate" industrial firms. Adams (1990) studied the impact of basic research by relating lagged measures of scientific output (as measured by counts of papers) to movements in productivity measures.   Jaffe et. al. (1993, 1996, 1998) have studied "knowledge spillovers" from academic science to industrial R&D by examining, among other things, trends in patenting by universities and the citations made to these university patents by other entities, including R&D-performing industrial firms.

A third stream of research has undertaken quantitative analysis of university-industry research collaboration. Zucker et. al. (1998) and Cockburn and Henderson

(1998, 2000) have used measures of direct collaboration (i.e., co-authored papers) between academic scientists and industrial R&D labs, finding that measures of firm research performance are correlated with measures of "connectedness" to academic science. A number of papers (Zucker et. al. 1998, Audretsch and Stephan, 1996) have studied "start-up" activity linked to academic research or academic researchers. Finally, several recent studies have examined university licensing of university-generated inventions (Barnes, Mowery, and Ziedonis, 1998, Mowery et. al., 1998, Thursby and Thursby, 2000).

This paper uses patent citations to academic papers to measure "knowledge spillovers" between academic science and industrial R&D. It is not the first paper to use such data – Francis Narin and his collaborators have pioneered the use of these data (see Narin et. al., 1997) in large-scale statistical analysis and, in fact, the patent citation data used in this paper was originally generated by Narin's firm, CHI Research. However, this paper takes a very different approach to the data than has Narin's work. On the one hand, I focus solely on patent citations to academic papers authored by scientists affiliated with the campuses and laboratories of the University of California (UC) system and Stanford University.[1] In contrast, Narin's work has looked at a much broader sample of citations to multiple universities, public R&D labs, and academic publications generated by private firms.

The limited scope of my analysis allows me to subject individual citations and the patents in which they appear to a high level of scrutiny. Using data on the residence of

---

[1] Data on citations to the University of Southern California have been acquired but not yet analyzed. In the future, I hope to acquire and utilize data on the California Institute of Technology as well. The inclusion of data from these institutions would cover almost all patent citations made to academic papers over my sample time period. I acknowledge that the scope of the present study does not quite live up to what is implied by its title!

the inventors named in the patent, I am able to examine issues of geographic localization

of knowledge spillovers. By matching citing patents to a control group of non-citing

patents, I am able to study aspects of individual patents which are correlated with

citations and the distribution of citations over time and "technology space" as well as

geographic space.   Finally, I am able to identify "highly cited" academic scientists and

"intensively citing" industrial firms whom are then interviewed to facilitate a richer

understanding of the kinds of interactions between academia and industry that generate

the observed citations. This complementary fieldwork is directly inspired by Jaffe,

Fogarty, and Banks (1999).[2]

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

Since the contribution of this paper lies, to a great extent, in the data being used, it

is worthwhile to point out at the outset both the advantages and disadvantages of the data.

The primary advantage is rather dramatically illustrated in Figure 1. This graph

illustrates the trends over the 1988-1997 period in several alternative indices of university

research output and knowledge spillovers for the University of California's 9 campuses

and affiliated laboratories, including university patents by issue year (patents), invention

disclosures by year of disclosure filing (disclosures), new licenses of university

technology by date of contract (licenses), the number of citations to previous university

patents by issue year of the citing patent (citations to UC patents), and the number of

citations to UC-generated academic papers by issue year of the citing patent (citations to

UC papers). The latter index towers over everything else, and it is growing (almost

exponentially) over time whereas the other indices are comparatively stagnant. The clear

implication is that there are far more data points to work with using this measure than any

---

[2] Future drafts will include a section which discusses the results of these field interviews.

of the examined alternatives. Figure 2 gives a similar graph for Stanford, driving home the point.

Despite the passage of the Bayh-Dole Act and despite the best efforts of eager university technology transfer officers to encourage this sort of activity, California's research universities still produce a relatively small number of patents.[3] While the numbers have increased over time, the results of Henderson et. al. (1998) on a national sample of university patents suggest that more marginal inventions are being patented, such that the average quality of university patents is actually *declining*. Likewise, few patented inventions are licensed, and only a handful of these ever generate substantial revenues. Thursby and Thursby (2000) suggest that, nationwide, university licensing efforts may also be running into diminishing returns.

However, a focus on patents or licensing may miss important channels by which academic institutions provide useful technological information to industrial innovators. Not all academic disciplines produce intellectual outputs that lend themselves to patent protection. It is also probably true that few academics possess the traits and skills that would enable them to bring an innovation from the initial "concept" stage all the way to successful licensing contract negotiations with a potential manufacturer. In this context, it is perhaps not surprising that attempts to expand university patenting and licensing from the initial low levels already bear evidence of diminishing returns.

---

[3] The number of patents is small relative to what a "patent-intensive" firm might be expected to produce with the same level of R&D spending. It is *not* small relative to the levels of patenting in other research university systems. It is worth pointing out that UC had instituted procedures for the disclosure and licensing of university inventions which preceded the Bayh-Dole Act by several decades, so the impact of the Act is, perhaps, less evident here. See Mowery et. al. (1998) for more detail and an institutional history of technology licensing at UC and Stanford.

On the other hand, the academic promotion system creates strong incentives for academic scientists to publish all results of scientific merit. Thus, UC and Stanford generate thousands of papers annually. If we wish to measure the impact of university research, then perhaps we should take as our starting point the broadest measure of university research output – academic publication. To the extent that citations to these publications reflect knowledge spillovers, Figures 1 and 2 would seem to imply that the spillovers are growing in importance in a way that is consistent with the widely held conviction that industrial research is increasingly building upon academic science.

I will argue throughout this paper that the available evidence suggests that these citations *do* reflect knowledge spillovers. This view is strongly supported by my initial field interviews of highly cited scholars and intensively citing firms. However, the reader may find large sample survey evidence more persuasive. Such evidence in presented in Cohen et. al. (1994), and Table 1 summarizes a larger table in their paper which conveys the same message. When asked by what means they obtain useful research results from academia which can serve as inputs to their own R&D process, industrial R&D directors across a wide range of industries *consistently listed the academic literature as one of the most important channels*. Averaging across industries, academic publications are the *single most important* source of such information. Note that publications play a particularly important role in the pharmaceutical industry. Faulkner and Senker (1995) present similar evidence from in-depth interviews of executives in the biotechnology industry, summarized in Table 2. Research by Cockburn and Henderson (1998) and Zucker et. al. (1998, 1999) has stressed the role of direct university-industry research collaboration in promoting information flows in the pharmaceutical and biotechnology

industries. While nothing in this paper calls into question the view that direct collaboration is important, the cited survey evidence and my own field work suggests that direct collaboration is not the only significant channel of such information flows in these sectors.

That being said, these citations do possess some disadvantages, some of which they share with the use of patent citations to other patents. As Jaffe et. al. (1998) have stressed, patent citations can appear for reasons that have little or nothing to do with knowledge spillovers. This general truth also applies to citations to the scientific literature, though with perhaps less force. The legal obligation to cite relevant "prior art" generates citations where there were no "spillovers," and citations can be added *ex-post* by parties other than the inventor. Tackling this problem head-on, Jaffe et. al. (1998) presents the results of a field study suggesting that, despite the presence of substantial "noise" in patent citation data, there is enough "signal" in them to support inference of knowledge spillovers, particularly when such inference is based on large numbers of patents from multiple organizations. My own fieldwork suggests that a similar result obtains with citations to the scientific literature, but caution in making that inference is clearly warranted.

In my view, the other primary disadvantage (again, a disadvantage shared with citations to previous patents) is that the appearance of citations does not directly lead to an accounting, in dollar value terms, of the economic benefits created by university research. University licensing revenues may be relatively small numbers, but at least they are dollar values. No research results presented in this paper will get us anywhere

close to such a dollar value accounting, but I do sketch out in section V how we might use citations as an intermediate step toward such an accounting.

The final disadvantage is somewhat unique to this study: I am, after all, only looking at citations to the publications of a single public university system and a single private university. I would like to make the argument that these are uniquely important, productive institutions, but I do not make any claim that the results presented in this paper are generalizable to all leading research universities, much less noncorporate scientific institutions in general.

The rest of the paper will proceed as follows. First, I describe the citations data along several dimensions in section III. I also briefly discuss the fieldwork component of this study. Next, I conduct some preliminary econometric analysis. This requires me to merge my citations data with other data on the geographic location and technology class of the citing patent, and combine this merged data with a matched set of "nonciting" control patents. I describe the data creation process, the statistical models used to analyze the resulting data, and my preliminary results in section IV. In section V, I go on to describe other ways in which these citations data might be used. Section VI presents some (very) preliminary conclusions.

III.   Describing the Data

Again, the reader is referred to Figures 1 and 2, which give the time series trends for several alternative indices of "knowledge flows" from UC and Stanford. Figures 3 and 4 aggregate citations over time, but break them out across seven major fields of science for UC and Stanford respectively. Biotechnology-related fields of science (biomedical research, clinical medicine) constitute a very large portion of the total. They

8

also account for much of the recent increase in citations to academic research, as is illustrated by Figure 5 for UC. The figure for Stanford is similar, unless one breaks off Stanford Medical School, in which case "engineering" and "physics" related technologies are the main drivers of changes over time, as is illustrated by Figure 6.

Figure 7 breaks down citations according to the campuses and UC-affiliated laboratories with which the cited scholar was affiliated at the time of paper publication. Clearly medical schools are important drivers of overall citations. However, "biotech-related" fields are important outside of medical schools in the UC system. Even at UC-Berkeley, which possesses one of the nation's strongest engineering faculties, "engineering" and "physics" related citations are less numerous than "biotech-related" citations. That being said, one sees increases over time in citations to fields such as "engineering" and "physics," particularly at Berkeley and Stanford. Many of these patents and the cited papers connected to developments in electrical engineering and computer science.

The relative dominance of biotech citations mirrors trends in the aggregate citations data noted by Narin et. al. (1997), so it is not an artifact of my sample. Interestingly, it also mirrors trends in the distribution of university patents and licenses across fields. Though the numbers involved are much smaller, the studies of UC patenting and licensing conducted by Mowery et. al. and Mowery and Ziedonis also demonstrate the dominant role of biotech-related invention. Biotech also plays a strong role in the patenting and licensing of Stanford and Columbia.

Mowery et. al. suggest several reasons for this. First, they point to the large share of federal (and state) research funding focused on the life sciences, a trend which

stretches back to the 1970s. In part, the output measures represent the return to decades of sustained public R&D investment in biomedical research at leading universities. These authors also suggest that industrial research in biotechnology and pharmaceuticals is now "closer" to academic research. In other words, the process of product invention and development now builds much more closely and directly on academic bioscience than it used to, a development discussed at length in the work of Zucker et. al., Cockburn and Henderson, and Gambardella (1998). Evidently, these same effects are reflected in my citations data.

It is also of interest to look at the (unconditional) distribution of citations in space and time. Space and time constraints prevent more than a cursory glance at this, but the reader is referred to Figure 8, which provides a 3-dimensional representation of the incidence of citation in geographic space for the entire United States. The height of the cones represents the number of patents citing UC-Berkeley generated papers within a particular U.S. county, as identified by the address of the first inventor listed on the patent application. Note the concentration of citations in California and in the Northeastern research/industrial corridor – a "bicoastal" pattern that Mowery and Ziedonis also find in citations to UC-generated patents. Figure 9 presents a similar distribution for California counties only. Of course, the underlying geographic distribution of research activity is also quite skewed, and any formal investigation of the geographic localization of knowledge spillovers would have to control for this.

Following Jaffe et. al. (1993), we conducted a formal test of the geographic localization of knowledge spillovers by matching each of our citing patents with a nonciting "control" patent issued on roughly the same date in the same patent class as the

cititing patent. Let p_c be the probability that a citation comes from the same county as that in which the cited university campus is located. Let p_0 be the corresponding probability for a randomly drawn control patent. We test for localization using the following test statistic:

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

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.

The (unconditional) distribution of time lags between the publication date of the paper and the application (filing) date of the patent is given in Figure 10.[4] The shape of this distribution resembles the double-exponential curves estimated by Jaffe and Trajtenberg. The modal citation lag is quite short, but the distribution is quite heavily skewed to the right, with nontrivial numbers of citations being made to fairly old papers. In a very small number of cases, the lag is negative, suggesting that academic references were added to a patent application *after* the initial filing.

Fieldwork

Since the "raw" citations data make references to specific scholars, it is relatively easy to identify on each UC campus "highly cited" scholars – that is, scholars whose work is frequently cited in patents, including patents assigned to firms. I have interviewed a number of these scholars about their cited research, showing them in the interview a comprehensive list of the patent citations to their work and asking them about the possible technological linkages between these patented inventions and their papers as well as any relationship or connection they might have to the citing organization.

---

[4] This is for citations to UC papers only.

While some citations arise in the context of a formal relationship between the cited scholar and the citing organization, most do not. Interviewed scholars are often surprised to find their work cited in the patents of particular industrial firms. However, upon a close reading of the patent abstract, they are often able to identify a plausible technological linkage between the patented invention and the cited research. There is considerable variance among highly cited scholars in terms of the extent to which they have attempted to profit financially from their own research. Some act as consultants to firms, some deliberately seek corporate funding for their labs, and a number of scholars have successfully issued one or more patents protecting an invention. However, many highly cited scholars have done none of those things. The common denominator among all cited scholars and cited papers is scientific quality. The scholars are frequently the leading intellectual lights in their departments, and the cited papers often represent either their most important scientific contributions or a methodological advance with widespread application.

I have also contacted industrial R&D managers of "frequently citing" firms, in order to examine the "knowledge flow" process from the perspective of the firm. The managers I have interviewed generally accept the view that patent citations reflect knowledge spillovers, although one corporate patent attorney emphasized that some references to the scientific literature are added *ex-post* or for defensive reasons. R&D managers emphasized the role of the scientific literature as a vehicle for knowledge flow, but they also stressed the importance of cultivating long-term relationships with key academic experts – which was often reflected in the citations patterns. Interestingly, co-authorship did not receive much emphasis in my discussions with R&D managers. The

universal view among interviewees on the corporate side was that the rise in the incidence of citation of academic research represents a real "convergence" of research in industry and academia rather than merely a change in citation practices or the computerization of scientific literature indices.

Econometric Analysis

Data Construction and Basic Approach

The parallels between this research project and the analysis of patent citations pursued by Jaffe, Trajtenberg, and Henderson suggests the use of similar methodology. While I would like to employ a citations function approach, data availability precludes it. I do not have good information on the number of potentially citable papers being generated by UC and Stanford by scientific field.[5] I only observe papers that are cited at least once. In principle, measures of academic publications could be created from publicly available sources, but obtaining the data and matching it to the correct campus is likely to be a long, expensive undertaking.

Instead, I take the following approach. After matching my initial data on citations to additional data on the citing patents, I then construct a random sample of nonciting patents drawn from the same set of issue years, 1988-1997.[6] The presence of this "control group" of nonciting patents enables me to conduct statistical analysis of the likelihood of a given patent making citations to UC (or Stanford) academic research as a function of the characteristics of the citing patent (including, by extension, the

---

[5] I thank Jim Adams for providing me with his data on paper counts, which includes information on some of the campuses I study. Unfortunately, there is only a limited overlap in the time series dimension of his data and the years of my sample of patent citations.

[6] This component of my research relied heavily on data from the REI data base at Case Western Reserve University. I am grateful to Adam Jaffe and Michael Fogarty for access to these data.

characteristics of its named first inventor) and the cited paper (including, by extension, the campus affiliation of the authors).

Taking this step involves collapsing my sample of nearly 40,000 citations down to the much smaller number of unique citing patents, many of which make more than one reference to academic publications from UC and/or Stanford. At the moment, econometric work has focused on UC-citing patents. Results for Stanford will be integrated at a later date. The integer nature of the number of citations to academic publications made per patent, which will be the dependent variable, calls for the use of count data models. Regression analysis based on the standard Poisson and Negative Binomial models has become increasingly familiar, so no derivation will be given for these models, and they will be estimated as a "benchmark."[7] However, key features of the data will require me to modify the likelihood functions of the benchmark models.

Conceptually, one can think of the likelihood of observing a citation to academic research as being a function of an unobserved latent variable – "proximity to academic science."[8] Conditioning on a patent being sufficiently "close" to academic research for a citation to take place, one may observe the patent making anywhere from 1 to as many as 38 citations to academic papers. The actual number of citations made will be a function of attributes of the citing patent (such as geographic and temporal distance from the relevant research) and a function of the cited research and the campus where it was conducted. This seems to call for a specification analogous to the "Tobit" model, but one set up to handle the "integer" or "count" nature of the dependent variable when citations are actually observed.

---

[7] The classic reference is Hausman, Hall, and Griliches (1984).
[8] I am grateful to Adam Jaffe for discussions on these issues.

A statistical model which comes close to meeting these requirements is the so-called hurdle Poisson model and its generalization, the hurdle negative binomial model. An alternative formulation with some similarities to the "hurdle Poisson" is the "zero-inflated Poisson" model, which we utilize in the current draft. Alternatively, I can conduct econometric analysis *using only those patents which cite a UC academic paper at least once*. This implies a Poisson (or Negative Binomial) distribution which is truncated from below (i.e., at zero). The basic features of these models are described in the next section.

Sketch Derivation of the Estimation Techniques

A complete derivation of these models is given in Cameron and Trivedi (1998). Here we present only the essential features of these derivations, beginning with truncated Poisson and negative binomial distributions, and the implications of truncation for empirical analysis. This brief description draws heavily on Cameron and Trivedi (1998) and uses their notation. Let

$$H(y_i,?) ? \Pr[Y_i ? y_i] \tag{1}$$

denote the CDF of the discrete random variable with PDF $h(y_i,?)$, where $?$ is a parameter vector. In my applications, of course, y_i will be the number of citations to academic research made by patent i. If realizations of y less than a positive integer r (in our case, 1) are omitted, the ensuing distribution is given by

$$f(y_i,? \mid y_i ? r) ? \frac{h(y_i,?)}{1 ? H(r ? 1,?)} \tag{2}$$

One special case would be the left-truncated negative binomial, for which

$$h(y_i,?) ? \frac{?(y_i ? ?^{?1})}{?(?^{?1})?(y_i ? 1)}(??_i)^{y_i}(1 ? ??_i)^{?(y_i ?? ?1)} \tag{3}$$

15

where $\xi \sim (\mu_i, \sigma)$. The truncated mean and variance are defined by

$$\mu_i^* = \mu_i + \sigma_i$$

$$\sigma_i^{*2} = \sigma_i^2 + \sigma\sigma_i^2 + \sigma_i(\sigma_i - r) \tag{4}$$

$$\sigma_i = \sigma_i[1 + \sigma(r+1)]\lambda(r+1, \mu_i, \sigma)$$

$$\lambda(r+1, \mu_i, \sigma) = h(r+1, \mu_i)/1 - H(r+1, \mu_i)$$

In a similar way, the truncated mean and variance of the Poisson distribution can be

derived as a limiting case of the above, where $\sigma \to 0$.

Skipping from the general to the specific, the mean and variance of the Poisson

distribution truncated at zero are

$$E[y_i \mid y_i > 0] = \frac{\lambda_i}{1 - e^{-\lambda_i}} \tag{5}$$

and

$$V[y_i \mid y_i > 0] = E[y_i \mid y_i > 0][1 - \Pr[y = 0]E[y_i \mid y_i > 0]$$

$$= \frac{\lambda_i}{1 - e^{-\lambda_i}}\left[1 - \frac{\lambda_i e^{-\lambda_i}}{1 - e^{-\lambda_i}}\right] \tag{6}$$

A more general negative binomial distribution truncated at zero would have the following

first two moments:

$$E[y_i \mid y_i > 0] = \frac{\lambda_i}{1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}} \tag{7}$$

and

$$V[y_i \mid y_i > 0] = \frac{\lambda_i}{1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}}\left[1 + (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}} - \frac{\lambda_i}{1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}}\right] \tag{8}$$

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

$$L(\beta) = \sum_{i=1}^{n} \left\{ y_i \ln(\lambda_i) - \lambda_i - \ln\left[1 - \exp(-\lambda_i)\sum_{j=0}^{r-1} \lambda_i^j / j!\right] - \ln(y_i!)\right\} \tag{9}$$

where the MLE of β is the solution of

$$\sum_{i=1}^{n} \left[y_i - \lambda_i - \lambda_i \delta_i^{r-1}\right]\frac{\partial \lambda_i}{\partial \beta} = 0 \tag{10}$$

where

$$\delta_i = \frac{\lambda_i h(r, \lambda_i)}{1 - H(r-1, \lambda_i)} \tag{11}$$

While the preceding models are appropriate for exploring patents which cite an academic paper at least once, it is possible that much could be learned from a model which could accommodate a sample of patents which never cite such papers (our "control" group) as well as patents which cite papers once or multiple times. Two such models exist in the received econometrics literature – the "hurdle" Poisson model and the "zero-inflated" Poisson model, both of which have a more general negative binomial version.

In essence, a hurdle model of either variety is a finite mixture generated by combining the zeroes generated by one density with the zeroes and positives generated by a second zero-truncated density. The moments are determined by the probability of crossing the zero "threshold" and by moments of the second density. To put this into mathematical notation

$$E[y \mid x] ? \Pr[y ? 0 \mid x] E_{y?0}[y \mid y ? 0, x] \tag{12}$$

For a concrete example, consider the negative binomial hurdle model, which I will estimate in *subsequent* drafts of the paper. Let $?_{1i} ? \exp(x_i^c ?_1)$ be the negative binomial mean parameter for the case of *zero* counts. Similarly, let $?_{2i} ? ?_2(x_i^c ?_2)$ for the positive set J={1, 2,…}. Further define the indicator function $1[y_i ? J] ? 1$ if $y_i ? J$ and $1[y_i ? J] ? 0$ if y_i=0. From the negative binomial with a quadratic variance function, the following probabilities can be obtained:

$$\Pr[y_i ? 0 \mid x_i] ? (1 ? ?_1 ?_{1i})^{?1/?_1} \tag{13}$$

$$1 ? \Pr[y_i ? 0 \mid x_i] ? \underset{y_i ? J}{?} h(y_i \mid x_i) ? 1 ? (1 ? ?_1 ?_{1i})^{?1/?_1} \tag{14}$$

$$\Pr[y_i \mid x_i, y_i ? 0] ? \frac{?(y_i ? ?_2^{?1}}{?(?_2^{?1}) ?(y_i ? 1)} \left\{ \frac{1}{(1 ? ?_2 ?_{2i})^{1/?_2} ? 1} \right\}^{??_2^{?1}} ? \left\{ \frac{?_{2i}}{?_{2i} ? ?_2^{?1}} \right\} \tag{15}$$

Equation (13) gives the probability of zero counts, equation (14) gives the probability of "crossing the threshold," and equation (15) is the truncated-at-zero distribution. The log-likelihood function splits into two components, such that:

$$L_1(?_1, ?_1) ? \sum_{i?1}^{n} \{(1 ? 1[y_i ? J]) \ln \Pr[y_i ? 0\} \mid x_i ?? \sum_{i?1}^{n} 1[y_i ? J] \ln(1 ? \Pr[y_i ? 0 \mid x_i]) \tag{16}$$

and

$$L_2(?_2,?_2) \ ? \ \sum_{i?1}^{n} 1[y_i \ ? \ J]\ln\Pr[y_i \mid x_i, y_i \ ? \ 0] \tag{17}$$

so that

$$L(?_1,?_2,?_1,?_2) \ ? \ L_1(?_1,?_1) \ ? \ L_2(?_2,?_2) \tag{18}$$

Note that this model contains a critical assumption – the component of the likelihood function which determines whether or not a nonzero realization of the dependent variable occurs is separable from and estimated independently of the component of the likelihood function which determines the count of the dependent variable, conditioning on that count being greater than zero. This conveys a practical advantage, as it makes estimation easier. However, in the context of my study, that assumption is potentially problematic. Intuitively, "proximity to academic science" could influence not only the likelihood of a citation occurring but also the number of citations actually made.

Lambert (1992), among others, has introduced an alternative to the hurdle approach. Consider the following:

$$\Pr[y_i \ ? \ 0] \ ? \ ?_i \ ? \ (1 \ ? \ ?_i)e^{??_i} \tag{19}$$

$$\Pr[y_i \ ? \ 0] \ ? \ (1 \ ? \ ?_i)\frac{e^{??_i}?_i^{r}}{r!}, r \ ? \ 1,2,... \tag{20}$$

where $?_i$ is the proportion of zeros. Lambert defines $?_i \ ? \ ?(x_i, ?)$ and proposed parameterizing $?_i$ as a logistic function of an observable vector of covariates z, thus ensuring nonnegativity of $?_i$; that is

y_i=0 with probability $?_i$

$$y_i \ \sim \ P[?_i] \text{ with probability } (1\text{-}?_i) \tag{21}$$

$$\pmb{?}_i \ ? \ \frac{\exp(z_i^{\varsigma}\pmb{?})}{1 \ ? \ \exp(z_i\pmb{?})}$$

I will follow Lambert in using the logistic functional form for $\pmb{?}_i$. Let $1(y_i=0)$ denote an indicator function that takes value of 1 if $y_i=0$ and zero otherwise. The joint likelihood function after omitting constants is given by

$$L(\pmb{?},\pmb{?}) \ ? \ \sum_{i?1}^{n} 1(y_i \ ? \ 0)\ln(\exp(z_i\pmb{?}) \ ? \ \exp(? \exp(x_i\pmb{?}))) \ ? \ \sum_{i?1}^{n} (1 \ ? \ 1(y_i \ ? \ 0))(y_i x_i\pmb{?} \ ? \ \exp(x_i\pmb{?}))$$
$$? \ \sum_{i?1}^{n} \ln(1 \ ? \ \exp(z_i\pmb{?})) \tag{22}$$

Here, too, I am assuming functional independence of the $\pmb{?}_i$ and $\pmb{?}_i$ components of the joint likelihood function. To the extent that this assumption is questionable, the results will need to be interpreted with appropriate caution.

Going further, it is clear that behind the data generating processes producing the patents and citations in my sample are inventors actually choosing where, in the technology space, to conduct research. Some inventors may deliberately choose to work in regions of the technology space where a rich foundation of prior academic research makes commercial R&D more productive. This suggests problems of endogeneity that neither the "hurdle" models nor the ZIP model would be able to handle. For these reasons, the econometric results contained herein are not presented as the last word, nor are the estimated coefficients given strong structural interpretations. At this stage, I am using these regressions to describe multivariate correlations in the data – no more than that.

Specifications Used

Recall that, in the preceding Poisson-based equations, $?_i$? $\exp(x_i^c?)$ defines the "exogenous" variables used and the regression parameters estimated; estimation of negative binomial models involves the estimation of the additional parameter, $?$. My "baseline" specification will be:

$$u_i \ ? \ \exp(?_1 \ ? \ ?_2 dist_i \ ? \ ?_3 dstate_i \ ? \ ?_4 dcnty_i \ ? \ ?\ ?_o O_i \ ? \ ? \ ?_c C_i \ ? \ ? \ ?_f F_i \ ? \ ? \ ?_L L_i) \quad (23)$$

where *dist* is a measure of linear distance between a cited campus (e.g., UC-Berkeley) and "location of invention" of the citing patent, which is presumed to be the county containing the address of the first inventor listed on the patent document. When a given patent document cites more than one California campus, the distance measure is an *average* measure of distance to each cited campus, weighted by the number of citations made to each campus. In specifications in which matched control patents are used, distance is measured between the location of the control patent and the campus cited by the citing patent to which our control patent is matched. Since the impact of geographic distance on knowledge spillovers is unlikely to be linear, I also include two dummy variables denoting significant geographic boundaries. *Dstate* is a dummy variable equal to 1 if the citing patent (or matched control) and the cited campus are located in the same state. *Dcnty* is a similarly constructed dummy variable equal to 1 if the citing patent and cited campus are located in the same county. When a given patent cites more than one campus, these dummy variables are set equal to one if any of the cited campuses is located in the same state or county. The coefficients on these three variables should provide some sense of the extent of geographic localization of knowledge spillovers, controlling for other attributes of the citing patents and cited campuses.

The *O*'s are a set of dummy variables corresponding to the organizational form of the assignee of the citing patent, based on an assignee classification system developed by Meg Fernando, former administrator of the REI patent data base. These include government, universities, non-profit non-university research labs, and private firms, and are incorporated into our specification on the grounds that different classes of assignees may have a differential propensity to cite academic papers. It would be particularly useful to control for the higher propensity of university patents to cite academic papers of the faculty inventor.

The *C*'s are dummy variables for the different cited campuses, with medical schools distinguished from the main campuses. A patent citing more than one UC campus will have more than one of these dummy variables equal to 1. Likewise, the F's are dummy variables for the scientific field of the cited paper, where I borrow a categorization of scientific disciplines developed by CHI Research, Inc.: biology, chemistry, biomedical research, clinical medicine, earth and space sciences, engineering and technology, physics, mathematics, and psychology. This allows me to control for field effects in estimating the differential "citedness" of different campuses, and also allows me to control for campus effects in estimating the differential "citedness" of different groups of academic disciplines. When more than one paper is cited, all relevant dummy variables are set equal to 1.

Finally, I want to get some sense of how temporal distance between the cited paper and the citing patent affect the probability of citation. In practice, I do this by defining a set of lag dummy variables (the *L*'s), each set equal to 1 when the citing patent cites a paper whose publication year preceded the patent application year by that lag

amount. Where more than one paper is cited, more than one lag dummy variable will be set equal to 1. Inference using these coefficients will need to keep in mind that the lag dummy variables are potentially picking up "cohort effects" of the citing patents and "cohort effects" of the citing papers as well as the impact of time lags, *per se*. In later drafts, I hope to use a more sophisticated specification to obtain a cleaner estimate of the effect of time lags.

The ZIP models require that I specify a set of variables determining the probability of "noncitation," that is, I need to define $\varphi_i \equiv \dfrac{\exp(z_i'\gamma)}{1 + \exp(z_i'\gamma)}$. A simple approach is to suppose that the propensity to cite academic science depends on where you are in the "patent space." Some industrial technologies are quite proximate to academic scientific research, others are less so. If I make the rather heroic assumption that the location of a given patent in the patent space is exogenous, or, at least think of it as predetermined, then I can use information based on the patent class assigned to the patent to define φ_i. As an initial step, I create a set of dummy variables (*dcat₁-dcat₃*) set equal to 1 if a patent is assigned to one of the classes which frequently cite biomedical research, clinical medicine, or electrical engineering, respectively. Thus

$$z_i = \gamma_0 + \gamma_1 dcat_{1i} + \gamma_2 dcat_{2i} + \gamma_3 dcat_{3i} \tag{24}$$

Description of Initial Results

Initial regression results are reported in Tables 3 and 4. As a benchmark, Table 3 presents results of estimates using Poisson and negative binomial models on the *truncated* sample (that is, the sample of citing patents). The first three columns of regression results present coefficients, standard errors, and z-statistics, respectively, from

a Poisson regression which includes only measures of geographic and temporal distance between the citing patent and the cited paper. The next three columns present the same information from a negative binomial regression which also includes campus, field, and organizational form effects. Due to space constraints, we only show results on the distance, time, and field coefficients.

Several aspects of these results merit comment. Distance seems to matter – a result that is very much line with previous work by Jaffe et. al. and by Mowery et. al. Being in the same state has a statistically significant impact on the probability of citation, as does being in the same county in the negative binomial results. On the other hand, our measures of linear distance do not have a significant impact on expected citation in either specification.

The coefficients on the lag dummy variables display a rather curious pattern. It seems that the "lag effects" peak at short lags between publication of the paper and application of the citing patent, then peak again at much longer lag lengths. One interpretation is that patented innovations benefits both from 1) recent publications which embody the latest results and 2) older publications which contain truly central, paradigm-shattering results. Another interpretation is that the estimates of the impact of lag length are confounded with cohort effects of the cited papers and citing patents. In future work, I hope to explore alternative specifications in order to unpack and separately measure these effects.

The field effects are strong and significant in the negative binomial estimates, and their pattern is what one might expect given the distribution of citations across fields. Biomedical research and clinical medicine have large, highly significant coefficients.

Engineering/technology and physics have smaller, but still significantly positive

coefficients. Once field effects are controlled for, the differential citedness of the

different UC campuses is much less pronounced. The "campus effect" coefficients are

not shown for reasons of space, nor are the organizational coefficients. It is clear from

the latter, though, that universities are more likely to cite academic papers in their patents

than other kinds of institutions.

In future drafts, I plan to present estimates from modified Poisson and NB models

which explicitly deal with the truncation in this subsample. At the time of this writing,

these results are not yet available.

In Table 4, I present results on the full sample of citing patents plus the matched

nonciting control patents. As a benchmark, I start with results from a zero-inflated

Poisson model, with the $?_i$ function defined as in equation (24). These results are

presented in the first three columns of Table 4, using the same format as in Table 3.

Interestingly, the inclusion of nonciting patents seems to increase the measured

impact of proximity on the level of citations. The *dstate* dummy variable's coefficient

doubles in this specification. Including nonciting controls also seems to shift the "peak"

of our estimated lag effects to longer lag lengths. The field effects are also much larger

in this expanded sample, though care needs to be taken in interpreting that coefficient, as

the "citing field" dummy variables of all control patents are set equal to zero.

The ZIP results include estimated coefficients on the variables used in the $?_i$

function. A Wald test of the restriction that these parameters are equal to zero is

resoundingly rejected, and a "straight" Poisson model is therefore rejected in favor of the

ZIP model. Similar results obtain with the ZINB model. Not surprisingly, the crude

measures of "location in patent space" turn out to be good predictors of the probability of citation, but given the way these data were generated we can hardly view that as confirmation of any explicit hypothesis.

In addition to the coefficients reported, the Poisson specification also includes as regressors the "organizational" dummy variables. Briefly summarized, the coefficients on these variables indicate that individual assignees and corporate assignees are systematically less likely to cite academic papers in their patents relative to other organizations. (That being said, corporate assignees do account for the majority of citing patents.)

The final three columns in this table present results from a zero-inflated negative binomial model. In this specification, I only use data on *corporate* citing and control patents, and I do not attempt to separately estimate code or field effects. This changes the sample, but also gives us a sense of the extent to which distance and lag effects vary for *corporate* citers of academic papers. By and large, the results are qualitatively similar to our other results. Note that, if anything, the impact of geographic proximity is even stronger here than elsewhere – here being in the same county has an additional positive impact of even greater magnitude than being in the same state. Note also that the lag effects have a pattern similar to those generated in previous specifications. The "lag effects" from several specifications are graphed out in Figure 11.

V.   Next Steps

I have demonstrated that industrial patent citations to academic papers are numerous and increasing, I have suggested that these citations are indicative of knowledge spillovers, and I have attempted to trace out the paths of these spillovers

26

across time, space, and technology class, controlling for various attributes of the citing patent and the cited academic research. What I have *not* done is give the reader any sense of the economic value, if any, created by these spillovers.

Taking that next step would require an examination of what the citing organizations (especially citing firms) *do* with the knowledge they extract from academic science. Cockburn and Henderson (1998) and Zucker et. al. (1998) have undertaken studies which examine how firm innovative performance is affected by measures of "connectedness" to academic science. Following the lead of these earlier papers, it should be possible to compare the innovative output of frequently citing firms to less frequently citing or non-citing firms. *Ceteris paribus*, does the incorporation of academic science into the industrial innovation process lead to better outcomes? Do firms which utilize UC or Stanford academic science generate more and better patents, obtain higher levels of revenue and profit, and generate more value for shareholders over sustained periods of time than firms which do not? A quantitative, econometric investigation of this question holds out the possibility of being able to both demonstrate and quantify the positive impact of public science. While it may never be possible to establish plausible "conversion factors" for turning paper or citation counts into dollar values, one might at least obtain some systematic sense of the difference university-industry research spillovers make through this sort of exercise.

Such an analysis would require a different approach to the data. One such approach would be to make the *assignee*, rather than the patent, the unit of analysis. In principle, I could obtain *all* the patents taken out by the assignees identified in our data base. For firms, assignee names could be linked to CUSIP or other firm identifier codes,

and patenting data could be linked with R&D input, sales, and stock market data. Not only could such data be used to explore the impact of citation on innovative output, but it could also be used to study patterns of citation by particular assignees over time. How do firms learn where the good academic science is? Do they zero on in on particular favored sources of academic science over time?

Armed with that knowledge, one can then return to the stream of research pursued by Adams and Griliches – the measurement of the output of science. A somewhat pessimistic observation of Adams and Griliches (1996) was that science itself, as measured by quality adjusted paper counts, seems to show some evidence of diminishing returns. However, if more recent cohorts of papers are increasingly cited by industrial inventions, and if that higher incidence of citation is leading to more or better products and services, then a broader measure of scientific output may actually yield evidence of constant or even increasing returns. The marriage of measures of academic resource input – federal, state, and local R&D dollars – with patent citation augmented measures of academic scientific output may better enable us to not only track but also optimize the transformation of research inputs into economic outputs.

VI.   Conclusions

At this early stage in the research process, of course, any "conclusions" must remain quite tentative. Certainly, I hope to produce a second draft in short order with contains estimates based on all the models sketched out in section IV. I also hope to incorporate results which use data on citations to Stanford University and, in the longer run, the California Institute of Technology and the University of Southern California.

That being said, I think several points can be made even at this early stage. 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 (much of generated by Francis Narin) has only begun this process.

Second, many of the patterns that others have discovered in citations to university patents in general (Jaffe et. al.) and citations to Stanford/University of California patents in particular are also reflected in citations to academic papers. That is, citations are concentrated in "biotech-related" fields, medical schools play a large role in generating cited research, and there is evidence of geographic localization of knowledge spillovers in our data.

Third, these data suggest that the temporal link between academic science and patented innovation is short. The modal lag in the raw data is only 2 years, and the pattern of "lag effects" in the econometric evidence also suggests that it is relatively recent science that is a driving force behind patenting. While the marginal university patent may be less "idea-rich" than it used to be, there is no evidence that the marginal paper, particularly in the highly cited disciplines, is any less "idea-rich" than it used to be. In fact, since citations are increasing must faster than papers, the crude numbers and estimates presented herein would seem to suggest that the marginal "quality" (or at least "marginal relevance") of papers in at least some disciplines is increasing.

Finally, the preceding observations would seem to indicate that the research agenda sketched out in this paper holds considerable promise. Given the availability (at a

price) of similar data for other major university systems, and the clear interest of university administrators in documenting and improving the rate at which they deliver useful technological information to the private sector, it is my hope that this paper will stimulate similar research by other scholars in other states. The creation of a "master" data set containing such data for the top 30 or so university systems would likely prove to be an extremely important and useful research tool. While such a data set may be beyond the reach of any individual scholar, it should be very much within the reach of the community of scholars involved in the NBER productivity program and similar groups.

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.

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.

Branstetter, L., 1999, "Is FDI a Channel of R&D Spillovers: Evidence from Japan's
FDI in the U.S." working paper, University of California, Davis.

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, 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.

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.

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., 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, 1997, "Stronger Protection or Technological Revolution: Which is Behind the Recent Surge in Patenting?" working paper.

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

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

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.

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.

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

Thursby, J. and M. Thursby, 2000, "Who is Selling the Ivory Tower? Sources of Growth in University Licensing," NBER Working Paper No. 7718.

Zucker, L., M. Darby, and M. Brewer, 1998, "Intellectual Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88: 290-306.
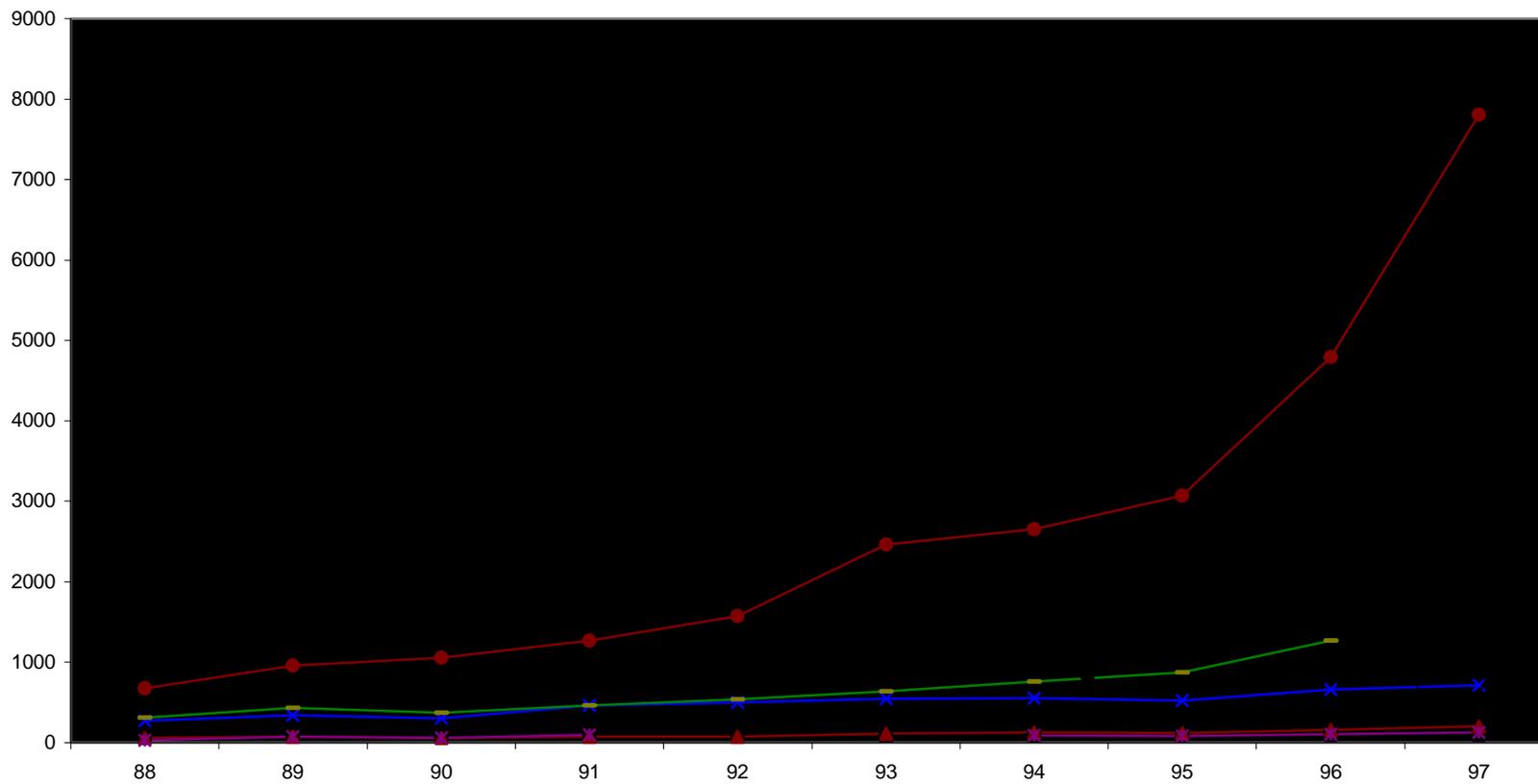
Figure 1   Citations to UC papers vs other indicators
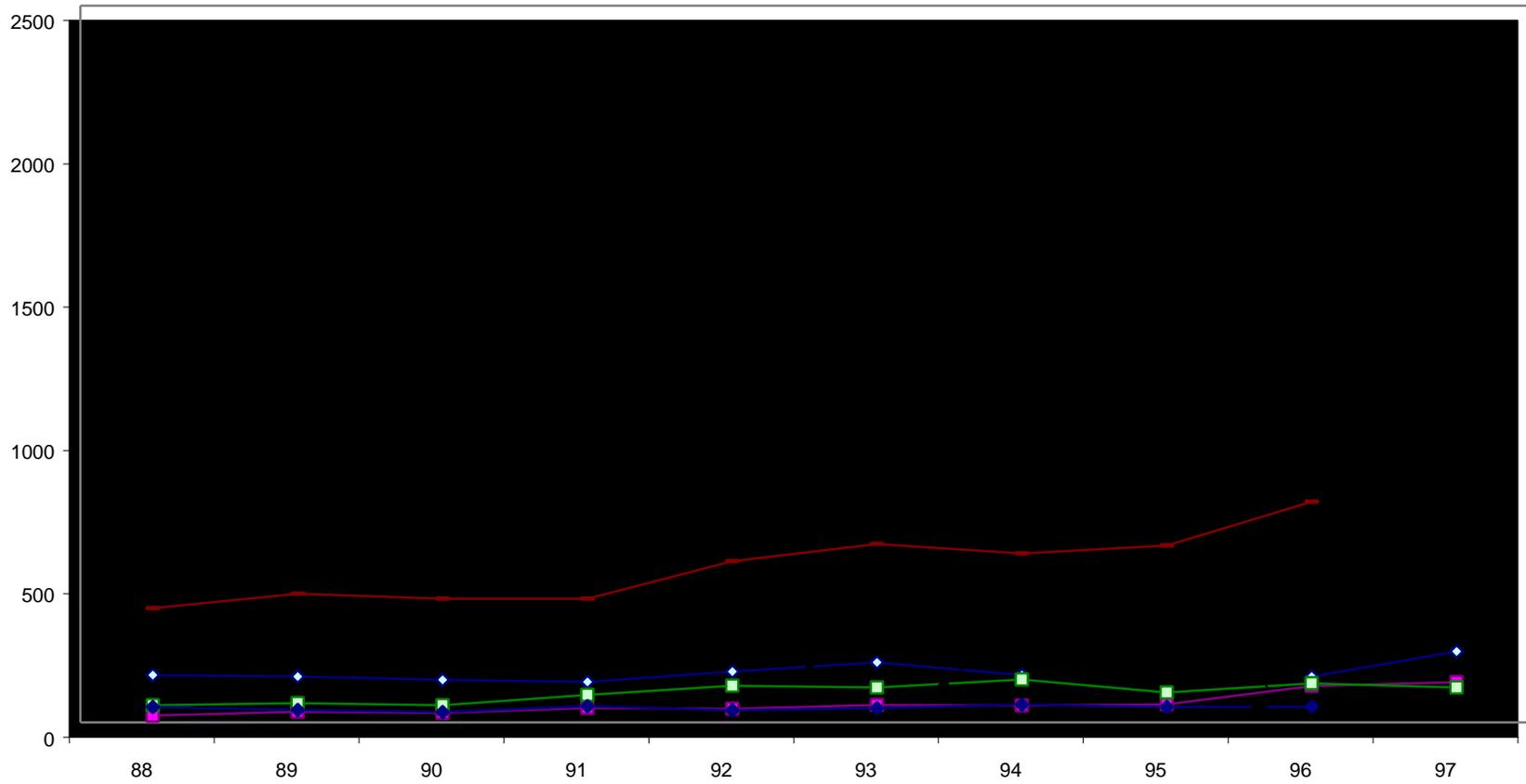
Figure 2 Citations of Stanford papers vs. other indicators

Table 1 Importance to Industrial R&D of Information Sources on University Research

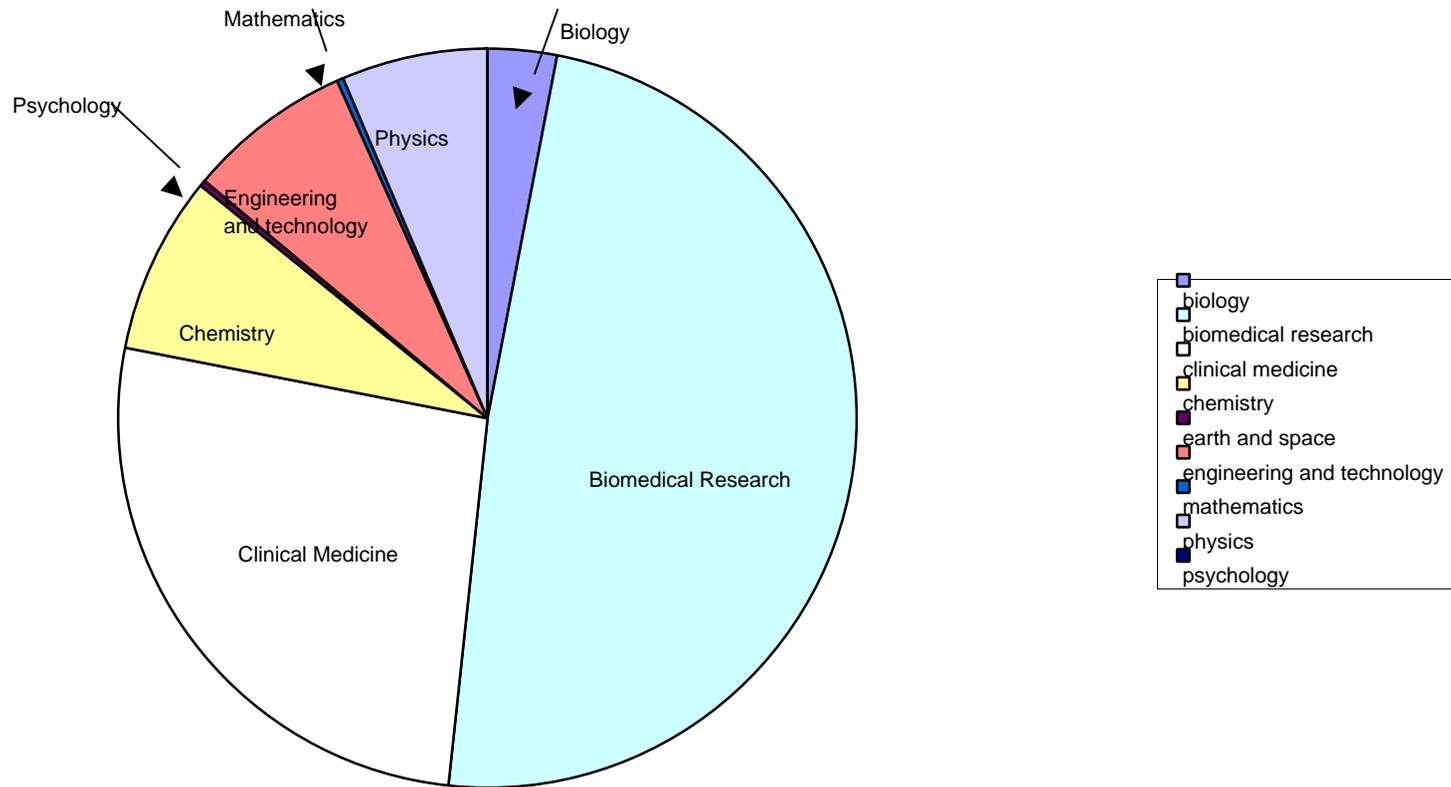| SIC	Industry	Patents	Publications	Conferences	Informal channels	Hires	Licenses	JVs	Contract Research	Consulting	Personal Exchange
2320	Petroleum	0	**46.67**	53.33	33.33	13.3	13.33	13	26.67	46.67	0
2400	Chemicals	25	**34.37**	28.12	18.75	18.8	7.81	16	20.63	26.56	9.37
2423	Drugs	56.86	**72.55**	60.78	60.78	31.4	35.29	41	54.9	54.9	7.84
2922	Machine Tools	10	**40**	40	40	20	0	10	20	40	0
3010	Computers	8.33	**41.67**	41.67	33.33	33.3	4.17	8.3	8.33	29.17	4.17
3100	Electrical Equipment	9.09	**31.82**	22.73	22.73	0	0	9.1	13.64	9.09	0
3210	Electronic Components	20	**36**	28	36	32	12	12	8	33.33	4
3211	Semiconductors	22.22	**61.11**	55.56	64.71	27.8	16.67	28	16.67	33.33	5.56
3220	Communications Equip.	5.88	**50**	32.35	32.35	29.4	8.82	8.8	17.65	29.41	20.59
3311	Medical Equip.	27.54	**37.68**	34.78	46.38	18.8	18.84	23	23.19	44.93	5.8
3312	Precision Instruments	25	**50**	44.44	44.44	11.1	13.89	19	8.33	36.11	5.56
3410	Car/Truck	33.33	**33.33**	11.11	33.33	11.1	11.11	22	33.33	22.22	11.11
3430	Autoparts	9.37	**43.75**	31.25	25	18.8	9.37	22	18.75	21.87	9.37
3530	Aerospace	14.58	**58.33**	50	54.17	18.8	6.25	40	35.42	39.58	4.17
	All Manufacturing	17.61	**40.91**	34.42	35.28	19.9	9.73	18	21.26	32.15	5.84

Source: Cohen, Florida, Randazzese, and Walsh, 1998, pp. 180-181

Table 2   Impact of Public Sector Research in Biotechnology

Activity	Overall	Literature	Contact	Recruitment
Future Innovations	9.1	4.5	2.3	2.3
Search	45.5	25	16	4.5
RD&D	22.7	13.6	9.1	
Instrumentalities	22.7	9.1	9.1	4.5
Overall		52.2	36.4	11.4

Source: Faulkner and Senker, 1995

Figure 3   Citations by Scientific Field, UC

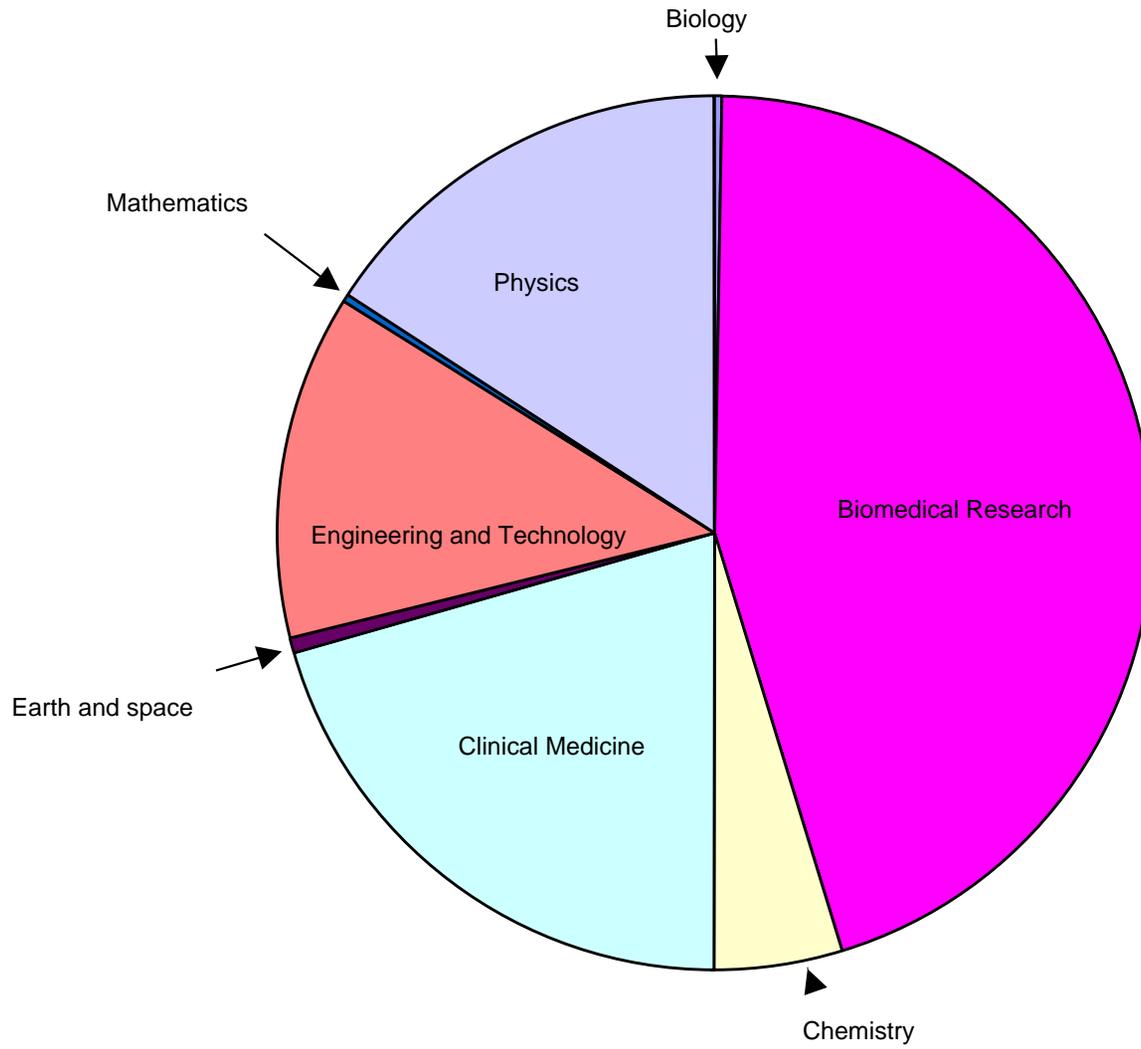**Figure 4   Stanford citations by field, including Stanford Medical School**

Biology

Mathematics

Physics

Engineering and Technology

Biomedical Research

Earth and space

Clinical Medicine

Chemistry

Figure 5   Biotech citations drive overall trends for UC
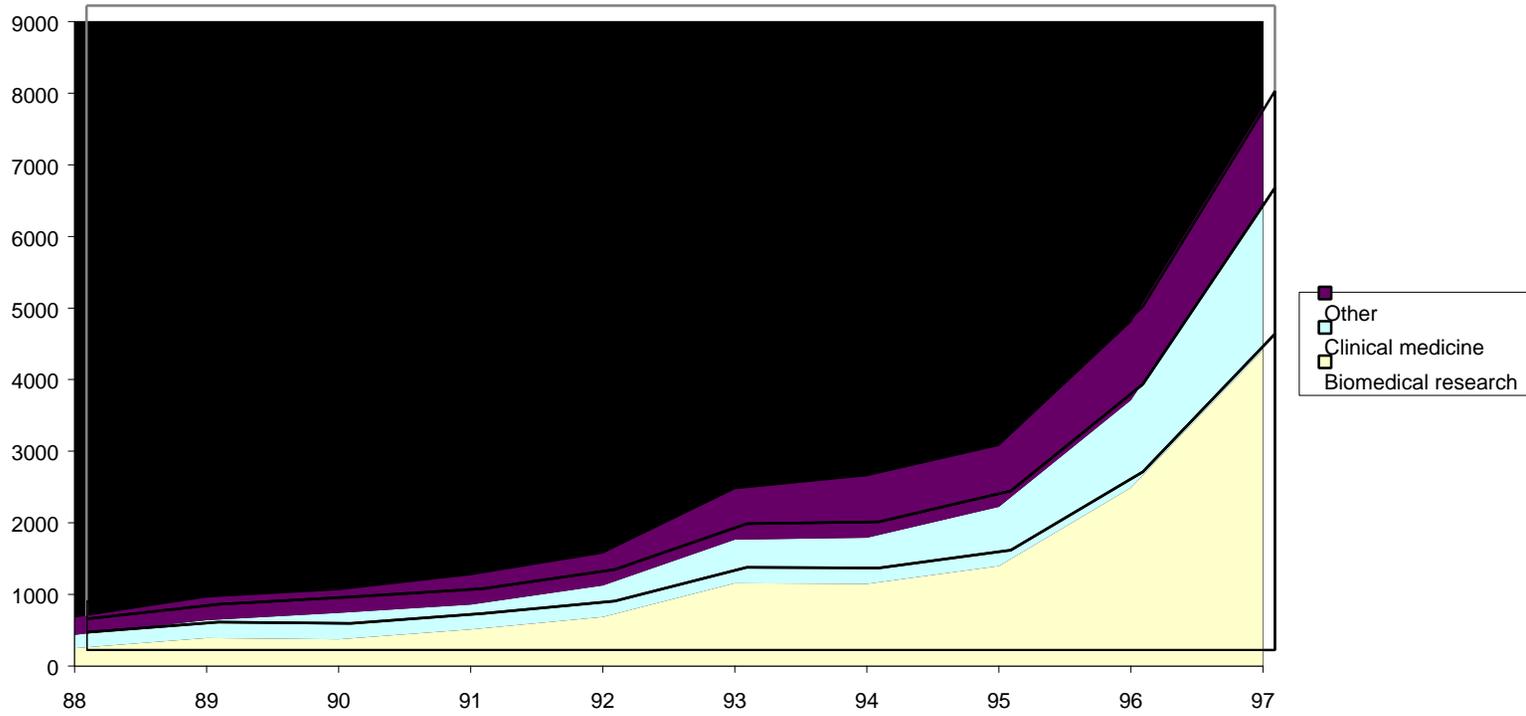
Figure 6 Stanford citations by field, 1988-1997
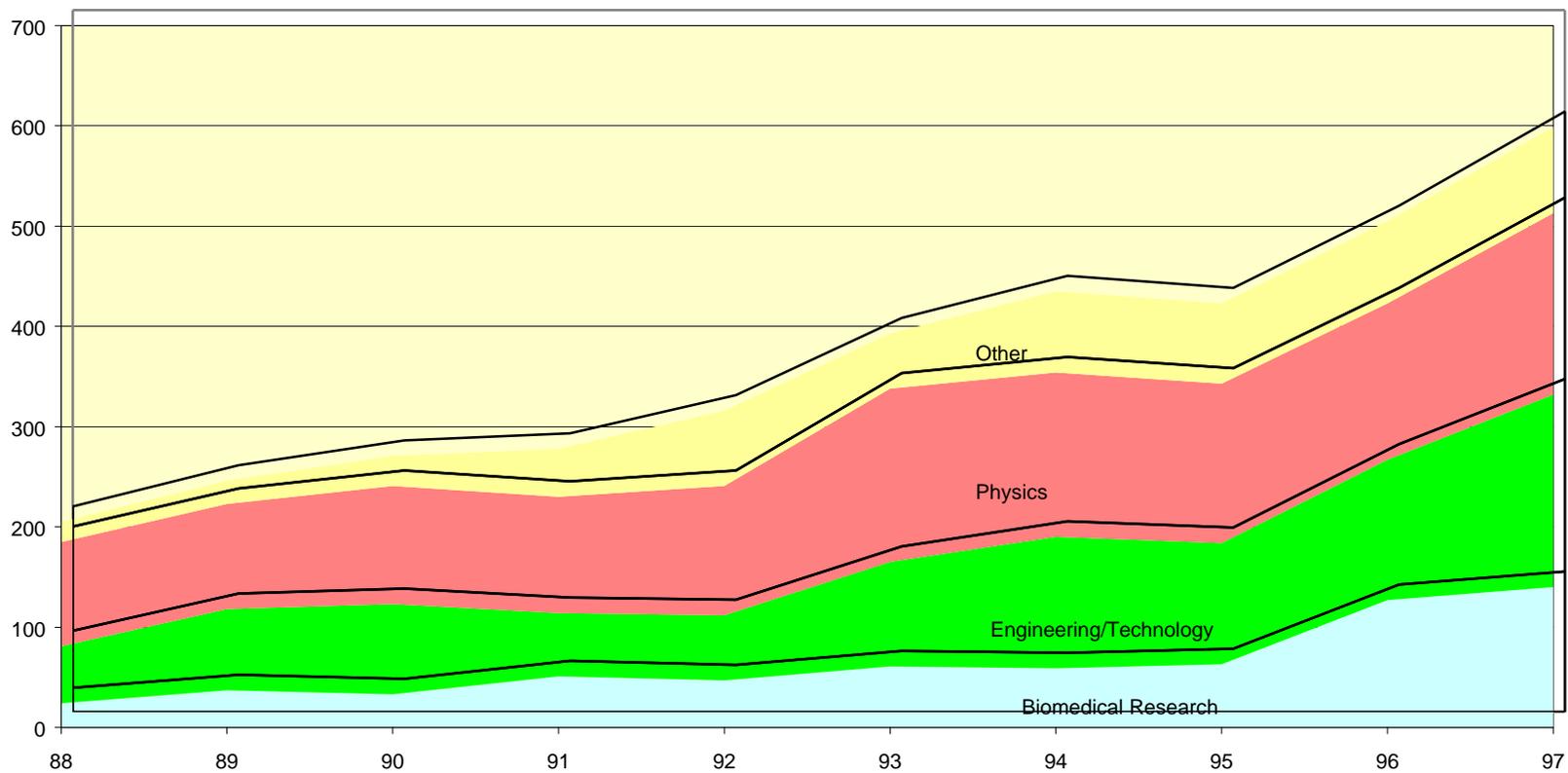**(Excluding Medical School)**

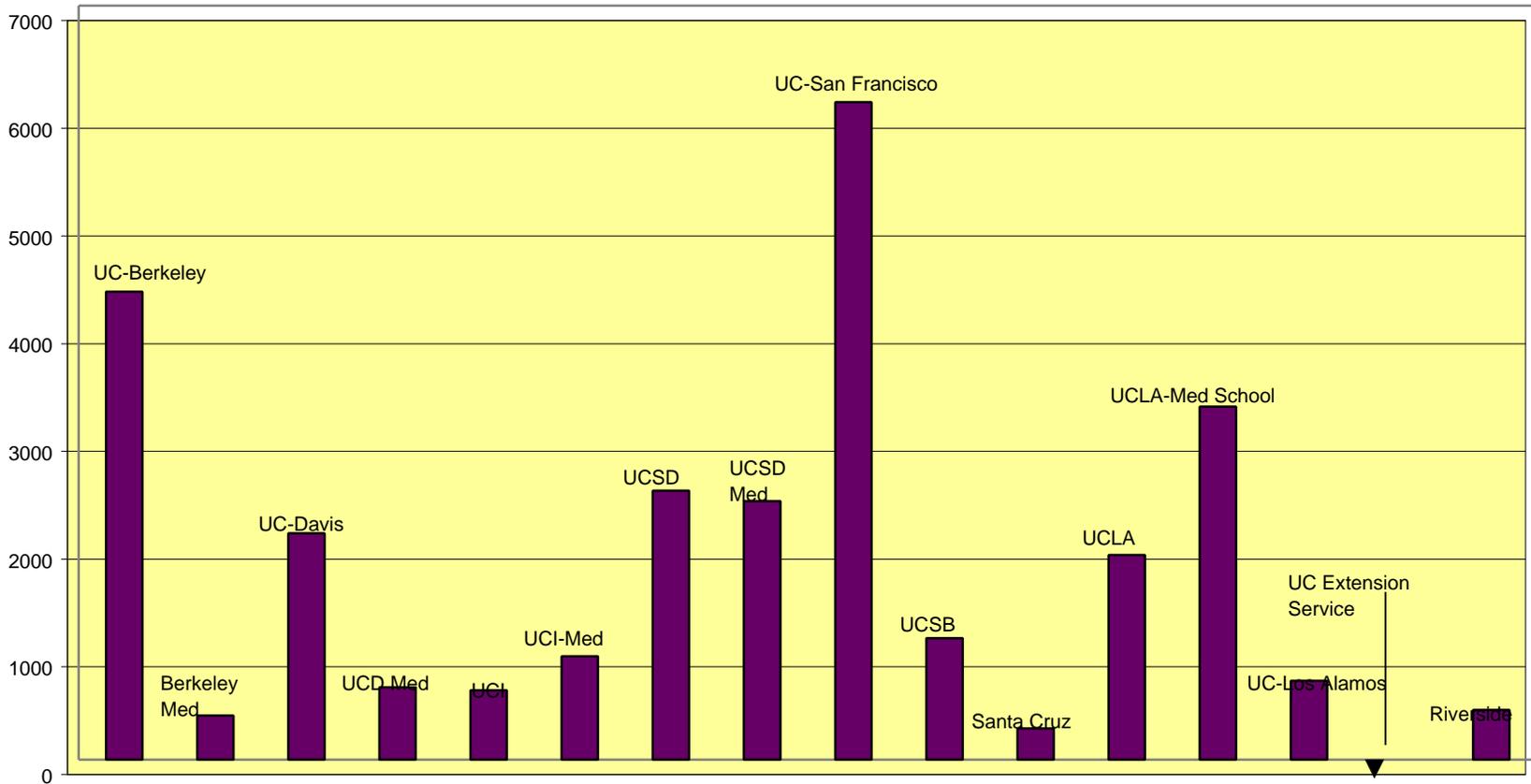**Figure 7   Total citations by UC campus/institution**

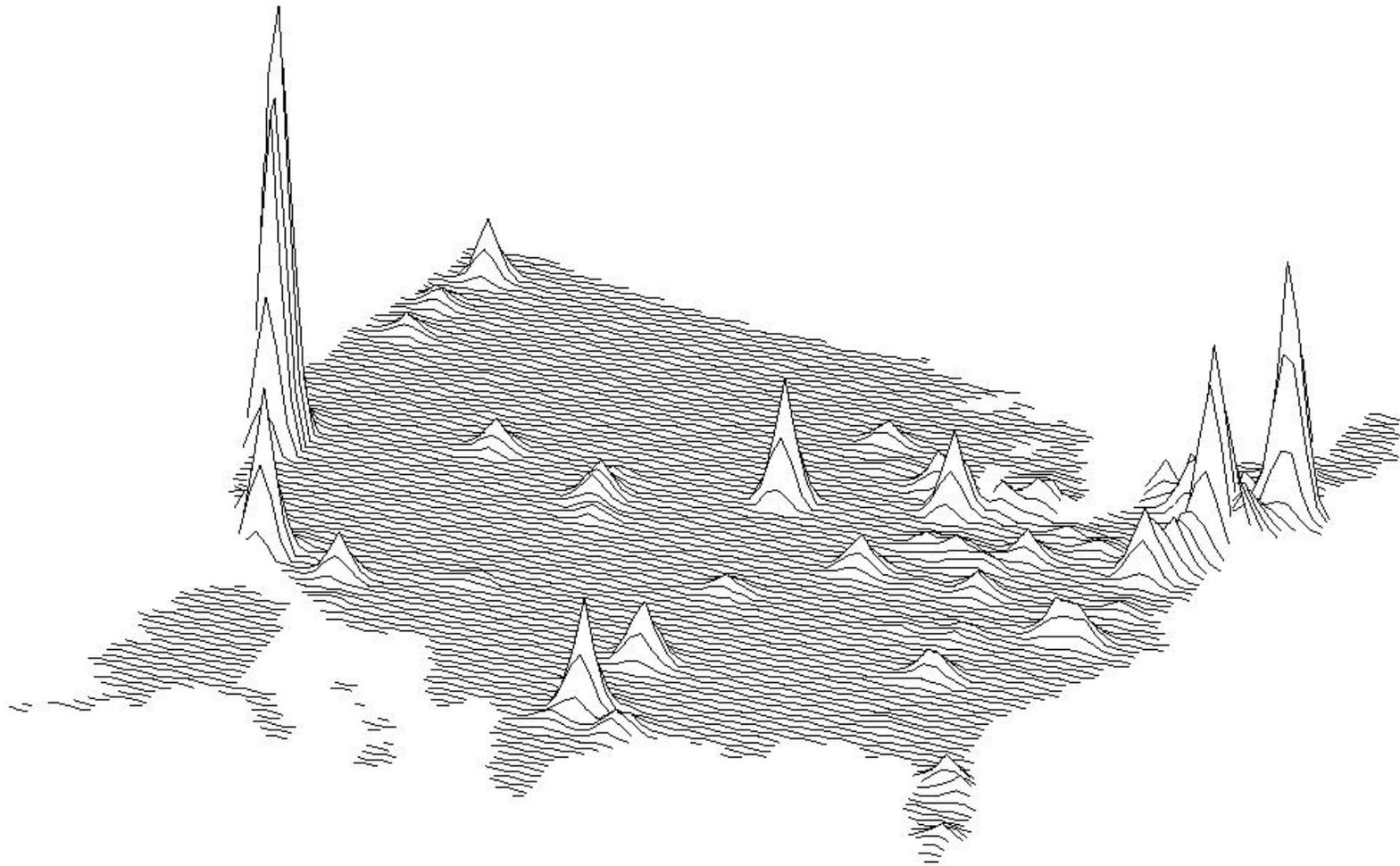Figure 8   Citations to UC Berkeley Papers, US
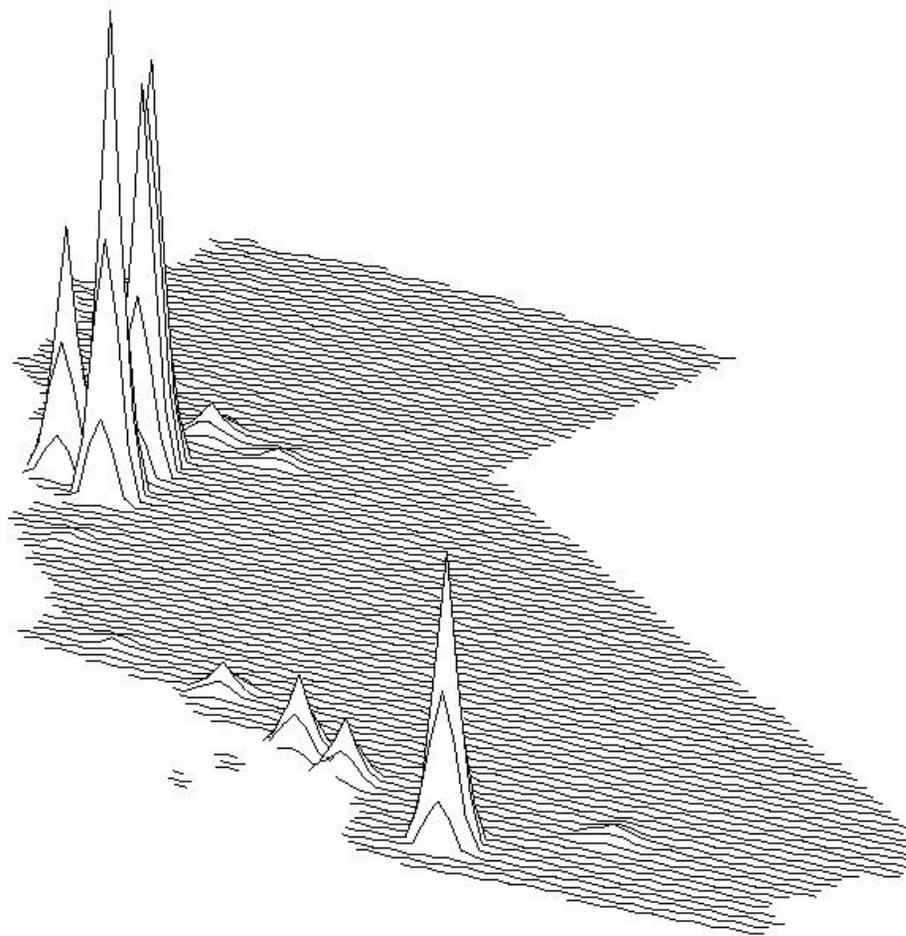
Figure 9 Citations to UC Berkeley Papers, California

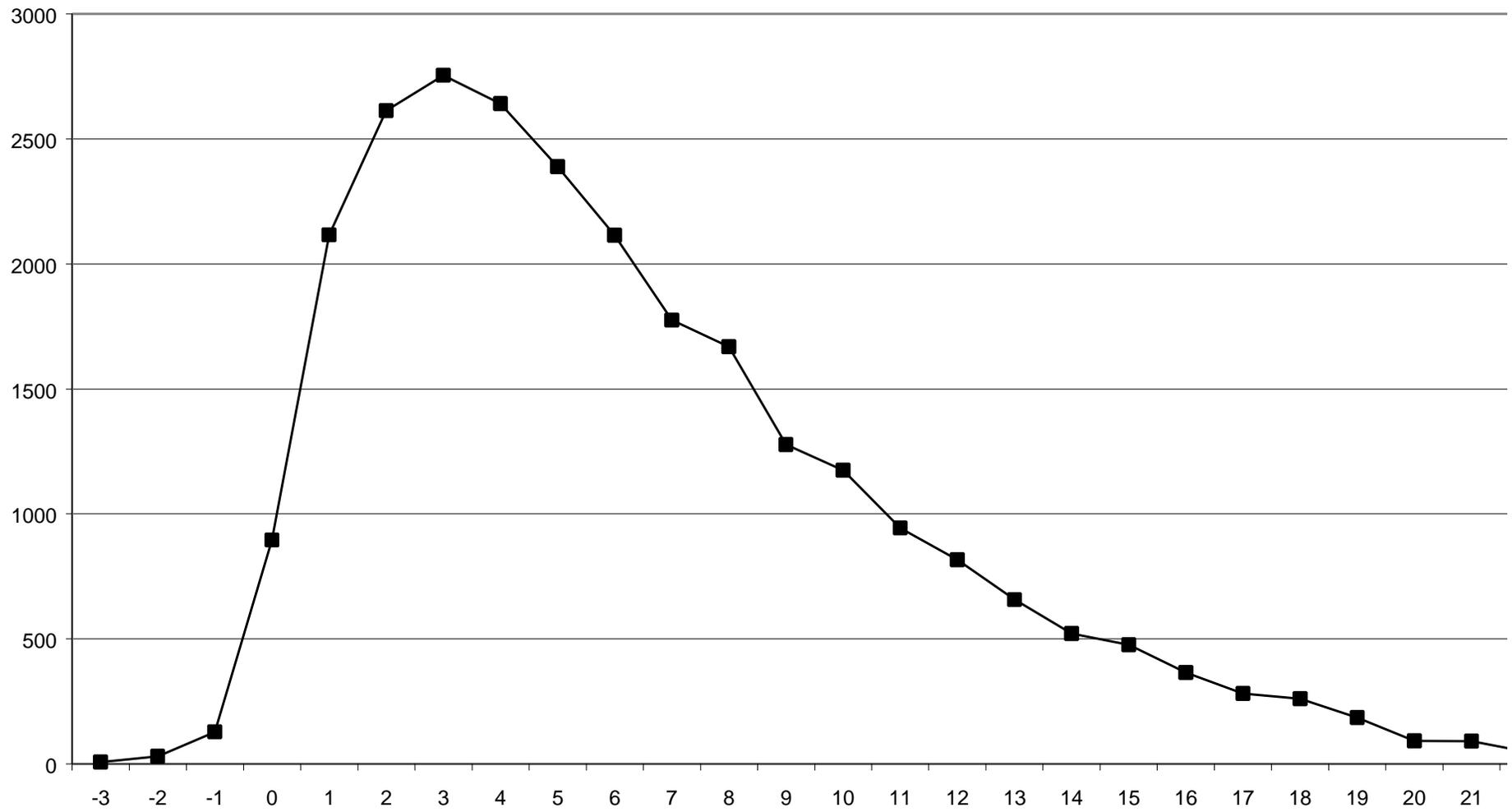**Figure 10   Lags between paper publication and patent application**

TABLE 3   Econometric Analysis of Citing Patents Only

Variable	Poisson			Negative Binomial		
	Coeff.	Std. Err.	z-stat	Coeff.	Std. Err.	z-stat
dist	0.052545	0.037272	1.41	-0.00203	0.038137	-0.053
dstate	0.021823	0.002642	8.26	0.010019	0.002966	3.378
dcnty	0.000947	0.003952	0.24	0.016167	0.004436	3.644
Dlag0	0.369132	0.029263	12.615	0.318366	0.029552	10.773
Dlag1	0.406983	0.021491	18.937	0.328933	0.022207	14.812
Dlag2	0.390081	0.020043	19.462	0.328275	0.020387	16.102
Dlag3	0.389842	0.019945	19.546	0.331016	0.020281	16.321
Dlag4	0.405211	0.020102	20.158	0.336577	0.020523	16.4
Dlag5	0.361611	0.02054	17.605	0.293899	0.020918	14.05
Dlag6	0.380266	0.021281	17.869	0.319121	0.021641	14.746
Dlag7	0.300711	0.02279	13.195	0.2408	0.023024	10.459
Dlag8	0.290672	0.025003	11.626	0.233868	0.025031	9.343
Dlag9	0.267624	0.0268	9.986	0.22207	0.026839	8.274
Dlag10	0.292398	0.02828	10.339	0.236958	0.028416	8.339
Dlag11	0.253135	0.031311	8.084	0.197521	0.032013	6.17
Dlag12	0.293841	0.032585	9.018	0.267821	0.033074	8.098
Dlag13	0.272878	0.035835	7.615	0.225961	0.036363	6.214
Dlag14	0.303293	0.040831	7.428	0.199679	0.042206	4.731
Dlag15	0.370119	0.042143	8.782	0.310247	0.043229	7.177
Dlag16	0.225049	0.048345	4.655	0.161185	0.049453	3.259
Dlag17	0.363246	0.051596	7.04	0.316077	0.051734	6.11
Dlag18	0.508345	0.063849	7.962	0.383934	0.065444	5.867
Dlag19	0.262838	0.076289	3.445	0.227468	0.077309	2.942
Dlag20	0.112348	0.091689	1.225	0.04834	0.095874	0.504
Dlag21	0.299037	0.132362	2.259	0.246146	0.134674	1.828
_cons	-0.09989	0.019069	-5.238	-0.43165	0.043481	-9.928
Biology				0.142797	0.040006	3.569
Biomed				0.243673	0.024646	9.887
Chemistry				0.186879	0.030365	6.154
Clin. Medicine				0.180109	0.023833	7.557
Earth/space				0.105041	0.136151	0.771
Engineering/tech				0.094863	0.037678	2.518
Mathematics				0.192904	0.379149	0.509
Physics				0.102372	0.03729	2.745

TABLE 4 Econometric Analysis of Citing Patents Plus Nonciting Controls

	Poisson			Negative Binomial		
Variable	Coeff.	Std. Error	z-stat	Coeff.	Std. Error	z-stat
dist	0.028518	0.03825	0.746	-0.02492	0.053027	-0.47
dstate	0.041533	0.002708	15.337	0.030714	0.003321	9.249
dcnty	0.005281	0.004156	1.271	0.039739	0.007057	5.631
Dlag0	0.183462	0.029213	6.28	0.415766	0.046894	8.866
Dlag1	0.144973	0.021573	6.72	0.438943	0.033029	13.29
Dlag2	0.182471	0.020064	9.094	0.412661	0.031047	13.292
Dlag3	0.176568	0.02002	8.82	0.487401	0.029599	16.467
Dlag4	0.197665	0.020032	9.868	0.459188	0.029682	15.47
Dlag5	0.229656	0.020326	11.299	0.430753	0.029691	14.508
Dlag6	0.247381	0.021513	11.499	0.443892	0.031693	14.006
Dlag7	0.133829	0.022685	5.9	0.314559	0.03458	9.097
Dlag8	0.168771	0.024382	6.922	0.334996	0.036051	9.292
Dlag9	0.118561	0.025947	4.569	0.279705	0.039065	7.16
Dlag10	0.121217	0.028229	4.294	0.281775	0.040235	7.003
Dlag11	0.105573	0.031759	3.324	0.310739	0.044123	7.043
Dlag12	0.043467	0.032944	1.319	0.294047	0.048249	6.094
Dlag13	0.166856	0.035612	4.685	0.235102	0.049194	4.779
Dlag14	0.106759	0.041106	2.597	0.365968	0.056556	6.471
Dlag15	0.207689	0.042265	4.914	0.387112	0.058335	6.636
Dlag16	0.01901	0.049637	0.383	0.357288	0.068931	5.183
Dlag17	0.230787	0.051975	4.44	0.343885	0.071315	4.822
Dlag18	0.181156	0.066885	2.708	0.476819	0.086675	5.501
Dlag19	0.118101	0.0778	1.518	0.222155	0.105675	2.102
Dlag20	0.217564	0.089988	2.418	0.132281	0.124251	1.065
Dlag21	0.363832	0.133659	2.722	0.560115	0.220336	2.542
Biology	0.684281	0.037331	18.33			
Biomed	1.413191	0.019157	73.769			
Chemistry	0.948693	0.026069	36.392			
Clin. Medicine	1.039707	0.019039	54.609			
Earth/space	1.38828	0.134705	10.306			
Engineering/tech	1.245991	0.032275	38.605			
Mathematics	1.274324	0.379088	3.362			
Physics	1.099169	0.031695	34.679			
apyear				0.027679	0.004683	5.911
_cons	-1.49231	0.041779	-35.719	-55.5148	9.324683	-5.954
inflate						
categ1	-19.819	536938.2	0	-5.70235	0.323109	-17.648
categ2	-4.18121	963.7202	-0.004	-1.50184	0.163473	-9.187
categ3	-3.48186	498.832	-0.007	-1.36812	0.107109	-12.773
_cons	-14.9426	97.82946	-0.153	0.853846	0.033085	25.808

Figure 11   Coefficients on Lag Terms