Is Academic Science Raising Innovative Productivity?
Theory and Evidence from Firm-Level Data

Lee Branstetter and Reiko Aoki

April 2005
Is Academic Science Raising Innovative Productivity?
Theory and Evidence from Firm-Level Data

Preliminary and Incomplete
Please Do Not Cite Without Permission *

Lee Branstetter
Daniel Stanton Associate Professor of Business
Columbia Business School
and NBER
lgb2001@columbia.edu

Reiko Aoki
University of Auckland
Hitotsubashi University
aokirei@ier.hit-u.ac.jp

March 2, 2005

*We thank CHI Research for generously providing the data on patent citations to academic papers at a generous discount over the rates charged to commercial users. We thank Delphine Irac and Yoshiaki Ogura for excellent research assistance. Aoki and Branstetter have both benefited from the hospitality of the Hitotsubashi University Institute of Economic Research, where some of this research was conducted. Branstetter’s stay at Hitotsubashi was supported in part by an Abe Fellowship. Branstetter also visited the Research Institute of Economy Trade and Industry, and he thanks that institution for its hospitality and financial support. We also gratefully acknowledge funding from the U.S. National Science Foundation, the California University-Industry Research Program, the Columbia Business School Center on Japanese Economy and Business, the National Bureau of Economic Research, and the Earth Institute at Columbia University and Project on Intergenerational Equity (PIE), funded by a scientific grant from Japan’s Ministry of Education, Culture, Sports, Science and Technology (grant
1 Introduction

Recent research points to an evident surge in innovative activity in the United States over the past fifteen years.¹ This is suggested by, among other things, a sharp rise in patent applications and patent grants that started in the late 1980s and has persisted through the end of the 1990s – a rise that has outpaced, by a considerable margin, 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 suggests that recent trends in patenting and related data are more consistent with 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.²

But if American R&D productivity has increased, then that raises the question of what factors are driving the increase.³ This paper attempts to assess the importance of one possible contributing factor – increased knowledge spillovers from U.S.-based academic science. Figure I 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.⁴ 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 13 times.⁵ Many at the Na-

³The work of Kortum and Lerner (2000) has stressed the potential role of venture capital-linked firms in improving U.S. R&D output.
⁴This 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).
⁵These 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.
tional 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. 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.”

But this is not the only interpretation of recent data trends, and it is not necessarily the correct one. Simply counting patent citations to science across technological and scientific fields, nations, and times, as prior researchers have done, tells us little about the impact the citations (and the knowledge flows that they trace out) are having on the inventive productivity of the citing firms and organizations. This is a serious limitation to our current knowledge, because the implications of increased citations for national technological progress, economic growth, and welfare will be a function of the impact of the underlying knowledge flows on the research productivity of the recipient inventors.

This paper seeks to remedy this gap in our knowledge by combining new theoretical work with new firm-level empirical analysis. The paper introduces a formal model of applied R&D, based on the pathbreaking work of Evenson and Kislev (1976). Although the model is quite simple, it generates a number of

---

6 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).
unambiguous predictions that can then be taken directly to our unique data set. In this model, applied R&D is represented as a search process. Breakthroughs in academic science can enrich the “search distributions” probed by firms, leading to an increase in the productivity of R&D spending. In our model, firms differ in their ability to benefit from these breakthroughs in academic science. Thus, academic breakthroughs have a differential impact on the cross-section of firms.

We measure the relative strength of the connection between academic science and firm invention by tracking the citations to scientific articles that appear in the U.S. patents of our sample firms. We possess such data for more than 1,200 firms over the period 1983-1999. We use measures of patent quality and total factor productivity, together with firm-level measures of R&D spending, to compute a number of different indices of R&D productivity. For firms in the pharmaceutical industry, broadly defined, we use firm-level data on successful product introductions to create an alternative measure of innovative output. We find support for the predictions of the model, and we find that the measured impact of academic science on research productivity is particularly strong in the biotechnology and pharmaceutical sectors.

2 The Link Between Academic Science and Industrial Innovation

Historical Perspective

>From their inception, publicly supported universities in the U.S. were focused on training students in the “practical arts.” In the late 19th and 20th 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. However, progress at the scientific frontier was still dominated by European institutions until the cataclysm of World War II.
The large U.S. postwar investment in basic research, much of it concentrated in universities, and the mass migration of leading European scientists to the United States quickly established America as the leading center of frontier scientific research (Rosenberg and Nelson, 1994). The infusion of federal funds 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 suggested that relationship between “frontier” academic science and industrial invention, while obviously important, was neither close nor direct.\footnote{See, for example, Derek De Solla Price (1965) and Lieberman (1978). This view was generally supported by the Defense Department’s ambitious “Project Hindsight” study of the impact of basic scientific research on weapons development, which concluded that the primary impact came not from science at the research frontier, but instead from “packed-down, thoroughly understood, carefully taught old science,” such as that typically presented in textbooks or university courses. See Sherwin and Isenson (1967), from which the quoted phrase is taken, for a review of Project Hindsight.}

**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.\footnote{For a comprehensive literature review that covers relevant research beyond the economics journals, see Agrawal (2001).} Case studies, manager interviews, and surveys have been used to assess the magnitude of this impact, the channels through which it flows, and changes in these factors over time.\footnote{Important recent studies relying primarily on case study techniques and surveys include Mansfield (1995), Cohen et. al. (1994), Faulkner and Senker (1995), Gambardella (1995) and Agrawal and Henderson (2002).} 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.\footnote{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.} 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) have used data on university patents and citations to these patents to quantify knowledge spillovers from academic science.\footnote{Barnes, Mowery, and Ziedonis (1998) and Mowery, Nelson, Sampat, and Ziedonis (1998) have undertaken a similar study for a smaller number of universities.} While patenting by universities has increased substantially in the United States over the last twenty years, there is evidence that as the num-
ber of university patents has grown, the marginal quality of those patents has declined.\footnote{See Jaffe, Trajtenberg, and Henderson (1998) and Hicks et al. (2001).}

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

This paper will use patent citations to academic papers to measure knowledge spillovers between academic science and industrial R&D.\footnote{In doing so, I am 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.} 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 in the U.S. and other public policy measures, universities have increased the extent to which they patent the research of university-affiliated scientists. They have also increased the extent to which
they license these patented technologies to private firms. Nevertheless, it is clear to observers that only a tiny fraction of the typical research university’s commercially relevant research output is ever patented, and only a fraction of this set of patents is ever licensed.\textsuperscript{17} 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 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.\textsuperscript{18}

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.

What our methodological approach clearly fails to measure is the contribution of “old science” to industrial invention. A significant component of the consulting work undertaken by university faculty consists of helping private industry understand and apply well-established – or, “old” – scientific techniques.

\textsuperscript{17}This result is also emphasized strongly in the interview-based evidence presented by Agrawal and Henderson (2002).

\textsuperscript{18}Other recent studies using data on patent citations to scientific papers include work by Fleming and Sorenson (2000, 2001) and Lim (2001). Neither of these studies focuses on the large change in citations to academic science over the course of the 1990s, which is the focus here.
and engineering principles, rather than helping firms incorporate the latest frontier science into their research agendas. Likewise, recent science and engineering graduates are often employed in functions that are quite far removed from the scientific frontier, but are nevertheless quite economically important to the financial success of their employers. This contribution will be completely missed by our approach. In such cases, there is no new patented invention incorporating recent science. But as the older literature on university-industry interaction has stressed, the propagation of “old” scientific and engineering knowledge to industry through training and consulting is a long-standing feature of the American university system. The new development stressed by the recent literature is the closer relationship between technology and relatively recent science. It is precisely this aspect of university-industry interaction that our methodological approach will most closely reflect.

Lessons of Previous Research Using Patent Citations to Academic Science

Since this paper will focus on the use of patent citations to scientific papers as an indicator of knowledge spillovers from academia, it is important to note the lessons that have been learned from previous studies using these data. This summary will necessarily be brief. For a more complete exploration of these issues, the reader is referred to the papers cited in this section.

Perhaps the strongest finding to emerge from these studies is the result that patent citations to science are highly concentrated in a set of academic disciplines and related technological fields that we might refer to as the “bio nexus.”\(^{19}\) Patents taken out in patent classes associated with pharmaceutical products, medical devices, and biotechnology display a much higher propensity to cite scientific papers than do patents in other fields. Perhaps not surprisingly, the academic disciplines within medicine and the life sciences traditionally most closely associated with these technological fields (molecular biology, various fields of clinical medicine, etc.) are the kinds of papers that are most likely to be cited in patent documents. This is true even when one controls for the large and growing number of “biotech” patents and the large and growing number of bioscience papers.\(^{20}\) If one looks outside the bio nexus, one sees evidence of a secondary concentration of patent-to-paper citation activity in information

---

\(^{19}\)See Narin et al. (1997), Hicks et al. (2001), Branstetter (2004), and Branstetter and Ogura (2004).

\(^{20}\)See Branstetter and Ogura (2004).
technology, but it is much less pronounced than that within the bio nexus. To
the extent that knowledge spillovers from academic science are actually increas-
ing research productivity, one might expect this effect to be concentrated in the
bio nexus. In the empirical work presented in this paper, we explicitly allow for
this impact to be differentially stronger for firms located in that nexus.

Second, the propensity of patents to cite science has risen substantially over
time. This is true even if one looks at patent-to-paper citation activity within
the bio nexus, controlling for changes in the volume and distribution across
fields of both patenting and publishing. Some of this increase is attributable
to discrete changes in U.S. patent law and citation practice, but a substantial
body of statistical and qualitative evidence strongly suggests that it also reflects
and increasing intensity of intellectual interaction between academic science and
industrial R&D, particularly within the bio nexus.

Third, citations to science are positively correlated with measures of patent
quality at the patent level. As we will note later, the micro literature on patents
has suggested several measures of patent “quality” – quantitative features of
the patent document – which have been demonstrated to be positively corre-
lated with the ex-post commercial and technological importance of the patent.
Three such measures include counts of ex-post (or “forward”) citations, counts
of claims contained in the patent document, and the measure of “generality”
proposed by Henderson, Jaffe, and Trajtenberg (1998). This latter measure is
a quantitative index of the diversity of technological fields across which ex-post
citations occur. An invention whose citations come from multiple technological
fields can be thought of as having a more “general” impact than an invention
whose citations come from a single technological field. The formal definition
of the index is

\[ \text{Generality}_i = 1 - \sum_{k=1}^{N_i} \left( \frac{N_{citingi,k}}{N_{citingi}} \right)^2 \]

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

Various studies, including Branstetter (2004), have shown that patent cita-
tions to science are positively and significantly correlated with all these measures
of quality. At the patent level, it is difficult to give this correlation a causal in-
terpretation, but it is suggestive nonetheless.\(^{21}\)

\(^{21}\)This finding is also consistent with the work of Nagaoka (2004) and Sorenson and Fleming
Finally, while the received literature has focused on the patent-to-paper citation activity of patents generated by U.S.-based inventors, some evidence suggests broadly similar patterns in the behavior of innovative firms based outside the U.S., particularly Japanese firms. Branstetter and Ug (2004) undertake a detailed study of the patterns of citations to academic science found in the U.S. patents of a sample of Japanese firms. They find patterns broadly similar to those in the patents of U.S.-based inventors, although the average propensity to cite science tends to be lower for Japan-based inventors than for U.S.-based inventors when one conducts comparisons within patent classes.\footnote{Nagaoka (2004) finds the same pattern.}

Relatively speaking, the same concentration in the “bio nexus” and the pronounced increase over time in the propensity to cite evidence is clear in the Japanese data. Complementing our own findings, Nagaoka (2004) finds a positive relationship between patent citations to science and the “value” of patents, as measured by ex-post citations, using a sample of U.S. patents granted to U.S. and Japanese inventors in the information technology sector. Work by Hicks (1993), Kobayashi (1998, 2003), Odagiri (1999), Pechter (2000, 2001), and Walsh and Cohen (2004) also suggests a strengthening link between Japanese industrial R&D and academic science in the 1980s and 1990s, even in the absence of formal mechanisms for technology transfer between academia and industry until the late 1990s. A preliminary study of the citations found in Japanese documents undertaken by Tamada et al. (2003a, 2003b) suggests that the patterns similar to those found in the U.S. patents of Japanese firms.

\section{Theoretical Model}

This assumes a firm deciding how many draws to make based only on basic science utilization, patentability requirement and cost of draws. There is no dynamic consideration. Draws are chosen to maximize marginal profit from expected number of patents. This has advantage of being consistent with pooled data. Patentability requirement only needs to be common to all firms within an industry, within a period.

Applied research undertaken by a firm $i$ in industry $j$ in period $t$ is $n$ independent draws from distribution $f(\cdot | \beta)$ (as in Evenson and Kislev (1976)). Parameter $\beta$ is the level of basic science utilized. This would depend on the level of basic knowledge available, $\theta$, and rate at which a particular firm uses basic
science, $\mu$, such as $\beta = \mu \theta$. The latter would be common across all firms while $\mu$ is firm specific and may even be endogenous. Distribution with $\beta$ first-order stochastic dominates distribution with $\beta' < \beta$ i.e.,

$$F(\xi|\beta) \leq F(\xi|\beta') \text{ for all } \xi.$$  

Each period, firm makes $n$ draws resulting in $n$ realizations, $x^1, x^2, \ldots, x^n$. A new technology (a draw) will become a patent if it is above the minimum requirement for patentability (novelty, non-obviousness) denoted $y$. The number of patents, $N$ given $n$ draws is equal to number of draws equal to or greater than $y$. The distribution of $N$ is,

$$\text{Prob}(N = k|y, \beta, n) = \frac{n!}{k!(n-k)!} [1 - F(y|\beta)]^k [F(y|\beta)]^{n-k}.$$  

The expected number of patents given $n$ draws is,

$$E[N|y, \beta, n] = \sum_{k=1}^{n} \frac{n!}{k!(n-k)!} [1 - F(y|\beta)]^k [F(y|\beta)]^{n-k} = n [1 - F(y|\beta)].$$  \hspace{1cm} (1)

**Lemma 1.** Expected number of patents is increasing in number of draws ($n$) and level of basic science utilization ($\beta$) and decreasing in patentability requirement ($y$).

**Proof.** $1 - F(y|\beta)$ is increasing in $\beta$ from first-order stochastic dominance assumption. Cumulative distribution $F(y|\beta)$ is increasing in $y$. \hspace{0.5cm} $\square$

**Optimal behavior**

A firm’s period $t$ instantaneous profit is,

$$\Pi_t = p_tQ(K_t, L_t, S_t) - w_tL_t - r_tK_t, \hspace{1cm} (2)$$

where $S_t$ is the stocks of patents accumulated by the firm up to period period $t - 1$. That is, using $N_\tau$ as the patents acquired in period $\tau$, stock available in period $t$ is,

$$S_t = \sum_{\tau \leq t-1} N_\tau.$$  

10
Output at time $t$, depends on patents on the stock of patent as well as labor and capital. A firm maximizes

$$V_t = \Pi_t + E_t[\Pi_{t+1}] - w_tL_t - r_tK_t - c_t n_t,$$

where $n_t$ is the number of draws that determine distribution number of patents to be acquired this year, $N_t$. The problem can be separated into the contemporaneous part, choosing $L_t$ and $K_t$ to maximized $\Pi_t$, and, investment for the future,

$$E_t[\Pi_{t+1}] - c_t n_t.$$

We can approximate the first term by,

$$E_t[\frac{\partial \Pi_{t+1}}{\partial S_{t+1}} N_t],$$

using $dS = S_{t+1} - S_t = N_t$. Note that since capital and labor are chosen optimally each period, the Envelope Theorem implies that the marginal profit from new stock of patents is equal to the marginal revenue. Furthermore if we assume $\frac{\partial \Pi_{t+1}}{\partial S_{t+1}}$ is independent of $S_{t+1}$ (i.e., $\Pi_{t+1}$ is linear in $S_{t+1}$) then the firm chooses $n_t$ to maximize,

$$\pi_t E[N_t] - c_t n_t,$$

where $\pi_t = \frac{\partial \Pi_{t+1}}{\partial S_{t+1}}$.

Now we are ready to analyze the optimal behavior of firm $i$ in industry $j$ in period $t$ taking into account how distribution of new patents depend on basic science access. Basic science utilization is determined by the amount of basic science available at that time, $\theta_{ijt}$ and how well the firm as able to access it, $\mu_{ijt}$, so that $\beta_{ijt} = \mu_{ijt} \theta_{ijt}$. The patentability requirement is $y_{ijt}$.

The cost of $n$ draws is $c_{ijt}(n)$. $E[N]$ patents correspond to $\pi_{ijt} E[N]$ dollars of profits where $\pi_{ijt}$ is the marginal profit in (3). Firm $i$ chooses $n^*_{ijt}$, the optimal number of draws, to maximize,

$$\pi_{ijt} E[N] [y_{ijt}, \beta_{ijt}, n] - c_{ijt}(n),$$

subject to a capacity constraint,

$$n \leq \kappa_{ijt}.$$

This is additional profit from the new patents since the total firm profit would
depend on all patents owned. Since we are counting a flow of patents (not accumulation of patents), $\pi E[N]$ is extra profit from these new patents.

We consider two different forms of the cost function:

(1) $c_{ijt}(n) = c_{ijt}n$ with $c_{ijt} < 1 - F(y_{jt}j|\beta_{ijt})$,

(2) $c'_{ijt}(n) > 0$, $c''_{ijt}(n) > 0$.

We immediately have the following for case (1),

**Lemma 2.** If $c_{ijt}(n) = c_{ijt}n$ and $c_{ijt} < 1 - F(y_{jt}j|\beta_{ijt})$, then

$$n^*_{ijt} = \kappa_{ijt}.$$  

The expected number of patents is,

$$E[N|y_{jt}, \beta_{ijt}, n^*_{ijt}] = K_{ijt} \left[1 - F(y_{jt}j|\beta_{ijt})\right].$$

The expected number of patents is increasing in basic science utilization, controlling for capacity constraint. In particular, even if R&D remains at the capacity level, basic scientific will be increasing patents and marginal profit. The effect on profit will be more significant when marginal profit of a patent $\pi_{ijt}$ is large, such as the bio-nexus.

With a more general case (2), the optimal number of draws satisfies the first-order condition,

$$\pi^{ijt} \left[1 - F(y_{jt}j|\beta_{ijt})\right] = c'_{ijt}(n^*).  \tag{4}$$

This is demonstrated in Figure 1. It is easy to show the following,

**Lemma 3.** When the cost function is strictly increasing and convex (case (2)), the optimal number of draws, $n^*_{ijt} = n^*_{ijt}(y_{jt}, \beta_{ijt})$ is decreasing in $y_{jt}$ and increasing in $\beta_{ijt}$ and $\pi_{ijt}$. The expected number of patents is,

$$E[N|y_{jt}, \beta_{ijt}, n^*_{ijt}] = n^*_{ijt}(y_{jt}, \beta_{ijt}) \left[1 - F(y_{jt}j|\beta_{ijt})\right].$$

The expected number of patents is decreasing in $y_{jt}$ and increasing in $\beta_{ijt}$ and $\pi_{ijt}$.

Generally, we can say the following,
**Proposition 1.**

1. Greater scientific research utilization implies greater expected number of patents. 

2. The effect will be larger for industries where each patent generates larger profits.

### 4 Data Description

We measure the relative strength of the connection between academic science and firm invention by tracking the citations to scientific articles that appear in the U.S. patents of our sample firms. We possess such data for more than 1,200 patent-generating firms over the period 1983-1999. This group of firms includes a substantial number of technology-intensive firms based outside the U.S. that are extensive users of the U.S. patent system. For example, the database contains information on more than 300 Japanese firms and more than 200 firms based in Western Europe. From the NBER Patent Citation database documented in Hall et al. (2001), we obtain information on the complete set of patents granted to these firms by the U.S. Patent and Trademark Office between 1983 and 1999.

The NBER Patent database allows us to construct, for each patent, measures of the quality and technological impact of the patented invention. The micro literature on patents has suggested several measures of patent “quality” – quantitative features of the patent document – which have been demonstrated to be positively correlated with the ex-post commercial and technological importance of the patent. Three such measures include counts of ex-post (or “forward”) citations, counts of claims contained in the patent document, and the measure of “generality” proposed by Henderson, Jaffe, and Trajtenberg (1998). We can thus construct for each firm a count of the number of patents taken out in each year which is weighted by the ex-post citations these patents receive. Because we only observe patents granted up to the end of 1999, we cumulate these patent citations over a four-year window from the date of grant. Earlier research has suggested that the number of citations received in the first four-to-five years after grant is indicative of the total number of citations eventually received over much longer windows.

In the same manner, we can construct annual counts of patents where the patents are weighted by the number of claims contained in the patent application. We can also measure the average “generality” of a firm’s cohort of patent
applications in a given year, and use this as an additional measure of patent quality. For pharmaceutical firms, an additional measure can be employed. It is possible to identify particular patents and groups of patents that are associated with medical treatments that have been formally approved by the FDA and are currently being sold in the market. While the linkage between patents and products is generally not available in most industries, it is available for pharmaceuticals and related sectors. Thus, we can present regressions in which the outcome measure is the number of patents per (application) year that eventually lead to actual approved products. This provides a measure of research output that is much more closely related to consumer welfare and social surplus than the other patent-based measures described above.\footnote{These data were provided to the author by Frank Lichtenberg. They are based on proprietary IMS data.}

The NBER Patent Citation database does not include information on patent citations to academic science. These data were purchased from CHI Research, which provided them to us under a contract that severely restricts data access. These data were generated in the following way. CHI Research scans the electronic records of the U.S. PTO, obtaining all citations to patent and non-patent prior art listed on the front page of the patent application. It then uses proprietary software to clean and standardize the citations to non-patent prior art, using, for instance, a standard set of scientific journal names. For the subset of scientific journals tracked by the Institute for Scientific Information’s Science Citation Index, we were able to recover information on the article cited, including the exact reference and the scientific field into which the article could be classified. For the majority of these articles, we were also able to obtain data on the authors and the authors’ institutional affiliations at the time of authorship.\footnote{Information on author affiliation was obtained separately by matching the full article citation to author information in the Science Citation Index for the years 1980-1997. The patent and paper years do not overlap exactly, due to differences in data availability. As a consequence, institutional affiliation data are only available for about 75% of the cited articles.} The information on scientific content and author identity is not used in the analysis presented in this paper, but it will be the focus of further work described in the concluding section.

Publicly traded firms in the United States have been required to disclose their annual R&D spending since the early 1970s. As a consequence, for each publicly traded U.S.-based firm in our data set, we can generally obtain reasonably high quality data on total R&D spending by firm fiscal year. The same accounting requirements do not generally exist elsewhere in the OECD. However, prior
research has suggested a number of publicly available data sources that seem to provide reasonably accurate R&D data for selected firms based in Japan and Western Europe.\textsuperscript{25} These data sources are employed in the current paper. We note that patent data exist for many entities for which reasonable R&D could not be obtained, limiting the cross-sectional dimension of our database and biasing the sample toward larger, publicly traded firms.

In addition to R&D spending, we can obtain from publicly available data sources on listed companies information on total sales, total employment, and capital stock. Data for U.S. firms is taken from COMPUSTAT. Data for Japanese firms is generally taken from the Development Bank of Japan Corporate Finance Database. Data on Western European firms comes from Datastream, Osiris, and Compustat Global. These data can be used in a production function framework to estimate the impact of changes in citation intensity on productivity.

5 Econometric Analysis at the Firm Level

Discussion of trends in citations data, such as that shown in Figures I and II, is of limited interest unless the knowledge spillovers indicated by these citations are actually enhancing the research productivity of the firms and other organizations that receive them. Are innovators learning from academic science in such a way that they are able to produce more inventions or better inventions than they otherwise could? To evaluate this question, we need quantitative measures of the “quality” of patented inventions. This harkens back to the two interpretations of the “patentability constraint” noted in the theory section. If the U.S. Patent and Trademark Office could be relied on to consistently enforce high standards of novelty, non-obviousness, and utility in making patent grants, then we could expect that only inventions that materially advanced the state of the art would be granted patents. In fact, a large body of evidence seems to demonstrate that the U.S. PTO has not enforced standards in this way. The typical patent grant tends to be of quite limited technological or commercial value, and there is enormous skewness in the \textit{ex-post} realized commercial value of patented inventions.\textsuperscript{26} Thus, for our purposes, it may be of interest to look

\textsuperscript{25} Data on Japanese firm R&D spending is taken primarily from the annual surveys of R&D spending conducted by Toyo Keizai and Nihon Keizai Shimbun and reported in the \textit{Kaisha Shiki Ho} and the \textit{Nikkei Kaisha Joho}, respectively. Data on the R&D spending of Western European firms is taken from Osiris, Datastream, and Compustat Global.

\textsuperscript{26} See Harhoff et al. (1999).
at a measure of patent output where patents are weighted by some measure of quality, in order to produce a meaningful measure of inventive output that is consistent with Proposition 1 derived in the theory section.

Because we observe the innovative output of the same firm at different points in time, we can, at least in principle, control for the “average research quality” of the firm’s R&D operation by using firm fixed effects. Conditioning on this and on the level of R&D investment, is it true that an increase in the incidence of citation of academic science in a firm’s patents is strongly correlated with an increase in the quality or quantity of that firm’s patented inventions? If the answer to this question is yes, we have evidence that the knowledge flows from academia to the sample firms, indicated by citation counts, are really having a positive impact on the firms’ research productivity. Using the approximations used in (3), a potential econometric specification for investigating this is the following:

\[ \text{Qual}_{it} = \beta_0 + \beta_1 \text{lrnd}_{it} + \beta_2 \text{Citing}_{it} + \beta_3 \text{lp}_{it} + \sum_t \delta_t T_t + \theta_i + \varepsilon_{it} \]

The dependent variable will generally be a measure of the number of patents filed by firm \( i \) in year \( t \), where the patent counts are adjusted in some manner for their quality. For the purposes of this estimation, firm patents are assigned a date based on their application date rather than their grant date. Quality-adjusted patent output is modeled as a function of R&D spending (\( \text{lrnd} \)), the overall level of patenting (\( \text{lp} \)), a firm “fixed effect” (\( \theta_i \)) and the number of citations made in a patent cohort (\( \text{Citing} \)). Because we simultaneously control for the level of patenting, the coefficient on the \( \text{Citing} \) variable will be, in effect, measuring changes in the intensity with which successive patent cohorts cite science. In our estimation, we also allow for a full set of year dummies (the \( \delta_t \)'s), in order to incorporate changes in citations practices that may have affected the entire sample of firms in a similar way. As an alternative dependent variable for firms in the “bio nexus,” we can also use our count of the number of patents that eventually became FDA-approved products.

Measures of patent quantity and quality and new product introductions are not the only measures of inventive output that could be used. Drawing upon the firm-level data in the COMPUSTAT, DBJ, and related databases databases, we can also construct firm level measures of total factor productivity. If we can identify a positive effect of citation intensity on these market-based measures
of firm innovative performance, we will have even stronger evidence that the knowledge flows tracked by patent citations are having a real, measurable impact on the research productivity of citing firms. To this end, we estimate a simple Cobb-Douglas production function based on the the formulation (2). We approximate $L_{t+1}$, $K_{t+1}$ by those of period $t$. As explained after equation (3), the regression is actually on sales. The estimated equation is

$$Q_{it} = \beta_0 + \beta_1 lkap_{it} + \beta_2 lemp_{it} + \beta_3 lrnd_{it} + \beta_4 Citing_{it-1} + \beta_5 lp_{it-1} + \alpha_i + \sum_{t=1}^{T} \delta_i T_t + \varepsilon_{it}$$

We regress the log of real sales of firm $i$ in year $t$ ($Q_{it}$) on a measure of the log of contemporaneous real capital stock ($lkap$), the log of employment ($lemp$), the log of contemporaneous R&D spending ($lrnd$), and our science citation measure ($Citing$).\(^{27}\) As before, we include as a control the size of the patent cohort that produced the count of science citations ($lp$). In this context, lagging the science citation measure and the corresponding patent measure may be important, for reasons that we discuss below. We will allow for short lags of length $l$. As in earlier regressions, we include firm and year fixed effects. In this context, a linear fixed-effects approach is appropriate. Here, the coefficient on our science citation term measures the contribution of learning from science to the firm’s productivity growth, because identification comes from the correlation between within firm changes in productivity levels and within firm changes in levels of science citation.

Results for preliminary firm level regressions are easily summarized. Coefficients are reported in Table I. The first two columns of this table report results using citation-weighted patent counts as the measure of innovative output. Because of the count nature of the dependent variable, the regression technique used is the fixed-effects negative binomial estimator developed by Hausman, Hall, and Griliches. When this estimator is used, log-likelihood rather than R-squared is reported as a measure of goodness of fit. As shown in column 1, for the full sample, measures of science citation are not strongly positively correlated with citation adjusted patent measures. On the other hand, the results for the pharma/biotech/chemical subsample, given in column 2, suggest that the measured impact of an increase in science citations, controlling for the number of patents, is positive and statistically significant. Because the log of science citation-weighted patent counts is the measure of innovative output, we will have even stronger evidence that the knowledge flows tracked by patent citations are having a real, measurable impact on the research productivity of citing firms. To this end, we estimate a simple Cobb-Douglas production function based on the formulation (2).

\(^{27}\)Ideally, an output specification should use an R&D stock measure rather than the flow measure introduced here. Preliminary investigations with an R&D stock measure, however, generate results very similar to those reported here.
citations is entered into the regression, the coefficient has an elasticity interpretation, and it implies that a 100% increase in science citations will result in a roughly 8.5% increase in quality-adjusted patent output, given the same level of R&D spending. This would seem to be a modest marginal impact, but the reader should recall the substantial increase in the number of patent citations to science that have been observed over the past fifteen years. The absolute number of such citations has increased thirteen-fold since the mid-1980s.

Measures of science citation are positively and significantly related to claims-adjusted total patent counts, as shown in columns 3-4. Again, the dependent variable here is a count variable – the number of patents weighted by the number of claims in each patent – so fixed effects negative binomial models are used to estimate the impact of science citations on inventive output. The estimated effect is considerably stronger in the pharma/biotech/chemical subsample. In fact, the marginal impact of science citations is more than twice as high.

In columns 5-6, the inventive output measure is the average “generality” of patents taken out by firm $i$ in year $t$. Linear, fixed effects regressions are used here. The dependent variable is a ratio, so the coefficient on the science citations measure does not have an elasticity interpretation, as it does elsewhere in the table. As the reader can see from the results shown in column 5, there is relatively little impact of science citations on inventive output in the full sample. However, as shown in column 6, there does appear to be a statistically significant positive effect in the pharma/biotech/chemical subsample.

Results of regressions using measures of “patents that led to products” show a very strong, very significant effect of science linkages on inventive outcomes. This is shown in column 7, where the count data nature of the dependent variable requires turning again to the fixed effects negative binomial specification. Due to data restrictions, these results are only available for the pharma/biotech/chemical subsample. The estimated coefficient is quite large. Controlling for the size of the patent cohort, a 100% increase in science citations raises the number of successful product introductions by more than 50%!

In the last two columns of Table I, production function regressions find a positive, significant effect of citations to science on firm total factor productivity.

A comparison of the coefficients in columns 8 and 9 suggests that the productivity boost is substantially stronger in the biotech/chemical/pharmaceutical subsample. These results need to be viewed with some caution, as capital expenditures and R&D expenditures have not been deflated with the appropriate industry-level price indices. In addition, questions could be raised about the
assumption, imposed here, that there is an essentially contemporaneous effect of invention quality on productivity. Where the patented invention protects a process, it could be believed that the process is implemented as soon as the patent is filed. However, many patents protect products, patents for new products tend to be taken out at a relatively early stage in the product development process, and one could easily imagine that it takes time for even highly successful new products to have a measurable impact on firms’ total revenue streams.

Taking that point seriously, Table II utilizes various lags of the science citation intensity measure in the production function specification, drawing upon data for the full sample. As can clearly be seen, the measured impact of an increase in science citation intensity remains statistically robust, regardless of which lag is used. We see this as useful confirming evidence of a real impact of knowledge spillovers from academic science on invention that translates into measurable gains in firm revenues.

Given the interest of participants in this conference in the relevance of these findings to Japanese firms, we offer evidence on a subsample of Japan-based R&D performing firms in Table III. A criterion for inclusion in our sample is that firms patent extensively in the U.S., so we cannot claim that this set of Japanese firms, which numbers just over 300, is representative of Japanese manufacturing. However, given that Japanese R&D spending and patenting tends to be dominated by a relatively small number of relatively large firms, our sample is likely to be reflective of trends in that subset of firms that contribute disproportionately to technological innovation in Japan.

Despite the widespread belief that Japanese firms are insufficiently connected to university research, either in Japan or elsewhere, that the lack of connection to university science has been a major factor in Japan’s alleged lack of competitiveness over the last ten years, our empirical results suggest a relationship between the intensity of citation of academic science and productivity growth within our sample of Japanese firms. The coefficients are positive, statistically significant, and robust to the use of alternative lags. However, we note that the point estimates are smaller than those obtained for the full sample, which is dominated by U.S. firms.
6 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 power the surge in innovative activity in the U.S. in the 1990s? This paper has sought to shed light on these questions.

Recent research cited in this paper provides support to the notion that the nature of U.S. inventive activity, and, perhaps, inventive activity elsewhere in the industrialized world, has changed over the sample period, with an increased emphasis on the use of the knowledge generated by university-based scientists in later years. The timing of this change corresponds closely to a marked increase in patenting by U.S.-based entities, suggesting that knowledge spillovers from academic science may have been a significant factor contributing to the surge in U.S. industrial innovation.

In this paper, we have endeavored to make two intellectual contributions to this literature. First, we have presented a simple model of the R&D process, based on the pioneering work of Evenson and Kislev (1976), which shows how the research productivity of individual firms can be affected by knowledge spillovers from academic science. This model delivers unambiguous predictions about the relationship between a firm’s use of academic science in its innovative activities and research productivity.

We then seek to test these predictions with firm level data for roughly 1,200 patent-generating entities. Taken as a whole, our results generally suggest the possibility of a strong link between citation to science and inventive productivity, even when controlling for R&D investment. This link appears to be significantly stronger within the set of technical fields and related scientific disciplines that we have dubbed the “bio nexus.”

More could be done to assess the extent to which the link in the data between citation of science and research productivity is a causal one. A particularly promising approach may one suggested by Azoulay, Ding, and Stuart (2004). These authors use recent advances in biostatistics to refine the “difference in differences” estimation that has become so popular in the recent econometrics literature. This method, called Inverse Probability of Treatment Weighted (IPTW) estimation, generalizes so-called “propensity score” techniques for comparing citing and non-citing firms, allowing for “selection” of firms into research areas proximate to academic science based on a number of observable charac-
teristics. We believe that implementation of this class of techniques will allow us to go beyond documentation of a suggestive statistical association to make stronger claims regarding causality.

While it is useful to show a statistically robust association between science citations and productivity across a broad range of firms and industries, we also believe that much could be learned by narrowing our focus. The original Even-son and Kislev model, and much of the work that followed from it, attempted to map out the impact of a discrete scientific breakthrough on both optimal R&D investment and its marginal productivity over time. This perspective allows for a richer, more dynamic model than the simple one employed here.

In this paper, we have abstracted from this potentially richer set up, in part because our patent-generating firms are being hit by multiple “scientific opportunity” shocks at once, making the identification of the dynamic effects problematic. However, given the size and diversity of our sample of innovating firms, and the fact that we can identify the individual scientific papers being cited, it is quite possible that we could focus our attention on a narrower subset of firms, over a shorter period of time, that are responding primarily or exclusively to a single major scientific breakthrough. Observation of this group of firms and their patenting activity before and after the shock would allow us to trace out the dynamics of the richer model and also more exactly identify the $\mu$ parameter stressed in our simpler version.
Bibliography


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.


National Science Foundation, 2002, National Science and Engineering Indicators.


