

**ON THE IMPORTANCE OF GEOGRAPHIC AND TECHNOLOGICAL  
PROXIMITY FOR R&D SPILLOVERS:  
AN EMPIRICAL INVESTIGATION**

**Michael J. Orlando**

**JULY 2000**

**RWP 00-02**

Research Division  
Federal Reserve Bank of Kansas City

Michael J. Orlando is an economist at the Federal Reserve Bank of Kansas City. The author is indebted to Bruce Petersen, Marcus Berliant, Steven Fazzari, Douglass North, Robert Parks, and Josh Lerner for their invaluable guidance and comments. The views expressed are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

Orlando e-mail: [michael.j.orlando@kc.frb.org](mailto:michael.j.orlando@kc.frb.org)

## **Abstract**

Empirical studies of the external effects of R&D suggest that both geographic and technological distance attenuate inter-firm spillovers from innovative activity. The results presented here indicate that the tendency for R&D spillovers to localize economic activity is conditional on the technological relation between spillover generating and receiving firms. The production function framework is generalized to control for correlation between measures of geographic and technological proximity. Coefficient estimates confirm that R&D spillovers are largest among technological neighbors. However, spillovers within narrowly defined technological groups do not appear to be attenuated by distance. Geographic proximity serves to attenuate only those inter-firm spillovers that cross narrowly defined technological boundaries.

*Key words:* R&D, spillovers, innovation, industrial agglomeration, geography, firm-level, empirical studies.

*JEL classification:* O3, L6

# 1. INTRODUCTION

The appropriability conditions associated with investments in research and development (R&D) activities distinguish them from normal factors of production. Service flows from conventional inputs are fully appropriable. In contrast, service flows from R&D expenditures have public good-like qualities. Ideas for new products and production processes are non-rival; firms regularly make multiple and simultaneous use of inventions from their own R&D laboratories. In addition, while patents and trade secrets are routinely used to exclude others from use of one's ideas, enforcement is costly and of limited effect.<sup>1</sup> Thus, over time a firm's productive capacity may be augmented as a consequence of spillovers from external R&D.

This simple observation regarding the appropriability conditions of innovative activity has profound implications for a range of frontiers of economic inquiry. Spence (1984) formalized the tradeoff between static and dynamic efficiency first identified by Schumpeter (1942). Static inefficiency associated with market power may be offset by improved dynamic efficiency associated with increased appropriation of returns to innovation. Thus, fundamentals that govern inter-firm spillovers ultimately determine appropriability conditions that govern industrial structure and performance.<sup>2</sup>

City formation is hypothesized to arise in part from externalities, including those related to knowledge creation. This mechanism results in localization, or the agglomeration of similar industrial activity, if knowledge spillovers are confined to firms in the same industry.<sup>3</sup> Urbanization, or the agglomeration of a diversity of industrial

---

<sup>1</sup> Levin et al. (1987) list a variety of ways a firm may learn of rivals' technologies including patent disclosures, technical publications and conferences, high-tech labor market turnover, and reverse engineering of others' products.

<sup>2</sup> Cohen and Levin (1989, pp. 1060-1061) argue that empirical efforts directed at the Schumpeterian hypothesis have been preoccupied with the correlation between innovation and firm size and concentration. They are encouraged, however, by a movement toward "a broader consideration of the fundamental determinants of technical change in industry." They group more recent empirical research according to source of inter-industry variation in innovative activity and performance: demand, technological opportunity, and appropriability conditions (p. 1095). The research presented here is concerned with this latter determinant of innovative activity.

<sup>3</sup> Localization externalities are sometimes referred to as MAR externalities after Arrow's (1962) and Romer's (1986) formalization of Marshall's (1890) view that concentration of an industry in a city facilitates inter-firm spillovers and, therefore, growth of the industry and city. Urbanization externalities are sometimes referred to as Jacobs externalities after the view advanced by Jacobs (1969) that economically

activity, will occur if knowledge spills across industry boundaries. In either case, agglomeration will only occur if distance attenuates knowledge spillovers.

Spillovers from innovative activity also play a central role in an influential class of new growth theoretic models.<sup>4</sup> These models exploit the dual private/public nature of knowledge created through investments in innovative activity. The output of research activity is understood to be non-rival and (at least partly) non-excludable. Thus, private investments in knowledge creation may augment others' productivity, allowing the economy to sufficiently avoid diminishing returns at the aggregate level for some set of parameter values. Understanding the nature of this externality to knowledge creation is central to understanding whether such models provide useful insights for explaining observed economic growth.

A body of empirical evidence does suggest spillovers are significant and typify the appropriability conditions of R&D.<sup>5</sup> The earliest analyses find social returns exceed private returns to investments in innovative activity.<sup>6</sup> Spillovers from R&D have been shown to exist at the inter-firm and inter-industry levels.<sup>7</sup> More recently, Saxenian (1994) attributes the differential performance of regional high-tech economies to the interaction of culture and spillovers. Audretsch and Feldman (1996) interpret disproportionate clustering of innovative activity in knowledge intensive industries as evidence of inter-firm spillovers. And Henderson and Cockburn (1996) report significant inter-firm spillovers from knowledge creation in the pharmaceutical industry.<sup>8</sup>

This paper examines the importance of geographic and technological proximity for spillovers from R&D, with particular attention to the interplay between these factors. Direct evidence of spillovers is obtained from a firm-level production function framework in which R&D stocks of geographically and technologically proximate firms

---

significant spillovers occur through application of knowledge beyond the industry of origin (Glaeser et. al. 1992 p. 1127).

<sup>4</sup> See Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992).

<sup>5</sup> Surveys are provided by Cameron (1996), Griliches (1992), and Nadiri (1993).

<sup>6</sup> See Griliches (1958), Minasian (1962), and Terleckyj (1958).

<sup>7</sup> See Jaffe (1986) and Scherer (1982), respectively.

<sup>8</sup> The authors attribute superior R&D productivity in large firms to economies of scope and scale, suggesting a role for spillovers in the theory of the firm as well.

are found to be correlated with output. In contrast with previous studies, which assume a particular spillover pool structure, this empirical examination allows the data to suggest a structure.

A preliminary analysis is consistent with earlier findings that greater distance in both the geographic and technological dimensions attenuate spillovers from R&D. This method, however, is subject to the criticism that it fails to control for the high degree of correlation expected between measures of geographic and technological proximity. The main results are obtained from a more general model in which a two-dimensional sorting of external R&D stocks is used to allow an independent assessment of the importance of geographic and technological proximity. The main results indicate that spillovers are most significant among technologically similar firms, regardless of inter-firm distance. Spillovers from R&D outside a firm's own narrow (four-digit) Standard Industrial Classification (SIC), while less economically important, appear to be sensitive to geographic proximity.

“Geographically near R&D”, defined as that conducted within a specified radius of each firm, is varied in the analysis from 25 to 800 miles. The preliminary analysis supports the conventional wisdom that distance attenuates spillovers. Spillovers are most significant among firms within a 50-mile radius of one another and this effect is attenuated with increasing distance. This conclusion is only conditionally supported by the general analysis in which external R&D stocks are sorted into pools spanning both geographic and technological dimensions. Evidence of the importance of geographic proximity is limited to spillovers from firms outside (but not too far outside) one's own narrowly defined industry group. These results are interpreted to reflect the propensity for firms engaged in similar industrial activity to agglomerate geographically for reasons exclusive of localized intra- or inter-industry knowledge spillovers.

In contrast, the findings of this analysis are consistent with the hypothesis that technological proximity generally magnifies R&D spillovers. “Technologically near R&D” is defined as that conducted within a firm's own industry group. “Industry group” is defined at first the four-digit and then the three-digit SIC. The hypothesis is supported

by the preliminary analysis in which external R&D stocks are sorted into technologically near and distant pools. This conclusion is also supported by the main results obtained within the more general framework. These findings suggest spillovers are greatest among technologically similar firms, regardless of geographic proximity, and are attenuated as technological distance increases.

The next section comments on the mechanisms that govern the flow of disembodied knowledge across firm boundaries and surveys several widely cited contributions to the measurement of inter-firm spillovers. In section three I follow Hall and Mairesse (1995) who recommend the production function framework to quantify the importance of R&D in the production process; this model is extended to include external stocks of R&D. In section four, county-specific latitude/longitude data is merged with a panel of annual financial variables to characterize the relative proximity of each pair of firms in both geographic and technological dimensions. Empirical results are summarized in sections five and six. First, spillover pools are constructed according to unidimensional measures of proximity to replicate established findings that both geographic and technological distance attenuate spillovers. Results are then reported from a more general analysis which simultaneously incorporates these two measures of proximity. The implications of these findings are discussed in the concluding section.

## **2. THE DETERMINANTS OF SPILLOVERS**

Firms that are geographic neighbors may exchange knowledge through a variety of channels. Knowledge may be transmitted through employee interaction in social, civic, and professional organizations, participation in which may be geographically constrained. Normal employee turnover can result in significant cross-pollination of knowledge stocks. And geographically near firms are likely to share buyers and suppliers who also may serve as conduits for information flow.

Jaffe et al. (1993) provide evidence of localization of knowledge spillovers from inventive activity. The authors find that a patent's citations are more likely to come from the same state and standard metropolitan statistical area than are a technologically similar control group of patents. Audretsch and Feldman (1996) suggest this localization of

knowledge spillovers applies to innovative activity as well. They find greater clustering of innovative activity in industries where knowledge externalities are likely to play a greater role. They interpret concentration of innovation in excess of geographic concentration of production as evidence in support of the hypothesis that geographic distance attenuates knowledge spillovers.

Regardless of geographic proximity, spillovers are also believed to be higher between “technological neighbors”. According to this view, the ability to make productive use of another firm’s knowledge depends on the degree of technological similarity between firms. Every technology has a somewhat unique set of applications and language. Researchers in similar technological fields will interact in professional organizations, publish in commonly read journals, and, increasingly, browse a common set of web pages. Reverse engineering may be employed to maintain parity with one’s rivals. And spying and corporate espionage are thought to be relatively common among information intensive industries.<sup>9</sup>

Jaffe (1986) uses firm panel data and patent histories to construct a technological proximity-weighted spillover pool for each firm.<sup>10</sup> The author finds circumstantial evidence of technologically dependent spillovers in that firms active in research intensive technology groups enjoy higher research productivity and higher returns to R&D, controlling for industry specific technological opportunities.

Adams and Jaffe (1996) estimate both the geographic and technological sensitivity of spillovers at the intra-firm level. Using plant-level data, the authors find evidence of attenuating effects associated with both measures of distance. Specifically, they find that a firm’s research lab activity has the greatest impact on its own plants within 100 miles and those working in a similar product class. Due to data limitations, however, the authors do not consider the two hypotheses simultaneously and are limited

---

<sup>9</sup> For an example, see “Oil Firms Still Rely on Corporate Spies to be Well Informed”, in *The Wall Street Journal*, December 7, 1998.

<sup>10</sup> The author defines a multi-dimensional technology space spanning 47 individual patent classes. A vector in this space summarizes a firm’s patent history and thereby defines their technological position. The weight of one firm’s R&D in the another’s spillover pool is inversely proportional to the angular separation between patent vectors.

to a series of partial analyses.<sup>11</sup> Furthermore, while we expect firms to act to maximize spillovers among their own plants, it is likely that they try to minimize inter-firm spillovers. Thus, it is not clear that the authors' intra-firm level findings can be generalized to spillovers across firm boundaries.

### 3. A SIMPLE MODEL OF PRODUCTION WITH SPILLOVERS

As a point of departure I follow Hall and Mairesse (1995) and assume firm output,  $Y_{it}$ , can be represented with a conventional Cobb-Douglas production technology as follows:<sup>12</sup>

$$Y_{it} = A\hat{K}_{it}^{\pi} e^{\lambda t} C_{it}^{\alpha} L_{it}^{\beta} e^{\varepsilon_{it}} \quad (1)$$

where  $C_{it}$  is capital,  $L_{it}$  is labor,  $\alpha$  and  $\beta$  are their respective output elasticities, and  $\varepsilon$  is an i.i.d. disturbance. The function  $A\hat{K}_{it}^{\pi} e^{\lambda t}$  represents a firm-specific 'state of technology' which summarizes all knowledge relevant to firm  $i$ 's production possibilities in time period  $t$ .

The state of technology is a function of an exogenous time trend,  $\lambda$ , and other exogenous factors,  $A$ .<sup>13</sup> Assuming R&D is the single determinable component of knowledge relevant to the production process,<sup>14</sup> then  $\hat{K}_{it}$  represents the effective stock of R&D contributing to firm  $i$ 's production in time period  $t$ . If the outcome of R&D efforts

---

<sup>11</sup> Adams and Jaffe (1996, p. 703) note that partial analyses which find, for example, that technological distance mitigates spillovers may be detecting the importance of geographic proximity.

<sup>12</sup> While one may argue that little can be learned about spillovers from the partial equilibrium relationship summarized by a firm production function, Hall and Mairesse (1995) find that the production function framework is preferred to the rate of return formulation. Furthermore, through comparison to the 'semi-reduced form' approach, the authors find the production function does not yield biased estimates of R&D elasticity when controls for permanent firm effects are included. One may argue still for something more behavioral, such as a location theoretic framework. But why firms are where they are is presumably the outcome of utility maximizing decisions of decisionmakers. The purpose of this paper is to say something about the way in which spillovers occur and so it seems reasonable, before complicating the analysis with complex behavioral stories, to ask just what insights a simple model of spillovers might yield.

<sup>13</sup> For ease of exposition, the assumption of a common intercept term,  $A$ , and time trend,  $\lambda$ , is maintained throughout the derivation. The estimation procedure relaxes this assumption by incorporating state, industry, and year specific intercept terms. The assumption of constancy in the other parameters should not be too offensive since the scope of the analysis will be limited to a single two-digit standard industrial classification.

<sup>14</sup> Cockburn and Griliches (1988) conclude "... [d]ata on R&D expenditures ... are stronger measures of input to the process by which firms produce technical innovation than patents are of its output."

are at least partly non-rival and non-excludable then  $\hat{K}_{it}$  will be an aggregation of R&D stocks accumulated through one's own research effort, as in Hall and Mairesse (1995), and those available through spillovers from other firms.

Construction of this effective stock of R&D has been the focus of many early attempts to identify the presence of inter-firm spillovers. As we turn our attention to understanding the nature of R&D spillovers, such measurement issues may be recast as empirical questions. In contrast to earlier studies which have assumed a weighting scheme consistent with spillover channels described in section 2, the strategy pursued here is to allow the data to reveal the structure of this weighting scheme.

Assume that the effective stock of R&D for firm  $i$  can be expressed as follows:

$$\hat{K}_{it} = rK_{it}^{-\rho} + \sum_{j \neq i}^I s_{ij} K_{jt}^{-\rho} \quad (2)$$

where  $K_{it}$  and  $K_{jt}$  represent R&D stocks employed by the own firm and parties external to the firm, respectively. Substitution of (2) into (1) and redefining  $\pi = -\frac{1}{\rho}$  leaves a production function in which individual R&D stocks are complimentary to conventional inputs but interact with one another in a familiar constant elasticity of substitution (CES) process:

$$Y_i = A e^{\lambda t} C_i^\alpha L_i^\beta (rK_{it}^{-\rho} + \sum_{j \neq i}^I s_{ij} K_{jt}^{-\rho})^{-\frac{1}{\rho}} e^{\varepsilon_i} \quad (3)$$

where the subscript  $t$  is suppressed for clarity. In principle, then, each firm,  $i$ , may have access to a unique state of technology. This technology is a combination of own R&D with weight  $r$  and the R&D stocks of other firms,  $j$ , each with weight  $s_{ij}$ .

It is well known that a special case of the CES production function is that where  $\rho = 0$ , corresponding to a Cobb-Douglas technology. A second order Taylor series expansion of (3) around that point may be written as

$$y_i = a + \lambda t + \alpha c_i + \beta l_i + r k_i + \sum_{j \neq i}^I s_{ij} k_j - \frac{\rho}{2} \sum_{j \neq i}^I r s_{ij} (k_i - k_j)^2 - \frac{\rho}{2} \sum_{h \neq i}^I \sum_{j \neq i}^I s_{ih} s_{ij} (k_h - k_j)^2 + \varepsilon_i \quad (4)$$

where lower-case variables represent natural logarithms of their upper-case counterparts.<sup>15</sup> If the  $\rho = 0$  assumption is not unreasonable then estimates of the spillover weights  $s_{ij}$  are interpreted as elasticities of output with respect to the R&D stock of the respective firm  $j$ . Unfortunately, direct estimation of this spillover weighting matrix is far too costly in terms of degrees of freedom. But we can appeal to *a priori* expectations regarding the pattern of variation in the  $s_{ij}$ 's to structure an estimable equation from the cumbersome reduced form of this most general case.

Consider the hypothesis that spillovers are a function of geographic distance between firms. For the purpose of this discussion, assume that the productivity of external R&D varies only with distance between the spillover sending and receiving firms. Specifically, nearby firms' R&D provides greater spillovers than do distant firms.<sup>16</sup> In order to make (4) empirically manageable, we need to limit the number of parameters requiring estimation. The technique used here is to draw a circle of arbitrary radius,  $R$ , around each firm  $i$  and define all firms inside that circle elements of the set  $G_{Ni}(R)$ . All firms outside of the circle are considered elements of the set  $G_{Zi}(R)$ . Equation (4) may then be expressed as

$$y_i = a + \lambda t + \alpha c_i + \beta l_i + r k_i + s_N \sum_{n \neq i}^{G_{Ni}} k_n + s_Z \sum_{z \neq i}^{G_{Zi}} k_z + \varepsilon_i \quad (5)$$

where the second order terms are dropped since  $\rho$  is assumed to be 0. Our principal interest is with the magnitude and sign of the near spillover pool coefficient relative to that of the distant spillover pool and relative to the definition of spillover pool boundaries. If spillovers are sensitive to the relative locations of other innovating firms

---

<sup>15</sup> Derivation included as Appendix A. The linear case ( $\rho = -1$ ) was also estimated, however, goodness of fit was reduced as compared to the Cobb-Douglas case and spillover pool coefficients were not significant. Admittedly, this could be a consequence of the noisier nature of the spillover pool variable in this case. However, it may also be interpreted as evidence against the linear functional form which assumes own and others' R&D serve as substitutes. This interpretation is consistent with Cohen and Levinthal (1989) who argue that incentives to engage in own R&D include development of absorptive capacity, or the ability to make use of technology developed by others. Furthermore, preliminary grid-search results from non-linear estimation of (3) suggest  $\rho$  converges to a value between -0.1 and 0.1.

<sup>16</sup> Alternatively, one could argue that spillovers from nearby firms are received *faster* than those from distant firms. But if this is the case then lead time advantages (Levin et al. 1987) will make later service flows from external R&D less valuable the spillover receiving firm and the empirical effect should be the same.

then the near R&D pool variable will be positively correlated with firm value added,  $s_N > 0$ . The near R&D pool coefficient will also be larger than that of the distant spillover pool, i.e.  $s_N > s_Z$ . Furthermore, increasing the distance to the near pool / distant pool boundary will include firms farther and farther from firm  $i$  into the set  $G_{Ni}$ , imposing a downward bias on  $s_N$ , i.e.  $\frac{\partial s_N}{\partial G_{Ni}(R)} < 0$ .

Of course, this same apparatus may be used to evaluate the sensitivity of interfirm spillovers to technological similarity. Firms engaging in different research, using dissimilar technical processes, or making distinct goods may be considered “technologically distant” from one another. Diffusion of R&D services from one such firm to the next should be lower relative to diffusion between firms that are “technologically near”. Therefore, tests of a technological proximity hypothesis analogous to those proposed in the preceding paragraph may be constructed by sorting all firms into technologically near and technologically distant pools.

As noted in section 1, if technological and geographic proximity are highly correlated then interpretation of results from the partial analyses described in the preceding paragraphs may be problematic. Therefore, the main analysis generalizes the pooling technique described above to simultaneously control for geographic and technological proximity. R&D stocks external to each firm  $i$  are sorted into pools of firms that are geographically near/technologically near, geographically near/technologically distant, geographically distant/technologically near, and geographically distant/technologically distant. The resulting reduced form, analogous to (5), is:

$$y_i = a + \lambda t + \alpha c_i + \beta l_i + r k_i + s_{NN} \sum_{n \neq i}^{G_N T_{Ni}} k_n + s_{NZ} \sum_{u \neq i}^{G_N T_{Zi}} k_u + s_{ZN} \sum_{v \neq i}^{G_Z T_{Ni}} k_v + s_{ZZ} \sum_{z \neq i}^{G_Z T_{Zi}} k_z + \varepsilon_i \quad (6)$$

where  $G_N T_{Ni}$ ,  $G_N T_{Zi}$ ,  $G_Z T_{Ni}$ , and  $G_Z T_{Zi}$  are the respective subsets of spillover generating firms for each firm  $i$  and  $n$ ,  $u$ ,  $v$ , and  $z$  are their respective indexes. It should therefore be noted that estimating (5) under the geographic (technological) proximity hypothesis is equivalent to estimating the more general model (6) with the restriction that  $s_{NN} = s_{NZ}$  and  $s_{ZN} = s_{ZZ}$  ( $s_{NN} = s_{ZN}$  and  $s_{NZ} = s_{ZZ}$ ).

The preceding derivation is based on two key assumptions. First, as stated explicitly above, the elasticities of own output with respect to external R&D, the  $s_{ij}$ 's, are assumed to vary only with distance between spillover producing and receiving firms. Returning to (4), this says that  $s_{ij}$  varies in  $j$  only with the distance between firms  $i$  and  $j$ . Second, implied in the derivation of (5) is the assumption that the  $s_{ij}$ 's are also constant across spillover receiving firms, i.e.  $s_{ij}$  is constant for all  $i$  for a given pool boundary radius  $R$ .

It should be noted that the reduced form presented in (5) yields a spillover pool elasticity of output that is not scale neutral in the way of conventional input elasticities. The scale neutrality of conventional input elasticities stems from the assumption that the scale of inputs is correlated with the scale of outputs for a given firm size. For example, an elasticity of output with respect to capital of 0.1 is interpreted to mean that a one-percent increase in capital will yield a 0.1 percent increase in output, *ceteris paribus*. For a small firm a one percent increase in capital will imply a 0.1 percent increase in output which is presumably small in absolute terms. Larger firms will experience a larger nominal change in output in proportion to their nominally larger one percent change in capital.

But firms share the same nominal size of spillover pool regardless of their own size. While the elasticity,  $s$ , is still technically defined as the percentage change in output attributable to a percentage change in spillover pool R&D, the scale of this R&D is constant across firms of varying size. This construction implicitly assumes that large firms receive a proportionately larger nominal benefit than small firms from the same size spillover pool.

Given that it is the non-rival nature of R&D that motivates this research in the first place, this assumption is intuitively appealing. Additional support is drawn from Jaffe (1986) who reports benefits from R&D spillovers increasing in own scale of R&D. Additional evidence in support of this implicit assumption is found in Henderson and Cockburn (1996) who conclude in an examination of size and research productivity in the pharmaceutical industry:

“[L]arger research efforts are more productive ... because they realize economies of scope by sustaining diverse portfolios of research projects that capture internal and external knowledge spillovers.”<sup>17</sup>

Thus, the assumption that large firms benefit more from the same nominal base of spillover pool R&D seems sensible.

#### 4. DATA SELECTION AND EMPIRICAL SPECIFICATION

This examination of inter-firm R&D spillovers is intended to illuminate the microfoundations presumed to drive models of urban/regional agglomeration and economic growth at the macro level. As such, SIC 35, Industrial, Commercial Machinery, and Computer Equipment, represents a prime candidate for analysis.<sup>18</sup>

Innovations in these investment goods manufacturing industries are thought to be prime sources of economy wide growth in both the Arrow learning-by-doing and Schumpeterian traditions. Furthermore, selection of SIC 35 incorporates 357, Computer and Office Equipment, a sector increasingly perceived to play a central role in the renewed trend of productivity growth.

<u>Table 1</u>	
SIC 35: three-digit standard industrial classifications	
<u>SIC</u>	<u>Name</u>
352	Farm & garden machinery and equipment
353	Construction, mining, and material handling machinery and equipment
354	Metalworking machinery and equipment
355	Special industry machinery, except metalworking machinery
356	General industrial machinery and equipment
357	Computer & office equipment
358	Refrigeration & service industry machinery
359	Miscellaneous industrial and commercial machinery and equipment

<sup>17</sup> Henderson and Cockburn 1996, p. 32.

<sup>18</sup> Griliches identifies this sector as one of the four most R&D intensive two-digit SIC's; the others being 36 (Electrical Equipment), 37 (Transportation Equipment), and 28 (Chemicals and Pharmaceuticals). A recent review of the National Science Foundation publication *Research and Development in Industry: 1996* confirms that these remain the top four R&D-conducting industries, with SIC 35 accounting for 11% of total privately funded R&D and the remaining top three industries accounting for an additional 49%.

SIC 35 is composed of the investment goods manufacturing industries listed in table 1.<sup>19</sup> Annual firm-level financial data used in the analysis are obtained from Standard & Poor's COMPUSTAT database. Annual data for 515 publicly-traded U.S. firms engaged in R&D activity in SIC 35 extend from 1972 through 1995 and spans 39 states and 29 four-digit (or eight three-digit) SIC's. Key financial variables used in the following regressions are summarized in table 2.

<u>Table 2</u> Summary of Variable Construction			
<u>Variable Symbol</u>	<u>Variable Name</u>	<u>Proxy</u>	
<i>Y</i>	output	Sales	
<i>C</i>	capital	constructed from capital spending variable following Salinger and Summers (1983) with correction for acquisitions and divestitures per Chirinko et. al. (1999)	
<i>L</i>	labor	number of employees	
<i>K</i>	R&D stock	constructed from R&D expenditure using perpetual inventory method assuming 15% depreciation rate and pre-sample period growth rate of 5%	

Estimation error imposed by use of sales as a proxy for output will be confined to the constant term if these charges are some fixed proportion of sales. This assumption will be valid in a panel data setting where a firm fixed-effects model is used. To the extent that variation in the materials and energy fraction of sales is an industry or region fixed effect, this assumption should be reasonable in the cross-section through use of industry- and state-specific dummies.<sup>20</sup>

The capital variable makes use of widely available accounting data on capital expenditures and is an improvement on the commonly used Salinger and Summers (1983) algorithm.<sup>21</sup> Following Chirinko et al. (1999), I begin with the reported value for net property, plant, and equipment and iteratively accumulate and depreciate annual capital

<sup>19</sup> A detailed listing of the four digit SIC which compose SIC 35 is included as Appendix B.

<sup>20</sup> Basu (1996) reports that material inputs are nearly perfectly correlated with output.

<sup>21</sup> Chirinko et al. (1999).

expenditures. Divestitures and acquisitions data is used to adjust each firm's year-end capital stock. Empirical studies which fail to make this adjustment typically exclude firms engaged in mergers over the sample period, reducing sample size and introducing selection bias.

The R&D stock variable is constructed with the perpetual inventory method.<sup>22</sup> The results reported below are invariant to a preliminary check of sensitivity to various depreciation and pre-sample period growth rates.<sup>23</sup>

Excepting labor, all variables are in nominal dollars. Nominal values are deflated with price indexes obtained from the Bureau of Labor Statistics website.<sup>24</sup> Sales are deflated with three digit SIC specific producer price indexes. Capital is deflated with a capital expense index. R&D stock is deflated with the occupational cost index for technical professionals.<sup>25</sup>

In order to examine the impact of geographic distance on spillovers, each firm in the sample must be located in space. To do so, each firm is assumed to exist at the geographic centroid of the county location of its corporate headquarters. A circle is effectively drawn around each firm and all other firms that fall inside the circle are defined as geographically near; the remaining firms are defined as geographically distant.<sup>26</sup>

---

<sup>22</sup> See Griliches (1979).

<sup>23</sup> Preliminary estimates of the results reported in the following sections were insensitive to depreciation rates ranging from five to 30 percent. Hall & Mairesse (1995) also observe "... the choice of depreciation rate in constructing R&D capital does not make much difference to the coefficient estimates ... This result has already been observed in a number of previous studies and arises from the basic fact that the time series of R&D expenditure *within firm* does not vary all that much."

<sup>24</sup> A detailed description of deflator construction is included as Appendix C.

<sup>25</sup> According to the "Advanced Release of Selected Tables from the Research and Development in Industry: 1995" report by the NSF, a majority of R&D expense is for salaried technical professionals with materials and supplies accounting for 10% to 20%. In the absence of an R&D specific price deflator I use the occupational cost index for technical professional. Also, see Grabowski (1968).

<sup>26</sup> Each firm's geographic location is defined with the state and county name as reported in the COMPUSTAT Name and Location datafile. The COMPUSTAT Names and Financial Variables datafiles are merged with data obtained from the U.S. Geological Survey (USGS) Geographic Names Information System (GNIS). Each observation in this dataset reports the latitude and longitude of the geographic centroid of a county in degrees, minutes, and seconds. Federal Information Processing Standard state and county codes are used to map a set of latitude/longitude coordinates from the USGS GNIS database to each observation in the COMPUSTAT Financial Variables datafile. The distance between any two firms in a given year is then computed as the distance between their respective county centroids. Assuming a spherical earth of actual earth volume, the arc distance in miles between any two firms  $i$  and  $j$  can be

Use of corporate headquarters to represent firm location may be questionable for the purpose of spillover detection. One may argue that our true interest is in the location of innovation, not necessarily in the location of corporate headquarters. However, if firms view R&D as their most strategically important investment they are likely to locate this activity close to corporate headquarters, suggesting little error will be introduced with the location assumption proposed here. Furthermore, while R&D may be a reasonable proxy of the scale of a firm's innovative activity, spillovers from this implied knowledge base may emerge from any of the locations that compose the firm; R&D facilities, production facilities, or corporate headquarters. Thus, corporate headquarters may be as good a proxy of firm location as we can hope to find.

Nevertheless, the convention in this literature is to define the location of R&D activity as the probable location of spillover generation and reception. The *Directory of American Research and Technology 1993* was consulted to establish the reasonableness of the claim that corporate headquarters may be a useful proxy for the source location of R&D spillovers. This volume catalogs the location of a firm's corporate headquarters and the location and composition of that firm's R&D activity. Approximately Half of the firms included in this analysis have entries in this directory. Of this sample, 87 percent conduct at least part of their R&D in their corporate headquarters city. And fully 73 percent of the sample conduct their R&D exclusively in the same city as their corporate headquarters.

Perhaps less obvious, if no less controversial, is how to define a firm's location in technological space. The Jaffe (1986) method (see section 2) requires detailed patent data for each firm, making it poorly suited for use with commonly available datasets such as COMPUSTAT. A less data-intensive method, albeit much less sophisticated, is to simply use each firm's SIC code. The *Standard Industrial Classification Manual: 1987* states that the SIC "... was developed for use in the classification of establishments by type of

---

derived as  $d_{ij} = 2 * 3,959 * \arcsin \sqrt{\sin^2 \left( \frac{lat_j - lat_i}{2} \right) + \left( \frac{\cos(lat_j) + \cos(lat_i)}{2} \sin \left( \frac{lon_j - lon_i}{2} \right) \right)^2}$  where 3,959 is the radius of the earth in miles and latitude and longitude values are in radians.

activity ...” and that “[e]ach establishment is to be classified according to its primary activity.”<sup>27</sup> COMPUSTAT assigns a single SIC to each firm by analyzing establishment industrial activity provided in firm *10K* reports and selecting the code considered to represent the firm’s major area of activity.<sup>28</sup> Thus, an admittedly ordinal measure of technological proximity is used here whereby firms in same SIC group are defined as technologically near and all others as technologically distant. The technological proximity hypothesis is first examined assuming the technologically near spillover pool is limited to one’s own four-digit SIC. This analysis is repeated at the three-digit SIC to examine the sensitivity of results to a broader definition of technological nearness.

Use of a firm’s primary SIC to characterize technological location is far from ideal. A firm’s true technological position is most certainly a multidimensional measure reflecting their diversity of innovative activity<sup>29</sup> or the diversity of their production operations or both. However, measurement error associated with use of the binary definition of technological proximity proposed here should impose a downward bias on regression estimates. Thus, any evidence in their favor is arguably a robust finding.<sup>30</sup>

The raw financial variables are retrieved from the COMPUSTAT industrial, full coverage, and research annual data files. Observations in the raw data panel extend from 1970 to 1998<sup>31</sup>. Observations are selected for those firms whose primary industrial activity is within SIC 35 and whose corporate headquarters are within the 50 United States. Observations are deleted if the raw sales, employment, capital investment, or R&D expenditure entries are missing or combined with other variables. Further data trimming excludes observations where sales, employment, constructed capital stock, or constructed R&D stock are non-positive, leaving 4,680 observations.

---

<sup>27</sup> *Standard Industrial Classification Manual: 1987*, p. 11.

<sup>28</sup> *COMPUSTAT User’s Guide* (1997), p. 249.

<sup>29</sup> Perhaps most properly computed in Jaffe 1986.

<sup>30</sup> Furthermore, Griliches (1992, p. S33) suggests that “[o]ne could argue that this is what the SIC classification is for. Presumably, the usefulness of somebody else’s research to you is highest if he is in the same four-digit SIC classification as you are; it is still high if he is in the same three-digit industry group; and, while lower than before, the results of research by a firm in your own two-digit classification (but not three-digit) are more likely to be valuable to you than the average results of research outside it.”

<sup>31</sup> Although the annual financial variables extend back to 1950, COMPUSTAT has only been tabulating R&D expenditures since 1970.

An additional 105 observations are excluded which represent the lowest two percent of observations ordered by sales. This censoring is imposed to restrict the panel from young firms, often high-tech, that “go public” with little or no sales and are poor candidates for modeling with the production function framework. Indeed, 87 of the 105 excluded observations come from the highest tech sector in the panel; SIC 357 - Computer and Office Equipment. Preliminary analyses to check the sensitivity of regression estimates to this censoring indicate that the only result is to improve the goodness of fit of the own R&D coefficient. All other coefficient estimates are unaffected.

The final, unbalanced data panel includes 4,575 observations on 515 firms extending from 1972 to 1995. These data span 39 states and 29 four-digit (or eight three-digit) SIC’s. Summary statistics are presented in table 3.

<b>Table 3</b>				
<b>Final Data Panel Summary Statistics; n = 4,575</b>				
<u>Variable</u>	<u>mean</u>	<u>std. dev.</u>	<u>min</u>	<u>max</u>
sales (\$M)	619	2,100	0.299	41,400
capital stock (\$M)	294	1,040	0.010	15,100
employment (k)	7.71	22.5	0.003	438
R&D stock (\$M)	143	607	0.001	10,100

Spillover pools are seeded with a value of one in order to ensure they are well defined under the log transformation for all firm observations. This is a problem with the geographically near/technologically near spillover pool variable in particular which will be empty for firms that are in a remote location from others in their industry. Table 4 reports the average spillover pool size by total R&D and number of firms in each pool assuming a geographically near spillover pool boundary radius of 50 miles and a technologically near spillover pool defined by a firm’s four-digit SIC.

Recall that a principal motivation of this research is to assess the impact of industrial agglomeration on our interpretation of evidence of inter-firm spillovers. It is presumed that firms in the same industry tend to cluster spatially which, to some degree, will make evidence of technologically mediated spillovers and geographically mediated spillovers observationally equivalent. Before we begin the regression analysis, therefore,

<b>Table 4</b>		
<b>Spillover Pool Summary:</b>		
<b>average pool R&amp;D stock / average number firms in pool</b>		
	geographically near (within 50 miles)	geographically distant (beyond 50 miles)
technologically near (in same 4-digit SIC):	84/0.6	857/9.2
technologically distant (outside own 4-digit SIC):	1,500/9.0	26,000/217

it is interesting to consider the extent to which industrial agglomeration is present in the particular data panel used in this study.

Table 5 presents a comparison of the probability that any pair of firms in the panel are in the same industry (column b) to the conditional probability that they are in the same industry given they are in the ‘same location’ (column c). Firms are considered to be in the ‘same location’ if the distance between them does not exceed the value specified in column a. Firms are considered to be in the same industry if they share the same four-digit SIC. Conditional probabilities are calculated assuming ‘same location’ is defined with a 25, 50, and 100 mile radius around each firm  $i$ .

<b>Table 5</b>				
<b>Industrial Agglomeration among 515 Firms in the Panel</b>				
(a)	(b)	(c)	(d)	(e)
maximum distance for any pair of firms $i$ and $j$ to be considered in the ‘same location’:	Pr[SIC <sub><math>j</math></sub> =SIC <sub><math>i</math></sub> ]	Pr[SIC <sub><math>j</math></sub> =SIC <sub><math>i</math></sub>   loc <sub><math>j</math></sub> =loc <sub><math>i</math></sub> ]	$d_m = (c) - (b)$	t-H <sub>0</sub> : $d_m=0$
25 miles	0.052 (0.022)	0.289 (0.158)	0.237 (0.158)	34.0
50 miles	0.052 (0.022)	0.217 (0.149)	0.165 (0.150)	25.0
100 miles	0.052 (0.022)	0.172 (0.132)	0.120 (0.132)	20.6

\* standard errors in parentheses.

Column (b) indicates that for each firm  $i$ , an average of 5.2 percent of the remaining firms in the panel are in the same four-digit SIC. Column (c) reports these values conditional on the definition provided in (a). For example, if we consider only the subset of firms

within 50 miles of each other, fully 21.7 percent are in the same narrowly defined industry group. That is, firms within 50 miles of one another are over four times as likely to be in the same four-digit SIC than are any pair of firms from the panel in general.

The mean difference between the baseline (column b) and conditional (column c) probabilities for each firm is computed (column d) to test the hypothesis that these samples are selected from the same population. Returning to our example, the probability that any firm within 50 miles of another is in the same industry is, on average, 16.5 percentage points higher than the baseline probability. This sample mean is highly statistically significant, revealing evidence of industrial agglomeration in these data and suggesting inter-firm measures of geographic proximity will be correlated with measures of technological proximity.

Equation 7 presents the restricted form of the model derived in section 3 in an error components framework.<sup>32</sup>

$$y_{it} = \Gamma_0' D_{it} + \gamma_1 c_{it} + \gamma_2 l_{it} + \gamma_3 k_{it} + \gamma_4 \sum_{n \neq i}^{G_{Ni}} k_{nt-1} + \gamma_5 \sum_{z \neq i}^{G_{Zi}} k_{zt-1} + (u_i + \varepsilon_{it}) \quad (7)$$

The model is restricted in the sense that external R&D stocks will be sorted into only two of four possible spillover pools.  $D_{it}$  is a vector of location, sector, and time-specific dummy variables for firm  $i$  in time  $t$ . State and four-digit SIC dummies are included to control for a broad range of technological opportunity shocks and other narrowly defined fixed-effects. Individual year dummies are included to control for non-linear time specific technological opportunities.

Section 5 begins by considering the hypothesis that R&D spillovers are greatest from those firms that are geographically proximate, defined as those firms within 50 miles of the spillover receiving firm. For each firm  $i$ , all other firms are sorted into either a geographically near pool,  $G_{Ni}$ , or geographically distant pool,  $G_{Zi}$ . Each spillover pool variable is constructed by summing over the natural log transformed R&D stocks of all the firms in the respective pool.

---

<sup>32</sup> Mairesse 1993, p. 433.

Capital, labor, and own R&D stocks are assumed to become productive contemporaneously with own output. Spillover pool stocks are assumed to become productive at a one-year lag to reflect the additional time it takes to internalize publicly available knowledge.<sup>33</sup> Evenson (1968) examines aggregate data for U.S. agriculture and concludes that the lag structure of R&D takes an inverted V shape. He concludes that the peak weight from R&D flows is at five to eight year lags and little contribution is received from R&D expenditure at lags in excess of 10 to 16 years. But Wagner (1968) provides survey evidence that these lags are much shorter for industrial R&D, perhaps reflecting the more applied nature of private R&D expenditures. Assuming a constant rate of R&D expenditure, the 15 percent depreciation rate assumed in this study corresponds to an average R&D stock vintage of six years.

Due to the serially correlated nature of R&D time series, selecting the proper lag structure for R&D stock is a commonly recognized problem in the R&D productivity literature. As is generally the case, point estimates of R&D pool coefficients reported in the following sections are sensitive to alternative lag assumptions. For obvious reasons, however, this sensitivity is restricted to the time-series dimension. Furthermore, the qualitative findings of the analysis are unaffected by varying assumptions on lag length.

Following Mairesse (1993), between- (8.a) and within-firm (8.b) regressions are estimated in place of the total panel model presented in (7). This approach allows explicit examination of the cross-sectional and time-series variation in the data.

$$\bar{y}_i = \Gamma_0^b \bar{D}_i + \gamma_1^b \bar{c}_i + \gamma_2^b \bar{l}_i + \gamma_3^b \bar{k}_i + \gamma_4^b \overline{\sum_{n \neq i}^{G_{N_i}} k_n} + \gamma_5^b \overline{\sum_{z \neq i}^{G_{Z_i}} k_z} + (u_i + \bar{\varepsilon}_i) \quad (8.a)$$

$$y_{it} - \bar{y}_i = \Gamma_0^w (D_{it} - \bar{D}_i) + \gamma_1^w (c_{it} - \bar{c}_i) + \gamma_2^w (l_{it} - \bar{l}_i) + \gamma_3^w (k_{it} - \bar{k}_i) \\ + \gamma_4^w \left( \sum_{n \neq i}^{G_{N_i}} k_{nt-1} - \overline{\sum_{n \neq i}^{G_{N_i}} k_n} \right) + \gamma_5^w \left( \sum_{z \neq i}^{G_{Z_i}} k_{zt-1} - \overline{\sum_{z \neq i}^{G_{Z_i}} k_z} \right) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (8.b)$$

---

<sup>33</sup> A preliminary sensitivity analysis of this lag structure suggests the results reported below are robust. Various lags of own capital, own R&D, and R&D pools were considered. The main result seems to be to affect the point estimates of own input coefficient. The magnitudes and, more importantly, relative magnitudes of spillover pool coefficients are unaffected by this sensitivity.

Following Bartlesman et al. (1994), between-firm variation coefficient estimates are interpreted as long-run elasticities as they reflect the response of output to permanent changes in variable values.

Estimates from the between-firm regression are often dismissed by econometricians as biased since the individual error term,  $u_i$ , is presumably correlated with the explanatory variables.<sup>34</sup> Estimates in 8.a will be biased to the extent that industry- and state-specific fixed effects do not control for firm specific variation in external (market and technological opportunities) and internal (personnel quality) fixed effects. Within-firm variation coefficient estimates are interpreted as short-run elasticities since they are driven by more transient changes in variable values. Mairesse (1993) argues that these estimates may also be biased to the extent that time dummies do not control for temporal variation in capacity utilization.<sup>35</sup> Therefore, both between- and within-firm regression estimates are reported in the following chapter.

A two stage least squares (2SLS) procedure is used to control for estimation bias due to simultaneous choice of capital, labor, and output. This estimation procedure employs lagged values of own variable inputs as instruments for capital, labor, and own R&D. The regressions proposed in equations 8.a and 8.b are repeated to test the technological proximity hypothesis and finally the combined geographic/technological proximity hypothesis.

## **5. THE RESTRICTED ANALYSES: TESTING THE GEOGRAPHIC AND TECHNOLOGICAL PROXIMITY HYPOTHESES INDEPENDENTLY**

This section presents between- and within-firm regression estimates from the restricted model proposed in section 3. These preliminary results are provided as a point of departure for the general results and sensitivity analyses presented in the following section. They are intended to illustrate that the framework proposed here can obtain

---

<sup>34</sup> Mairesse 1993, p. 427. For arguments in favor of within estimates due to biases expected in between estimates, see Mundlak (1961), Hoch (1962), and Mundlak (1978).

<sup>35</sup> Ibid., p. 429. For arguments in favor of between estimates due to biases expected in within estimates, see Mairesse (1978) and Griliches and Mairesse (1984).

findings consistent with the conventional wisdom that R&D spillovers are enhanced by both geographic and technological proximity.<sup>36</sup>

Table 6 reports regression estimates for equations 8.a and 8.b where “near” is defined as “within a 50 mile radius.” All firms 50 miles or nearer to each firm  $i$ , regardless of their technological proximity, are pooled into the set  $G_{Ni}$ . The R&D of these firms compose the near spillover pool for firm  $i$ . The remaining set of firms, set  $G_{Zi}$ , define the distant spillover pool for firm  $i$ .

Goodness of fit is high in both the between- and within-firm regressions. The cross-sectional model exhibits nearly constant returns to scale with respect to conventional inputs, though the capital coefficient is not statistically significant. In the

<b>Table 6</b>		
<b>Geographic Proximity Hypothesis: estimates assuming spillover pool boundary = 50 miles<sup>a</sup></b>		
	<u>Between<sup>b</sup></u>	<u>Within<sup>c</sup></u>
$c$	0.040 (0.035)	0.116** (0.022)
$L$	0.975** (0.040)	1.026** (0.021)
$k$	0.027** (0.011)	0.068** (0.010)
$\Sigma k_{Gn}$	0.012** (0.002)	0.005** (0.001)
$\Sigma k_{Gz}$	0.004** (0.001)	0.000 (0.001)
adj R <sup>2</sup>	0.96	0.69
n	515	4,575

<sup>a</sup> standard errors in parentheses.

<sup>b</sup> intercept, state, and industry specific fixed effect coefficients not reported.

<sup>c</sup> intercept and year specific fixed effect coefficient estimates not reported.

\* significant at the 10 percent level.

\*\*significant at the 5 percent level.

time-series regression, all own input coefficient estimates are significant at the five-percent level. The time-series model exhibits significant increasing returns to scale with

<sup>36</sup> The following analysis was initially conducted including the 2<sup>nd</sup> order terms derived in equation (4). However, these coefficient estimates were not significant. Also, see footnote 15.

respect to conventional inputs which is likely a result of measurement bias due to variation in capacity utilization over the business cycle.<sup>37</sup>

Turning to the spillover pool coefficients, the geographically near pool ( $\Sigma k_{Gn}$ ) coefficient is significant in both the between- and within-firm regressions. In the between case, own output is three times more responsive to changes in R&D stocks less than 50 miles away than it is to changes in R&D beyond 50 miles. The near pool coefficient is larger than that of the distant spillover pool ( $\Sigma k_{Gz}$ ) in the within case as well, a finding consistent with the claim the geographic distance attenuates inter-firm spillovers.

It is also interesting to compare the magnitude of the geographically near spillover pool coefficient to that of own R&D. Consider first the time-series dimension. A one percent shock to own R&D will result in an increase in own output of 0.068 percent in the short run. The impact of a change in R&D by the average firm in the nearby spillover pool is roughly 1/14 this magnitude. Recall from table 4 that, on average, approximately 10 firms are in the geographically nearby spillover pool. Therefore, a synchronous increase of R&D by firms in one's near spillover pool will result in an increase in output that is nearly as significant as the excess returns attributable to own R&D.<sup>38</sup>

Changes to external R&D appear even more important when we consider estimates obtained from the between-firm regression. In this case, changes in external R&D are nearly half as important as changes in own R&D. In the long run, a synchronous change in R&D of nearby firms may increase output by a magnitude greater

---

<sup>37</sup> Since the production function is primarily a technical relation, the correct measurement concept for independent variables is 'utilized inputs.' If capital, labor, or R&D stocks are costly to adjust then a firm will choose to overutilize (underutilize) inputs in the early stages of an economic expansion (contraction). Measurement error is introduced by use of purchased inputs as a proxy for utilized inputs. For example, consider a demand shock for a firm's product. Since capital is costly to adjust, the firm will increase both purchased inputs and utilization of existing stocks to service the increased demand. Total capital utilization will increase in proportion to output. But since input purchases understate this increase in utilization, the relatively large increase in output will be attributed to the relatively small increase in purchased inputs, imposing an upward bias on the capital coefficient estimate.

<sup>38</sup> The own R&D coefficient estimates presented here are understated due to failure to control for double counting of R&D expenditures. Schankerman (1981) showed that since these charges are also included in capital and labor variables, conventional input coefficient estimates account for normal returns to R&D inputs. The coefficient presented here is typically given an "excess returns interpretation." That is, the coefficient is interpreted to reflect only the risk premium associated with R&D expenditures rather than the total return.

than that of a comparable change in own R&D. This may seem an extreme finding but it is consistent with the view that, over time, returns to knowledge generating activities are difficult to contain.

Table 7 reports between- and within-firm regression spillover pool coefficient estimates as the definition of the nearby spillover pool is varied from 25 miles to 800 miles.<sup>39</sup> The geographically near pool coefficient exceeds that of the distant pool in nearly all cases. This difference, however, appears to diminish as the geographically near/distant spillover pool boundary is expanded. As the distance to the geographically near spillover pool boundary is increased, firms from the distant pool are shifted into the

<b>Table 7</b>							
<b>Geographic Proximity Hypothesis: sensitivity of R&amp;D coefficients to varying spillover pool boundary<sup>a</sup></b>							
	miles:	<u>25</u>	<u>50</u>	<u>100</u>	<u>200</u>	<u>400</u>	<u>800</u>
B	$k$	0.028**	0.027**	0.030**	0.027**	0.029**	0.026**
E		(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
T	$\Sigma k_{Gn}$	0.015**	0.012**	0.009**	0.006**	0.007**	0.005**
W		(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
E	$\Sigma k_{Gz}$	0.004**	0.004**	0.004**	0.004**	0.004**	0.004**
E		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N							
W	$k$	0.067**	0.068**	0.069**	0.069**	0.070**	0.067**
I		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
T	$\Sigma k_{Gn}$	0.008**	0.005**	0.002*	-0.000	0.001**	0.001*
H		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
I	$\Sigma k_{Gz}$	0.000	0.000	0.000	0.000	0.000	-0.000
N		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

<sup>a</sup> standard errors in parentheses.

\* significant at the 10 percent level.

\*\*significant at the 5 percent level.

near pool. These firms are ‘least near’ relative to the composition of firms in the near spillover pool. This reallocation also removes the ‘least distant’ subset of firms from the distant pool. In both cases, external R&D stock coefficients will be biased downward if geographic distance attenuates spillovers.

<sup>39</sup> Conventional input coefficients and goodness of fit statistics are not reported since they are largely unchanged as compared with those reported in table 6.

This sensitivity condition is satisfied for the geographically near spillover pool coefficient in both the cross-section and time series dimensions. In the between case, the nearby spillover pool coefficient drops by two thirds when the near/distant boundary is expanded from 25 to 800 miles. This reduction in coefficient value is even more extreme in the within case. Indeed, the nearby pool coefficient appears unimportant for distances in excess of 100 miles. The geographically distant spillover pool coefficients, however, do not appear to be diminished in the between case as this pool is restricted to only the most distant firms.

Comparison of nearby and distant spillover pool coefficients are consistent with the hypothesis that geographic proximity enhances spillover transmission. The output response to an increase in R&D from an average firm within 50 miles is three times as large as that for an average firm beyond this radius. This relation obtains in the time-series as well where distant spillover pool coefficients are insignificant. Furthermore, sensitivities of spillover pool coefficients to variation in the distance to the spillover pool boundaries are also consistent with the hypothesis that geographic distance attenuates R&D spillovers.

Table 8 reports regression estimates for equations 8.a and 8.b where “near” is defined as “within the same four-digit SIC.” All firms within one’s own narrowly defined industry group, regardless of geographic proximity, are pooled into the set  $T_{Ni}$ . The R&D of these firms compose the near spillover pool for firm  $i$ . The remaining set of firms,  $T_{Zi}$ , define the distant spillover pool for firm  $i$ .

Goodness of fit is high in both the between- and within-firm regressions. All own input coefficients are significant at the five percent level in both cases. This reflects an increase in magnitude and significance for the own capital coefficient which was insignificant in the between case in table 6. The cross-sectional model exhibits slightly increasing returns to scale with respect to conventional inputs. The time-series model exhibits significantly increasing returns to scale. The latter finding is again interpreted to reflect estimation bias due to variation in capacity utilization over the business cycle.

Turning to the spillover pool coefficients, the technologically near spillover pool

<b>Table 8</b>		
<b>Technological Proximity Hypothesis: estimates assuming spillover pool boundary = 4-digit SIC<sup>a</sup></b>		
	<u>Between<sup>b</sup></u>	<u>Within<sup>c</sup></u>
<i>c</i>	0.068** (0.034)	0.109** (0.021)
<i>l</i>	0.953** (0.038)	1.033** (0.021)
<i>k</i>	0.036** (0.010)	0.048** (0.010)
$\Sigma k_{Tn}$	0.042** (0.003)	0.011** (0.001)
$\Sigma k_{Tz}$	0.002** (0.001)	-0.000 (0.001)
adj R <sup>2</sup>	0.96	0.70
n	515	4,575

<sup>a</sup> standard errors in parentheses.

<sup>b</sup> intercept, state, and industry specific fixed effect coefficients not reported.

<sup>c</sup> intercept and year specific fixed effect coefficient estimates not reported.

\* significant at the 10 percent level.

\*\*significant at the 5 percent level.

( $\Sigma k_{Tn}$ ) coefficient is significant in both the between- and within-firm regressions. In both cases this coefficient is significantly larger than that of the distant spillover pool ( $\Sigma k_{Tz}$ ), a finding consistent with the hypothesis that technological closeness accentuates inter-firm spillovers.

It is also interesting to compare the magnitude of the technologically near spillover pool coefficient to that of own R&D. The time-series dimension regression suggests a one percent shock to own R&D will result in an increase in own output of 0.048 percent in the short run. The impact of a change in R&D by the average firm in the nearby spillover pool is nearly one-fourth this magnitude. Recall from table 4 that, on average, there are approximately 10 firms are in the technologically near spillover pool. Therefore, a synchronous increase of R&D by all firms in this pool will result in an increase in output that is larger than the response from an increase in one's own R&D. This estimate is also larger than that from the prior section regarding short-run response to changes in the R&D stocks of geographic neighbors.

Changes in external R&D appear even more important when we consider estimates obtained from the between-firm regression. In this case, changes in the R&D of technological neighbors are at least as important as changes in own R&D. While this estimate may appear extreme, it is important to keep in mind that the own R&D coefficient is biased downward due to double counting of R&D expenditures in labor and capital variables.<sup>40</sup> Nonetheless, this result is not inconsistent with earlier findings in the literature of economically important returns from external R&D stocks.<sup>41</sup> Furthermore, the finding of large and significant spillovers from technological neighbors is consistent with the view advanced by Cohen and Levinthal (1989) that the objective of much R&D investment is to develop a firm's absorptive capacity; i.e. the ability to make use of rivals' ideas.

Table 9 reports spillover pool coefficient estimates from the between- and within-firm regressions where the definition of the nearby spillover pool is relaxed from the four-digit to the three-digit SIC level.<sup>42</sup> This sensitivity results in reallocating the most technologically similar firms from the distant spillover pool to the near spillover pool. If technological similarity accentuates spillovers the external R&D stock coefficients will be biased downward as the definition of the spillover pool boundary is expanded. This condition is satisfied for both the near and distant spillover pools in both the cross-sectional and time-series regressions. In the between-firm regression, the near pool coefficient is larger than that of the distant pool for technological pool boundaries defined at both the four-digit and the three-digit levels. Furthermore, as we move from the four-digit to the three-digit specification, effectively sweeping technologically dissimilar firms into the nearby pool, this coefficient is biased downward. The technologically distant spillover pool coefficient is also biased downward as this pool is increasingly limited to the most technologically dissimilar firms in the panel.

Time-series estimates are also consistent with the hypothesis that spillovers from

---

<sup>40</sup> See footnote 37. In contrast, while the within coefficient is subject to the same downward bias, it is also subject to an upward bias associated with procyclical variation in capacity utilization.

<sup>41</sup> See footnote 5.

<sup>42</sup> Conventional input coefficients and goodness of fit statistics are not reported since they are largely unchanged by this sensitivity

<b>Table 9</b>			
<b>Technological Proximity Hypothesis: sensitivity of R&amp;D coefficients to varying spillover pool boundary<sup>a</sup></b>			
	pool boundary:	<u>4-d SIC</u>	<u>3-d SIC</u>
B	$k$	0.036**	0.028**
E		(0.010)	(0.009)
T	$\Sigma k_{Tn}$	0.042**	0.013**
W		(0.003)	(0.001)
E	$\Sigma k_{Tz}$	0.002**	-0.002*
E		(0.001)	(0.001)
N			
W	$k$	0.048**	0.055**
I		(0.010)	(0.009)
T	$\Sigma k_{Tn}$	0.011**	0.007**
H		(0.001)	(0.001)
I	$\Sigma k_{Tz}$	-0.000	-0.001**
N		(0.001)	(0.000)

<sup>a</sup> standard errors in parentheses.

\* significant at the 10 percent level.

\*\*significant at the 5 percent level.

technologically similar firms are greater than are those from technologically distant firms. Nearby pool coefficients are larger than those of the distant spillover pool. In addition, the within-firm estimates of the technologically near spillover pool coefficient are biased downward as the pool boundary is expanded.

The combined results from tables 6 through 9 suggest that both technological and geographic proximity accentuate R&D spillovers. The strength of spillovers does appear to vary, with stronger spillovers estimated from a firms technological neighbors as compared to those from geographic neighbors. Nearby spillover pool coefficients from the relatively weakest definition of technological closeness (Table 9, last column) are roughly equal to those from the most restrictive definition of geographic closeness employed in these analyses (Table 7, first column). Recall from the previous chapter, however, that evidence of industrial agglomeration is present in this data panel (Table 5). This finding suggests that inter-firm measures of geographic proximity will be correlated with measures of technological proximity. Consequently, industrial agglomeration makes interpretation of the results presented in this section problematic.

## 6. THE GENERAL ANALYSIS: TESTING THE GEOGRAPHIC AND TECHNOLOGICAL PROXIMITY HYPOTHESES SIMULTANEOUSLY

This section presents estimates from the general spillover model proposed in section 3 which simultaneously accounts for the roles of geographic and technological proximity. Equations 8.a and 8.b are respecified to contain four spillover pools that vary in both the geographic and technological dimensions.

Table 10 reports between- and within-firm regression estimates obtained by defining R&D stocks within 50 miles as geographically near and those within a firm's own four-digit SIC as technologically near.  $\Sigma k_{GnTn}$ ,  $\Sigma k_{GnTz}$ ,  $\Sigma k_{GzTn}$ , and  $\Sigma k_{GzTz}$  represent R&D spillover pools that are geographically near / technologically near, geographically near / technologically distant, geographically distant / technologically near, and

<b>Table 10</b>		
<b>Combined Geographic and Technological Proximity</b>		
<b>Hypothesis: geographic boundary = 50 miles;</b>		
<b>technological boundary = 4-digit SIC<sup>a</sup></b>		
	<u>Between<sup>b</sup></u>	<u>Within<sup>c</sup></u>
<i>c</i>	0.056* (0.033)	0.105** (0.021)
<i>l</i>	0.963** (0.037)	1.030** (0.021)
<i>k</i>	0.037** (0.010)	0.048** (0.010)
$\Sigma k_{GnTn}$	0.032** (0.007)	0.010** (0.004)
$\Sigma k_{GnTz}$	0.009** (0.002)	0.005** (0.001)
$\Sigma k_{GzTn}$	0.030** (0.003)	0.011** (0.001)
$\Sigma k_{GzTz}$	0.002** (0.001)	-0.000 (0.001)
adj R <sup>2</sup>	0.96	0.70
n	515	4,575

<sup>a</sup> standard errors in parentheses.

<sup>b</sup> intercept, state, and industry specific fixed effect coefficients not reported.

<sup>c</sup> intercept and year specific fixed effect coefficient estimates not reported.

\* significant at the 10 percent level.

\*\*significant at the 5 percent level.

geographically distant / technologically distant, respectively.

Goodness of fit is high in both the between- and within-firm regressions. In the between case, capital and labor coefficients are significant at the 10 and five percent levels, respectively. Own R&D is also significant at the five percent level. The cross-sectional model exhibits constant or slightly increasing returns to scale with respect to conventional inputs. All own input coefficient estimates are significant at the five percent level in the time-series regression. Again, this model exhibits significant increasing returns to scale with respect to own factor inputs.

Turning to the spillover pool coefficients, it is now possible to isolate the effects of geographic proximity from technological proximity. First, consider the importance of technological proximity for spillovers *after* controlling for geography. To do so we compare coefficients from technologically near and distant R&D that meet the same geographic criteria. In the between case, consider spillovers from R&D stocks within 50 miles of the spillover receiving firm. The output response to R&D from the same four-digit SIC is over three times larger than that from outside of a firm's own four-digit SIC (0.032 vs. 0.009). This relation holds for geographically distant R&D stocks as well: spillovers from R&D beyond 50 miles are greatest from that within a firm's own four-digit SIC (0.030 vs. 0.002).

Qualitatively similar results are obtained from the within-firm regressions. Spillovers from geographically near R&D are twice as large from that within a firm's own four-digit industry. However, spillovers from geographically distant R&D are *only* observed from within a firm's technology group.

The independent role of geographic proximity can be assessed by comparing spillover pool coefficients that meet common technological criteria. In the between case, a comparison of  $k_{GnTn}$  to  $k_{GzTn}$  pool coefficients (0.032 v. 0.030) indicates that spillovers from within a firm's own four-digit SIC do not appear to be attenuated by distance. In contrast, our intuition regarding the importance of geographic proximity is supported if we consider only spillovers from technologically dissimilar firms. Within this subset,

spillovers are greatest from geographic neighbors (0.009 vs. 0.002). This same pattern is observed from the within-firm regression results.

These results present a characterization of R&D spillovers that is much richer than that provided by the restricted analysis reported in the previous section. The general analysis results are consistent with evidence of the importance of technological proximity obtained in the restricted analysis. Technological proximity appears to enhance inter-firm spillovers regardless of geographic proximity. But evidence regarding the importance of geographic proximity obtained from the general analysis is mixed. Geographic distance does attenuate spillovers between technologically dissimilar firms. However, spillovers among firms in the same four-digit SIC do not appear attenuated by distance. Thus, at least part of the correlation between output and the R&D of geographic neighbors appears to be driven by the tendency of technologically similar firms to agglomerate in space.

As in the previous section, additional evidence regarding the role of geographic and technological factors for R&D spillovers can be obtained by varying the spillover pool boundaries. Tables 11 and 12 report spillover pool coefficient estimates for the between- and within-firm regressions, respectively. The ‘geographically near’ pool specification is varied from 25 to 800 miles and the technologically near specification is relaxed from the four-digit to the three-digit SIC.<sup>43</sup>

In the between case, coefficients of the ‘technologically near’ spillover pools ( $\Sigma k_{GnTn}$  and  $\Sigma k_{GzTn}$ ) are not diminished as the near pool radius is expanded from 25 to 800 miles.<sup>44</sup> This finding is counterintuitive, suggesting that geographic distance has little attenuating effect on spillovers among technologically similar firms.

The geographically near/technologically distant spillover pool coefficient behaves in a more intuitively appealing way. This coefficient suggests that spillovers from R&D stocks within a firm’s own narrowly defined, four-digit SIC and within 25 miles are

---

<sup>43</sup> Conventional input coefficients and goodness of fit statistics are not reported since they are largely unchanged by this sensitivity

<sup>44</sup> However, the  $k_{GnTn}$  pool coefficient is diminished as the spillover pool radius is contracted from 50 to 25 miles. This latter effect is puzzling but may reflect heightened output and factor market competition among geographically proximate direct competitors. If so, this effect should be exacerbated in the time-series dimension. Indeed, this is the case as will be observed in table 12.

significant. Furthermore, as the definition of geographic proximity is relaxed to include more distant R&D stocks into this pool, the coefficient is attenuated. These findings provide a richer characterization of the R&D spillover process, suggesting a departure between spillovers from other firms within one's own SIC and those from firms without.

The lower half of table 11 relaxes the technological boundary of the near pool to include firms in the broader, three-digit SIC. As expected, all of the spillover pool coefficient estimates are biased downward by an expansion of the technological boundary from the four-digit to the three-digit level. The most interesting result, however, is again with the  $\Sigma k_{GnTz}$  pool coefficient. This is the only coefficient that shows any evidence of

<b>Table 11</b>							
<b>Between-firm Regression Estimates: combined geographic and technological proximity hypothesis<sup>a</sup></b>							
	miles:	<u>25</u>	<u>50</u>	<u>100</u>	<u>200</u>	<u>400</u>	<u>800</u>
4 - d. S I C	k	0.039** (0.010)	0.037** (0.010)	0.038** (0.010)	0.036** (0.010)	0.038** (0.010)	0.036** (0.010)
	$\Sigma k_{GnTn}$	0.026** (0.009)	0.032** (0.007)	0.038** (0.007)	0.032** (0.006)	0.034** (0.004)	0.033** (0.004)
	$\Sigma k_{GnTz}$	0.012** (0.003)	0.009** (0.002)	0.005** (0.002)	0.002* (0.001)	0.004** (0.002)	0.002* (0.001)
	$\Sigma k_{GzTn}$	0.031** (0.003)	0.030** (0.003)	0.030** (0.003)	0.031** (0.003)	0.030** (0.004)	0.031** (0.004)
	$\Sigma k_{GzTz}$	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)
	k	0.030** (0.010)	0.029** (0.009)	0.028** (0.009)	0.027** (0.009)	0.029** (0.009)	0.027** (0.009)
	$\Sigma k_{GnTn}$	0.017** (0.003)	0.022** (0.002)	0.021** (0.002)	0.016** (0.002)	0.016** (0.002)	0.014** (0.001)
	$\Sigma k_{GnTz}$	0.007* (0.004)	0.001 (0.002)	-0.001 (0.002)	-0.004** (0.002)	-0.001 (0.002)	-0.002* (0.001)
	$\Sigma k_{GzTn}$	0.012** (0.001)	0.012** (0.001)	0.012** (0.001)	0.012** (0.001)	0.012** (0.001)	0.012** (0.001)
	$\Sigma k_{GzTz}$	-0.002** (0.001)	-0.002** (0.001)	-0.002** (0.001)	-0.002** (0.001)	-0.002* (0.001)	-0.001 (0.001)

<sup>a</sup> standard errors in parentheses.

\* significant at the 10 percent level.

\*\*significant at the 5 percent level.

attenuation as the geographic radius is expanded. Indeed, this coefficient is completely attenuated when the pool boundary is expanded beyond 25 miles. This estimate suggests that the only spillovers received from very technologically dissimilar firms (those outside one's own three-digit SIC) are from very geographically proximate R&D stocks.

It is also interesting that some evidence of the importance of geographic proximity emerges in the  $\Sigma k_{GnTn}$  pool as the technological boundary is expanded to the three-digit SIC. The  $\Sigma k_{GnTn}$  pool coefficient, while insensitive to expansion of pool boundaries in the top half of table 11, appears to be declining, even if only slightly, in the bottom half of the table. Given the measurement problems in our proxy for firm location, this result would at least recommend additional research into the localization of spillovers among firms in the same three-digit SIC. The bias reported in table 11 is consistent with the claim that spillovers from R&D outside of a firm's own four-digit SIC but within one's own three-digit SIC are attenuated by distance.

The within-firm regression results are qualitatively similar to those obtained in the cross section. However, a few exceptions can be observed. First, under the most restrictive definition of the  $\Sigma k_{GnTn}$  spillover pool, 25 mile radius and four-digit SIC, this coefficient is negative. If the within-firm regression emphasizes short-run variation in the data, this coefficient is not inconsistent with the interpretation offered for the diminished coefficient in the between case: i.e. the geographically near/technologically near spillover pool may be proxying for competitive pressure in output and factor input markets by a firm's closest competitors.<sup>45</sup>

Another difference from the between results is that the  $\Sigma k_{GnTn}$  coefficient now appears to be biased downward as the geographic boundary is expanded in the top panel of table 12 rather than in the bottom panel. That is, in the short-run, spillovers from R&D in a firm's narrowly defined four-digit SIC are sensitive, if only slightly so, to geographic distance while spillovers from within a firm's own three-digit SIC are not.

Lastly and perhaps most interesting is the finding of the attenuating effect of distance on the  $\Sigma k_{GnTz}$  pool, a finding observed in the between regression results as well.

---

<sup>45</sup> See footnote 44.

<b>Table 12</b>							
<b>Within-firm Regression Estimates: combined geographic and technological proximity hypothesis<sup>a</sup></b>							
	miles:	<u>25</u>	<u>50</u>	<u>100</u>	<u>200</u>	<u>400</u>	<u>800</u>
4	k	0.045**	0.048**	0.048**	0.049**	0.050**	0.047**
-		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
d.	$\Sigma k_{GnTn}$	-0.011**	0.010**	0.009**	0.004*	0.004**	0.005**
		(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
S	$\Sigma k_{GnTz}$	0.009**	0.005**	0.002**	-0.000	0.001**	0.001**
I		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
C	$\Sigma k_{GzTn}$	0.011**	0.011**	0.011**	0.012**	0.014**	0.016**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	$\Sigma k_{GzTz}$	-0.000	-0.000	-0.000	-0.000	-0.000	-0.001*
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
3	k	0.055**	0.056**	0.058**	0.057**	0.058**	0.053**
-		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
d.	$\Sigma k_{GnTn}$	0.005**	0.006**	0.004**	0.003**	0.004**	0.005**
		(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
S	$\Sigma k_{GnTz}$	0.004**	0.001	0.001	-0.001	0.000	0.000
I		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
C	$\Sigma k_{GzTn}$	0.007**	0.007**	0.007**	0.007**	0.007**	0.007**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	$\Sigma k_{GzTz}$	-0.001**	-0.001**	-0.001**	-0.001**	-0.002**	-0.003**
		(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)

<sup>a</sup> standard errors in parentheses.

\* significant at the 10 percent level.

\*\*significant at the 5 percent level.

This coefficient remains sensitive to distance under the four-digit SIC boundary assumption. Under the three-digit SIC boundary assumption, spillovers again are only significant from R&D stocks within 25 miles of the spillover receiving firm. The short-run output response to changes in geographically near R&D appear largely to be driven by the subset of firms within one's own three-digit SIC.

The conclusions from this general analysis reveal subtleties masked by the partial analyses provided in section 5. While geographic distance appears to play some role in attenuating R&D spillovers, the effect appears to be limited to spillovers from outside one's own narrowly defined, four-digit SIC. It does not appear to matter whether R&D in a firm's own four-digit SIC is conducted geographically near or distant. Spillovers from

R&D outside a firm's own four-digit SIC do appear to be attenuated by distance. When a spillover pool is constructed to contain all R&D outside a firm's own four-digit SIC, that coefficient is attenuated as the pool boundary is expanded to include increasingly distant firms. If firms from one's own three-digit SIC are then excluded from this pool, spillovers appear to be significant from only the most geographically proximate firms.

## 7. CONCLUDING DISCUSSION

This analysis improves our understanding of the importance of geographic and technological factors for inter-firm spillovers from R&D activity. Parameter estimates obtained in a production function framework indicate that spillovers are significant and important from geographically and technologically proximate R&D stocks. This study demonstrates, however, that R&D spillover estimates obtained by methods that fail to control for industrial agglomeration may not reveal the conditional nature of the importance of geography.

The restricted analysis results presented in section 5 are consistent with intuition and existing empirical evidence which suggests that both geographic and technological distance attenuate knowledge spillovers. Results from the general analysis reported in section 6, however, suggest that the importance of geographic proximity is conditional on the technical relation between spillover sending and receiving units. Spillovers from R&D outside a firm's own narrowly defined industry group are increasing in geographic proximity. However, R&D spillovers from within a firm's own industry are insensitive to distance. Conversely, evidence that technological similarity accentuates spillovers is insensitive to distance between spillover sending and receiving units.

Table 10 indicates that the short-run return from R&D stocks within a firm's own four-digit SIC is, on average, about one-fifth as important as changes in a firm's own R&D. A one percent increase in the R&D stock of a single firm in the same industry implies a 0.01 percent increase in own output. In the longer-run, these spillovers are nearly as important as the contribution of own R&D stocks.<sup>46</sup> These magnitudes are

---

<sup>46</sup> Recall, however, that the own R&D coefficient is underestimated due to double counting with no countervailing bias, as in the within case, associated with procyclical variation in capacity utilization.

unaffected by geographic distance between spillover sending and receiving firms—the same average elasticity is reported for R&D within and beyond a 50 mile radius around each firm.

In contrast, returns from the R&D of technologically distant firms are sensitive to geographic proximity to the spillover receiver. In the short-run, the returns from the R&D of each technologically dissimilar firm within 50 miles are half as large as those from technologically similar firms (0.005 vs. 0.010). In the long-run the factor is three (0.032 vs. 0.009). But long-run returns from the R&D of technologically dissimilar firms beyond 50 miles are less than one-fourth the magnitude of their geographically proximate counterparts. And short-run returns are insignificant altogether.

These findings are not inconsistent with existing evidence of the importance of geographic and technological proximity for knowledge spillovers. While Jaffe et al. (1993) present compelling evidence that spillovers from inventive activity are localized, these results suggest that spillovers from innovative activity may behave differently. They find that patents have a higher likelihood of proximity to their own citations than to patents in a technologically similar control group. Their findings suggest that scientists who are defining the ‘state of the art’ benefit from proximity to potential collaborators. The evidence presented in this study, however, suggests that within-industry spillovers from industrial research activity that result in marketable innovations are not localized. To the extent that localization externalities from industrial R&D do exist they appear to be attributable to the application of knowledge from outside a firm’s own narrowly defined technology group. This interpretation suggests that the broader and presumably more applied set of activities summarized by private R&D expenditures result in spillovers that are quickly transmitted through firms in the same industry.<sup>47</sup>

The significance and magnitude of R&D spillovers documented here support continued research into the role of public policy intervention in the R&D arena. Positive

---

<sup>47</sup> If academic scientists and institutions are less likely to capitalize residual claims on inventions, one may also argue that the decision to collaborate is the outcome of a cost minimization calculation. In contrast, firms who may be more effective at appropriating rents from innovation may be more accurately modeled with a profit maximization objective and more likely to justify the cost of travel for the purpose of collaboration and knowledge transfer.

externalities to innovative activity suggest that a decentralized system of investment may not provide sufficient incentives to generate the socially optimal level of R&D activity. However, several studies have documented the superior profitability of private R&D investment over its public counterpart, arguing against government funded research. While this return premium may be attributable to the more applied nature of private R&D expenditures, it also reflects a difference of incentives directing employment of resources in each of these sectors.

The geographic and technological pattern of spillovers documented here has implications for our assessment of the efficiency of concentrated market structures in knowledge intensive industries. The finding that R&D spillovers are largest among firms in the same narrowly defined industry may support arguments in defense of increased concentration in particular industries. Presumably, to the extent that dominant firms internalize a larger fraction of total returns to innovative activity they will invest in more of it. Among technologically similar firms, the partial spillover enhancing effect of geographic proximity is much less significant. A defense of mergers between firms in a particular geographic region therefore may not be justified by the ‘internalization of knowledge spillovers’ argument.

The evidence presented in this study is particularly relevant to the current debate regarding the role of knowledge spillovers in city formation. These findings are consistent with those of Feldman and Audretsch (1999) who report “a tendency for innovative activity in complementary industries to cluster together in geographic space.”<sup>48</sup> This evidence is consistent with the argument advanced by Jacobs (1969) that firms in urban agglomerations are more productive because they are able to make use of ideas from other industries. This is not to say MAR externalities, knowledge spillovers among firms in the same industry, are unimportant. Indeed, the estimates presented in chapter five indicate that they are more important and more likely to be statistically significant than are Jacobs externalities. However, MAR ‘localization’ externalities do not appear to serve a localizing function. These findings suggest that knowledge

---

<sup>48</sup> Feldman and Audretsch 1999, p. 411.

spillovers may provide the centripetal force necessary for industrially diverse agglomerations. However, the set of activities that compose industrially homogenous cities are likely drawn together by something other than geographically dependent knowledge spillovers. Other potential agglomerating forces such as natural advantages or shared intermediate inputs deserve increased attention in the empirical literature.

Finally, these results may serve to guide future empirical research related to the new growth theory and also have implications for understanding the process of economic development and the effectiveness of development policy. The puzzling empirical regularity of conditional convergence is contrary to the predictions of basic knowledge-spillover growth models. In these models, the public goods nature of knowledge suggests economies should share a common set of production possibilities. The foregoing analysis would suggest convergence clubs might be defined by similarity of members' technological bases, regardless of geographic proximity.<sup>49</sup> If knowledge is transferred among technologically similar firms then our understanding of conditional convergence may be served by an analysis of institutional arrangements that facilitate industrial convergence. In contrast, countries without moderately advanced technology intensive sectors may receive little benefits from knowledge spillovers from more highly developed economies.

---

<sup>49</sup> For a recent theoretical contribution along these lines, see Basu and Weil, 1998.

## References

- Adams, James and Adam Jaffe, 1996, "Bounding the Effects of R&D: an investigation using matched establishment firm data," *RAND Journal of Economics*, 27: 700-721.
- Aghion, Philippe and Peter Howitt, 1992, "A Model of Growth Through Creative Destruction," *Econometrica*, 60(2): 323-351.
- Arrow, Kenneth J., 1962, "The Economic Implications of Learning by Doing," *Review of Economic Studies*, 29: 155-73.
- Audretsch, David and Maryann Feldman, 1996, "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86(3): 630-640.
- Bartlesman, Eric J., Ricardo J. Caballero, and Richard K. Lyons, 1994, "Customer- and Supplier-Driven Externalities," *American Economic Review*, 84(4): 1075-1084.
- Basu, Susanto, 1996, "Procyclical Productivity: Increasing Returns or Cyclical Utilization?," *Quarterly Journal of Economics*, 111(3): 719-51.
- \_\_\_\_\_, and David N. Weil, 1998, "Appropriate Technology and Growth," *Quarterly Journal of Economics*, 113(4): 1025-54.
- Bowker, R. R., Database Publishing Group, 1992, *Directory of American Research and Technology 1993: Organizations Active in Product Development for Business*, 27<sup>th</sup> edition, New Providence, New Jersey: R. R. Bowker.
- Cameron, Gavin, 1996, "Innovation and Economic Growth," *Centre for Economic Performance Discussion Paper no. 277*.
- Chirinko, Robert S., Steven M. Fazzari, and Andrew P. Meyer, 1999, "How Responsive is Business Capital Formation to its User Cost? An Exploration with Micro Data," *Journal of Public Economics*, 74: 53-80.
- Cockburn, Iain and Zvi Griliches, 1988, "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents," *American Economic Review*, 78(2): 419-423.
- Cohen, Wesley M. and Richard C. Levin, 1989, "Empirical Studies of Innovation and Market Structure," in Richard Schmalensee and Robert D. Willig (eds.), *Handbook of Industrial Organization*, Amsterdam: North-Holland.
- \_\_\_\_\_, and Daniel A. Levinthal, 1989, "Innovation and Learning: The Two Faces of R&D," *Economic Journal*, 99(3): 569-96.

- Cooper, Christopher, 1998, "Oil Firms Still Rely on Corporate Spies to be Well Informed," *Wall Street Journal*, December 7: A, p. 1, col. 4.
- Evenson, R. 1968, "The contribution of agricultural research and extension to agricultural production," Unpublished Ph.D. thesis, University of Chicago.
- Feldman, Maryann P. and David B. Audretsch, 1999, "Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition," *European Economic Review*, 43(2): 409-29.
- Glaeser, Edward L., Hedi D. Kallal, Josè A. Scheinkman, and Andrei Shleifer, 1992, "Growth in Cities," *Journal of Political Economy*, 100(6): 1126-1152.
- Grabowski, H. G., 1968, "The Determinants of Industrial Research and Development: a study of the chemical, drug, and petroleum industries," *Journal of Political Economy*, 72:292-306.
- Griliches, Zvi, 1958, "Research Cost and Social Returns: Hybrid corn and related innovations," *Journal of Political Economy*, 66(5): 419-431.
- \_\_\_\_\_, 1979, "Issues in Assessing the Contribution of R&D to Productivity Growth," *Bell Journal of Economics*, 10(1): 92-116.
- \_\_\_\_\_, 1992, "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, 94:s29-s47.
- \_\_\_\_\_, and Jacques Mairesse, 1984, "Productivity and R&D at the Firm Level," in Zvi Griliches, ed., *R&D, Patents and Productivity*, Chicago: Chicago University Press.
- Grossman, Gene and Elhanan Helpman, 1991, *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.
- Hall, Bronwyn and Jacques Mairesse, 1995, "Exploring the Relationship between R&D and Productivity in French Manufacturing Firms," *Journal of Econometrics*, 65: 263-293.
- Henderson, Rebecca and Iain Cockburn, 1996, "Scale, scope, and spillovers: the determinants of research productivity in drug discovery," *The RAND Journal of Economics*, 27(1): 32-59.
- Hoch, I., 1962, "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data," *Econometrica*, 30: 34-53.
- Jacobs, Jane, 1969, *The Economy of Cities*, New York: Random House.

Jaffe, Adam B., 1986, "Technological Opportunity and Spillovers of R&D: Evidence from Firms, Patents, Profits and Market Value," *American Economic Review*, 76: 984-1001.

\_\_\_\_\_, Manuel Trajtenberg, and Rebecca Henderson, 1993, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent citations," *The Quarterly Journal of Economics*, August 1993: 577-598.

Levin, Richard C., Alvin K. Klevorick, Richard R. Nelson, and Sidney G. Winter, 1987, "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3: 783-820.

Mairesse, Jacques, 1978, "New Estimates of Embodied and Disembodied Technical Progress," *Annales de l'INSEE*, 30-31: 681-791.

\_\_\_\_\_, 1993, "Time-Series and Cross-Sectional Estimates on Panel Data: why they are different and why should they be equal?," in G. S. Maddala (ed.) *The Econometrics of Panel Data*, Volume 2. Elger Reference Collection Series. *International Library of Critical Writings in Econometrics*, Volume 1. UK: Aldershot.

Minasian, J., 1962, "The Economics of Research and Development," in Richard R. Nelson (ed.) *The Rate and Direction of Inventive Activity*, NBER Special Conference Series, vol. 13, Princeton, NJ: Princeton University Press.

Marshall, Alfred, 1890, *Principles of Economics*, London: Macmillan.

Mundlak, Yair, 1961, "Empirical Production Function Free of Management Bias," *Journal of Farm Economics*, 43: 44-56.

\_\_\_\_\_, 1978, "On the Pooling of Time-Series and Cross-Section Data," *Econometrica*, 46: 49-85.

Nadiri, M. Ishaq, 1993, "Innovations and Technological Spillovers," *National Bureau of Economic Research Working Paper No. 4423*.

National Science Foundation, 1998, *Research and Development in Industry: 1996*, Arlington, VA: NSF.

Office of Management and Budget, Executive Office of the President, 1987, *Standard Industrial Classification Manual*, Springfield, VA: National Technical Information Service.

Romer, Paul M., 1986, "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 94: 1002-37.

- \_\_\_\_\_, 1990, "Endogenous Technological Change," *Journal of Political Economy*, 98(5,2): S71-101.
- Salinger, M. and L. Summers, 1983, "Tax Reform and Corporate Investment," in Martin Feldstein (ed.), *Behavioral Simulation Methods in Tax Policy Analysis*, Chicago: University of Chicago Press.
- Saxenian, AnnaLee, 1994, *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, Massachusetts: Harvard University Press.
- Schankerman, Mark, 1981, "The Effects of Double-Counting and Expensing on the Measured Returns to R&D," *Review of Economics and Statistics*, 63(3) 454-58.
- Scherer, F., 1982, "Inter-industry Technology Flows and Productivity Growth," *Review of Economics and Statistics*, 64: 627-634.
- Schumpeter, J., 1942, *Capitalism, Socialism, and Democracy*, New York: Harper.
- Spence, A. Michael, 1984, "Cost Reduction, Competition, and Industry Performance," *Econometrica*, 52:101-121.
- Standard & Poor, 1997, *COMPUSTAT User's Guide*, Englewood, CO: McGraw-Hill.
- Terleckyj, N., 1958, "Factors Underlying Productivity: some empirical observations," *Journal of the American Statistical Association*, 53:593.
- Wagner, L. U., 1968, "Problems in estimating research and development investments and stock," In *1968 Proceedings of the Business and Economic Statistics Section*, Washington, DC: ASA, 189-97.

## Appendix A

Derivation of second order Taylor series expansion of equation (3) at  $\rho=0$

Recalling equation (3)

$$Y_i = Ae^{\lambda t} C_i^\alpha L_i^\beta (rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij} K_j^{-\rho})^{-\frac{1}{\rho}} e^{\varepsilon_i} \quad (3)$$

we begin by taking logs and rearrange to write

$$y_i - a - \lambda t - \alpha c_i - \beta l_i - \varepsilon_i = -\frac{1}{\rho} \ln(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij} K_j^{-\rho}) \quad (A.1)$$

where lower case variables represent natural logarithms of their upper case counterparts. The problem is that the right hand side (RHS) of (A.1) is undefined at  $\rho=0$ . If, however, this expression exists in the limit as  $\rho \rightarrow 0$  then we can take the approximation. Assuming

$r + \sum_{j \neq i}^I s_{ij} = 1$ ,<sup>50</sup>  $\lim_{\rho \rightarrow 0} (RHS)$  tends to  $\frac{0}{0}$ . Applying L'Hopital's rule,

$$\lim_{\rho \rightarrow 0} \frac{-\ln(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij} K_j^{-\rho})}{\rho} \text{ converges to } \lim_{\rho \rightarrow 0} \frac{(-rK_i^{-\rho} k_i - \sum_{j \neq i}^I s_{ij} K_j^{-\rho} k_j)}{(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij} K_j^{-\rho})} \text{ or } rk_i + \sum_{j \neq i}^I s_{ij} k_j.$$

This tells us that the RHS of (A.1) is defined at  $\rho=0$  so we can approximate the function at this point.

It was observed through experience that the Taylor series expansion is greatly simplified if we separate the RHS into two functions,  $f = -\frac{1}{\rho}$  and

$g = \ln(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij} K_j^{-\rho})$ . Of course,  $f$  is not well defined at  $\rho=0$  but  $g$  is so we will approximate it.

---

<sup>50</sup> This condition can be achieved with a simple renormalization. Suppose we begin in (A.1) with

$$\hat{r} + \sum_{j \neq i}^I \hat{s}_{ij} \neq 1. \text{ But then } \frac{\hat{r}}{\hat{r} + \sum_{j \neq i}^I \hat{s}_{ij}} + \frac{\sum_{j \neq i}^I \hat{s}_{ij}}{\hat{r} + \sum_{j \neq i}^I \hat{s}_{ij}} = 1 \text{ and can define } r = \frac{\hat{r}}{\hat{r} + \sum_{j \neq i}^I \hat{s}_{ij}} \text{ and } s_{ij} = \frac{\hat{s}_{ij}}{\hat{r} + \sum_{j \neq i}^I \hat{s}_{ij}} \forall j \neq i$$

and add  $\frac{1}{\rho} \ln(\hat{r} + \sum_{j \neq i}^I \hat{s}_{ij})$  to the left hand side of (A.1).

Taking derivatives,  $g' = \frac{(-rK_i^{-\rho}k_i - \sum_{j \neq i}^I s_{ij}K_j^{-\rho}k_j)}{(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij}K_j^{-\rho})}$  and

$$g'' = \frac{(rK_i^{-\rho}k_i^2 + \sum_{j \neq i}^I s_{ij}K_j^{-\rho}k_j^2)(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij}K_j^{-\rho}) - (-rK_i^{-\rho}k_i - \sum_{j \neq i}^I s_{ij}K_j^{-\rho}k_j)^2}{(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij}K_j^{-\rho})^2}$$

$$= \frac{\sum_{j \neq i}^I rs_{ij}(K_iK_j)^{-\rho}(k_i - k_j)^2 + \sum_{h \neq i}^I \sum_{j \neq i}^I s_{ih}s_{ij}(K_hK_j)^{-\rho}(k_h - k_j)^2}{(rK_i^{-\rho} + \sum_{j \neq i}^I s_{ij}K_j^{-\rho})^2}$$

Now, we can write the second order Taylor series expansion of  $g$  at  $\rho=0$  as

$$\ln(rK_i^{-\rho_0} + \sum_{j \neq i}^I s_{ij}K_j^{-\rho_0}) \Big|_{\rho_0=0} + \frac{(-rK_i^{-\rho_0}k_i - \sum_{j \neq i}^I s_{ij}K_j^{-\rho_0}k_j)}{(rK_i^{-\rho_0} + \sum_{j \neq i}^I s_{ij}K_j^{-\rho_0})} (\rho - \rho_0) \Big|_{\rho_0=0}$$

$$+ \frac{\sum_{j \neq i}^I rs_{ij}(K_iK_j)^{-\rho_0}(k_i - k_j)^2 + \sum_{h \neq i}^I \sum_{j \neq i}^I s_{ih}s_{ij}(K_hK_j)^{-\rho_0}(k_h - k_j)^2}{(rK_i^{-\rho_0} + \sum_{j \neq i}^I s_{ij}K_j^{-\rho_0})^2} \frac{(\rho - \rho_0)^2}{2} \Big|_{\rho_0=0}$$

$$= \ln(1) + (-rk_i - \sum_{j \neq i}^I s_{ij}k_j)\rho + [\sum_{j \neq i}^I rs_{ij}(k_i - k_j)^2 + \sum_{h \neq i}^I \sum_{j \neq i}^I s_{ih}s_{ij}(k_h - k_j)^2] \frac{\rho^2}{2}$$

Multiplying through by  $f = -\frac{1}{\rho}$  and substituting back in for the RHS of (A.1) leaves

$$y_i - a - \lambda t - \alpha c_i - \beta l_i - \varepsilon_i = rk_i + \sum_{j \neq i}^I s_{ij}k_j - \frac{\rho}{2} \sum_{j \neq i}^I rs_{ij}(k_i - k_j)^2 - \frac{\rho}{2} \sum_{h \neq i}^I \sum_{j \neq i}^I s_{ih}s_{ij}(k_h - k_j)^2 \quad (\text{A.2})$$

Appendix B

<u>Table B.1</u>	
Detailed Listing of Four-Digit SIC's of Interest	
<u>SIC</u>	<u>Name</u>
3520	Farm & garden machinery and equipment
3523	Farm machinery and equipment
3524	Lawn, garden tractors, equipment
3530	Construction, mining, material handling equipment
3531	Construction machinery and equipment
3532	Mining machinery, equipment, excluding oil field
3533	Oil & gas field machinery, equipment
3537	Industrial trucks, tractors, trailers
3540	Metalworking machinery and equipment
3541	Machine tools, metal cutting
3550	Special industry machinery
3555	Printing trades machinery, equipment
3559	Special industry machinery, nec.
3560	General industrial machinery and equipment
3561	Pumps and pumping equipment
3562	Ball and roller bearings
3564	Industrial commercial fans, blowers, other equipment
3567	Industrial process furnaces, ovens
3569	General industrial machinery and equipment, nec.
3570	Computer & office equipment
3571	Electronic computers
3572	Computer storage devices
3575	Computer terminals
3576	Computer communications equipment
3577	Computer peripheral equipment, nec.
3578	Calculators, accounting machinery, excluding computers
3579	Office machines, nec.
3580	Refrigeration & service industrial machinery
3585	Air-conditioning, heating, refrigeration equipment
3590	Miscellaneous industrial and commercial machinery and equipment

## Appendix C

### Note on Deflator Construction

Price indices for deflating nominal financial variables were obtained from the Bureau of Labor Statistics (BLS) website (<http://www.bls.gov>). Value added for each three digit industry is deflated with the producer price index (PPI) specific to that industry. Capital is deflated using a capital expense index. As a majority of R&D expense is for salaries of technical professionals, R&D stock is deflated using the occupational cost index (OCI) for technical professionals. A capital expense index was available for my entire time period of interest. The remainder of this note documents the construction of the PPI deflator and the OCI for technical professionals.

In constructing industry specific PPI's, my objective was to use as industry specific data as possible in order to obtain the most representative data for each industry panel and to preserve as much variability within and across panels as possible. Few would argue with the claim that there is likely considerable variation in deflators across two digit standard industrial classifications. Similarly, one might argue that within a particular two digit industry there is considerable variation across deflators of three digit subdivisions. For instance, in SIC 35 (Industrial, commercial machinery, computer equipment) we would expect different price indices for output from SIC 357 (Computer & office equipment) and SIC 358 (Refrigeration & service industrial machinery). However, it becomes questionable that the same holds generally within three digit industries. So, given the considerable effort it would take to incorporate heterogeneity in sales deflators at the four digit level, I do not do so here.

Discussions with BLS staff reveal that SIC specific producer price indices (PPI's) were not generated prior to the late 1970's through early 1980's. In the case of SIC 35, PPI's go back to 1985. For my data set, I need values back to 1973.

To construct such indices, I visually examined the value of the wholesale price index (WPI) in log-linear form over my time period of interest (1973-1992). There is a noticeable trend break point at 1981, with the growth rate larger in the 1973-1981 time period. The objective here, then, is to construct price indices for each three digit SIC industry over the unknown time period (1973-1984) which are representative of what we know about those industry output prices vis a vis aggregate wholesale prices.

To do so I regress  $\ln(\text{WPI})$  and  $\ln(\text{price index})$  against time from 81 to 92 and 85 to 92, respectively. This gives me a measure of the trend growth rate after the break point. For each three digit industry I set

$$PPI_{t-1} = \frac{PPI_t}{\exp(g_{t-1 \rightarrow t}^{WPI} - (g_{81 \rightarrow 92}^{WPI} - g_{85 \rightarrow 92}^{PPI}))}$$

where  $g_{t-1 \rightarrow t}^{WPI}$  is the year to year growth rate in the WPI.

The OCI for technical professionals is only available back through 1979. There was no obvious break point in this price index (as was the case with WPI) so I simply straight line extrapolated this index in semi-log scale back to 1973.