DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 07-05

Academic Knowledge Spillovers Re-examined: a Look at the Effect of Exogenous Federal Funding

Eric T. Stuen

University of Colorado

October 30, 2007

Center for Economic Analysis
Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

ACADEMIC KNOWLEDGE SPILLOVERS RE-EXAMINED: A LOOK AT THE EFFECT OF EXOGENOUS FEDERAL FUNDING

ERIC T. STUEN*

Abstract

Discoveries in basic scientific research at universities are useful in applied research in industry and sometimes lead to commercially valuable innovations. Many empirical studies have documented a positive relationship between academic research and innovation by firms. However, interpreting this relationship as a causal spillover from academia to industry is difficult since a substantial share of academic R&D is funded by industry. Proximity to industry also influences the quality of professorial talent at a university, and location decisions of both industry and academics may be correlated with other unobservables. Given the presence of such difficult identification issues, this paper uses a novel empirical method to re-examine the effect of academic research in particular metropolitan areas on commercial innovation produced in those locations. I exploit the fact that members of certain appropriations sub-committees within the U.S. Congress can influence the process of allocating federal research funds in favor of their constituents, which leads to 'exogenous' variation in research funding at particular universities that is plausibly uncorrelated with factors that affect industrial innovation.

I construct a detailed panel of micro-data on patent counts, publication counts, doctorates granted and industry R&D expenditures at the metropolitan area/technology area/year level with which I find evidence that measures of academic scientific knowledge are positively related to industrial patenting. Using a city-year panel dataset of industrial patents and academic publications, I find an elasticity of patents with respect to publications from universities in that metropolitan area of 1.17, but that this elasticity is reduced to 0.95 when relying only on variation in publications attributable to "congressional favors". This translates into an extra patent produced by industry for every 7 extra academic publications produced by universities located in that city owing to the extra research funds diverted to those universities. An elasticity of patents with respect to citations to academic publications of 0.55 is also reduced in the IV set up. These results provide evidence of spillovers unrelated to ordinary market transactions between firms and universities. Data on Ph.D. recipients from each university is used to examine another channel through which academia affects industry – the employment of students with frontier- level technical knowledge. My results show that the employment of new doctoral graduates in science and engineering by industry is also positively related to industrial patenting.

Oct 30, 2007

Preliminary draft, please do not cite or quote without permission.

*Doctoral candidate at Dept. of Economics, University of Colorado at Boulder. Contact: eric.stuen@gmail.com. I thank A. Mushfiq Mobarak, Keith Maskus, Terra McKinnish, Wolfgang Keller, David Heinlein and participants in the WEAI 2007 conference for helpful comments. I am grateful to Jeffrey Brainard at the *Chronicle of Higher Education* and James Savage for providing data on direct academic appropriations, and to Richard Beaudoin for excellent programming assistance.

I Introduction

Universities and research institutes are frequently touted as sources of knowledge that enhance innovation in their local economies. To casual observers, the decision of high-tech firms to locate near world-class research universities is evidence of local economic benefits stemming from academic scientific research. While access to talented new graduates is a major consideration, access to new ideas – new knowledge at the "frontier" – is another important benefit. Spillovers from universities are especially important due to the role of academic science, which is more basic or "upstream", in fostering the development of applied technology by firms. While many studies have examined the linkages between academic science and industrial innovation, econometric identification of causal spillover effects is difficult due to unobservable factors affecting the co-location of faculty research talent and industry research. This study makes use of the fact that members of the U.S. Congress influence the distribution of federal research funding, creating variation in academic science that is plausibly uncorrelated with local industrial innovation.

Evidence abounds of collaboration and interaction between firms and universities: Intel maintains labs in Berkeley, Pittsburgh and Seattle that are staffed by faculty from UC Berkeley, Carnegie Mellon University, and the University of Washington, respectively.[1] The biotech company Lucigen is one of several firms working with the University of Wisconsin to create biocatalysts that can convert crops into ethanol,[2] and Dupont has a program with MIT - the Dupont-MIT Alliance - in which Dupont has a $60

[1] www.intel.com/research/network
[2] University of Wisconsin-Madison News "Industry Partners Bring Vital Applied Knowledge to the GLBRC Project", June 26, 2007

million commitment to MIT over the period 2000-2010.[3] Indeed, over the period of 1970 to 2000, fully 43% of academic R&D funding came from non-federal sources, with a large share from industry.[4]

Many studies have examined academic knowledge spillovers as the geographic co-location of academic R&D and measures of industrial innovation. This approach does not distinguish, however, between sources of knowledge external to firms, or independent academic research, and academic research that is essentially an extension of a firm's own R&D activity. Such a distinction is important when considering the economic benefits of academic research – independent research may be more aligned with the broad interests of society than "in-house" research, and may be used by a greater number of firms. Yet examining spillover effects from independent academic research is not as simple as considering only federally funded research – federal grants allocated through peer-review go to highly talented researchers, and the presence of nearby industry may draw professorial talent to a university. For example, the University of Rochester has cited its relationship with Kodak for the return of Henry A. Kautz, a "national leader in artificial intelligence", to its faculty.[5] Geographic and other unobserved factors, such as the talent of new graduates in science and engineering, may also influence the co-location of academic and industrial R&D. Therefore, estimates of the correlation between academic R&D and industrial innovation are not identifying a spillover externality per se, but rather measuring the total role of academic R&D in applied research by industry.

---

[3] MIT News Office "Dupont Backs MIT Research with Additional $25M", May 19, 2005
[4] Calculated from the NSF's *Survey of R&D Expenditures at Universities and Colleges*, available at webcaspar.nsf.org, for Universities sampled in this paper
[5] University of Rochester News Bulletin "New Academic-Industry Collaboration Brings Talent to Rochester", October 4, 2006

PhD graduates in science and engineering are also sources of new knowledge for firms. They spend years gaining hands-on experience with frontier-level knowledge, so their knowledge contribution to their future employers is different than a general contribution of human-capital. Enrollment levels may depend on the presence of industry as well, and an analysis of a spillover effect should account for the number of new graduates.

This paper uses the political determination of federal research funding for academia to isolate the causal effect of basic academic research on the innovative activity of firms located in the same cities as those universities. Members of Congress on certain appropriation sub-committees have control over the budgets of granting agencies, and since different agencies have different propensities to support research by location, research funds can be channeled to their own constituents. This creates variation in research funding that is unassociated with commercial innovation and other local characteristics linking academic and industrial R&D, which can then be used to measure the knowledge spillover effect of academic R&D on local innovation. Using a panel dataset of counts of industrial patents and academic scientific publications at the metropolitan/year level, I find a high elasticity of patents with respect to publications of 1.17, but that this is reduced to 0.95 when only variation in publications due to "congressional favors" is used. Similarly, an elasticity of patents with respect to academic citations of publications of 0.55 is reduced to a negative, but insignificant, 0.55. This suggests that in the aggregate, knowledge spillovers take place at a lower rate than transfer of knowledge through market transactions, but this difference varies by technology area. Using data on the actual counts of new PhD graduates per metropolitan/year observation, I find that separately controlling for this channel of influence from academia to industry does not affect the finding above, although the number of new PhDs is related to the level of industrial patenting. In summary, the link between academic and industrial innovation is likely multi-dimensional and bi-directional, and my empirical methodology allows me to isolate a specific component of this link, which is the influence of independent basic academic research on the innovation activities of firms located in the same cities.

The paper is structured as follows: section II discusses prior literature in academic knowledge spillovers. Section III offers a discussion of the economics of academic knowledge spillovers, including a rationale for local spillovers. In section IV the empirical methodology and method of identification is described, while section V describes the data, section VI presents empirical results, and section VII concludes.

II Prior Literature

This paper follows a stream of econometric studies that measure spillovers indirectly, as the geographic coincidence of innovation with academic resources. This is often styled as the estimation of an innovation production function in which firms use their own research and development (R&D) expenditures combined in some way with an academic input (academic R&D). Such an innovation production function was first proposed for individual firms by Pakes and Griliches (1980, 1984), and later adapted to include academic R&D by Jaffe (1989). Jaffe's study estimated the relationship between industrial patent counts and both industrial R&D and academic R&D, at the level of state and year. Recognizing that the large portion of academic R&D funded by industry would bias the OLS estimate of the spillover coefficient, Jaffe used state demographics such as population and the number of universities as instruments for academic R&D. The central findings included IV estimates of the elasticity of industrial patenting with respect to university R&D equal to 0.191 for drugs and medical technology, and 0.125 for electronics technology.

Subsequently, many studies have looked at different aspects of academic knowledge spillovers. Acs, Audretsch and Feldman (1992) criticized Jaffe's use of patent counts as a measure of innovation, since the quality and novelty of innovation documented by each patent varies widely, and the propensity to patent also varies widely across industries. The use of surveys of corporate managers and the examination of citations by patents became more frequently used methods than geographic analysis[6]; the urban and regional literature focused on academic spillovers as a source of agglomeration economies. Although geographic econometric analysis has been employed by some studies, most have either ignored the confounding of academic and industrial R&D or used an identification method similar to Jaffe (1989); I am unaware of studies that have considered identification with respect to the location of faculty talent and other unobservables. Two of the more important studies are discussed below.

Anselin, Varga and Acs (1997) examined spillovers at the level of metropolitan/technology area, for one year. They used similar instruments as Jaffe to isolate variation in academic R&D that is unrelated to industry R&D: student enrollment, and academic expenditures. Their core instrumented finding was an elasticity of innovation with respect to academic R&D of 0.259. Consistent with the critique of

[6] See Appendix B for a short summary of this literature; detailed reviews are contained in Agrawal (2001) and Salter and Martin (2001)

Jaffe's use of patent counts, their measure of innovation was a count of innovations constructed by the U.S. Small Business Innovation Database, for 1982[7]. This unique innovation measure was limited to that single year, which prohibited use of variation over time. They were also able to investigate the effect of distance on spillovers by using "spatial lags" of the explanatory variables, which turn out to have less of a positive effect than R&D conducted near the center of an MSA.

A more recent study is from Agrawal and Cockburn (2003). By limiting their study to three narrow technological areas in electrical engineering, they were able to closely link publishing in those areas (creation of new knowledge) to patents in those areas. They find a high degree of geographic co-location of patenting and publishing, and that the presence of an "anchor tenant"[8] increases the degree of co-location. While the above studies interpreted this co-location as evidence of a causal spillover, this study explicitly acknowledged the difficulties in assigning causality. As Agrawal and Cockburn put it,

> "Though there are good reasons to believe that papers "cause" patents in the sense that downstream industrial R&D activity relies on upstream science, it is quite possible that causation runs in the opposite direction. We have not specified a production function technology for R&D nor made any assumptions about the behavior of actors in this process." (Agrawal and Cockburn, 2003, pp. 1243-45)

The present study is not completely agnostic about the mechanisms for spillovers. Much is known about the channels of academic knowledge spillovers, and the applicability of an econometric approach to measure them is considered in section III.

[7] This was a one-time assessment of innovations listed in trade journals.
[8] Agrawal and Cockburn define an "anchor tenant" as a company with both at least one patent in a particular narrow tech. area and at least one thousand patents in general.

Identification with a different approach than that of Jaffe (1989) and Anselin, Varga and Acs (1997) is detailed in section IV.

III Discussion of Tacit Knowledge and Local Spillovers

In the jargon of recent papers in economics and management, tacit knowledge refers to a "component" of knowledge that is "costly or impossible to codify" (Agrawal, 2001, pp. 291). This is a simplification of the original meaning of tacit knowledge, conceived by Polanyi (1958) as knowledge subjective to the bearer, of which the bearer may not even be conscious. Cowan, David and Foray (2000) provide a detailed discussion of tacit knowledge in the context of economics, including a nuanced discussion of what it means for knowledge to be codified. Because of the existence of such tacit components of knowledge, researchers have a greater ability to communicate their knowledge through personal interaction than through written forms, which require a common jargon to understand. Academic researchers, in their unique standing at the frontier of knowledge, may have to be physically present in the development of new products to share their tacit knowledge. Jensen and Thursby (2001, pp. 240-241) acknowledge this fact in their analysis of university licensing:

> "Perhaps the most striking result of the survey is that when they are licensed, most university inventions are little more than a 'proof of concept.' No one knows their commercial potential because they are in such an early stage of development. Indeed, they are so embryonic that additional effort in development by the inventor is required for a reasonable chance of commercial success."

Due to travel costs, such collaboration will occur near where the researcher lives, and hence cause local knowledge spillovers. But what constitutes "locality"? If knowledge dissemination comes about through repeated personal interaction or requires an employment relationship, a reasonable definition of locality would be commutable distance from a city center. In this study, the geographic unit of observation will be referred to as a “metropolitan area”, which is either simply the Metropolitan Statistical Area (MSA) in which a university or research institute is located, or a Core-Based Statistical Area (CBSA) of the Census Bureau, which consist of several contiguous Metropolitan Statistical Areas. In a few cases, the official MSA or CBSA was modified to include neighboring counties or MSA, if such counties had a population center and were not part of another MSA or CBSA. This was to keep an approximate two-hour driving rule - one can argue that some researchers will commute further, but this definition captures the idea of separated, regional labor markets. Figure 1 shows the MSA and CBSA in the state of Virginia used in this study in heavy dotted lines, as an example of these geographic definitions[9]. In the upper-right corner is the Washington-Baltimore-Northern Virginia CBSA. Bordering to its south is the Richmond MSA, itself bordered by the Charlottesville and Virginia Beach-Norfolk-Newport News MSAs. To the left is the Blacksburg-Christiansburg-Radford MSA combined with the Roanoke MSA. In the far lower-left, the Johnson City-Kingsport-Bristol MSA contains counties in Virginia, but was not included in the sample because it lacked a research institution.

Although in the following analysis production functions of industrial patents will be estimated, these are clearly reduced-form equations that do not estimate structural parameters. Intuitively, and supported by evidence from surveys and case studies, the mechanisms for academic knowledge spillovers seem clear: discoverers of new knowledge (professors and graduate students) are hired by firms to assist with the

[9] Map for Figure 1 from “Virginia – Core Based Statistical Areas, District of Columbia, Counties and Independent Cities”, U.S. Dept. of Commerce, Economics and Statistics Administration, U.S. Census Bureau

development of new products and processes. Although some knowledge may be transferred through publishing, due to the tacit component of knowledge on the frontier of research, employment or consulting relationships must be a factor causing local spillovers.

IV Identification and Empirical Method

Identification of a causal effect of academic research on industrial innovation has been difficult because of the complexity of the links between academic science and industrial R&D. Estimates of the relationship between academic R&D and measures of industrial innovation confirm what is widely known: there is a high degree of collaboration. More interesting, in terms of its economic implications, is isolation of an academic knowledge spillover effect that is independent of industry influence. Knowledge created independently, from federal or university funding, may be disseminated more broadly than knowledge supported by a single firm. For example, a new technique could be licensed to many firms instead of granting an exclusive license to a single firm that could extract a monopolistic rent from development of the technology.

Ideal instrumental variables in this context would be related to the level of academic knowledge but would be orthogonal to variation in the error term caused by unobserved factors connected with both academic science and industrial innovation. The instruments used by Jaffe (1989)[10] and Anselin, Varga and Acs (1997)[11] to isolate a knowledge spillover effect are clearly positively correlated with the level of academic

[10] Jaffe's instruments for academic R&D include the number of public and private universities per state, the number of federally funded R&D centers and state population.
[11] The instruments employed by Anselin, Varga and Acs for academic R&D include student enrollment and higher education expenditures per MSA.

knowledge, but reflect conditions that may be influenced by the level of industrial innovation. The number of private universities in a state, for example, could be correlated with the quality of faculty in science and engineering in the same state, and hence be correlated with the quantity of federal research grants and overall academic R&D. Instruments reflecting demand for education, such as population or educational expenditures, could also be influencing the quality distribution. Indeed, Jaffe admitted this possibility, writing "Despite the attempt to control for unobserved 'quality' of universities, one cannot really interpret these results structurally, in the sense of predicting the resulting change in patents if research spending were exogenously increased. (Jaffe, 1989, pp. 968)"

More satisfactory instruments would be variables that are plausibly unrelated to the industrial funding of academic R&D, the talent of researchers in a regional labor market and other unobserved concomitant factors, but yet have a connection to the level of research performed there. One such channel is in the allocation of federal research funding. Although 95-99% of federal research funding is done through the peer review system[12], members of Congress can "rig" the process so that more peer-reviewed funds are diverted to institutions in their states or districts of representation. In an empirical study of the effect of congressional representation on academic R&D, Payne (2003, pp. 326) claims that "as much as 39% of federal research funding is diverted for reasons associated with the representation of one's constituents." Savage (1991) found that members of the appropriations committees are especially influential, since appropriation subcommittees manage the size of the budget of federal agencies. Since certain federal

[12] A back of the envelope calculation from dividing of Academic Earmark Grant totals reported in Savage (1999) by the NSF's reported values for total research support.

agencies are more likely than others to fund research in particular locations, this allows Congress to direct research funds to agencies likely to aid institutions connected to their own constituents.

The makeup of congressional appropriations sub-committees is plausibly unrelated to local factors influencing academic and industrial research. Membership on such committees is based on the sharing of power within Congress, and many members retain their positions for decades. Loss of a committee seat is usually the result of a lost election or a retirement, which is based on local political conditions or personal conditions. Also, research funding is only a small portion of the federal budget, and members of congress seek seats on the appropriations committee with an eye for distributing the whole budget. Given such facts, it seems unlikely that committee membership would be related to industrial research and local labor market conditions. Indeed, Aghion, Boustan, Hoxby and Vandenbussche (2005) use congressional representation dummy variables as instruments for federal research funding in their study of the effects of educational spending on state economic growth, under the premise that such variables are unrelated to unobservable conditions affecting state economic growth.

Therefore, to identify an academic knowledge spillover effect, this study will make use of the connection between appropriations committee membership and academic R&D by using indicators of congressional representation as instruments for measures of scientific knowledge: publication counts and publications citations counts. An additional instrument included is an actual measure of politically designated research funds, "academic earmarked grants". These are grants to universities and research institutes that are written directly into the federal budget by the members of the appropriation committees, bypassing the peer-review system[13]. This instrument is not as clearly unrelated to the unobservables as representation dummies: lobbying effort by universities plays a role in their distribution, and such lobbying may be influenced by industrial interests. However, there is substantial arbitrariness in the distribution of academic earmarked grants. Such grants tend to go to institutions with congressional representation, and their magnitude depends on the size of specific projects targeted by federal agencies. In an analysis of earmarked grants by Savage (1999), the University of Hawaii, University of Pittsburgh and Iowa State University were revealed to be the recipients of the most earmarked dollars over the 1980-1996 period. The University of West Virginia, with powerful Senator Robert Byrd as Senate Appropriations Chair for many of those years, came in fourth. In contrast, MIT received a relatively small $21,625,000 in independent earmarks over the 1990-2003 period, $20 million of which was from a single defense-related grant[14]. This is evidence that the pattern of academic earmarked grants is highly influenced by appropriation committee membership as well. This instrument is interesting because the quantity of federally influenced research funds is known; its inclusion or exclusion does not greatly affect the overall power of the instruments or the two-stage least squares results, as shown in Table 3.

The discussion above is centered on the use of instrumental variables to isolate a causal spillover effect, but its identification also depends crucially on the measurement and empirical method used. The general form of empirical model that will be estimated is given in equation 1:

[13] The size of earmarked grants has grown considerably in the past twenty-five years: Savage (1999) documented an increase from $16 million in 1980 to $727 million in 1993.
[14] Data source: *Chronicle of Higher Education*

$$I_{i,j,t} = \beta * K_{i,j,t} + X_{i,j,t}'\phi + \alpha_{ij} + \gamma_j * t + \varepsilon_{i,j,t} \qquad (1)$$

Here, *I* represents private-sector innovation, *K* is a measure of academic knowledge, and *X* is a vector of control variables for innovative characteristics (such as private sector R&D or employment of researchers) that vary by metropolitan area (i), technological area (j) and year (t). The measure of private-sector innovation used is a count of patents assigned to U.S. corporations, while two measures of academic knowledge are alternatively used: academic publication counts and counts of citations to those publications. The fixed effects, alpha, are included to control for differences in size between metro areas and technology areas, while the technology trends, gamma, are included to control for heterogeneity in the rate of patenting activity across technological areas. Here beta is the parameter of interest: the rate of knowledge spillovers or the direct local marginal effect of knowledge discovery on innovation.

This empirical model allows estimation of spillover effects at the (local) metropolitan level, which adheres to the logic of tacit knowledge discussed in section III. This is the first study to use publication counts in all fields of science and engineering as a measure of academic knowledge, a measure which is an observed outcome of the discovery process. By using a relatively long panel dataset, it is possible to include the fixed effects and trends. The fixed effects control for size differences in patenting across metropolitan areas that can't be controlled for with estimates of industry R&D alone. Controlling for trends in patenting activity in different technological areas allows different technological categories to be included in the same regression. Such trends may arise for reasons unrelated to R&D: the scope of patenting may change as patenting strategy changes within an industry, or patent protection may be extended to previously unpatentable inventions. Hall, Jaffe and Trajtenberg (2001) note that in the aggregate, the fields of “Computers and Communications” and “Drugs and Medical” have seen a rapid increase in the share of patents attributed to them in the 1990’s. This suggests that in studying patenting activity across all technologies it is important to control for technology-specific trends.

V. Data[15]

The empirical analysis makes use of a panel dataset comprising 108 metropolitan areas, five technological areas and years from 1977 to 1999, and an aggregate panel that combines the technological areas. Since congressional appropriation committee representation cannot be linked to any one technological area, the instrumental-variables analysis was done with the aggregate panel and a separate panel for each technology area. The reasons for the dimensions of the panel will be made clear in the following descriptions of the individual variables.

Corporate patent counts serve as the measure of innovation and the dependent variable. Data on patent counts was obtained from the NBER patent dataset, as detailed in Hall, Jaffe and Trajtenberg (2001)[16]. Patents assigned to US businesses and indicating a US residence of the first inventor were used; they were distributed geographically by the county of the city of the first inventor, the counties being assigned uniquely to metropolitan areas. Patents assigned to counties outside of any included metropolitan area were ignored. Year assignments were based on application year of eventually

[15] See appendix A for details of construction of variables

[16] The current NBER Patent data covers patents granted from 1963 to 1999, but additional data on patents to 2002 were available from Bronwyn Hall’s website. Inventor location data for 2000-2002 was compiled by this author from the USPTO’s Cassis database and is available upon request.

granted patents, to avoid undue lag issues associated with grant dates. Citation weighted patents would have been preferable to better capture the quality of innovations, but due to the long lag in citations and the shortness of the panel, it was not feasible.[17] Patents were assigned to five technological areas, “Drugs and Medical”, “Chemical and Synthetic Materials”, “Electrical, Sensing and Computing”, “Mechanical and Transport” and “Agricultural, Mining, Environmental and Other”, based on their NBER assigned subcategory. These technological areas were kept broad in order to facilitate matching of the explanatory variables.

A caveat is in order when using patent data to measure innovation. Yes, patents vary widely in their import and generality, but limiting the set of patents to U.S. corporate patents in specific high-tech industries may serve to reduce this variation. In addition, with a mean of 50 patents per observation in the metro/technology area/year panel and 274 patents per observation in the aggregate metro/year panel, much of this variation in quality will be averaged out. As for variation in propensity to patent by technological area, as mentioned above, tech-specific trend variables in the disaggregated panel should help mitigate this variation. In the aggregate, a separate panel was constructed for each technology area, to observe differences in the spillover rate.

Publication counts, and the counts of the number of citations received from other publications, serve as measures of scientific knowledge[18]. This data was obtained from automated searches on *Web of Science*, using a Perl script[19]. Publication counts were collected for 218 research universities and medical schools, 98 non-profit research

[17] The average length from application year to year of citation is roughly eight years.
[18] Knowledge being what it is, we can not expect to quantify it exactly, nor expect a one-to-one match between publishing and patenting.
[19] With grateful acknowledgement of Richard Beaudoin for programming assistance.

institutes and 5 federally funded research and development centers, all within the U.S[20]. This set of institutions included those that had received at least $10 million in federal research grants in 2003.

Metropolitan areas were included in the sample if they contained at least one included institution, and it is important to note the wide geographic distribution of these research institutions. Every state contains at least one university in the sample, and most are located in non-contiguous metropolitan areas. Large cities, however, benefit from containing most independent non-profit research institutions and medical schools. Where metropolitan areas are contiguous and small geographically, they are combined into CBSA, which should limit measurement error due to overlapping regional labor markets[21].

Academic knowledge spillovers should be at best only a small factor contributing to industrial innovation. Two measures were used to control for the primary factor, industrial research and development expenditures. Payroll in Scientific R&D services was compiled from the Census Bureau's *County Business Patterns* (CBP) data. This measure is proportional to the level of subcontracted R&D. While it measures a part of R&D expenditures accurately, that part is relatively small compared to the "in-house" component of R&D. However, the "in-house" component is not available, to the best of my knowledge, at the establishment level due to strategic needs for secrecy. A second R&D control variable, which will be called "Industry R&D (estimate)", was constructed by combining information on R&D at the state level from the NSF's *Survey of Industrial*

[20] See appendix A for a correspondence between fields of science and the five technological areas used in the estimation.
[21] In most cases, the universities and institutes themselves were located near the geographic center of the metropolitan area.

Research and Development (SIRD) and information at the county level from the (CBP)[22]. Both the SIRD and the CBP are useful data sources as they have a time dimension – the CBP was accessible annually back to 1977 and the SIRD biennially to 1973. This variable provides a comprehensive estimate of R&D by field, but it does so with significant measurement error, as the state level SIRD data is hampered by many firms choosing not to self-report their R&D expenditures and it relies on strong assumptions about the within-state and within-industry distributions of R&D. The two proxies for R&D are positively correlated however, as shown in Table 1.

Another important control variable is the number of new Ph.D. graduates. This variable was constructed through licensed access to the NSF's *Survey of Earned Doctorates*, which is an extensive survey of both the graduate education and post-doctoral plans of all people who earn doctorates in the U.S. Recipients were assigned to the same metropolitan area as the university of matriculation. Although it is common for new Ph.D. recipients to take jobs outside of the regional labor market in which they graduated, in an analysis of the SED done by Stephan, Sumell, Black and Adams (2004, pp.162) it was shown that more than 40% take jobs in the same state as graduation. This suggests that the count of new Ph.D. recipients will be positively correlated with the count of new Ph.D. recipients hired locally. This source of human capital may conceivably have a different effect than researchers hired in general (i.e. not be redundant to the R&D measures), as new Ph.D. recipients may have been engaged in the creation of new knowledge at their academic departments and be themselves mechanisms for spillovers.

[22] Details of construction of this variable are reported in appendix A.

The main instrumental variables used were indicators for whether an observation had congressional representation on subcommittees of the Appropriation committees, subcommittees that are linked to the funding of academic research. Lists of appropriation committee members were compiled from volumes of the *Congressional Staff Directory*, plus information on their state and district of representation, party affiliation and status as chairperson or ranking minority member. Savage (1991) found that members of the appropriation subcommittees for Agriculture and Defense play a significant role in determining where research funding is allocated, and Payne (2003) notes that certain subcommittees oversee the budget for the major agencies that fund research: the subcommittee on VA, HUD and Independent Agencies oversees the NSF and National Institutes of Health. Payne (2003) also documented statistically a relationship between subcommittee membership and federal research funding at the university level. For each of four subcommittees[23], four variables created include indicators for House and Senate general membership, and House and Senate chair status. Every metro area in a state was considered to be represented simultaneously by a Senator, while metro areas were considered to be represented by a House member if the district of that member either overlapped with or bordered that metro area.

As a final instrumental variable, data on academic earmarked grants comes from two sources: the Chronicle of Higher Education, for the period 1990-2003, and James Savage for the period 1980-1996. Neither source is an exhaustive list of earmarked grants, but the Chronicle lists 11,161 distinct appropriations, while Savage found 3,788 for the earlier period. In the overlapping period, 1990-1996, the listed grants are

[23] Subcommittees included were Agriculture, Defense, "VA, HUD and Independent Agencies" and "Labor, Heath and Human Services and Education"

generally similar, with the Chronicle reporting 10-20% more per year[24]. Grants were assigned to metropolitan and year by the location of the grantee institution, and metro/year aggregates were used in the following analysis[25].

VI. Empirical Results

OLS regressions measuring the general rate of collaboration at the most disaggregated level, that of metropolitan, technology area and year, are reported in Table 2. Since the congressional representation instruments do not vary by technology area, that analysis did not include the instrumental variables. They were used in two-stage least squares regressions estimated at the metro/year level and reported in Tables 4-6. Table 3 presents the results of the first stage of instrumental variables regression, to show the power of the instruments under different conditions.

In column 1 of Table 2 we see OLS estimates of the production function in equation (1), but without fixed effects or trends. The estimate of beta shows a general collaborative effect of one additional corporate patent for an increase of 42.9 publications. At the mean levels of patenting and publishing, the elasticity of patents with respect to publishing is 0.13. All of the control variables have a positive and statistically significant effect on patenting as well: an extra doctorate earned is associated with an extra 0.7 patents, the marginal effect of a million dollars of "in-house" R&D is 0.037 patents, while that of a million dollars of subcontracted R&D is 0.14 patents. Inclusion of the fixed effects, in column 2, causes all of these estimated effects to be

[24] The difference was due to different definitions of an "earmark". I use the Chronicle data from 1990, excluding appropriations not related to science and engineering. The difference is not quantitatively important to the present study.

[25] An attempt was made to assign earmarked grants to technological areas. A rough assignment was possible for the Chronicle data, but it was ultimately unnecessary for the analysis.

diminished. Inclusion of the trend variables in column 3 serves to diminish the effect of the control variables, while increasing the effect of publications. The size of the trend coefficients are interesting to note: relative to patenting in Drugs and Medicine, only patenting in the Chemical technology area declined over the period 1977-1999. On the other hand, Electrical and Computing patents saw an average increase of 4.09 patents per metropolitan and year over the increase in Drugs and Medicine. This is after controlling for industrial R&D, and so represents increased patenting for other reasons, most likely shifting incentives for patenting due to strategic or legal changes. Column 4 omits the variable “Payroll in Scientific R&D Services”, and we see the coefficient on publications increase dramatically, showing that the measured spillover effect is particularly sensitive to this control variable.

The simple regression without fixed effects or trends is again estimated in column 5, but this time with citation-weighted publication counts. The estimated effect shows that an increase in citations of two thousand is associated with an increase of one patent, or an elasticity at the means of 0.09. Although two thousand citations may seem like a lot for a one patent gain, it is not uncommon for one seminal paper in a major field of science to have one thousand or more citations. This estimate explicitly shows that the quality of academic research matters for knowledge spillovers. This effect disappears in column 6 with the inclusion of fixed effects, but re-emerges in column 7, although diminished, with the additional inclusion of trends.

The results in Table 2 confirm a connection between academic research and industrial innovation, which persists even after the inclusion of controls for trends and the quality of the research. However, this is not evidence that the direction of the effect is from academic research to industrial innovation. Table 3 shows the results of first-stage regressions, with publication and publication citation counts regressed on the other control variables and the set of instruments[26]. From the F-statistics on exclusion of the instrument set in the range of 5.77 to 8.08, we see that they are collectively powerful in predicting publications and publication citations. The instruments associated with the Defense and Agriculture subcommittees are particularly powerful (the presence of a general member on the Agriculture subcommittee is associated with an extra 118 publications in column 1), while chair status on the HUD and Independent Agencies subcommittee is associated with a decrease in publishing activity[27]. This differential effect by subcommittee was noted, in terms of its effect on research funding, by both Savage (1999) and Payne (2003). Here, it surfaces again in an effect on publishing, at the level of metropolitan area. In columns 2 and 5, we see that exclusion of the "new PhD" variable causes the individual coefficients on the instruments to change, but does not highly affect their overall power. We see that when the new PhD count is excluded, the earmarked grants variable is positive and statistically significant in both columns 2 and 5, so that an additional million dollars is associated with 6.3 extra publications. Column 3 is included to show that the power of the instrument set is little affected by the exclusion of the earmarked grants variable.

In Table 4, the full instrument set is used to predict the knowledge measures in two-stage least squares regressions. Column 1 confirms that the basic OLS coefficients have the same sign and statistical significance in this aggregate panel as when using the disaggregated panel. The estimated coefficient on publications in column 2 shows a

[26] All instruments have been lagged two years to reflect the time between grant funding and publication.
[27] This negative effect was documented by Savage (1991) as an effort by "saintly" chairs to restrain pork-barrel spending.

marginal effect of 0.206, or 1 patent for every 5 publications, an elasticity of 1.17. This effect is much higher than the disaggregated result, and is probably due to inability to control for differences in patenting across technology areas as in Table 2. After implementing two-stage least squares, the marginal effect is reduced to 0.168. While the estimate is not precise enough to rule out the possibility that the true coefficient is the same as the OLS estimate, it is suggestive of a positive bias in the OLS estimate, showing the effect of endogeneity to be inflating the estimate of spillovers by almost 25%.

With citation-weighted publications, a quite different result emerges. While the OLS fixed effects coefficient of 0.003 in column 6 shows that an extra 333 academic citations are associated with an extra patent, an elasticity at the means of 0.55, the identified spillover effect in column 7 is negative and marginally statistically significant. This spillover estimate is precise enough to rule out parity with the general effect, as the estimate in column 6 is not within the 95% confidence interval of the IV estimate. Another feature of the regression in column 7 is that the estimated effect of a new PhD has risen dramatically. Since this variable may also be endogenous, column 8 shows the regression without the new PhD variable, and the estimated coefficient on citation-weighed publications is now indistinguishable from zero. Collectively, these results suggest that the local innovative benefit of independent academic research is less than academic research in general.

What conclusion should be drawn from this disparity in the results between publications and citations? Examining the coefficients on the new PhD count, another potentially endogenous variable, we see that the PhD variable is positive and significant in columns 6 and 7, but not statistically significant in columns 2 and 3. Publication counts may be in effect measuring the same thing as the new PhD count, the "quantity" of new ideas. Also we see that the estimate of the marginal effect of a new PhD increases dramatically when going from OLS to 2SLS (columns 2 to 3, and 6 to 7). This emphasizes the idea that the knowledge measures are highly related to the new PhD count, as its effect is greater when only exogenous variation in publications and citations is used.

Table 5 presents regressions done with panels constructed using each technological area separately (i.e. the number of patents in Drugs and Medical regressed on publications in Drugs and Medical, etc). Here the full set of non-technology specific instruments is used to predict the number of publications in a particular technological area. The first thing to note is that the magnitudes of the OLS coefficients on publications vary markedly by technological area. The general collaborative effect is largest for Chemical and Mechanical industries. The estimated spillover effect from implementing two-stage least squares varies as well, with that of Chemical, and Electrical and Computing, actually increasing relative to the OLS. This increase is consistent with Jaffe (1989), which also found an increase in the IV spillover effect for those areas, relative to the OLS. It may be true that the marginal effect of independent research is greater than that of collaborative research for those technology areas. Only in Drugs and Medical do we see any evidence of attenuation of the estimated coefficient as in the aggregate regression. The 2SLS coefficients in columns 8 and 10 are too imprecisely measured to tell the direction of the bias for those areas.

Table 6 again shows that the spillover effect is greatest for Chemicals. Again for Drugs and Medical we see attenuation in the effect towards zero, but now this is also true for Agriculture, Mining and Environmental. For Electrical and Computing, and Mechanical, the 2SLS coefficient is too imprecisely measured to make an assessment of the bias.

VII. Conclusion and Future Research

This study examines the link between academic science and industrial innovation, and isolates a causal spillover component of the general relationship with the application of novel instrumental variables to reduced form production functions. Unlike previous studies, it uses measures of scientific knowledge that are more direct than R&D – publication counts and publication citation counts. Even after implementing metropolitan/technology area fixed effects and technology area trends, evidence of a positive association between academic knowledge and industrial innovation is found. Using instrumental variables, I identify a spillover effect that is generally smaller than the general effect. For industries involved with chemical and electrical and computing technologies, the measured spillover effect is larger than the general effect, suggesting that independent, basic research has a greater marginal effect on local patenting than academic research funded by industry. For other areas of technology, especially "Drugs and Medical" the measured spillover effect is substantially smaller than the general effect, suggesting that the local innovative effect of academic research in those areas is largely generated by collaboration between academia and industry. New PhD graduates are also shown to be positively related to local rate of innovation, although this effect is diminished after controlling for the number of academic publications, suggesting that the number of new PhD graduates may be another measure of academic knowledge, and a channel for academic spillovers.

These results are preliminary, as further robustness checks need to be done in considering the results. The identifying restriction that the effect of academic knowledge is limited to the metropolitan area in which it was discovered could be relaxed by creating publication measures at the state or census division level. The sample could also be split by the population of metropolitan areas, to determine whether large urban areas exhibit a different relationship than smaller, isolated metros. Non-linear effects of publishing could also be examined.

Further research needs to be done on the role of new PhD graduates as a channel of spillovers, as their numbers may also be endogenous to the local rate of innovation. Subsequent drafts of this paper may include instrumentation of the new PhD count, since enrollment may also respond to diversion of federal research funding.

A companion paper will model academic-industry collaboration and spillovers as a search between researchers embodying new knowledge and collaborating firms. Hopefully such an effort will lead to greater understanding of the channels of knowledge transfer and guide future empirical research in academy-industry research collaboration.

References

Acs, Z., Audretsch, D. and M. Feldman "Real Effects of Academic Research: Comment" *The American Economic Review*. V. 82, n. 1 (Mar. 1992), pp. 363

Adams, J. D. "Comparative Localization of Academic and Industrial Spillovers" *Journal of Economic Geography*, V. 2 (Jul 2002), pp. 253

Aghion, P., Boustan, L., Hoxby, C. and J. Vandenbussche "Exploiting States' Mistakes to Identify the Causal Impact of Education on Growth" Working Paper, UCLA Economics Online Papers (August, 2005)

Agrawal, A. "University-to-Industry Knowledge Transfer: Literature Review and Unanswered Questions" *International Journal of Management Reviews* V. 3, I. 4 (December, 2001), pp. 285

Agrawal, A. and I. Cockburn "The Anchor Tenant Hypothesis: Exploring the Role of Large, Local, R&D-intensive Firms in Regional Innovation Systems" *International Journal of Industrial Economics* V. 21 (2003)

Anselin, L., Varga, A. and Z. Acs "Local Geographic Spillovers between University Research and High Technology Innovations" *Journal of Urban Economics*, V. 42 p. 422 (1997)

Audretsch, D. and M. Feldman "R&D Spillovers and the Geography of Innovation and Production" *American Economic Review* V. 86, n.3 (June 1996), pp. 630

Branstetter, L. and Y. Ogura "Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations" Working Paper 11561, NBER (2005)

Cohen, W. M., Nelson R. and J. P. Walsh "Links and Impacts: The Influence of Public Research on Industrial R&D" *Management Science* V. 48, n. 1 (Jan 2002), pp. 1

Cowan, R., David, P., and D. Foray "The Explicit Economics of Knowledge Codification and Tacitness" *Industrial and Corporate Change* V. 9, n. 2 (2000)

Hall, B., Jaffe, A., and M. Tratjenberg "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." Working Paper 8498, NBER (2001)

Henderson, R., Jaffe, A. and M. Trajtenberg "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988" *Review of Economics and Statistics* V. 80, n.1 (Feb. 1998) pp. 119

Jaffe, A. "Real Effects of Academic Research" *American Economic Review* V. 79, n.5 (December 1989), pp. 957

Jaffe, A., M. Trajtenberg and R. Henderson "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations" *Quarterly Journal of Economics* V. 108 n. 3 (August, 1993) pp. 577

Jensen, R. and M. Thursby "Proofs and Prototypes for Sale: The Licensing of University Inventions" *American Economic Review*, V. 91 n.1 (Mar. 2001)

Kim, J., Lee, S. J. and G. Marschke "The Influence of University Research on Industrial Innovation" Working Paper 11447, NBER (2005)

Mansfield, E. "Academic Research Underlying Industrial Innovations: Sources, Characteristics and Financing" *Review of Economics and Statistics*, V. 77, n.1 (Feb., 1995), pp. 55-65

Pakes, A. and Z. Griliches "Patents and R&D at the Firm Level: A First Look" in Z. Griliches, ed., R&D, Patents and Productivity, Chicago: University of Chicago Press, 1984, ch. 3.

Pakes, A. and Z. Griliches "Patents and R&D at the Firm Level: A first Report" *Economics Letters* V. 5 pp. 377 (1980)

Payne, A. "The Effects of Congressional Appropriation Committee Membership on the Distribution of Federal Research Funding to Universities" *Economic Inquiry* V. 41, n.2 (April 2003)

Polyani, M. Personal Knowledge: Towards a Post-Critical Philosophy, Chicago: University of Chicago Press, 1958

Salter, A. and B. Martin "The Economic Benefits of Publicly Funded Basic Research: A Critical Review" *Research Policy* V. 30 (2001) p. 509

Savage, J. "Saints and Cardinals in Appropriations Committees and the Fight Against Distributive Politics" *Legislative Studies Quarterly* V. 16 n. 3 (1991) pp. 329

Savage, J. Funding Science in America: Congress, Universities and the Politics of the Academic Pork Barrel, Cambridge, United Kingdom: Cambridge University Press, 1999

Varga, A. "Local Academic Knowledge Transfers and the Concentration of Economic Activity" *Journal of Regional Science* V. 40, n. 2 (2000) pp. 289

Figure 1.

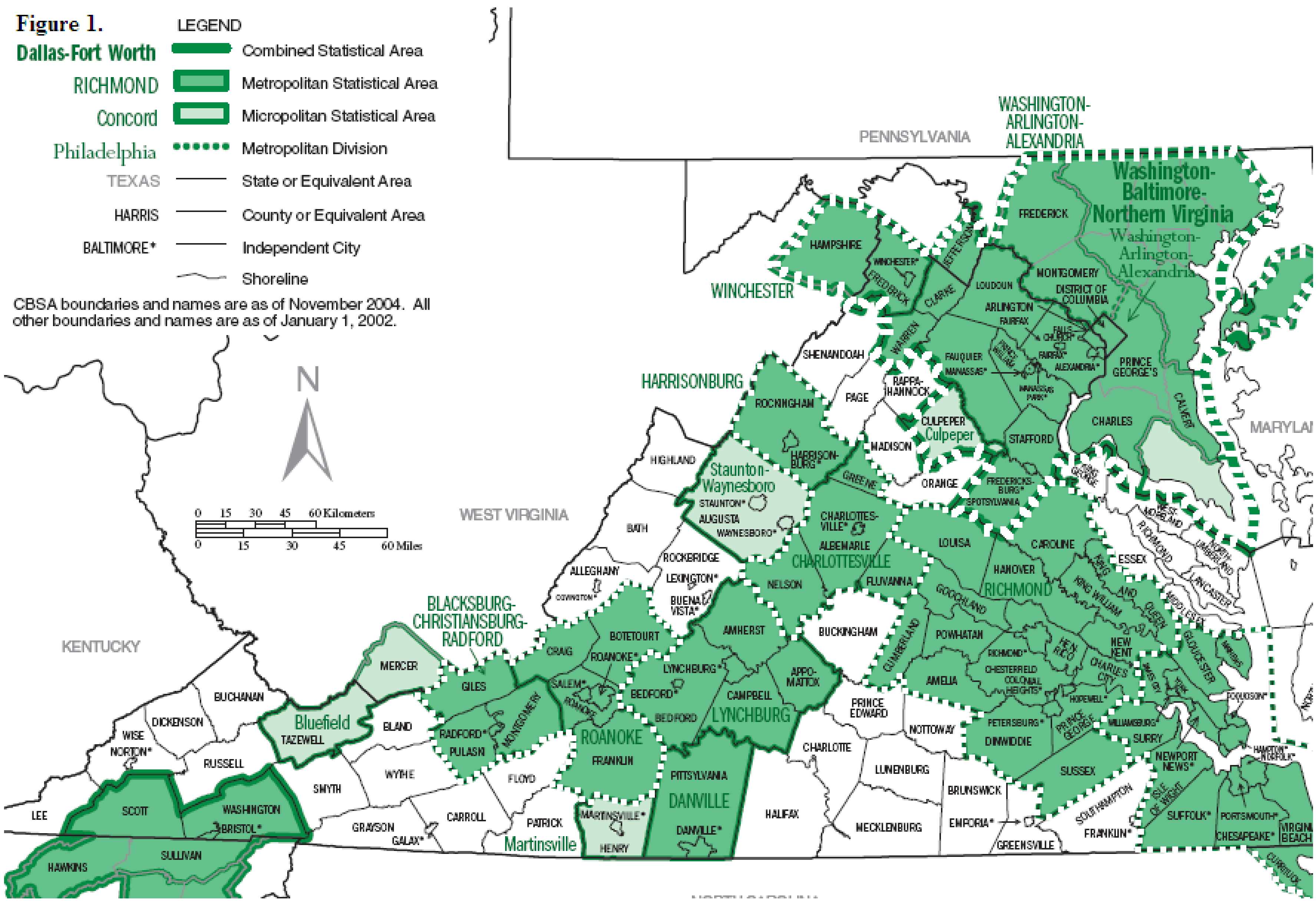
LEGEND
Dallas-Fort Worth
Combined Statistical Area
RICHMOND
Metropolitan Statistical Area
Concord
Micropolitan Statistical Area
Philadelphia
Metropolitan Division
TEXAS
State or Equivalent Area
HARRIS
County or Equivalent Area
BALTIMORE*
Independent City
Shoreline
CBSA boundaries and names are as of November 2004. All other boundaries and names are as of January 1, 2002.
N
0 15 30 45 60 Kilometers
0 15 30 45 60 Miles
PENNSYLVANIA
WEST VIRGINIA
KENTUCKY
MARYLAND
WASHINGTON-ARLINGTON-ALEXANDRIA
Washington-Baltimore-Northern Virginia
Washington-Arlington-Alexandria
WINCHESTER
HARRISONBURG
Staunton-Waynesboro
Culpeper
CHARLOTTESVILLE
RICHMOND
BLACKSBURG-CHRISTIANSBURG-RADFORD
Bluefield
ROANOKE
LYNCHBURG
DANVILLE
Martinsville

Table 1: Sample Statistics and Simple Correlations

	5 tech. area panel (Used in Table 2)					Metro-year Aggregate panel (used in Table 4)				
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Patent count	12190	50.00	142.67	0	4136	2462	247.54	608.99	0	8934
Publications	12190	284.43	566.08	0	8929	2462	1408.29	2178.50	0	18763
Publication Citations	12190	9197.40	24842.97	0	406502	2462	45538.72	90773.11	0	775727
New PhD count	12190	28.80	48.94	0	556	2462	142.59	196.60	0	1305
Industrial R&D (est.)	12130	159.06	857.81	0	25308.3	2462	783.66	3252.29	0	46522.56
Payroll in Scientific R&D Services	12130	72.96	286.05	0	5882.59	2462	359.44	1420.68	0	29412.95

Simple Correlations	Pat. Count	Pub. Ct.	Cit.-wgt.	PhDs	Ind. R&D	Sci. R&D	Pat. Count	Pub. Ct.	Cit.-wgt.	PhDs	Ind. R&D	Sci. R&D
Patent count	1.00						1.00					
Publications	0.52	1.00					0.77	1.00				
Publication Citations	0.48	0.96	1.00				0.77	0.96	1.00			
New PhD count	0.54	0.85	0.79	1.00			0.70	0.86	0.84	1.00		
Industrial R&D (est.)	0.51	0.51	0.46	0.50	1.00		0.65	0.59	0.57	0.50	1.00	
Payroll in Scientific R&D Services	0.49	0.43	0.33	0.40	0.48	1.00	0.55	0.53	0.45	0.44	0.63	1.00

Table 2: OLS Regressions at level of Metropolitan Area, Technology area, Year

Dependent Variable: Patent Counts	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Publication Count	0.0233***	0.0165***	0.0293***	0.0748***				
	(0.0033)	(0.0057)	(0.0059)	(0.0059)				
Citation-weighed Publication Count					0.0005***	-0.0001	0.0003**	0.0005***
					(0.0001)	(0.0001)	(0.0001)	(0.0001)
New PhD Count	0.7033***	0.6609***	0.4679***	0.6213***	0.7180***	0.7949***	0.5724***	0.9954***
	(0.0379)	(0.0697)	(0.0723)	(0.0751)	(0.0338)	(0.0655)	(0.0703)	(0.0723)
Industrial R&D (estimate)	0.0370***	-0.0272***	-0.0266***	-0.0005	0.0370***	-0.0254***	-0.0252***	0.0072***
	(0.0014)	(0.0025)	(0.0025)	(0.0025)	(0.0014)	(0.0025)	(0.0025)	(0.0024)
Payroll in Scientific R&D Services	0.1409***	0.1203***	0.1096***		0.1443***	0.1228***	0.1137***	
	(0.0040)	(0.0036)	(0.0035)		(0.0040)	(0.0034)	(0.0034)	
Trend: Chemical			-1.3034***	- 1.6797***			-1.0551***	- 0.9128***
			(0.2524)	(0.2625)			(0.2474)	(0.2590)
Trend: Electrical and Computing			4.0881***	4.9485***			4.0967***	5.0282***
			(0.2298)	(0.2375)			(0.2304)	(0.2394)
Trend: Mechanical			1.1528***	1.3461***			1.2570***	1.6210***
			(0.2447)	(0.2547)			(0.2447)	(0.2558)
Trend: Mining, Ag. and Env. Services			1.3219***	1.9402***			1.3403***	1.9992***
			(0.2310)	(0.2396)			(0.2322)	(0.2421)
Observations	12010	12010	12010	12010	12010	12010	12010	12010
Number of Cross-Sections	530	530	530	530	530	530	530	530
Metro/Tech. area fixed effects?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
R-squared	0.45	0.15	0.18	0.11	0.45	0.15	0.18	0.10

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: Trends are relative to patenting in the excluded trend, Drugs and Medical

Table 3: First stage Instrumental Variables Regressions

		(1)	(2)	(3)	(4)	(5)
Dependent Variable:			Publication Counts		Citation-Weighed Publication Counts	
F-test of Excluded Instruments		6.46	5.77	6.88	6.58	8.08
(P-value)		0.00	0.00	0.00	0.00	0.00
Payroll in Scientific R&D Services		0.1139***	0.2646***	0.1139***	-1.4810***	3.9499***
		(0.0096)	(0.0122)	(0.0096)	(0.3698)	(0.4570)
Industrial R&D (estimate)		0.0888***	0.1480***	0.0888***	5.0493***	7.1829***
		(0.0094)	(0.0126)	(0.0094)	(0.3609)	(0.4718)
New PhD Count (in Science and Engineering)		8.9683***		8.9722***	323.1454***	
		(0.2021)		(0.1986)	(7.7525)	
Academic "Earmarked Grants"		0.0000	0.0000063***		-0.0000	0.00020***
		(0.0000)	(0.0000010)		(0.0000)	(0.00004)
Appropriations Subcommittee: Agriculture	House member has district in metro/year	118.4269***	75.0615**	118.3161***	4,078.5128***	2,515.9732**
		(24.2719)	(32.9218)	(24.2444)	(930.9658)	(1,228.3250)
	House subcommittee chair has district in metro/year	-91.0490	11.6188	-90.9735	-4,174.5662	-475.2434
		(70.3369)	(95.4289)	(70.3185)	(2,697.8211)	(3,560.4914)
	Senate member represents State of obs.	59.5036**	88.2356**	59.3459**	2,393.8530**	3,429.1249**
		(27.8829)	(37.8401)	(27.8375)	(1,069.4677)	(1,411.8301)
	Senate subcommittee chair represents State of obs.	52.1770	-20.4476	52.3380	1,203.0514	-1,413.7555
		(59.7926)	(81.1365)	(59.7607)	(2,293.3851)	(3,027.2349)
Appropriations Subcommittee: Defense	House member has district in metro/year	98.3092***	175.8563***	98.2937***	2,227.7046**	5,021.8778***
		(24.7670)	(33.5368)	(24.7613)	(949.9550)	(1,251.2698)
	House subcommittee chair has district in metro/year	-17.3072	-95.2226	-17.3102	-1,600.5334	-4,407.9774**
		(44.0492)	(59.7481)	(44.0398)	(1,689.5357)	(2,229.2251)
	Senate member represents State of obs.	-5.1953	110.2829***	-5.1616	936.0321	5,096.9380***
		(23.7661)	(32.0678)	(23.7590)	(911.5644)	(1,196.4638)
	Senate subcommittee chair represents State of obs.	41.1144	-38.9112	41.3107	2,720.2079	-163.2705
		(80.4910)	(109.2371)	(80.4528)	(3,087.2876)	(4,075.6788)
Appropriations Subcommittee: Labor, Health and Human Services, and Education	House member has district in metro/year	-155.2739***	-47.6566	-155.4383***	-4,309.6065***	-431.9457
		(29.6589)	(40.1264)	(29.6123)	(1,137.5872)	(1,497.1303)
	House subcommittee chair has district in metro/year	25.7736	-108.1709	25.6761	2,141.2977	-2,684.9859
		(63.9891)	(86.7669)	(63.9690)	(2,454.3466)	(3,237.3078)
	Senate member represents State of obs.	53.2722*	40.7056	53.3371*	345.4293	-107.3678
		(28.2036)	(38.2837)	(28.1910)	(1,081.7664)	(1,428.3792)
	Senate subcommittee chair represents State of obs.	47.7764	-9.4449	47.8505	-1,915.6681	-3,977.4614
		(51.5536)	(69.9607)	(51.5379)	(1,977.3714)	(2,610.2628)
Appropriations Subcommittee: VA, HUD and Independent Agencies (NSF, NIH)	House member has district in metro/year	-33.7023	7.9418	-33.9319	772.1147	2,272.6338*
		(26.0325)	(35.3154)	(25.9372)	(998.4933)	(1,317.6335)
	House subcommittee chair has district in metro/year	-168.3129***	-414.3174***	-167.8339***	-16,635.2842***	-25,499.3129***
		(57.6839)	(77.9419)	(57.4955)	(2,212.5063)	(2,908.0449)
	Senate member represents State of obs.	-60.2909**	-75.6371**	-60.3034**	-3,601.9144***	-4,154.8683***
		(23.9096)	(32.4532)	(23.9042)	(917.0677)	(1,210.8431)
	Senate subcommittee chair represents State of obs.	-109.3529*	-161.1864**	-109.1299*	2,496.1620	628.4956
		(58.7413)	(79.7240)	(58.6914)	(2,253.0614)	(2,974.5347)
Observations		2462	2462	2462	2462	2462
Number of cross-sections		108	108	108	108	108
R-squared		0.70	0.45	0.70	0.61	0.32

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Instrumental Variables Regressions with Aggregate Panel (metro/year)

Dependent Variable: Patent Counts	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
Publication Counts	0.1288***	0.2058***	0.1681***	0.1531***				
	(0.0087)	(0.0128)	(0.0624)	(0.0416)				
Citation-Weighed Publication Counts					0.0028***	0.0030***	-0.0031*	0.0009
					(0.0002)	(0.0003)	(0.0018)	(0.0011)
Payroll in Scientific R&D Services	0.0636***	0.0596***	0.0635***	0.0735***	0.0834***	0.0871***	0.0753***	0.1104***
	(0.0062)	(0.0061)	(0.0089)	(0.0123)	(0.0061)	(0.0062)	(0.0074)	(0.0073)
Industry R&D (estimate)	0.0463***	-0.0162***	-0.0130*	-0.0084	0.0452***	-0.0137**	0.0168	0.0077
	(0.0029)	(0.0058)	(0.0078)	(0.0084)	(0.0028)	(0.0062)	(0.0108)	(0.0100)
New PhD Count	0.3311***	0.0029	0.3398		0.4338***	0.8699***	2.8587***	
	(0.0909)	(0.1685)	(0.5713)		(0.0766)	(0.1706)	(0.5873)	
Observations	2462	2462	2462	2462	2462	2462	2462	2462
R-squared	0.71	0.37			0.72	0.32		
Fixed effects?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Number of cross-sections	108	108	108	108	108	108	108	108

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: IV regressions on publication counts, metro/year panel, by technology category

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Patent Counts	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	Drugs and Medical		Chemical		Electrical and Computing		Mechanical		Agriculture, Mining and Environmental	
Publication Count	0.0380***	0.0087	0.9462***	1.1004***	0.2021***	0.3645***	0.3129***	-0.0659	0.0358***	0.0442
	(0.0032)	(0.0171)	(0.0709)	(0.3597)	(0.0097)	(0.0617)	(0.0326)	(0.2064)	(0.0056)	(0.0278)
Industrial R&D (estimate)	-0.0306***	-0.0288***	-0.0383**	-0.0442**	0.1456***	0.1165***	0.0007	0.0116	0.0033	0.0013
	(0.0013)	(0.0017)	(0.0155)	(0.0205)	(0.0068)	(0.0131)	(0.0062)	(0.0086)	(0.0083)	(0.0105)
Payroll in Scientific R&D Services	0.0077**	0.0149***	0.2139***	0.2019***	-0.0229***	-0.0442***	0.0867***	0.1163***	0.0252***	0.0230***
	(0.0038)	(0.0056)	(0.0133)	(0.0306)	(0.0055)	(0.0099)	(0.0074)	(0.0176)	(0.0030)	(0.0077)
New PhD Count	0.1483***	0.4563**	-1.7141***	-2.1392**	0.0368	-0.7759**	0.1064	1.5014*	-0.1495	-0.1475
	(0.0533)	(0.1846)	(0.3823)	(1.0448)	(0.0728)	(0.3142)	(0.1972)	(0.7771)	(0.1259)	(0.1261)
Observations	2402	2402	2402	2402	2402	2402	2402	2402	2402	2402
R-squared	0.25		0.30		0.61		0.25		0.11	

All regressions with metro fixed effects, 106 cross-sections
Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: IV regressions on citation-weighed publication counts, Metro/year panel, by technology category

Dependent Variable: Patent Counts	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS
	Drugs and Medical		Chemical		Electrical and Computing		Mechanical		Agriculture, Mining and Environmental	
Citation-weighed Publication Count	0.0008***	-0.0002	0.0113***	0.0361*	0.0032***	0.0015	-0.0091***	-0.0018	0.0007***	0.0004
	(0.0001)	(0.0004)	(0.0026)	(0.0212)	(0.0002)	(0.0012)	(0.0018)	(0.0121)	(0.0001)	(0.0006)
Industrial R&D (estimate)	-0.0315***	-0.0275***	-0.0110	-0.0306	0.1635***	0.1733***	0.0134**	0.0104	0.0000	0.0056
	(0.0013)	(0.0021)	(0.0159)	(0.0231)	(0.0070)	(0.0097)	(0.0063)	(0.0079)	(0.0085)	(0.0135)
Payroll in Scientific R&D Services	0.0185***	0.0167***	0.2836***	0.2742***	0.0086	0.0059	0.1091***	0.1107***	0.0304***	0.0324***
	(0.0037)	(0.0039)	(0.0125)	(0.0150)	(0.0056)	(0.0059)	(0.0070)	(0.0075)	(0.0027)	(0.0045)
New PhD Count	0.1971***	0.6177***	0.3101	-0.9759	0.4675***	0.7771***	1.6291***	1.3302**	-0.1653	-0.1617
	(0.0515)	(0.1797)	(0.3663)	(1.1505)	(0.0693)	(0.2205)	(0.1743)	(0.5190)	(0.1263)	(0.1266)
Observations	2402	2402	2402	2402	2402	2402	2402	2402	2402	2402
R-squared	0.25		0.25		0.58		0.23		0.10	

All regressions with metro fixed effects, 106 cross-sections

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix A: Data Compilation Details

Patent Data: Patent data were compiled from the NBER U.S. Patent Citations Datafile (see www.nber.org). An extension for patents granted in years 2000-02 was compiled with data from Bronwen Hall and inventor data from the USPTO Cassis database. Varying length of time from application to grant date meant that counts for recent years were truncated. Years with truncated patent counts were not kept in the panel, except for 1999, for which undercounting was limited to 10-15%. Assignment of patents to technology areas were made according to the following table:

Table A1

NBER Subcategory	Subcategory Name	Tech Area	Tech Area Name
11	Agriculture, Food, Textiles	5	Agriculture, Mining, Env. and other
12	Coating	2	Chemical
13	Gas	2	Chemical
14	Organic Compounds	2	Chemical
15	Resins	2	Chemical
19	Misc. Chemical	2	Chemical
21	Communications	3	Electrical and Computing
22	Computer Hardware & Software	3	Electrical and Computing
23	Computer Peripherals	3	Electrical and Computing
24	Information Storage	3	Electrical and Computing
31	Drugs	1	Drugs and Medical
32	Surgery & Medical Instruments	1	Drugs and Medical
33	Biotechnology	1	Drugs and Medical
39	Misc. Drugs and Medical	1	Drugs and Medical
41	Electrical Devices	3	Electrical and Computing
42	Electrical Lighting	3	Electrical and Computing
45	Power Systems	3	Electrical and Computing
46	Semiconductor Devices	3	Electrical and Computing
49	Misc. Electrical	3	Electrical and Computing
51	Materials Processing & Handling	4	Mechanical
52	Metal Working	4	Mechanical
53	Motors & Engines	4	Mechanical
54	Optics	3	Electrical and Computing
55	Transportation	4	Mechanical
59	Misc. Mechanical	4	Mechanical
61	Agriculture, Husbandry, Food	5	Agriculture, Mining, Env. and other
63	Apparel and Textile	5	Agriculture, Mining, Env. and other
64	Earth Working and Wells	5	Agriculture, Mining, Env. and other
65	Furniture, House Fixtures	5	Agriculture, Mining, Env. and other
66	Heating	5	Agriculture, Mining, Env. and other
67	Pipes and Joints	4	Mechanical

Publication Data: Counts of academic publications, and citations to those publications, were constructed from automated searches of Thompson Inc.'s *Web of Science* using a Perl script. All publications in journals tracked by Web of Science associated with 218 U.S. universities, 98 non-profit research institutes and five federally funded R&D centers, from 1973-2001 were included. Assignments of publications to fields of science were made using an algorithm that combined information from the "subject category" and "address" fields of the publication record[28]. Assignment from fields of science to technology areas was made according to the following table:

Table A2

Field of Science and Engineering	Tech. Area	Technology Area Name
Mathematics	3	Electrical and Computing
Computer Science	3	Electrical and Computing
Statistics / Biostatistics	1	Drugs and Medical
Chemistry	2	Chemical
Physics	3	Electrical and Computing
Astrophysics / Astronomy	3	Electrical and Computing
Geosciences	5	Agriculture, Mining, Env. and other
Oceanography	5	Agriculture, Mining, Env. and other
Biochemistry / Molecular Biology	1	Drugs and Medical
Genetics	1,5	Drugs and Medical, Agriculture, Mining, Env. and other
Neurosciences	1	Drugs and Medical
Pharmacology	1	Drugs and Medical
Physiology	1	Drugs and Medical
Cellular and Development Biology	1	Drugs and Medical
Ecology, Evolution and Behavior	5	Agriculture, Mining, Env. and other
Aerospace Engineering	3,4	Electrical and Computing, Mechanical
Biomedical Engineering	1	Drugs and Medical
Chemical Engineering	2,3	Chemical, Electrical and Computing
Civil Engineering	4	Mechanical
Electrical Engineering	3	Electrical and Computing
Industrial Engineering	4	Mechanical
Materials Engineering	2	Chemical
Mechanical Engineering	4	Mechanical

Publications (and the number of citations to them) were assigned to the institution of each author, so coauthored papers (and their citations) were essentially weighted by the number of authors.

[28] See Stuen, Maskus and Mobarak, "Foreign PhD Students and Knowledge Creation: Evidence from Enrollment Fluctuations" Working Paper, 2007 for a precise specification of this algorithm

Industrial research and development control variable: A control variable for private-sector R&D was constructed from several data sources. The primary source, the NSF's *Survey of Industrial R&D*, provides estimates of aggregate corporate R&D at the state level, for most odd years and some even years. See www.nsf.gov for data availability. In order to create a panel variable at the level of metro/tech-area/year, additional sources were needed. SIRD table 37, from 2003, provides a breakdown of R&D over broad (3-digit) industries. The aggregate payroll measure from the Census Bureau's *County Business Patterns* data was used to weight R&D of a state/industry group by the percent of aggregate payroll of a particular metro-area within that state. The CBP data was available by 3 digit NAICS industry more recently than 1997, but by 2 digit SIC industry for 1997 and previous years.

The aggregate payroll measure also varied by time, unlike the R&D decomposition by industry. Hence the necessary assumptions in using this variable include a) that the distribution of R&D among industries is roughly constant across states and time, and b) that R&D is positively correlated with aggregate payroll. This second assumption should be helped by the fact that R&D estimates and aggregate payroll estimates were matched at the industry level. Given all of these assumptions, and the fact that the original aggregate is itself an estimate, this variable should be treated as no more than a proxy for actual R&D.

The actual variable is created as so:

$$R\&D_{st} * \%R\&D_i * \%AP_{mt} = R\&D_{mit}$$

where *s* indicates state, *i* indicates tech-area, *m* indicates metropolitan area and *t* indicates year. Here, $\%R\&D_i$ is the percent of total corporate R&D performed in 2003 that was performed in industry *i*. Likewise, $\%AP_{mt}$ is the percent of aggregate payroll in a state that was performed by metro *m* in year *t*. Summing over industries in each tech-area creates metro/tech-area/year aggregates. Four metro areas included areas in several states (New York, Philadelphia, DC-Baltimore and Kansas City), so that "states" used in the creation of the variable were aggregates of the states spanned by these metropolitan areas.

One complication in such a construction is that the industries used by the SIRD and the CBP do not have a direct correspondence with the broad tech-areas in the panel. Also, the SIRD industries sometimes refer to 4 or 5 digit classifications, with which the CBP 3 digit NAICS and 2 digit SIC codes will not overlap with on a one-to-one basis. The following table lists the industries used in the SIRD and corresponding tech-areas, with a rationale for the match.

Table A3

Industry	NAICS	SIC	Technology Area	Explanation
Aerospace Products and Parts	336	372	4 (mechanical devices, machines)	
Architectural, engineering services	5413	87	4	
Basic chemicals	325	28, 3087, 3861	2 (chemical)	not 3254 or 283 (pharma)
Beverage and Tobacco Products	312	20,21	5 (agricultural and environmental techniques)	
Communications equipment	334	366	3 (electronics and computing)	
Computer systems design and	5415	737	3	

related services				
Computers and peripheral equipment	334	367, 357	3	
Construction	23	15,16,17	4	
Drugs and druggists' sundries	4242		0 (non-science)	for the sale, not manufacture
Electrical equipment, appliances, and components	335	36	3	
Electrical goods	4216		0	for the sale, not manufacture
Fabricated Metal Products	332	34	4	
Finance, insurance, and real estate	52		0	
Food (production)	11		5	
Furniture and related products	337	25	5	for wood products
Health care services	62	80	1	
Machinery	333	35	4	
Management of companies and enterprises	55		0	
Medical equipment and supplies	3391	384	1 (drugs and medical)	
Mining, extraction and support activity	21	10, 12, 13	5	excluding non-metallic
Motor vehicles, trailers and parts	336	371,373, 379	4	
Navigational, measuring, electromedical, manufacturing	3345	382	3	
Newspaper, periodical, book, and database; publishing	514	823,735,737	3	for innovations in databases
Nonmetallic mineral products	2123	14	2	for chemical-based mining processes
Other broadcasting and telecommunications	5179		0	
Other chemicals	3259	28, 30, 3861	2	
Other computer and electronic products	3344	3679	3	
Other information services	519	7375	3	
Other misc. manufacturing	3399	34, 39	4	
Other professional, scientific, and technical services	5419		0	
Other transportation equipment	3369	3751, 3711, 3795, 3799	4	
Paper, printing, and support activities	323	27	2	
Petroleum and coal products	324	29	2	
Pharmaceuticals and medicines	3254	283	1	
Plastics and rubber products	326	26, 30	2	
Primary metals	331	33	4	
Professional and commercial equipment and supplies	4234		0	for the sale, not manufacture
Resin, synthetic rubber, fibers and filament	3252	28	2	
Retail trade	44-45		0	
Scientific R&D services	5417	873	1,2,3,4,5	should cover all disciplines
Semiconductor and other electronic components	3344	367	3	
Software reproducing material	3346	3652, 3695	3	
Telecommunications	513	48	3	
Textiles, apparel, and leather	313	22	5	
Transportation and warehousing	48-49	45	4	
Utilities	22	49	3	
Wood Products	321	24	5	

New PhD Control Variable: Data on recipients of Doctorates in science and engineering from U.S. universities was compiled from the NSF's *Survey of Earned Doctorates* under a licensing agreement. The variable constructed is the number of Doctorates granted at all of the 218 universities in the sample. PhD's were assigned to metropolitan areas according to the location of their doctoral institution. They were assigned to fields of science (see Table A2) based on the specialized area of their dissertation, and these fields assigned to broad technology categories in the same way as publications. A matching of dissertation areas and fields is available upon request.

Appendix B: Related Literature

One alternate approach has been to study patterns and location of patent citations. Jaffe, Trajtenberg and Henderson (1993) found that academic patents cited by industrial patents were more likely to be cited by firms located near to the university. Other studies that look at citation trails are Henderson, Jaffe and Trajtenberg (1998), Branstetter and Ogura (2005) and Kim, Lee and Marschke (2006).

Yet another approach has been to use surveys and case studies to gain insight into the connection between academic and industrial research. Mansfield (1995) surveyed corporations and found that 10% of industrial innovation can be attributed to academic research. Other studies that have used surveys and case studies to document the mechanism of spillovers are Cohen, Nelson and Walsh (2002) and Jensen and Thursby (2001).

A final approach has been to consider academic research as one source of agglomeration economies, leading to clustering of innovative industry near universities. Audretsch and Feldman (1996) use Gini coefficients of industrial production and innovation directly as their dependent variables, and examine the correlations between these measures of clustering and factors affecting agglomeration such as academic R&D. Other studies that have looked at the relationship between knowledge spillovers and agglomeration are Varga (2000) and Adams (2002).