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 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

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 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 oRn/tehapeantechnoi agtly used hnology,g used h7 Ta(4 an2 0 0 1eso

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

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.

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

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

Here, I

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 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].

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

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 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 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.

regressiD6 Tun83 Tw0001 Tmn 8 show97 thtationcounts 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

number of new PhD graduates may be another measure of academic knowledge, and a channel for academic spillovers.

These results are preliminary, as furt

Figure 1.

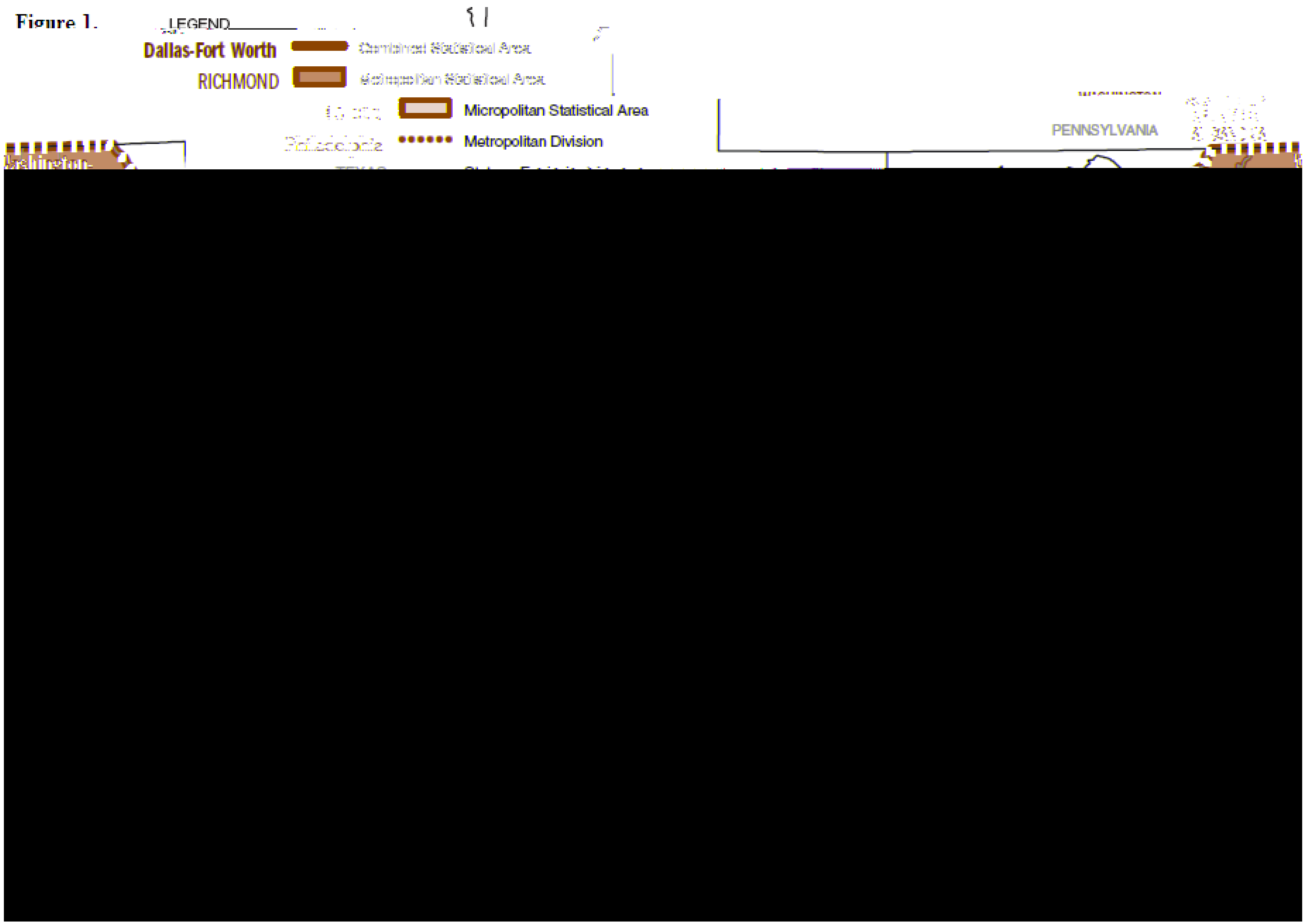
LEGEND
Dallas-Fort Worth
Combined Statistical Area
RICHMOND
Metropolitan Statistical Area
Micropolitan Statistical Area
Metropolitan Division
PENNSYLVANIA

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				

Table 3: First stage Instrumental Variables Regressions

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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	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 s,oMeMe(estiSs,

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:

$$R \quad D_{st} \qquad R \quad D_{i} \qquad AP_{mt} \qquad R \,\&\, D_{mit}$$

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 [illegible]	3344	3679	3	
Other information services	519	7375	3	[illegible]

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