

8-2005

# Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations

Lee G. Branstetter

*Columbia University*, [branstet@andrew.cmu.edu](mailto:branstet@andrew.cmu.edu)

Yoshiaki Ogura

*Kyoto University*

Follow this and additional works at: <http://repository.cmu.edu/sds>

---

This Working Paper is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Social and Decision Sciences by an authorized administrator of Research Showcase @ CMU. For more information, please contact [research-showcase@andrew.cmu.edu](mailto:research-showcase@andrew.cmu.edu).

NBER WORKING PAPER SERIES

IS ACADEMIC SCIENCE DRIVING A SURGE  
IN INDUSTRIAL INNOVATION? EVIDENCE  
FROM PATENT CITATIONS

Lee Branstetter  
Yoshiaki Ogura

Working Paper 11561  
<http://www.nber.org/papers/w11561>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
August 2005

We thank seminar participants at the NBER Summer Institute, Columbia Business School, the UCLA Anderson Graduate School of Management, Hitotsubashi University, the University of Tokyo, and the NBER-CREST-CRIW Zvi Griliches Memorial Conference for useful comments and suggestions. We are grateful to Jim Adams, Reiko Aoki, Pierre Azoulay, Rebecca Henderson, Adam Jaffe, Joshua Lerner, and Frank Lichtenberg for detailed comments on an earlier draft. We 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. By prior agreement, they will remain anonymous. We are indebted to Masami Imai, Hiau-Looi Kee, Changxiu Li, and Kaoru Nabeshima for excellent research assistance. We would like to thank Tony Breitzman, Francis Narin, Adam Jaffe, and Marie and Jerry Thursby for their help in obtaining the data used in this study. The Institute of Economic Research at Hitotsubashi University provided a hospitable environment in which part of this paper was written – Lee Branstetter thanks the faculty and staff for their support. 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, and the National Science Foundation (grant no. 0097941). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation or any other funding agency. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2005 by Lee Branstetter and Yoshiaki Ogura. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations  
Lee Branstetter and Yoshiaki Ogura  
NBER Working Paper No. 11561  
August 2005  
JEL No. O31, O38

**ABSTRACT**

What is driving the remarkable increase over the last decade in the propensity of patents to cite academic science? Does this trend indicate that stronger knowledge spillovers from academia have helped power the surge in innovative activity in the U.S. in the 1990s? This paper seeks to shed light on these questions by using a common empirical framework to assess the relative importance of various alternative hypotheses in explaining the growth in patent citations to science. Our analysis supports the notion that the nature of U.S. inventive activity has changed over the sample period, with an increased emphasis on the use of the knowledge generated by university-based scientists in later years. However, the concentration of patent-to-paper citation activity within what we call the "bio nexus" suggests that much of the contribution of knowledge spillovers from academia may be largely confined to bioscience-related inventions.

Lee Branstetter  
Columbia Business School  
815 Uris Hall  
3022 Broadway  
New York, NY 10027  
and NBER  
lgb2001@columbia.edu

Yoshiaki Ogura  
Kyoto University Institute of Economic  
Research  
Center for Advanced Policy Studies  
Yoshida Honmachi, Sakyo-ku, Kyoto-shi  
606-8501 JAPAN  
ogura@kier.kyoto-u.ac.jp

## I. Introduction

Recent research points to an apparent surge in innovative activity in the United States over the past fifteen years. This is suggested by, among other things, a sharp rise in patent applications and patent grants that has substantially outpaced increases in public and private R&D spending. While a large fraction of U.S. patent grants are awarded to foreign inventors, the fraction obtained by domestic inventors has risen – and this fraction has risen particularly rapidly in fields where patenting has grown most sharply. The recent patent surge could potentially be explained by an increase in the propensity of Americans to patent inventions, rather than an increase in the productivity of American research and development, but the recent research of Kortum and Lerner [1998, 2000, 2003] strongly supports the latter interpretation. If this conclusion is correct, then it could help explain the widely observed increase in U.S. TFP growth in recent years.<sup>1</sup>

But if American R&D productivity has increased, then that raises the question of what factors are driving the increase.<sup>2</sup> This paper attempts to assess the importance of one possible contributing factor – increased knowledge spillovers from U.S.-based academic science. In essence, this paper is an attempt to explain the phenomenon graphed out in Figure I. This figure shows that citations made by patents granted in the United States to articles in the scientific literature increased very rapidly from the mid 1980s through the late 1990s.<sup>3</sup> Over this period, the number of patents granted by the U.S. Patent and Trademark Office to U.S. residents more than doubled, real R&D expenditures in the United States rose by almost 40%, and global output of scientific articles increased by about 13%, but patent citations to science *increased more than*

---

<sup>1</sup> See Gordon [2002].

<sup>2</sup> The work of Kortum and Lerner [2000] has stressed the potential role of venture capital-linked firms in improving U.S. R&D output.

<sup>3</sup> This graph does not break down growth in citations by the nationality of the inventor, but data from the 2002 *National Science and Engineering Indicators* shows that the majority of these citations are made by domestic patent applicants, and U.S.-based academic science is disproportionately likely to be cited. The fraction of citations to science made to U.S. authors has increased over this period. See also Narin et. al. [1997] and Hicks et. al. [2001].

13 times.<sup>4</sup> Many at the National Science Foundation and other U.S. science policy agencies find this graph extremely interesting, because it seems to suggest – at least in some broad sense – that academic science and industrial technology are “closer” than they used to be. This could mean that publicly funded science is generating more spillovers to industrial innovation than in the past.<sup>5</sup> This, in turn, may have contributed in important ways to the apparent surge of innovative activity in the United States in the 1990s.

This positive interpretation of recent trends in the data is influenced by the theoretical contributions of Evenson and Kislev [1976] and the more recent analysis their work inspired, such as Adams [1990] and Kortum [1997]. In this general class of models, applied research is a search process that eventually exhausts the technological opportunities within a particular field. However, basic science can open up new “search distributions” for applied researchers, raising the productivity and the level of applied research effort – at least temporarily. Viewed through this theoretical lens, the concurrence of rapid growth in U.S. private R&D expenditures, even more rapid growth in patenting, mounting evidence of an acceleration in TFP growth, and still more rapid growth in the intensity with which U.S. patents cite academic science would all seem to suggest a response to new technological opportunities created by academic research. Not surprisingly, other advanced industrial nations are deliberately trying to foster closer connections between university-based scientific research and industrial R&D in conscious imitation of the “U.S. model.”

However, increasingly strong knowledge spillovers from academic science to industrial R&D are only one of several factors that could be driving the changes illustrated in Figure I. Furthermore, even if such knowledge spillovers are growing in strength, this could be happening in a number of different ways, which have different implications for public policy.

---

<sup>4</sup> These data come from the 2002 edition of the *National Science and Engineering Indicators*. The data on scientific article output may understate the growth in articles, but even a substantial correction of the official statistics would leave the basic message of Figure 1 essentially unchanged.

<sup>5</sup> This interpretation has been stressed in recent editions of the *National Science and Engineering Indicators* and in the recent work of Narin et. al. [1997].

In order to better understand these implications, we seek to accomplish two related goals in this paper. First, we conduct a systematic analysis of the growth in U.S. patent citations to the scientific articles generated by a particular subset of American research universities across space, time, cited fields of academic science, and citing fields of technology. We consider several possible sources of growth, including:

- (1) *Changes in the nature of academic science*, such that more recent science has become more relevant to inventors
- (2) *Changes in the nature of industrial invention*, such that inventors have become more likely to rely on academic research in their corporate R&D programs
- (3) *Changes in the industrial composition of patent activity*, such that there is more patenting in fields of technology that have historically been closely linked to science
- (4) *Changes in citation activity that are driven by legal considerations*, but have no connection to real “knowledge spillovers” from academic science to industrial innovation

As we will show below, we find that patent citations to science are overwhelmingly concentrated in a nexus of academic disciplines and technological fields that we dub the “bio nexus”; much of the growth in patent citations to science can be explained by the growth of patenting in this nexus, in which a close connection between science and invention has existed for decades. At first glance, it seems that changes in the industrial composition of patenting are the primary driving factor of increased patent citations to science.

Having established this fact, we then examine patent citation activity *within* the bio nexus more carefully, in order to explore the possibility that the expansion of patenting in the bio nexus is itself partly driven by increasingly powerful knowledge spillovers from academic science. We find suggestive evidence that this is the case. We also provide an abbreviated examination of patent-to-paper citation activity outside the bio nexus and show that this activity is concentrated in a secondary “information technology” nexus. In the concluding section, we outline some policy implications of our results and directions for future research. The main message of this paper is that increased knowledge flows from academia may have contributed significantly to the

innovation surge reflected in the U.S. patent statistics, but most of that impact appears to be confined to a narrow locus of technologies and scientific fields.<sup>6</sup>

## **II. The Link Between Academic Science and Industrial Innovation**

### Historical Perspective

In the late 19<sup>th</sup> and 20<sup>th</sup> centuries, the search for commercial applications of the preceding decades' scientific discoveries led to the early creation within American universities of new engineering disciplines, including chemical engineering, electrical engineering, and aeronautical engineering, as documented by Rosenberg and Nelson [1994]. However, progress at the scientific frontier was still dominated by European institutions through the 1930s. World War II and its aftermath prompted a substantial migration of European scientists to the U.S., where they received an unprecedented level of financial support from the federal government. The large U.S. postwar investment in basic research was predicated on the notion that investment in basic science would eventually lead to useful technological invention for use in both industry and in national defense. However, early attempts to assess the strength of this connection in the postwar era, including De Solla Price [1965], Lieberman [1978], and the Defense Department's ambitious "Project Hindsight," suggested that relationship between frontier academic science and industrial invention, while obviously important, was neither close nor direct.

### Lessons from the Recent Literature

Drawing upon a wide range of data sources and methodological approaches, the recent economics literature suggests that the linkage between frontier science and industrial technology is stronger and more direct than in the past.<sup>7</sup> Case studies, manager interviews, and surveys have been used to assess the magnitude of this impact, the channels through which it flows, and

---

<sup>6</sup> Our reliance on patent statistics means that we will miss the impact of academic research on certain domains of invention – such as “open source” software – that do not utilize patents at all.

<sup>7</sup> For a comprehensive literature review that covers relevant research beyond the economics journals, see Agrawal [2001].

changes in these factors over time.<sup>8</sup> These studies suggest that firms perceive academic research to be an important input into their own research process, though this importance differs widely across firms and industries.<sup>9</sup> A second stream of recent 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], Barnes et al. [1998], Mowery et al. [1998], and Kim et al. [2005] have used data on university patents and citations to these patents to quantify knowledge spillovers from academic science. While patenting by universities has increased substantially in the United States over the last twenty years, there is evidence that as the number of university patents has grown, the marginal quality of those patents has declined.<sup>10</sup>

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] and 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], Thursby and Thursby [2002], Shane [2000, 2001], and Lach and Schankerman [2003]. While the counts of licensed inventions have grown over time, there is also evidence that, like patents, the marginal value of licenses has declined as their number has increased [Thursby and Thursby, 2002]. Furthermore, this stream of literature suggests that inventions generated by universities are typically quite “embryonic” – bringing such inventions to the market requires extensive additional investment by private firms.

#### Using Patent Citations to Academic Science as Measures of Knowledge Spillovers

---

<sup>8</sup> Important recent studies relying primarily on case study techniques and surveys include Mansfield [1995], Cohen et. al. [1994], Faulkner and Senker [1995], and Agrawal and Henderson [2002].

<sup>9</sup> While the channels by which firms absorb the results of academic research vary across industries, the Cohen et. al. [1994] study suggests that the formal scientific literature is, on average, an important channel.

<sup>10</sup> See Jaffe, Trajtenberg, and Henderson [1998] and Hicks et al. [2001].



This paper will use patent citations to academic papers to measure knowledge spillovers between academic science and industrial R&D.<sup>11</sup> 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, regardless of discipline, to publish all research results of scientific merit. As a consequence, the top-ranked research universities generate thousands of academic papers each year. Similarly, inventors have an incentive to patent their useful inventions, and a legal obligation under U.S. patent law to make appropriate citations to the prior art – including academic science.

The recent research discussed in previous paragraphs indicates that, 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 and 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 commercially relevant research output is ever patented, and only a fraction of this set of patents is ever licensed.<sup>12</sup> To illustrate this, Figure II shows the trends over the 1988-1997 period in several alternative indices of university research output and knowledge spillovers for one of the university systems in my data set, the University of California, which includes nine separately managed campuses and a number of affiliated laboratories. The figure graphs university patents by issue year (patents), invention disclosures by year of disclosure filing (invention disclosures), new licenses of university technology by date of contract (licenses), the number of citations to previous university patents by issue year of the

---

<sup>11</sup> In doing so, we are building on the work of Francis Narin and his collaborators, who have pioneered the use of these data in large-sample “bibliometric” analysis. See Narin et al. [1997] and Hicks et al. [2001] for recent examples of this work.

<sup>12</sup> This result is also emphasized strongly in the interview-based evidence presented by Agrawal and Henderson [2002]. Thursby and Thursby [2004], in their study of 3,342 faculty members in science and engineering departments at Cornell, MIT, University of Pennsylvania, Purdue, Texas A&M, and the University of Wisconsin-Madison find that only 7.1% of their (person-year) observations show invention disclosures (the first step in the patenting process) by a faculty member. The overwhelming majority of professors never patent. Azoulay et al. [2004] show that this is true even in their study of 4,270 academic life scientists. Fewer than 4% of their observations represent patenting by a faculty member, and most patenting faculty produce only a handful of patents over the course of a career.

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). Clearly, citations to papers are far more numerous than any other indicator. This figure suggests that patent citations to academic papers may provide a much broader window through which to observe knowledge spillovers from academic science to inventive activity than any available alternative.<sup>13</sup>

Citations to scientific articles can reflect learning on the part of industrial inventors through multiple channels. For instance, a firm may learn about a useful scientific discovery through an informal consulting relationship with an academic scientist or through the hiring of graduate students trained by that scientist rather than through a systematic and regular reading of the professional scientific literature. Even in these cases, the confluence of academic scientists' interest in rapid publication of significant discoveries combined with firms' legal obligation to cite relevant prior art means that citations to scientific articles will often show up in patent documents, providing a "paper trail" of knowledge diffusion, even when the particular means of knowledge diffusion was something other than the publication itself.

It has long been recognized that patent citations are inserted into patent documents by both the patent applicants and patent examiners. Exploiting a change in patent data in 2001 that identifies the origin of citations, Sampat [2005] has examined the distribution of examiner and applicant-generated citations across patent classes and types of applicants. He finds that patent examiners collectively account for a surprisingly large fraction of total citations to previous patents (62%), calling into question the earlier interpretation of these citations as indicators of knowledge spillovers to the inventor. On the other hand, he finds that examiners account for a low fraction, and applicants a high fraction (90%), of citations to non-patent prior art.

### **III. Examining Patent Citations to Science: A Citations Function Approach**

#### Explaining the Growth in Patent Citations to Science: Four Alternative Views

---

<sup>13</sup> Other recent studies using data on patent citations to scientific papers include work by Fleming and Sorenson [2000, 2001] and Lim [2001]. None of these studies focuses on the large change in citations to academic science over the course of the 1990s, which is the focus here.

In the introduction, we noted that one of the goals of this paper is to identify the factors that best explain the growth in patent citations to science over time, and we listed four possible sources of variation that will receive particular scrutiny. Each of these can be associated with a particular hypothesis of what is driving the increase in patent citations to science. Before describing our data and empirical approach, we think it is useful to outline these hypotheses in greater detail.

The first is the **“increasing scientific fertility”** hypothesis, which posits that more recent cohorts of scientific papers contain more discoveries that are directly applicable to industrial research and development, and that this trend holds across many fields of science. A less positive view of the same phenomenon would be that academic scientists are doing less “fundamental science” and are deliberately pursuing more work with (potential) commercial applications, partly in response to financial incentives. Under this hypothesis, knowledge spillovers from academia to industry are increasing primarily *because of a qualitative change in the nature of the science being conducted at universities.*<sup>14</sup> If this is the primary driving factor, then a significant component of the growth in patent citations to science will be explained by an increasing propensity for more recent cohorts of scientific papers to be cited by patents.

The second is the **“changing methods of invention”** hypothesis, which posits that industrial inventors have changed the way they create new technology. The new approach to R&D draws more heavily on academic science than in the past, though it does not necessarily draw exclusively on the most recently published articles. The point being stressed is that it is the inventors themselves who are generating the increased citations as they *alter the direction and nature of their R&D programs to probe the new opportunities for industrial research created by*

---

<sup>14</sup> We note that here and elsewhere, we are being a bit loose in our use of the term “knowledge spillover.” The knowledge flows from academia to industry are only pure spillovers to the extent that industrial inventors receive them for free. In fact, conversations with industry-based R&D managers suggest that investments on the part of the firm (of various kinds) are necessary in order to effectively learn from these knowledge flows – so that they are not pure spillovers. See Cohen and Levinthal [1988], Zucker et. al. [1998], and Cockburn, Henderson, and Stern [1999].

*basic science*. Like the first hypothesis, this implies that knowledge spillovers from academic science are increasing over time, but the mechanism driving this increase is different. This would be reflected in an increasing propensity for more recent cohorts of patents across a wide range of technical fields to cite science.

The third is the “***changing composition of invention***” hypothesis, which posits that invention in certain areas of technology has been closely linked to science for some time, and, likewise, some fields of science have always been frequently cited by industrial patents. Under this hypothesis, there has been disproportionate growth in patenting in frequently citing patent classes. Similarly, growth in academic publications has been biased towards those fields of science which have historically been more closely linked to industrial R&D. In other words, at the level of individual technology classes and scientific fields, there has been little change in the relationship between science and technology *per se* – rather there has been a change in the distribution of patents and papers that generates the observed increase in citations.

The fourth hypothesis is the “***attorney-driven***” hypothesis, which posits that the change in patent citations is entirely driven by changes in citations practices. For various strategic reasons connected to the desire to impress patent examiners, the fear of subsequent litigation, or both, patent lawyers have instructed their clients to increase the number of citations made to the scientific literature. The increasing availability of data on the scientific prior art in electronic form has lowered the costs of such citations, further contributing to their growth. This hypothesis, in its extreme version, suggests that little can be learned about the changing relationship between science and technology from patent citation data.

These hypotheses are not mutually exclusive, but they have quite different implications for the appropriate interpretation of the growth in patent citations to papers. In order to understand what Figure I really means, how it relates (or not) to the recent American innovation surge, and what the appropriate policy response is, it is necessary to sort out the relative importance of these hypotheses in explaining the trend illustrated in that graph.

## Tracking Patent Citations to California-based Academic Science

If we are to do this, then it is essential that we examine changes in patent citations to papers while controlling for growth and changes in the distribution across fields of the population of potentially cited papers, growth and changes in the distribution across fields of the population of potentially citing patents, and differences in the historical propensity for different categories of patents to cite science. While it would be impractical to do this for the universe of academic publications and U.S. patents, it has been possible for us to obtain and link the requisite data for the campuses and affiliated research units of the University of California, Stanford University, the California Institute of Technology (Caltech), and the University of Southern California. These are the principal sources of academic research in the state of California. Our inference will be based on U.S. patent citations made to scientific articles generated by these institutions. The location of the inventor of the citing patent can be anywhere within the territory of the United States. Because the focus in this paper is on the impact of knowledge flows from academic science on U.S. invention, we restrict ourselves to the subset of U.S.-granted patents with U.S.-based inventors. However, we note that extending the sample to all patents granted by the U.S. PTO, regardless of location of the inventor, does not qualitatively alter any of our key empirical results.

Related research strongly suggests that the patterns in the citation data used in this study closely mirror national trends. Branstetter [2004] analyzes the *complete* set of nonpatent citations made by a random sample of 30,000 U.S. patents granted over the 1987-1997 period. While he uses a completely different statistical approach to these data than the one employed here and cannot control for changes in the volume and distribution of potentially citing papers, he finds the growth rate of patent citations to science and their distribution across fields of science and technology in that random sample to be similar to that indicated in the current paper. Nevertheless, one must be sensitive to the potential difficulties involved in generalizing from our

results to the entire American research university system. Wherever such difficulties arise, they are noted in the discussion of empirical results in sections IV, V, and VI.

From the University Science Indicators database generated by the Institute for Scientific Information, we have obtained comprehensive data on the publication of scientific articles by our sample of California research universities, by institution, year, and scientific field, from 1981-1997. These data are matched to data on patent citations made to these publications over the 1983-1999 (grant year) period, which were provided by CHI Research. CHI Research, which has since been acquired by the consulting firm ipIQ, developed a comprehensive data base of “non-patent references” made in U.S. patent documents. The focus in this paper is on the subset of references made to articles appearing in peer-reviewed scientific journals. In the CHI Research database, these references are put into a standardized format that can then be matched to data on papers published in the more than 4,000 journals covered by the Science Citation Index (SCI). To these data we match data on the universe of potentially citing U.S. utility patents granted over that same period, which is available from the NBER Patent Citation Database documented in Hall et. al. [2001].<sup>15</sup>

Trends in scientific publications generated by California research universities for a subset of scientific disciplines are plotted in Figure III. Particularly strong growth can be observed in biomedical research, “physics” (an aggregate which includes materials sciences fields connected to semiconductors), and “engineering and technology.”<sup>16</sup> Trends in U.S. patenting across different categories of technologies are similarly plotted in Figure IV. While patenting in all fields has increased over the sample period, particularly sharp increases can be seen in “drugs and medicine” and “computers and communications.”<sup>17</sup>

#### A Citation Function Approach to the Data

---

<sup>15</sup> Further details on data construction are provided in the Data Appendix.

<sup>16</sup> Comparison of these data with similar data for all major American research universities shows that California academic publication closely mirrors national trends.

<sup>17</sup> This graph does not break down patent trends by nationality of the inventor, but the fraction of patent grants awarded to domestic inventors has risen sharply in these two rapidly growing fields.

The empirical framework we use for analyzing these data borrows from the work of Jaffe and Trajtenberg [1996, 2002]. In this framework, we model the probability that a particular patent,  $p$ , applied for in year  $t$ , will cite a particular article,  $a$ , published in year  $T$ . This probability is determined by the combination of an exponential process by which knowledge diffuses and a second exponential process by which knowledge becomes superceded by subsequent research.

This probability is referred to in the work of Jaffe and Trajtenberg [1996, 2002] as the *citation frequency*. It is a function of the attributes of the citing patent (P), the attributes of the cited article (a), and the time lag between them (t-T). It can be rendered in notation as

$$(1) \quad p(a, P) = \alpha(a, P) \exp[-\beta_1(t - T)][1 - \exp(-\beta_2(t - T))]$$

Attributes of the citing patent that we incorporate into our analysis include the application year, the technical field (based on the primary technology class assigned by the patent examiner), the type of entity owning the patent (based on the identity of the assignee), and the geographic location of the patent, based on the address of the inventor. Attributes of the cited article that we consider include the publication year, the scientific field of the article, and the institution with which the authors were affiliated at the time of publication.

Given these data, one could sort all potentially citing patents and all potentially cited articles into cells corresponding to the attributes of articles and patents. The expected value of the number of citations from a particular group of patents to a particular group of articles could be represented as

$$(2) \quad E[c_{icelTSL}] = (n_{TSL})(n_{icel})\alpha_{icelTSL} \exp[-(\beta_1)(t - T)][1 - \exp(-\beta_2(t - T))]$$

where the dependent variable measures the number of citations made by patents in the appropriate categories of grant year (t), technology class (c), institutional type (e), and location of the citing patent's inventor (I) to articles in the appropriate categories of publication year (T), scientific field (S), and particular campus (L). The  $\alpha$ 's are multiplicative effects estimated relative to a

benchmark or “base” group of patents and articles. In this model, unlike the linear case, the null hypothesis of no effect corresponds to parameter values of unity rather than zero. Equation (2) can easily be rewritten as

$$(3) \quad \frac{E[c_{icelTSL}]}{(n_{TSL}) * (n_{icel})} = \alpha_{icelTSL} \exp[-\beta_1(t - T)][1 - \exp(-\beta_2(t - T))]$$

This is what Jaffe and Trajtenberg [1996] refer to as a *citations function*. If one adds an error term, then this equation can be estimated using nonlinear least squares. The estimating equation is thus

$$(4) \quad p_{icelTSL} = \alpha_i \alpha_c \alpha_e \alpha_l \alpha_T \alpha_S \alpha_L \exp[-\beta_1(t - T)][1 - \exp(-\beta_2(t - T))] + \varepsilon_{icelTSL}$$

where the dependent variable now measures the likelihood that a particular patent in the appropriate categories (grant year, technology class, institution type, and location) will cite an article in the appropriate categories (science field, source campus, and publication year).

Patents are placed into one of the following categories: computers and communications, chemicals, drugs and medicine, electronics, mechanical inventions, and a catch-all “other” category. These are the same categories for which patent growth is depicted in Figure III. Scientific articles are classified into the following fields: biology, biomedical research, chemistry, clinical medicine, engineering and technology, physics, and “other science.” Patent assignees are classified into the following institutional types: public science institutions (predominantly universities, research hospitals, and government laboratories), firms, and other institutions. The division of patents on the basis of location of the inventor and the assignment of patents and papers into groups based on grant and publication year, respectively, are discussed below.

We estimate various versions of (4) using the nonlinear least squares estimation routine of the STATA software package. When doing so, we weight the observations by the square root of the product of potentially cited articles and potentially citing patents corresponding to the cell, that is



$$(5) \quad w = \sqrt{(n_{tcel}) * (n_{TSL})}$$

This weighting scheme should take care of possible heteroskedasticity, since the observations correspond to “grouped data,” that is, each observation is an average (in the corresponding cell), computed by dividing the number of citations by  $(n_{tcel}) * (n_{TSL})$ .

#### IV. Empirical Evidence from Citations Functions

##### Localization in Time and Geographic Space

Regression results from a version of (4) are given in Table I. Using the parameter values from this regression, it is also possible to graph out the double exponential function implied by our parameter estimates, giving us a sense of how the “citedness” of a particular group of articles by a particular group of patents changes over time. This is graphed out for our “base case” in Figure V. The base case in this regression corresponds to patents assigned to firms, where the first inventor resides in the U.S. outside the state of California. The base patent grant period is 1983-1987, and the base publication period is 1981-1985. The base science category is biology, the base patent category is chemistry, and the base institution is Stanford University.<sup>18</sup>

The shape of the curve graphically demonstrates the first key result of this section – namely that citations to academic science are, to some extent, localized in time. Citations to science appear almost immediately after article publication, and the citation function peaks at a lag of about eight years after article publication. These lags are measured here with respect to the grant date of the patent. An alternative specification measuring patents by *application* date finds a modal lag between publication and application of five to six years, implying fairly rapid spillovers of knowledge from science into industrial invention. While the estimated lag structure

---

<sup>18</sup> As commonly understood, biology is an aggregate that contains components closely associated with the bio nexus (molecular biology) and components that are arguably not closely connected (such as population ecology). In this paper, however, we have classified the subdisciplines of biology closely connected to the bio nexus as “biomedical research.” Subdisciplines that remain within the biology aggregate used in this paper include such fields as ecology and “aquatic sciences.” They are not closely connected to the bio nexus and, defined this way, “biology” would seem to be a reasonable base category. Note also that the institutional boundary of campuses like Stanford is drawn to include affiliated medical schools.

demonstrates that papers continue to receive some citations even at relatively long lags, the citation frequency declines steadily after the peak lag.

These results also provide evidence of concentration in geographic space. Citing patents are assigned to two categories based on the recorded addresses of the inventor: California inventors and U.S. inventors outside California. U.S. inventors outside California are the base category, so the coefficients imply that California-based inventors in a given technology class are nearly three times more likely to cite California academic science.

The intranational localization of knowledge spillovers implied by the California effect seems large. However, the current specification arguably does *not* control well for regional clustering of industrial R&D within the particular niches of the broad technology categories we have employed. A finer disaggregation of patent classes would likely attenuate the measured degree of localization. Furthermore, as can be seen in Figure VII, it is still the case that large numbers of citations are made by inventors far from California. In fact, one sees a “bicoastal” concentration of citations, reflecting the clustering of U.S. innovative activity in the Northeast and the West Coast.<sup>19</sup>

#### Localization of Knowledge Flows in Technology Space and the “Changing Composition of Invention” Hypothesis

We find striking differences in the incidence of citation across fields of technology. Relative to the base category (chemicals), drug/medicine patents are 2.4 times more likely to cite science, whereas all other categories are substantially less likely to cite science. The typical patent in the least likely-to-cite category, mechanical patents, is only about 1% as likely to cite science as the typical chemical patent. The estimated gap between technology categories in citation propensity is quite substantial. Note that these estimated propensities control for the

---

<sup>19</sup> The recent work of Thompson and Fox-Keane (2005) suggests that earlier evidence of intranational geographic localization of knowledge spillovers may have exaggerated the real degree to which knowledge spillovers are regionally localized. Our cautious interpretation of our own findings of a “California effect” is based partially on this recent work.

number of patents in these categories over time, so that these coefficients are properly interpreted as an estimate of the differential “per-patent” propensity to cite science.

Continuing in this theme, we can also allow different categories of science to display different propensities to be cited by patented technologies. Note that the citation function specification controls for the number of “citable papers” within these science categories over time, as well as the number of potentially citing patents across fields of technology, so the coefficients on science categories are akin to a “per-paper” measure of technological fertility. The coefficients in Table I suggest that a paper in the “biomedical research” field is *about 41 times* more likely to be cited in a patent than a paper in the base category of biology. Papers in “chemistry” and “clinical medicine” are about five times as likely to be cited as a biology paper, while papers in the other science categories are substantially less likely to be cited than biology papers.<sup>20</sup> The gap between the most and the least intensely cited science categories is quite substantial.

As one can see in Figure IV, “biomedical” patenting has risen sharply over our sample period, both in absolute terms and relative to patenting in other technology categories. In fact, patenting in this area has risen more than four-fold. Likewise, as Figure III indicates, there has been substantial growth in publishing in bioscience areas by California research institutions. Even controlling for this growth, biotech patents are much more likely to cite science throughout the sample period, and bioscience papers are much more likely than other categories to be cited. This suggests that the aggregate trends in patent citations to science are driven largely by “biotech” patents citing “bioscience” papers. While there is growing citation activity outside this “bio nexus,” patent citations to science have, to date, been highly concentrated within it.

---

<sup>20</sup> In results available upon request, we estimated an “academic production function” for the university systems studied in this section of the paper, in which the output measure was the count of publications generated in a scientific field by a particular campus in a particular year. This was regressed on measures of “inputs” to the research process. The results suggest that the higher “productivity” of the biomedical sciences is not driven purely by the increase in R&D funding in that field.

In another take on the “composition hypothesis,” we have also looked at patenting by different categories of assignees: firms, public science institutions (universities, research institutes, and research hospitals), and a grab-bag category of “other institutions” in the non-profit sector. Assignment of a patent to one of these categories is based on the typography of assignees developed in the NBER patent citation database. Relative to the base category of firms, public science institutions are nearly four times as likely to cite academic science, and “other institutions” are almost twice as likely to cite academic science, according to Table I. This is unsurprising, given the connection that is likely to exist between academic science and academic patenting. Because these institutional categories accounted for a small fraction of total U.S. patenting, even by the end of our sample period, it is still the case that the vast majority of patent citations to California academic science are made by the patents of industrial firms, not universities.<sup>21</sup>

#### Evidence on “Changes in Methods of Invention”

Having incorporated fixed effects associated with the citing field of technology, the cited field of science, the cited institution, and characteristics of the citing inventor/assignee, we can also make some inference about average changes in citation patterns over time across fields. Perhaps the most interesting finding here is that the propensity to cite academic science is evidently growing over time. This can be seen by examining the pattern of coefficients on the citing patent grant year cohort terms. We group patents into categories corresponding to the years in which they were granted: 1983-87, 1988-90, 1991-93, 1994-96, and 1997-99. For these cohorts, we estimate an average propensity to cite science *relative* to the base category. These estimated propensities increase substantially from the “base category” of 1983-87, more than

---

<sup>21</sup> This statement requires some qualification. University patenting is highly concentrated in a small number of fields. By the end of our sample period, university patenting accounted for roughly 15% of health care-related patenting. That being said, the overall results in Table I are robust to the removal of patents granted to “public science institutions” (primarily universities and research hospitals) from the sample. In fact, in some ways, they become even stronger. See Table III and the discussion on page 23.

doubling by the end of the sample period.<sup>22</sup> Note that we have explicitly controlled for the fact that academic publications in the heavily cited branches of science have become more numerous and that there has been an increase in patenting in fields that heavily cite academic science. These results are consistent with the view that there has been *a change in the nature of invention* such that inventors now draw more heavily on academic science.

#### Evidence on Attorney-Driven Changes in Patent Citations to Scientific Papers

These results could also be driven, at least in part, by an “attorney-driven” change in citation practice, and, in fact, interpretation of the measured increase in the per-patent propensity to cite academic papers is clouded by the issue of the so-called “spike patents.”<sup>23</sup> In 1995, the effective period of monopoly granted to U.S. patent holders changed from 17 years after the grant date to 20 years from the filing date, in order to bring U.S. patent law more fully into compliance with the provisions of the TRIPs Agreement. This change took effect for patents filed after June 8, 1995. Patents filed prior to this deadline benefited from a “grandfather” provision – they were granted a monopoly of either 17 years from date of grant or 20 years from date of application, whichever was longer. Rejected patents re-filed after this deadline would also be subject to new evaluation criteria.

Applications submitted to the U.S. PTO more than doubled in May and June of 1995, and these applications, referred to as the “spike patents,” carried an unusually large number of citations to science. This surge in patenting seems to have been driven in part by a rush to file in order to benefit from the “grandfather” timing provision. The increase in citations to science seems to have been driven in part by a desire to avoid having to refile rejected patents under the new rules. Applicants thus erred on the side of caution by making far more than the usual number of citations to scientific material. Patents applied for in this period were issued gradually

---

<sup>22</sup> This pattern is quite robust to alternative aggregations of grant years into categories. Regression results demonstrating this are available from the author upon request.

<sup>23</sup> This issue is also discussed in the 2002 issue of *Science and Engineering Indicators* and in Hicks et. al. [2001].

over the next few years – dramatically increasing the average citations to science per patent in the overall data. Once the last of these applications was processed, average science citations per patent actually *fell*, as is illustrated in Figure VI. This kind of simple data tabulation might suggest that the connection between science and technology is weakening, after nearly a decade of rapid growth. That conclusion would be unwarranted, but it is likely that some of the movement in the aggregate data in the mid-to-late 1990s was “attorney-driven.”

Within the context of our empirical approach, one potential remedy for this problem is to remove the spike patents from my data set and re-run the citations function. The results are shown in Table II, and it can be seen here (and in all subsequent tables, where the spike patents have been removed), that the basic qualitative features of the previous empirical results remain. In particular, the finding of an increase in per-patent propensity to cite scientific papers is robust to the removal of these patents.

While the spike patents are the “attorney-driven” change that is easiest to identify in the data and to associate with a particular change in legislation, removal of the spike patents does not necessarily purge our data of changes in citations that reflect factors other than changes in knowledge spillovers. Over the course of our sample period, on-line databases have emerged that have made it steadily easier for inventors and their attorneys to identify the relevant non-patent prior art, including related scientific papers. One might be concerned that the measured increase in propensity to cite papers reflects declining search costs more than an increasing tendency to utilize academic science in commercial innovation.

There are two responses to this concern. First, the emergence of on-line databases has taken place across all scientific disciplines. However, the actual incidence of patent citations to science has remained remarkably concentrated in precisely those disciplines where qualitative evidence and practitioner accounts suggest that the actual knowledge spillovers have been the strongest and grown the most over time. Second, we provide some evidence in section V of a link between the propensity to cite science and inventive productivity at the firm level, measured

in various ways. If citations are purely defensive, and if the increase in citations purely reflects the decline in the cost of a citation, there is no reason to expect that firms which increase their citations will improve their research productivity. On the other hand, the interpretation that the increase in citations reflects an increase in spillovers would predict such an association.

A final note on “attorney-driven” changes in patent citations to science relates to the interesting recent work by Murray [2002] and Murray and Stern [2005] on so-called “patent-paper pairs.” The authors suggest that some academic scientists are publishing their work in the academic literature while *simultaneously* applying for patent protection for essentially the same material. To the extent that an increasingly large fraction of scientific output is destined to become patent-protected inventions, inventors may feel increasingly constrained to cite papers defensively, even if there are no direct knowledge spillovers.

But it is unclear how important the patent-paper pair phenomenon has been over the course of our sample period. First, open publication of a paper *prior* to the issue of the patent could undermine the ability of the author-inventor to secure intellectual property rights.<sup>24</sup> This creates a strategic incentive for inventors to defer publication until their patent rights are securely granted, in which case subsequent inventors could simply cite the relevant patent. Second, other studies show the fraction of publishing scientists who also patent is small, and that the patent counts of the patenting scientists are small relative to their publication counts (Thursby and Thursby, 2004; Azoulay et al., 2005). While interesting objects of study, patent-paper pairs may not be numerous enough to affect our results.

#### Evidence on Changes in Scientific Fertility

---

<sup>24</sup> U.S. patent law allows for a window of time during which inventors can seek protection for an invention whose essence has already been disclosed in, among other things, an academic publication. Nevertheless, this sort of prior disclosure increases the risks associated with patenting, particularly if there are disagreements with the Patent and Trademark Office about particular claims or patent scope. These concerns could be even more relevant for inventors seeking to obtain patent protection outside the United States – something that would be sought as a matter of course for potentially important inventions. It is worth pointing out that the patent-paper pair profiled in Murray [2002] was one in which the paper was published *after* the patent had been granted.

Another explanation of the increase in patent citations to papers would suggest that “academic” research has become steadily more practical, that is, relevant to inventors. In its simplest form, this change in the nature of academic science could show up as a tendency for more recent cohorts of academic papers to be more highly cited by patents than older cohorts of papers, when controlling for changes in the volume and distribution across fields of potentially cited papers and potentially citing patents. Our data allow us to test for precisely that tendency, by estimating a set of coefficients that correspond to “cohort effects” for the potentially cited papers. As we did with the patents, we group papers into categories corresponding to the year of publication: 1981-1984 (the base category), 1985-1988, 1989-1992, and 1993-1997. Our data source on papers ends in 1997, but there are very few patents granted by 1999 that cite papers with publication dates after 1997, so little is lost due to this data restriction. Recall that, because we control for the number of papers in a given cohort, the regression coefficient on the cohort effect measures average *per-paper* “citedness.”

The exact pattern revealed by the regression coefficients on these cohort effects varies a bit depending on the data used. In the full sample, measures of per-paper “citedness” increase, relative to the base period, in the late 1980s and early 1990s, peaking in the 1989-92 period. They then seem to decline somewhat in the most recent period, but estimated per-paper citedness remains higher than in the base period. This finding would seem to provide reasonably strong evidence for the “changes in scientific fertility” hypothesis. However, this finding is *not* robust to the exclusion of university patents from the sample. The latter point is illustrated in Table III, which presents results based on a sample that excludes both spike patents and patents assigned to universities and to other “public science institutions,” a category including research hospitals that often have links to universities. As can be seen, the apparent increase in per-paper citedness evaporates with this sample restriction. Instead, we see an apparently steady *decline* in per-paper citedness. If we narrowly define the scientific fertility hypothesis to mean an increase in the



average citedness of scientific papers over time, then there does not appear to be robust evidence for this in our data. If anything, citedness may be trending downward.

However, we need to emphasize that rejection of this *narrow* version of the hypothesis does *not* rule out the possibility that there has been a stream of scientific discoveries over time, at least in some academic disciplines, that inventors in industry have been able to build upon in their R&D activities. We will argue that there is evidence this has happened within the bio nexus. Instead, our results on scientific fertility suggest that the growth in patent citations over time has not been driven primarily by a broad-based change in the nature of academic science, *per se*, over time, but by a change over time in the degree to which private inventors use this science and a shift in the composition of invention across technology classes.

#### What Happens if We Exclude University and Research Hospital Patents?

Given the plausibly strong connection between university-based science and university-generated patents, it is obviously of interest to examine the robustness of our other empirical results to the exclusion of patents granted to “public science institutions” (a category which includes primarily universities and research hospitals) from the sample. Table III shows that, with the exception of results pertaining to scientific fertility, all of our main findings are robust to this sample restriction. In particular, the measured localization of spillovers within the bio nexus remains after dropping university patents. In fact, it becomes even more pronounced, as does the measured increase over time in per-patent propensity to cite science.

#### Summarizing the Lessons from the Full Sample

Once we exclude spike patents, it seems that trends in the data are best explained by a combination of the “changing composition” story and the “changing methods of invention” story.<sup>25</sup> At first glance, it would appear that the changing composition story is the more important

---

<sup>25</sup> The evidence in Tables I-III comes from a version of the citation function in which we constrain the obsolescence parameter to be the same across categories of technology. Following Jaffe and Trajtenberg [2002], we can allow this parameter to differ across patent technology categories. Results from such a

of that two. To see this, it is useful to examine Table IV, which presents results from a series of hypothesis tests. The data strongly reject the imposition of the constraint that methods of invention have not changed – or, more precisely, that per-patent propensities to cite science have not changed broadly across fields of technology. But the degradation in model fit generated by this constraint is fairly small.<sup>26</sup> On the other hand, imposing the constraint that the relative propensity of different patent classes to cite science is the same causes the adjusted R-squared to fall by about 67%, and imposing this constraint *and* the constraint that the relative citedness of different categories of science is the same causes the adjusted R-squared to fall by about 85%, relative to the unrestricted model. In other words, changes in the distribution of patenting across technologies and changes in the distribution of publications across fields would appear to explain much more of the total variance in patent citations to science than does average changes across fields in per-patent citation behavior over time.

But it is likely that a significant part of the substantial expansion in biotech patenting has been driven by increasing knowledge spillovers from university-based science. This is certainly the impression we have received from conversations with both scientists and corporate R&D managers in this field. According to these practitioners, methods of invention have changed dramatically *within* the bio nexus, and these changes have contributed to the relative expansion of patenting in this field. If this is the case, then the variance decomposition exercise presented in Table IV misses an important part of the story, because the bio nexus “field effect” is presumed to be time invariant. The developments described by practitioners suggest a change in the

---

regression are omitted for reasons of space. Allowing this parameter to vary does not change the qualitative patterns in the other results.

<sup>26</sup> The value of the Wald test parameter associated with this parameter restriction (see the second column, third row) easily exceeds the critical value of the Chi-Square distribution at the appropriate degrees of freedom. But the degradation in model fit generated by this constraint is relatively small. Relative to the unrestricted model, the adjusted R-squared of the restricted model declines by only about one percent. This can be inferred by comparing values in the third column – the adjusted R-squareds associated with the restricted models – with the adjusted R-squared of the restricted model given on the next to last row.

composition of invention *that is itself driven, in part, by field-specific changes in methods of invention.*

To explore this more complicated alternative hypothesis, we will look at changes in the patterns of patent citations to science within the bio nexus itself. We will find three pieces of evidence consistent with the practitioners' accounts. First, we will find evidence of a *substantial increase over time* in the propensity of bio nexus patents to cite bio nexus papers, even accounting for the strong growth in potentially cited patents and potentially citing papers. Second, we will note that total patenting in the bio nexus has grown much more rapidly than NSF estimates of total R&D investment – public and private – suggesting an increase in the productivity of research in the nexus as a whole. Finally, we will relate some preliminary findings from related research that suggests a positive association between patent citations to science and research productivity at the firm level.

#### **V. Are Knowledge Spillovers from Science Driving Patent Growth in the Bio Nexus?**

We first address this question by presenting evidence from a citations function using only data from the bio nexus.<sup>27</sup> The results are given in Table V. This table maintains the same aggregation scheme across patent classes as Tables I-III, but uses only data from the bio nexus in estimating the citations function. The “biomedical research” cluster of scientific fields is broken up into the “core fields” of biochemistry, biophysics, and molecular biology on the one hand and the remaining fields of biomedical research on the other. This sample excludes spike patents, but includes patents assigned to universities, research institutes, and research hospitals.

The qualitative results are similar to those estimated in the full sample. In particular, one finds, even within the bio nexus, statistically significant evidence of an increase in the per-patent propensity to cite science over time. In other words, even within this nexus, where citation

---

<sup>27</sup> We have also investigated what patent citations to science look like when we exclude the bio nexus from consideration. The data identify a secondary nexus clustered around the information technology disciplines. The results of this investigation are discussed in section VI.

activity has always been strongest and where the number of patents has been growing rapidly, the connection between industrial research and academic science seems to have grown substantially over time. The estimates on the grant year coefficients suggest that per-patent citation propensities had increased by more than 70% (relative to the base period) by 1997-99. In this particular sample of the data, the measured increase in per-paper citedness – our measure of changes in scientific fertility – remains roughly what it was in the overall sample including public science patents. That is, it suggests an increase, albeit non-monotonic, in scientific fertility over time. While this result does not survive the exclusion of public science patents, the result of an increase in per-patent citation propensity does.<sup>28</sup> Looking carefully within the bio nexus itself, one finds strong evidence of a change in the method of invention over our sample period.

The finding of an increase over time in the per-patent intensity to cite science is supported by a number of studies of the pharmaceutical and biotech industries. From its inception, the biotechnology industry has been closely linked to university-based science.<sup>29</sup> But over the course of the 1980s and 1990s, established pharmaceutical companies have increasingly drawn on recent scientific developments in their efforts to improve the efficiency of their drug discovery programs.<sup>30</sup> While the received literature has not yet tried to *quantify* the changes in this intensity of industrial borrowing from academic science over time, the basic trends in our data are reasonably consistent with the qualitative descriptions of changes over time in the existing literature.

Have these spillovers from academic science actually raised inventive productivity in the bio nexus? A casual examination of the aggregate evidence would suggest an affirmative answer. According to NSF data, total real R&D outlays from both public and private sources associated with the life sciences nearly doubled between 1985 and 1995. However, U.S. patenting in the bio

---

<sup>28</sup> Results from this robustness check are available from the authors upon request.

<sup>29</sup> See Kaplan and Murray [2003], Kenney [1986], Lichtenberg [forthcoming], and Gambardella [1995].

<sup>30</sup> See, among others, Cockburn, Henderson, and Stern [1999], who provide a useful qualitative description of these changes, then go on to document their implications for relative firm performance over time within the pharmaceutical industry.

nexus *more than tripled* over this period, which would seem to imply a considerable increase in R&D productivity for the nexus as a whole.<sup>31</sup>

Related research by Aoki and Branstetter [2004] has begun to explore more carefully the statistical association between the intensity of patent citations to science and research productivity at the firm level, using a formal theoretical model of the interaction between academic science and industrial research to guide empirical analysis. Results from this research suggest a positive, statistically significant correlation between increases in the intensity with which a particular firm cites science in its patents and measures of productivity and new product introductions.

We repeat two pieces of this evidence here. First, we can estimate a simple Cobb-Douglas production function in which the “science citation intensity” of firm patenting enters as an argument. The estimated equation is

$$(6) Q_{it} = \beta_0 + \beta_1 lkap_{it} + \beta_2 lemp_{it} + \beta_3 lrnd_{it} + \beta_4 lciting_{it-l} + \beta_5 lp_{it-l} + \alpha_i + \delta_t + \varepsilon_{it}$$

Where  $Q$  is output as measured by deflated sales of firm  $i$  in year  $t$ ,  $lkap$  is the log of the deflated capital stock, as measured by firm accounts,  $lemp$  is the log of the number of workers, and  $lrnd$  is the level of real expenditure on R&D.<sup>32</sup> The intensity of the firm’s use of academic science is measured by the log of the number of citations to academic science made by the cohort of patents applied for in year  $t$ ,  $lciting$ ; we also experiment with various lags of this measure. This measure includes patent citations made to *all* scientific papers, not just those generated by California-based research universities. Since there is a clear relationship between the number of citations and the number of patents generated by the firm in a year, we include the log of the number of patent applications (or various lags thereof)  $lp$ , as an additional control. We also allow for firm ( $\alpha_i$ ) and year ( $\delta_t$ ) fixed effects. Following a long tradition in the productivity literature, we

---

<sup>31</sup> On the other hand, studies by others have suggested that the patent to real R&D ratio has fallen substantially for large U.S.-based pharmaceutical companies – an important component of the nexus – over much of our sample period. See, among others, Henderson and Cockburn [1996].

<sup>32</sup> Alternative specifications estimated with an R&D stock produced results qualitatively similar to those reported in the paper. Our results are also robust to the inclusion of multiple lagged patent terms.

interpret the coefficients on the R&D and science intensity terms as measuring the marginal contribution of these measures to total factor productivity – output, adjusted for changes in standard tangible inputs.<sup>33</sup>

Columns (1) through (4) of Table VI illustrate the results we get from various lags of our science citation intensity measure. Given the inevitable lag between the generation of a patent and the implementation of patented technology into new products or processes on a sufficient scale to affect overall sales, it is important to demonstrate that the results are not purely contemporaneous.<sup>34</sup> The sample of firm-level data on which specification (6) is estimated, described in detail in Aoki and Branstetter [2004], includes firms across a full range of manufacturing industries, not just those linked to the bio nexus. The statistically significant coefficients imply that a 100% increase in science citations generates a 2%-3% increase in productivity. This would appear to be a modest effect, but the reader should recall that, in the aggregate economy, such citations have increased *13-fold* over our sample period. Given the various issues of measurement and causal inference, we would not want to overemphasize the power of this evidence, but it is clearly consistent with a positive relationship between productivity and science citations – and this relationship is not purely limited to firms in the bio nexus.<sup>35</sup>

The second piece of evidence for this relationship we take from Aoki and Branstetter [2004] uses a measure of inventive output that is only available for the “bio nexus”: data on new product introductions. The IMS LifeCycle Patent Focus Database connects FDA-approved drugs to the specific patents protecting the innovations incorporated into these drugs. By matching

---

<sup>33</sup> A more detailed description of these data sources is provided in the Appendix, included with this paper.

<sup>34</sup> While the coefficients of lagged values of the science intensity term are consistently positive and significant, this is not true for leading values of the science intensity term, providing an epsilon of reassurance on the issue of reverse causality. These results, omitted for reasons of space, are available upon request.

<sup>35</sup> Implausibly low estimates for the output elasticity of capital – which likely reflect measurement error in firm-level capital stocks – are only one of the problems that arise in this setting, hence our caution. See Griliches and Mairesse (1998).

information in this data base with our sample of patenting entities, we are able to identify the number of approved drugs that can be associated with the patents generated by a particular firm in a given year. Do increases in the science citation intensity of a firms' patents result in an increase in the number of approved drugs?<sup>36</sup>

The answer provided in column (6) of Table VI appears to be affirmative. In this column we present results based on the following specification:

$$(7) \quad \text{Approv}_{it} = \beta_0 + \beta_1 \text{lrnd}_{it} + \beta_2 \text{lciting}_{it} + \beta_3 \text{lp}_{it} + \alpha_i + \delta_t + \varepsilon_{it}$$

where the right hand side variables have the same definition as in specification (6) and the dependent variable counts the number of approved new products that can be linked to the patents generated by firm  $i$  in year  $t$ . Controlling for the size of the patent portfolio and the level of contemporaneous R&D investment, the coefficient on *lciting* measures the relationship between the “science citation intensity” of a patent cohort and its ability to yield successful products. Because the dependent variable is a count variable, we employ the fixed effects negative binomial estimator pioneered by Hausman, Hall, and Griliches [1984]. The estimated coefficient is roughly .7, implying that a 100% increase in science citations yields 70% more approved products. As is the case with the other results in Table VI, issues of measurement and causal inference necessitate caution in interpreting these results. Nevertheless, they are clearly consistent with a fairly strong positive link between science citations and the generation of inventions that can survive the formidable screening procedure erected by the U.S. FDA. We recently learned of related research by Markiewicz [2004], who has independently found a positive connection between citations to academic science and various measures of economic performance of drug and biotechnology companies, underscoring the point made here.

## VI. Evidence from the IT Nexus

---

<sup>36</sup> Lichtenberg [forthcomingb, 2003a, 2003b] has established a statistical association between the introduction of new drugs and measures of longevity, mortality reduction, and economic growth.

For every year of our sample period, roughly 70-90% of the total citations made by patents to scientific articles are made within the bio nexus. For this reason, we have concentrated much of our analysis on the patent-to-paper citation activity within this nexus. Before concluding the paper, however, we provide results of a citation function that deliberately *excludes* the bio nexus, in order to provide readers with a sense of how patent-to-paper citation activity is distributed across time and technological fields outside the bio nexus.<sup>37</sup> The pattern of knowledge diffusion from science to invention may be quite different outside the bio nexus; partitioning the data in this way allows us to quantify those differences.<sup>38</sup>

#### Evidence from the IT Nexus

Indeed, we find that the non-biotech subsample generates a significantly different pattern of results. The aggregate patent classes used are computers and communications (IT), chemistry, general electronics, mechanical inventions, and a catch-all “other” category. Science aggregates are engineering and technology, chemistry, physics, and a catch-all “other science” category. The other categories remain as before. Note that we are estimating roughly the same number of parameters for our non-biotech subsample as in the full sample, even though we have only a small fraction of the number of observations of patent-article citations. The relative thinness of the data here means that our parameter estimates need to be regarded with an extra measure of caution, even when they are statistically significant according to the conventional thresholds. Results are given in Table VII.

Patent-article citation activity outside the bio nexus is clearly concentrated in a secondary “IT” nexus. The IT patent classes cite science most frequently, displaying a propensity to cite

---

<sup>37</sup> The discipline of chemistry is somewhat unique in that it includes subdisciplines that are closely connected to the bio nexus and other subdisciplines that are completely unrelated. Given this dual nature, we include chemical patents and chemistry papers in both subsamples.

<sup>38</sup> A potential downside to this partition is that we lose “cross-nexus” citations, such as citations made by biotech patents to papers in mathematics and computer science. It is true that we observe an increase in such cross-nexus citations over time, probably reflecting the increasing importance of such fields as “bioinformatics,” but the aggregate numbers of these cross-nexus citations remain small, even in the most recent periods.



that is nearly 18 times as high as the base category of chemistry. General electronics patents are more than 7 times as likely to cite science, while mechanical patents are three times as likely. Articles in the physics fields are more than 44 times more likely to be cited than base category (chemistry) articles. The physics aggregate includes fields that relate to semiconductors and advanced materials. The engineering/technology aggregate (which includes computer science) is the next most highly cited, with a citedness per paper that is about 8 times greater than the base category. The rest of the sciences are significantly less likely to be cited. Incidentally, these results suggest that much of the citation activity involving chemical patents and chemistry papers comes from the bio nexus. Once chemistry is separated from those technologies and disciplines, it ceases to stand out in terms of patent-article citation activity.

In a striking contrast with earlier results, geographic localization seems to be much higher in this subsample. California-based inventors display a much higher likelihood of citing California science than the base (non-California U.S.) category of inventors. This pattern of results could very well reflect the increasing geographic concentration of the U.S. information technology industry in California. Another contrast with earlier results is a much higher propensity (relative to industrial firms) for patents generated by public science institutions to cite science. Public science institutions are nearly 39 times as likely to cite science as are firm patents, controlling for patent category. Patents held by “other institutions,” are less likely than firms to cite science in these fields, corresponding to the less significant role played by this category of assignee in non-biotech patenting.<sup>39</sup> The pattern of campus coefficients also highlights the unique role played by Stanford University within the sample. While, within the full sample and the bio nexus, a number of other institution’s “campus effects” were nearly as high, or even higher, than Stanford’s, in the non-bio subsample no other institution comes

---

<sup>39</sup> However, dropping “public science institution” patents from the sample does not qualitatively change the nature of the other results.

remotely close to Stanford's implied relative level of academic fertility. As with the estimates of geographic localization, it seems these data reflect the "Silicon Valley" phenomenon.

The patterns suggested by the coefficients on patent grant year cohorts and paper publication year cohorts also differ from those in previous regressions. Controlling for changes in the volume and distribution of publications and patents, all periods display a substantially greater per-patent propensity to cite science than the base period. Rather than the steadier increase one saw in the full sample, the pattern here looks more like a step function, with a sharp increase in the late 1980s. Although the increase relative to the base period is higher than in the bio nexus, one has to keep in mind that the absolute numbers of citations in this category remains *much* smaller than in the bio nexus. The increase in per patent propensity to cite science, combined with a sharp increase in patenting in the IT-related classes, explains most of the aggregate increase in citations to science outside the bio nexus.

While much of the recent qualitative literature on university-industry interaction has focused on the extensive borrowing from science taking place in the bio nexus, this activity is less well documented outside that nexus.<sup>40</sup> Nevertheless, the timing of the increase in per-patent propensity to cite science noted above is roughly coincident with two major changes in patenting – a substantial increase in patenting by semiconductor firms, especially the so-called "fabless" IC design firms, and a sharp increase in software patenting. The semiconductor industry has always had strong links to science [Hicks et al., 2001, Lim, 2003], but, as Hall and Ziedonis [2001] have showed, firms in this industry began sharply increasing their patenting in the 1980s and 1990s. Furthermore, new entrants (the so-called "fabless" design firms) emerged that were often closely linked to university engineering departments.

The increase in software patenting followed changes in U.S. patenting law and practice which expanded the ability of software inventors to patent, rather than copyright, their inventions

---

<sup>40</sup> Arora and Gambardella [1994] present many examples of what might be called a more "scientific" approach to industrial research outside the bio nexus, and stress the centrality of information technology in driving this shift.

[Bessen and Hunt, 2003]. There was little patent “prior art” to cite, so patents in this area have tended to cite more nonpatent prior art, including the relevant academic work in computer science and related fields. Software patents can be difficult to track, and the exacting timing of the measured increase depends on one’s definition of a software patent. Nevertheless, some observers suggest that there was a sharp increase in software patenting at the end of the 1980s.<sup>41</sup>

The final result to note from our exploration of citation activity outside the bio nexus is that recent cohorts of papers are not more likely to be cited. In fact, the per-paper citedness measures have sharply plummeted, even when one includes public science patents in the sample. One can also see that the estimated obsolescence coefficient is *substantially* higher than the overall sample, while the diffusion parameter is lower. These results need to be viewed together. On average, the gap between paper publication and patent citation is *much* shorter than it is in the bio nexus, such that very recent science is much more likely to get cited. Controlling for this short gap, however, there is no evidence that the most recent cohorts of papers generate more knowledge spillovers. In fact, the estimated decline in per-paper citedness is so sharp that the substantial increase in publications in these disciplines fails to make a positive contribution to total citations. In general, it seems that citations to science in these categories arrive more quickly, decay more rapidly, and peak at a lower level.

Framing these results in light of the alternative hypotheses stated in the introduction, it seems clear that the increase in citations outside the bio nexus has been driven almost entirely by composition effects – both in terms of fields of technology, fields of science, and institutions -- and “changing methods of invention.” In that sense, results here are broadly consistent with those discussed earlier. However, it must be stressed that citation activity in the secondary IT nexus identified in these data is substantially lower than that within the bio nexus – so much so that the IT nexus does not even show up in the full sample. The explosion of IT patenting in recent years

---

<sup>41</sup> See Bessen and Hunt [2003] who discuss the problems involved in measuring software patenting and provide alternative counts of software patents over time. Some of these series increase quite sharply in the late 1980s.

has been even more dramatic than that of bio nexus patenting, but the relative paucity of citations to science among these patents suggests that knowledge spillovers from academia have almost certainly *not* played the primary role in generating this patent explosion.

## VII. Conclusions and Extensions

What is driving the remarkable increase over the last decade in the propensity of patents to cite academic science? Does this trend indicate that stronger knowledge spillovers from academia have helped drive the surge in innovative activity in the U.S. in the 1990s? This paper has sought to shed light on these questions by using a common empirical framework to assess the relative importance of various alternative hypotheses in explaining the growth in patent citations to science. Our analysis supports the notion that the nature of U.S. inventive activity has changed over the sample period, with an increased emphasis on the use of the knowledge generated by university-based scientists in later years. That being said, knowledge flows from academia to industry, as they are measured in this paper, appear to have been overwhelmingly concentrated in the bio nexus throughout the sample period. While scientific breakthroughs generated by academic researchers, particularly in the life sciences, have generated new “technological opportunities,” these new opportunities are evidently limited in scope.

In our introduction, we laid out four alternative hypotheses that could possibly explain the sharp increase over time in the number of patent citations to science: the “increasing scientific fertility hypothesis,” the “attorney-driven” hypothesis, the “changing method of invention hypothesis,” and the “changing composition of invention” hypothesis. It is clear from Section IV that increased patenting in the bio nexus is, in a mechanical sense, the single most important driver of the growth in patent citations over time. This would seem to lead one to the conclusion that changing composition of invention is the hypothesis validated by the data.

But this begs the question of what has driven the expansion of patenting in the bio nexus. Conversations with practitioners suggest that part of the increase in bio nexus patenting has itself been generated by an increase in knowledge spillovers from academic science. Three pieces of

evidence support this view. First, within the bio nexus, there has been a pronounced increase over time in the propensity of patents to cite scientific articles, even controlling for the enormous increase in potentially citing patents and potentially cited articles. Second, the aggregate statistics on R&D spending and patenting suggest a rise in inventive productivity for the nexus as a whole. Third, we cite evidence at the firm level suggesting a positive association between inventive productivity and patent citations to science – one that is particularly strong in the bio nexus. This latter evidence is interesting in its own right, and is the subject of ongoing research.

## Bibliography

- Adams, James, 1990, "Fundamental Stocks of Knowledge and Productivity Growth," *Journal of Political Economy* 98: 673-702.
- Adams, James and Zvi Griliches, 1996, "Research Productivity in a System of Universities," NBER working paper no. 5833.
- Agrawal, Ajay, 2001, "Research on University-to-Industry Knowledge Transfer: Framework of Existing Literature and Future Questions," *International Journal of Management Reviews*.
- Agrawal, Ajay and Rebecca Henderson, 2002, "Putting Patents in Context: Exploring Knowledge Transfer from MIT," *Management Science*, Vol. 48, No. 1.
- Aoki, Reiko, and Lee Branstetter, 2004, "Is Academic Science Raising Innovative Productivity? Theory and Evidence from Firm-Level Data," working paper, Columbia Business School.
- Arora, Ashish and Alfonso Gambardella, 1990, "Complementarity and External Linkages: the Strategies of the Large Firms in Biotechnology," *Journal of Industrial Economics*, Vol. 38, pp. 361-379.
- Arora, Ashish and Alfonso Gambardella, 1994, "The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labor," *Research Policy*, Vol. 23, pp. 523-532.
- Audretsch, David and Paula Stephan, 1996, "Company-Scientist Locational Links: The Case of Biotechnology," *American Economic Review*, Vol. 86, No. 3.
- Azoulay, Pierre, Waverly Ding, and Toby Stuart, 2004, "The Impact of Academic Patenting on Public Research Output," working paper.
- Barnes, Michael, David Mowery, Arvids Ziedonis, 1998, "The Geographic Reach of Market and Nonmarket Channels of Technology Transfer: Comparing Citations and Licenses of University Patents," working paper.
- Bessen, James and Robert Hunt, 2003, "An Empirical Look at Software Patents," unpublished working paper.
- Branstetter, Lee, 2004, "Exploring the Link Between Academic Science and Industrial Innovation," unpublished working paper.
- Cockburn, Iain and Rebecca Henderson, 2000, "Publicly Funded Science and the Productivity of the Pharmaceutical Industry," paper prepared for the NBER Conference on Science and Public Policy.
- Cockburn, Iain, Rebecca Henderson, and Scott Stern, 1999, "The Diffusion of Science Driven Drug Discovery: Organizational Change in Pharmaceutical Research," NBER Working Paper No. 7359.
- Cockburn, Iain and Rebecca Henderson, 1998, "The Organization of Research in Drug

- Discovery,” *Journal of Industrial Economics*, Vol XLVI, No. 2.
- Cohen, Wesley, Richard Florida, Louis Randazzese, and John 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
- Cohen, Wesley and Daniel Levinthal, 1989, “Innovation and Learning: The Two Faces of R&D,” *Economic Journal*, no. 99, pp. 569-596.
- Evenson, Robert and Yoav Kislev, 1976, “A Stochastic Model of Applied Research,” *Journal of Political Economy*, Vol. 84, pp. 265-282.
- Faulkner, Wendy and Jennifer Senker, 1995, *Knowledge Frontiers: Public Sector Research and Industrial Innovation in Biotechnology, Engineering Ceramics, and Parallel Computing*, Oxford: Clarendon Press.
- Fleming, Lee and Olav Sorenson, 2001, “Science as a Map in Technological Search,” working paper.
- Gordon, Robert J., 2002, “Technology and Economic Performance in the American Economy,” NBER Working Paper no. 8771.
- Griliches, Zvi and Jacques Mairesse, 1998, “Production Functions : The Search for Identification,” in *Practicing Economics: Essays in Method and Application*, Cheltenham, Edward Elgar, London.
- Hausman, Jerry, Bronwyn Hall, and Zvi Griliches, 1984, “Econometric Models for Count Data with an Application to the Patents-R&D Relationship,” *Econometrica* 52, pp. 909-938.
- Henderson, Rebecca and Iain Cockburn, 1996, “Scale, Scope, and Spillovers: the Determinants of Research Productivity in Drug Discovery,” *Rand Journal of Economics* 27, pp. 32-59.
- Henderson, Rebecca, Adam B. Jaffe, and Manuel 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, Diana, Tony Breitzman, Dominic Olivastro, and Kim Hamilton, 2001, “The Changing Composition of Innovative Activity in the US – A Portrait Based on Patent Analysis,” *Research Policy* 30, pp. 681-703.
- Jaffe, Adam, 1989, “The Real Effects of Academic Research,” *American Economic Review*, 79 [5], pp. 957-70
- Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, 1993, “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *Quarterly Journal of Economics*, Vol. CVIII, No. 3.
- Jaffe, Adam and Manuel Trajtenberg, 2002, *Patents, Citations, and Innovations: A Window on the Knowledge Economy*, MIT Press.
- Jaffe, Adam and Manuel Trajtenberg, 1996, “Flows of Knowledge from Universities and Federal

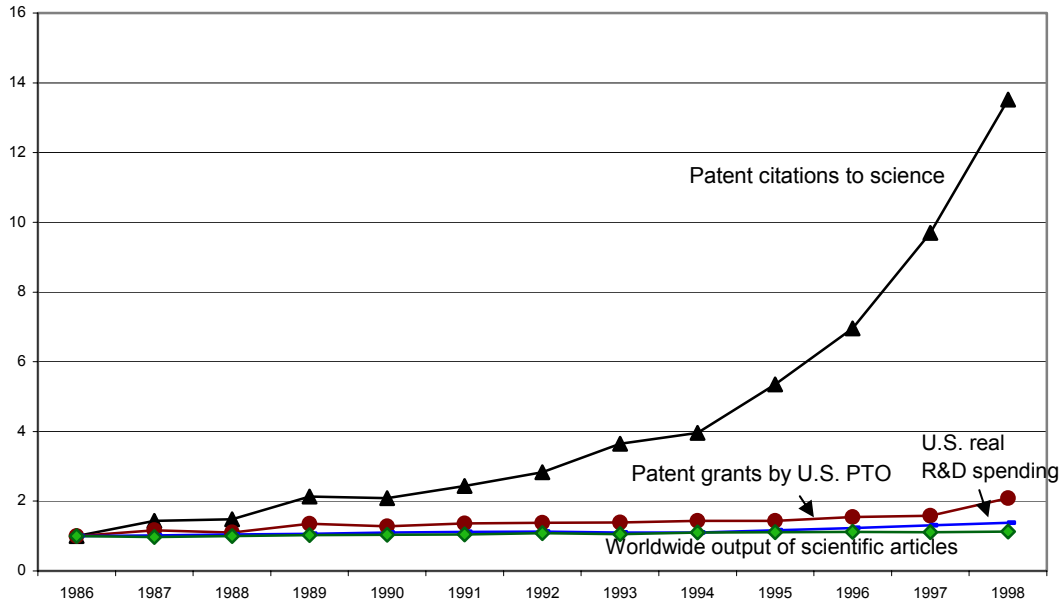
- Labs: Modeling the Flow of Patent Citations over Time and across Institutional and Geographic Boundaries,” NBER working paper no. 5712.
- Jaffe, Adam, Michael Fogarty, and Bruce Banks, 1998, “Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation,” *Journal of Industrial Economics*, Vol. XLVI, No. 2.
- Jensen, Richard and Marie Thursby, 1999, “Proofs and Prototypes for Sale: The Licensing of University Inventions,” *American Economic Review*, 91, pp. 240-259.
- Kenney, Martin, 1986, *Biotechnology: The University-Industrial Complex*, New Haven: Yale University Press.
- Kim, Jinyoung, Sangjoon John Lee, and Gerald Marschke, 2005, “The Influence of University Research on Industrial Innovation,” NBER working paper no. 11447.
- Kortum, Samuel, 1997, “Research, Patenting, and Technological Change,” *Econometrica*, Vol. 65, pp. 1389-1419.
- Kortum, Samuel and Joshua 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.
- Kortum, Samuel and Joshua Lerner, 2000, “Assessing the Contribution of Venture Capital to Innovation,” *Rand Journal of Economics*, 31, pp. 674-692.
- Kortum, Samuel and Joshua Lerner, 2003, “Unraveling the Patent Paradox,” unpublished working paper.
- Lach, Saul and Mark Schankerman, 2003, “Incentives and Invention in Universities,” unpublished working paper.
- Lichtenberg, Frank, forthcominga, “Public Policy and Innovation in the U.S. Pharmaceutical Industry,” in *Public Policy and Entrepreneurship*, by Douglas Holtz-Eakin, ed., MIT Press.
- Lichtenberg, Frank, forthcomingb, “Pharmaceutical Knowledge-Capital Accumulation and Longevity,” in *Measuring Capital in the New Economy*, by Carol Corrado, John Haltiwanger, and Dan Sichel, eds., University of Chicago Press.
- Lichtenberg, Frank, 2003a, “The Effect of New Drugs on HIV Mortality in the U.S., 1987-1998,” *Economics and Human Biology* 1 (2003), pp. 259-266.
- Lichtenberg, Frank, 2003b, “Pharmaceutical Innovation, Mortality Reduction, and Economic Growth,” in *Measuring the Gains from Medical Research: An Economic Approach*, ed., by Kevin Murphy and Robert Topel, University of Chicago Press.
- Lieberman, Marvin, 1978, “A Literature Citation Study of Science-Technology Coupling in Electronics,” *Proceedings of the IEEE*, 36 [1], pp. 5-13.
- Lim, Kwanghui, 2001, “The Relationship between Research and Innovation in the



- Semiconductor and Pharmaceutical Industries,” working paper.
- Markiewicz, Kira, 2004, “Absorptive Capacity and Innovation: Evidence from Pharmaceutical and Biotechnology Firms,” working paper, UC-Berkeley.
- Mansfield, Edwin, 1995, “Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing,” *The Review of Economics and Statistics* 77: 55-65.
- Mowery, David, Richard Nelson, Bhavin Sampat, and Arivds 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
- Murray, Fiona, 2002, “Innovation as Co-evolution of Scientific and Technological Networks: Exploring Tissue Engineering,” *Research Policy* 31: 1289-1403.
- Murray, Fiona and Sarah Kaplan, 2001, “The Role of Social Evidence in Shaping a Technological Discontinuity: The Evolution of Biotechnology,” unpublished working paper.
- Murray, Fiona, and Scott Stern, 2005, “Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis,” NBER Working Paper no. 11465.
- Narin, Francis, Kimberly Hamilton, and Dominic Olivastro, 1997, “The Increasing Linkage Between U.S. Technology and Public Science,” *Research Policy* 197: 101-121.
- National Science Foundation, 2002, *National Science and Engineering Indicators*.
- Office of Technology Transfer, University of California, 1997, Annual Report: University of California Technology Transfer Program. Oakland, CA: University of California.
- Price, Derek, 1965, “Is Technology Historically Independent of Science? A Study in Statistical Historiography,” *Technology and Culture*, 6, pp. 553-568.
- Rosenberg, Nathan and Richard Nelson, 1994, “American Universities and Technical Advance in Industry,” *Research Policy*, 23, pp. 323-348.
- Sampat, Bhaven, 2005, “Determinants of Patent Quality: An Empirical Analysis,” working paper.
- Shane, Scott, 2000, “Prior Knowledge and the Discovery of Entrepreneurial Opportunities,” *Organization Science*, 11, pp. 448-469.
- Shane, Scott, 2001, “Technological Opportunities and New Firm Creation,” *Management Science*, 47, pp. 205-220.
- Sorenson, Olav and Lee Fleming, 2001, “Science and the Diffusion of Knowledge,” working paper.

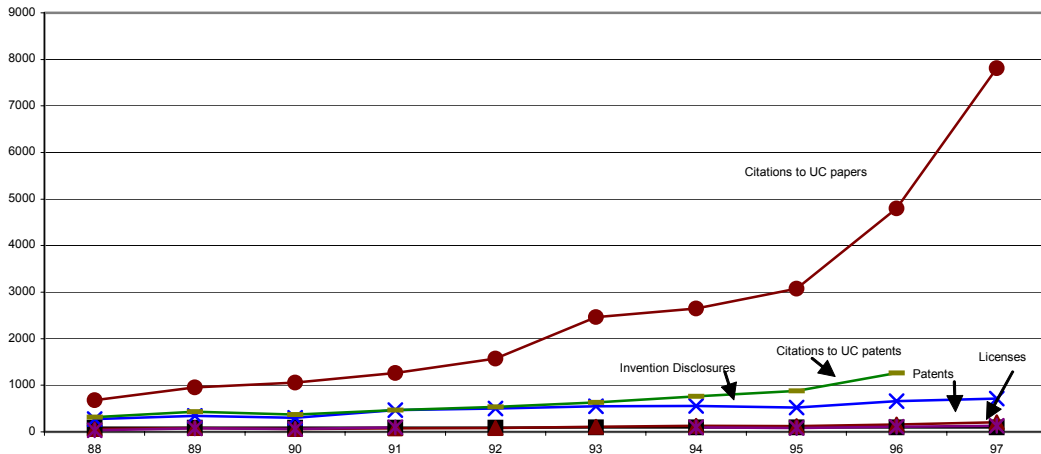
- Stephan, Paula, 1996, "The Economics of Science," *Journal of Economic Literature* 34: 1199-1235.
- Stern, Scott, 1999, "Do Scientists Pay to Be Scientists?" NBER Working Paper No. 7410.
- Thompson, Peter and Melanie Fox-Kean, 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review* 95 (1), pp. 450-460.
- Thursby, Jerry and Marie Thursby, 2002, "Who is Selling the Ivory Tower? Sources of Growth in University Licensing," *Management Science*, 48, pp. 90-104.
- Thursby, Jerry and Marie Thursby, 2004, "Patterns of Research and Licensing Activity of Science and Engineering Faculty," working paper.
- Zucker, Lynne, Michael Darby, and Michael Brewer, 1998, "Intellectual Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88: 290-306.

**Figure I Patent Citations to Academic Science**  
 Series are scaled so that 1986 values are equal to 1



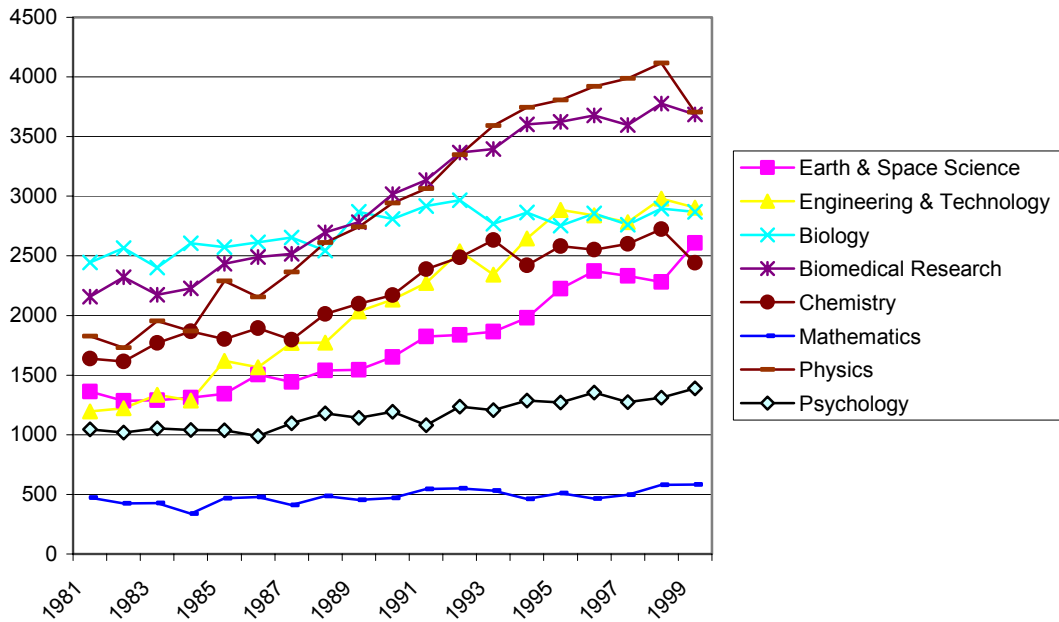
Source: National Science and Engineering Indicators, 2002

**Figure II Citations to UC papers vs other indicators**



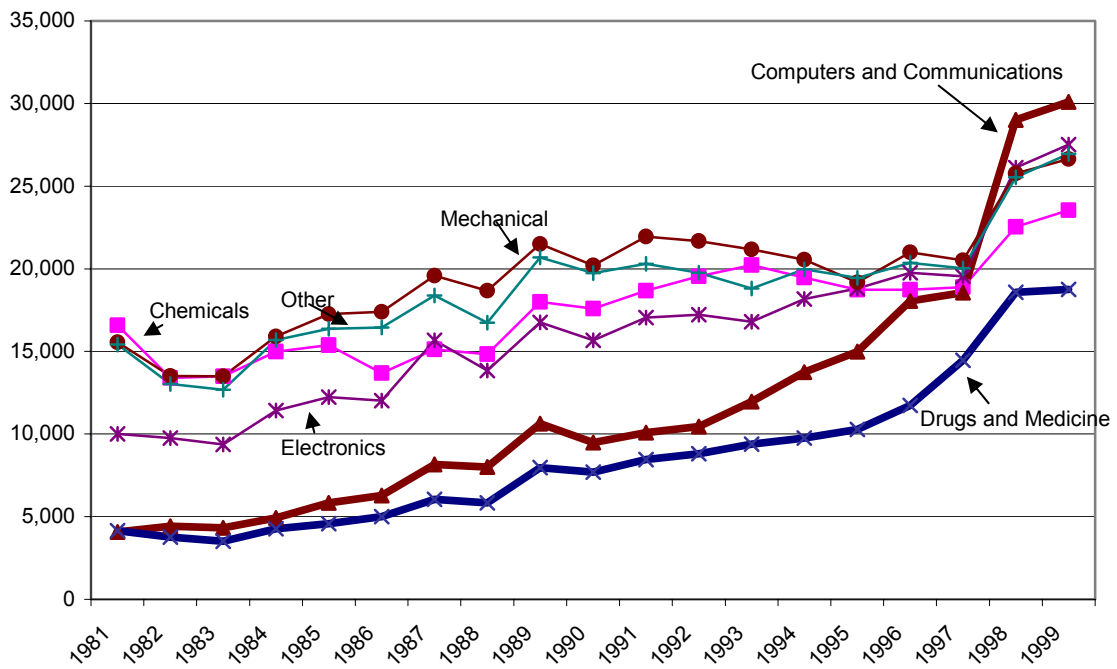
Source: Authors' calculations based on data from the University of California Technology Transfer Office annual report, AUTMN, the NBER Patent Citation Data Base, and CHI Research.

**Figure III**  
**Growth in California Academic Publishing, Excluding Clinical Medicine**



Source: University Science Indicators, 1999

**Figure IV Patent Grants by Technology Category, 1981-1999**



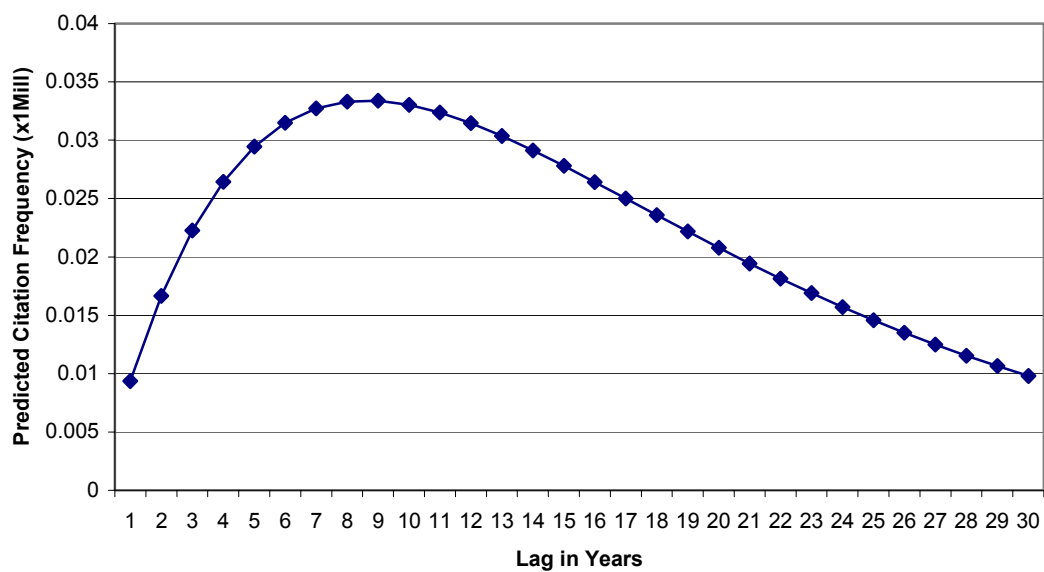
Source: NBER Patent Citation Database

**Table I Citation Function Results, Full Sample**

<b>Variable</b>	<b>Coefficient</b>	<b>T-statistic for H<sub>0</sub>: Parameter=1</b>
<b>Computers and Communications</b>	0.04	-126.82
<b>Drugs/medicine</b>	2.39	75.67
<b>Electronics</b>	0.05	-132.54
<b>Mechanical</b>	0.01	-110.78
<b>Other</b>	0.05	-100.33
<b>Biomedical research</b>	40.87	8.37
<b>Chemistry</b>	4.83	6.68
<b>Clinical Medicine</b>	5.36	6.92
<b>Eng/Technology</b>	0.25	-6.55
<b>Other Science</b>	0.37	-5.76
<b>Physics</b>	0.50	-4.32
<b>Caltech</b>	1.20	20.61
<b>Berkeley</b>	0.57	-67.43
<b>Davis</b>	0.42	-93.97
<b>Irvine</b>	0.44	-78.03
<b>Los Angeles</b>	0.39	-106.46
<b>Riverside</b>	0.26	-91.93
<b>Santa Barbara</b>	0.28	-74.35
<b>Santa Cruz</b>	0.26	-71.51
<b>San Diego</b>	1.02	2.89
<b>Santa Francisco</b>	0.85	-21.82
<b>USC</b>	0.54	-61.09
<b>US-CA</b>	2.67	105.69
<b>Other Institutions</b>	1.75	38.09
<b>Public Science</b>	3.68	99.18
<b>Grant year 88-90</b>	1.04	1.13
<b>Grant year 91-93</b>	1.03	0.96
<b>Grant year 94-96</b>	1.36	8.55
<b>Grant year 97-99</b>	2.10	15.56
<b>Paper pub year 85-88</b>	1.29	21.7
<b>Paper pub year 89-92</b>	1.39	17.4
<b>Paper pub year 93-97</b>	1.10	4.08
$\beta_1$ (obsolescence)	0.12	62.66
$\beta_2$ (diffusion)	1.07E-08	8.33
<b>Adjusted R-squared</b>	0.224	
<b>Number of observations</b>	556416	

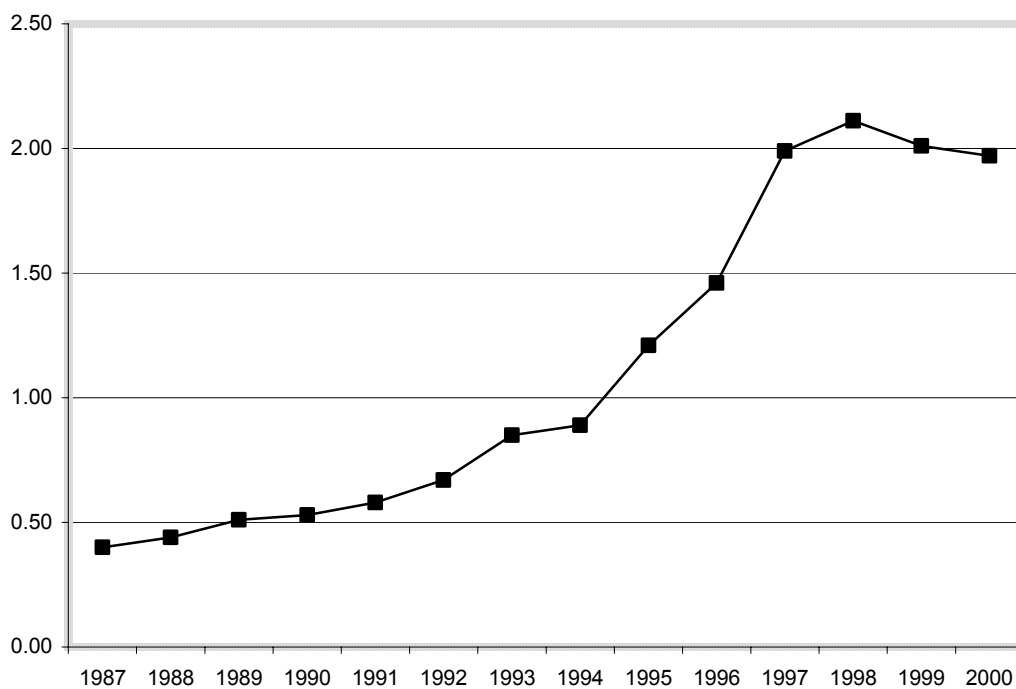
Base categories: Patent technology category=chemicals, scientific field=biology, academic institute=Stanford, patent assignee location=U.S./non-California, patent assignee type=firm, patent grant year=[1983, 1987], paper publication year=[1981, 1984].

Figure V Fitted Citation Frequency (Base Category)



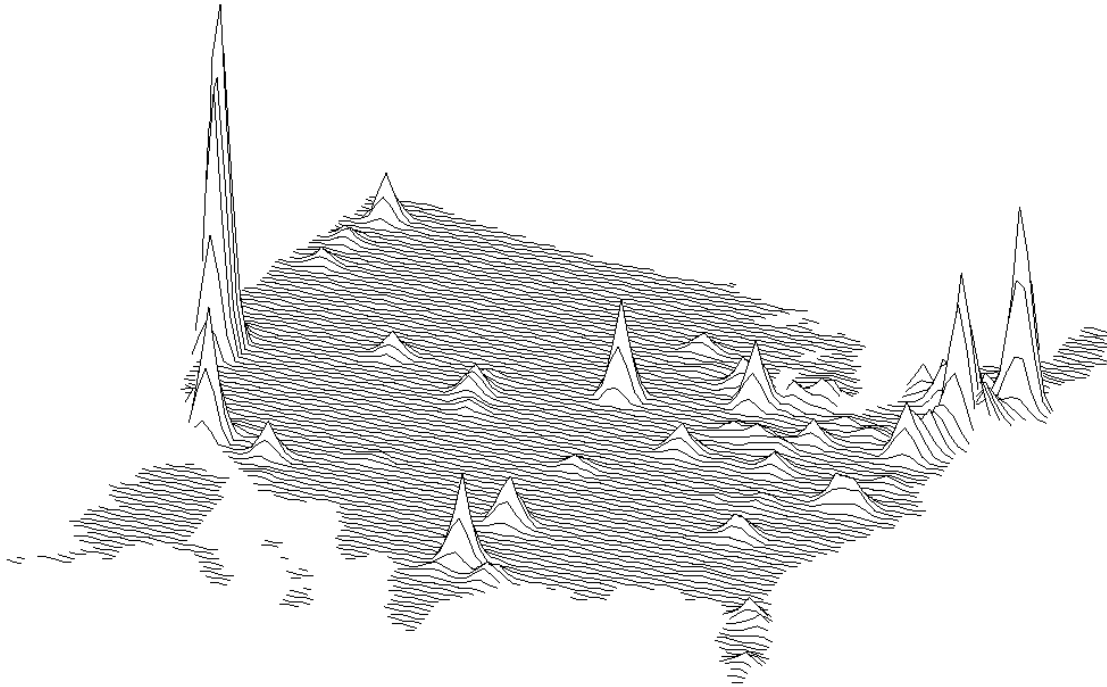
Source: Authors' calculations.

Figure VI Average Science Citations Per Patent



Source: National Science and Engineering Indicators, 2002, National Science Foundation

**Figure VII Citations to UC Berkeley Papers, US**



**Source: Authors' calculations based on NBER Patent Citation Database**

**Table II Citation Function Results, Excluding “Spike Patents”**

Variable	Coefficien	T-statistic for H <sub>0</sub> : Parameter=1
<b>Computers and Communications</b>	0.04	-126
<b>Drugs/medicine</b>	2.24	70.87
<b>Electronics</b>	0.06	-133.03
<b>Mechanical</b>	0.01	-110.84
<b>Other</b>	0.05	-100.01
<b>Biomedical research</b>	40.44	7.82
<b>Chemistry</b>	5.69	6.52
<b>Clinical Medicine</b>	5.23	6.43
<b>Eng/Technology</b>	0.29	-5.78
<b>Other Science</b>	0.38	-5.26
<b>Physics</b>	0.62	-2.88
<b>Caltech</b>	1.16	16.03
<b>Berkeley</b>	0.59	-59.86
<b>Davis</b>	0.35	-98.53
<b>Irvine</b>	0.47	-68.19
<b>Los Angeles</b>	0.39	-98.7
<b>Riverside</b>	0.28	-82
<b>Santa Barbara</b>	0.29	-68.85
<b>Santa Cruz</b>	0.22	-70.14
<b>San Diego</b>	1.04	5.17
<b>San Francisco</b>	0.85	-20.88
<b>USC</b>	0.53	-57.49
<b>US-CA</b>	2.68	98.85
<b>Other Institutions</b>	1.76	31.67
<b>Public Science</b>	4.30	87.8
<b>Grant year 88-90</b>	1.04	1.43
<b>Grant year 91-93</b>	1.07	2.15
<b>Grant year 94-96</b>	1.43	9.59
<b>Grant year 97-99</b>	1.98	14.38
<b>Paper pub year 85-88</b>	1.30	20.51
<b>Paper pub year 89-92</b>	1.37	15.55
<b>Paper pub year 93-97</b>	1.17	5.77
$\beta_1$ (obsolescence)	0.12	61.21
$\beta_2$ (diffusion)	1.01E-08	7.81
<b>Adjusted R-squared</b>	0.199	
<b>Number of observations</b>	556416	

Base categories: Patent technology category=chemicals, scientific field=biology, academic institute=Stanford, patent assignee location=U.S./non-California, patent assignee type=firm, patent grant year=[1983, 1987], paper publication year=[1981, 1984].



**Table III Citation Function Results, Excluding “Spike Patents” and University Patents**

	<b>Coefficient</b>	<b>T-statistic for H<sub>0</sub>: Parameter=1</b>
<b>Computers and Communications</b>	0.039	-52.65
<b>Drugs/medicine</b>	4.331	36.21
<b>Electronics</b>	0.072	-48.02
<b>Mechanical</b>	0.025	-44.74
<b>Other</b>	0.015	-47.16
<b>Biomedical research</b>	87.037	2.23
<b>Chemistry</b>	2.576	1.31
<b>Clinical Medicine</b>	9.232	2.01
<b>Engineering and Technology</b>	0.414	-1.38
<b>Other Science</b>	0.531	-1.11
<b>Physics</b>	0.709	-0.62
<b>Caltech</b>	1.211	15.03
<b>Berkeley</b>	0.394	-69.92
<b>Davis</b>	0.191	-95.16
<b>Irvine</b>	0.236	-73.62
<b>Los Angeles</b>	0.328	-81.43
<b>Santa Barbara</b>	0.159	-72.75
<b>Riverside</b>	0.640	-24.43
<b>Santa Cruz</b>	0.178	-54.48
<b>San Diego</b>	0.528	-51.70
<b>San Francisco</b>	0.505	-59.54
<b>USC</b>	0.589	-36.50
<b>US-CA</b>	2.721	55.94
<b>Other Institutions</b>	2.278	62.61
<b>Grant year 88-90</b>	1.210	5.28
<b>Grant year 91-93</b>	1.018	0.47
<b>Grant year 94-96</b>	1.232	4.59
<b>Grant year 97-99</b>	2.158	11.34
<b>Paper pub year 85-88</b>	0.717	-23.82
<b>Paper pub year 89-92</b>	0.520	-32.13
<b>Paper pub year 93-97</b>	0.407	-34.20
$\beta_1$ (obsolescence)	0.123	2.25
$\beta_2$ (diffusion)	5.26E-09	-52.65
<b>Adjusted R-squared</b>	0.119	
<b>Number of observations</b>	370944	

Base categories: Patent technology category=chemicals, scientific field=biology, academic institute=Stanford, patent assignee location=U.S./non-California, patent assignee type=firm, patent grant year=[1983, 1987], paper publication year=[1981, 1984]

**Table IV Wald Tests of Restrictions**

Hypotheses:

- (1)  $H_0$ : All coefficients of patent technology categories are the same.
- (2)  $H_0$ : All coefficients of paper fields are the same.
- (3)  $H_0$ : All coefficients of patent grant years are the same.
- (4)  $H_0$ : All coefficients of paper publication years are the same.

<b>Hypothesis</b>	<i>Test results</i>	
	<i>Chi-Sq.</i> <i>(p-value)</i>	<i>Adj. R<sup>2</sup> (rest.)</i>
(1)	63273.3 (0.000)	0.083
(2)	448.7 (0.000)	0.052
(1) and (2)	63721.9 (0.000)	0.029
(3)	2923.8 (0.000)	0.196
(4)	29429.4 (0.000)	0.197
<i>Adj. R<sup>2</sup> (unrest).</i>		0.199
<i># of obs.</i>		556416

**Table V Bio Nexus Results**

<b>Variable</b>	<b>Coefficient</b>	<b>T-statistic for H<sub>0</sub>: Parameter=1</b>
<b>Drugs &amp; Medical</b>	2.326	30.16
<b>Chemistry</b>	0.196	-56.39
<b>Clinical Medicine</b>	0.184	-80.2
<b>Other Biotech</b>	2.321	34.8
<b>Caltech</b>	0.911	-4.69
<b>Berkeley</b>	0.478	-37.35
<b>Davis</b>	0.297	-51.5
<b>Irvine</b>	0.367	-38.88
<b>Los Angeles</b>	0.364	-47.29
<b>Riverside</b>	0.221	-42.09
<b>Santa Barbara</b>	0.260	-33.1
<b>Santa Cruz</b>	0.243	-30.29
<b>San Diego</b>	0.985	-0.82
<b>San Francisco</b>	0.782	-14.3
<b>USC</b>	0.498	-28.45
<b>US-CA</b>	2.548	41.57
<b>Other Institutions</b>	1.614	11.49
<b>Public Science</b>	4.174	37.72
<b>Grant year 88-90</b>	1.040	0.58
<b>Grant year 91-93</b>	1.043	0.62
<b>Grant year 94-96</b>	1.377	3.88
<b>Grant year 97-99</b>	1.729	5.38
<b>Paper pub year 85-88</b>	1.296	8.81
<b>Paper pub year 89-92</b>	1.307	5.76
<b>Paper pub year 93-97</b>	1.105	1.63
$\beta_1$ (obsolescence)	0.115	24.2
$\beta_2$ (diffusion)	3.28E-07	15.03
<b>Adjusted R-squared</b>	0.189	
<b>Number of observations</b>	105984	

Base categories: Patent technology category=chemicals, scientific field=biochemistry, biophysics, and molecular biology, academic institute=Stanford, patent assignee location=U.S./non-California, patent assignee type=firm, patent grant year=[1983, 1987], paper publication year=[1981, 1984]

**Table VI The Link Between Science Citations and Inventive Productivity**

Variable	(1)	(2)	(3)	(4)	(5)
Outcome Measure	<b>Real Sales</b>	<b>Real Sales</b>	<b>Real Sales</b>	<b>Real Sales</b>	<b>New product introductions</b>
Science Citations	<b>.0302</b> <b>(.0045)</b>				<b>.6961</b> <b>(.0240)</b>
Science Citations t-1		<b>.0264</b> <b>(.0048)</b>			
Science Citations t-2			<b>.0238</b> <b>(.0051)</b>		
Science Citations t-3				<b>.0276</b> <b>(.0056)</b>	
R&D	.1657 (.0085)	.1730 (.0089)	.1846 (.0095)	.1977 (.0101)	.0865 (.0183)
Patenting	-.0049 (.0062)	-.0002 (.0065)	.0043 (.0068)	.0058 (.0072)	.4080 (.0328)
Capital	.0716 (.0060)	.0734 (.0062)	.0768 (.0065)	.0806 (.0068)	
Labor	.8387 (.0127)	.8257 (.0134)	.8049 (.0142)	.7872 (.0151)	
Observations	14,495	13,799	13,086	12,352	2,483
R-squared	.9743	.9747	.9753	.9758	
Log Likelihood					-4788

Standard errors are listed in parentheses. Columns (1)-(4) provide results of regression specification (6) run on a firm-level panel data set of 1,279 publicly traded companies based in the United States, Western Europe, and Japan. The science citation coefficient effectively provides the elasticity of TFP with respect to changes in the level of citations to academic science, controlling for the size of the cohort of patent applications. The data set measures all citations made by the patents of these companies to all scientific papers appearing in the peer-reviewed journals indexed by the Science Citation Index, not just those generated by California-based research universities. Patents are dated according to the year of application rather than the year of grant. Column (5) provides the results of regression specification (7) run on a firm-level panel data set of 273 publicly traded companies in the drug and biotechnology industries and related sectors. The sample includes most major manufacturers based in Western Europe and Japan.

**Table VII Results from the “IT Nexus”**

<b>Variable</b>	<b>Coefficient</b>	<b>T-statistic for H<sub>0</sub>: Parameter=1</b>
<b>Computers &amp; Communications</b>	17.757	7.16
<b>Electronics</b>	7.175	6.49
<b>Mechanical</b>	4.568	5.75
<b>Other</b>	0.205	-4.56
<b>Eng/Technology</b>	9.467	2.5
<b>Other Science</b>	0.527	-1.22
<b>Physics</b>	45.498	2.75
<b>Caltech</b>	0.140	-111.24
<b>Berkeley</b>	0.087	-116.36
<b>Davis</b>	0.457	-44.23
<b>Irvine</b>	0.011	-88.11
<b>Los Angeles</b>	0.044	-109.79
<b>Riverside</b>	0.013	-71.51
<b>Santa Barbara</b>	0.038	-111.09
<b>Santa Cruz</b>	0.005	-67.68
<b>San Diego</b>	0.024	-101.04
<b>San Francisco</b>	0.002	-116.04
<b>USC</b>	0.065	-94.57
<b>US-CA</b>	21.964	8.08
<b>Other Institutions</b>	0.436	-1.56
<b>Public Science</b>	38.980	6.94
<b>Grant year 88-90</b>	16.606	17.27
<b>Grant year 91-93</b>	7.166	11.97
<b>Grant year 94-96</b>	16.022	11.07
<b>Grant year 97-99</b>	16.255	7.16
<b>Paper pub year 85-88</b>	0.207	-153.46
<b>Paper pub year 89-92</b>	0.079	-173.74
<b>Paper pub year 93-97</b>	0.022	-320.19
$\beta_1$ (obsolescence)	0.544	88.66
$\beta_2$ (diffusion)	5.41E-11	2.36
<b>Adjusted R-squared</b>	0.074	
<b>Number of observations</b>	264960	

Base categories: Patent technology category=chemicals, scientific field=chemistry, academic institute=Stanford, patent assignee location=U.S./non-California, patent assignee type=firm, patent grant year=[1983, 1987], paper publication year=[1981, 1984]

## **Data Appendix**

This appendix provides additional details on the construction of the data sets used in the paper.

### **Construction of the Data Set Used in Tables I-IV**

From the University Science Indicators database generated by the Institute for Scientific Information and purchased by the authors, we have obtained comprehensive data on the publication of scientific articles by our sample of California research universities, by institution, year (of publication), and scientific field, from 1981-1997. The raw data are counts of publications, and they are assigned to a large number of narrowly defined scientific fields.

These data are matched to data on patent citations made to these publications over the 1983-1999 (grant year) period, which were provided by CHI Research. CHI Research, which has since been acquired by the consulting firm ipIQ, developed a comprehensive data base of “non-patent references” made in U.S. patent documents. These references include citations to scientific journals, industrial standards, technical disclosures, engineering manuals, etc. The focus on this paper is on the subset of references made to articles appearing in peer-reviewed scientific journals. In the CHI Research database, references to scientific journals are put into a standardized format, and these data can then be matched to data on papers published in the more than 4,000 journals covered by the Science Citation Index (SCI).<sup>42</sup> Through this matching process, we obtain data on patent citations to peer-reviewed scientific articles generated by California research universities.

The raw data thus obtained consisted of a patent-paper match – an observation that linked a specific U.S. patent grant with a specific paper for which at least one of the authors was affiliated with a California-based research university at the time of publication. The papers were

---

<sup>42</sup> For a more detailed description of the database developed by CHI Research, see Narin et. al. [1997]. Further details are also available from the authors upon request.

assigned to a set of scientific field categories developed by CHI Research based on the journal in which they appeared. The citing patents were also assigned to a set of technology categories, also developed by CHI Research, based on the primary patent class assigned by the U.S. patent examiner. For the purposes of our analysis, we only considered U.S. patent grants awarded to inventors with a U.S. address. However, inclusion of the full set of U.S. patent grants, including those made to foreigners, did not qualitatively change our results.

Finally, we obtained data on the universe of potentially citing U.S. utility patents granted over that same period, which is available from the NBER Patent Citation Database documented in Hall et. al. [2001]. These data use information on inventor address to identify “domestic” versus “foreign” patents. These data also provide information on the nature of the patent owner, so that we can distinguish between patents owned by universities, public research institutes, government agencies, and private firms. The geographic and institutional categories used in this data set are incorporated into our own analysis.

Given the nature of our analysis, we aggregated our raw paper, patent, and patent-to-paper citation data into cells, based on the scientific field, generating institution, and publication date of the cited paper, and the technological field, grant date, inventor location, and institutional category of the citing patent. This required us to construct a concordance between the disaggregated fields used by the University Science Indicators database and the more aggregated fields used by the CHI Research data. Papers with authors from more than one California institution are “credited” to each institution. The large number of categories implies a large number of cells and, hence, a large number of observations.

#### **Construction of the Data Set Used in Table VI**

The data set used in Table VI is at the firm-level. The NBER Patent Citation Database was used to identify the patent assignee codes assigned by the U.S. Patent and Trademark Office to the most significant “patent-generating” industrial firms and not-for-profit organizations. This set of mostly large enterprises was supplemented with a set of patent assignee codes assigned to

smaller firms in technology-intensive industries, including biotechnology, semiconductors, and information technology. While the majority of these assignee codes were associated with U.S.-based organizations, the codes associated with roughly 300 Japanese and 200 West European firms were also included.

Using these assignee codes, information was downloaded from the NBER Patent Citations Database on *all* patents granted to these firms and organizations from 1982-1999. With the help of CHI Research, information on all of the citations made by these patents to scientific papers published in journals tracked by the Science Citation Index was obtained. For more than 70% of the papers cited, the institutional affiliation of the author at the time of publication can be identified, although these data are not incorporated in the empirical analysis shown in the paper. Instead, counts of the number of citations to science references that appear in each year's cohort of patents were constructed, where patents are dated by the year of application rather than the year of grant. By looking at the number of science citations while controlling for the size of the patent cohort, one can obtain a firm-specific, time-varying measure of the science citation intensity of the firm's patented inventions.

The IMS LifeCycle Patent Focus Database maintains a record of approved drugs and the U.S. patents associated with them. This database was employed to create a count of (eventually) approved drugs that can be associated with the cohort of a firm's patent applications in a given year. This provides a way of linking the science citation intensity of a patent cohort with the number of drugs that patent cohort eventually produces. This measure of product innovations is only available for pharmaceutical and biotechnology firms, but it provides a direct connection between patents and products.

Most of the largest firms generating patent applications are publicly traded corporations, allowing for the linkage of patent data with data on R&D investment, labor and capital inputs, and sales. This kind of linked data was obtained for nearly 1,300 firms. Data on accounts for U.S. firms were taken from Compustat. Data on Western European firms were taken from



Compustat Global, Datastream, and Osiris. Data on Japanese firms were taken from the Development Bank of Japan Corporate Finance Database, with supplementary R&D data taken from firm-level R&D surveys published in the Japan Company Handbook series distributed by Toyoko Keizai. Corporate output was deflated by producer price indices associated with the firm's primary industry. This is obviously problematic in the case of highly diversified manufactures. Measures of the capital stock were deflated "book value" measures taken directly from firm accounts. Measures of R&D stocks were obtained by applying the perpetual inventory method to our R&D expenditure series. Regressions run with these R&D stock measures yielded results qualitatively similar to those using R&D flow measures. However, these stocks are subject to considerable measurement error due to the absence of "pre-sample" data on R&D expenditure with which to calculate an initial stock. Obtaining better measures of capital and R&D stock is the focus of ongoing research efforts.