NBER WORKING PAPER SERIES

BOUNDING THE EFFECTS OF R&D: AN INVESTIGATION USING MATCHED ESTABLISHMENT-FIRM DATA

James D. Adams
Adam B. Jaffe

Working Paper 5544

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 1996

We thank Mark Doms and Su Peck Troske of the Center for Economic Studies, Bureau of the Census, for advice concerning the data on plants and firms used in this paper. Manjula Singh helped match the Compustat, LRD, and auxiliary establishment files discussed below. The comments of two referees significantly improved the paper. We benefitted from suggestions by Lawrence Kenny, William Long, David Sappington and seminar participants at the Bureau of the Census, Florida, Maryland, and the National Bureau of Economic Research. The paper also benefitted from presentations at the Fourth International Symposium on the Management of Technology, Sainte Adele, Quebec, October 1993, and Meetings of the American Economic Association, Boston, January 1994. This paper is part of NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper 5544
April 1996

BOUNDING THE EFFECTS OF R&D: AN INVESTIGATION USING MATCHED ESTABLISHMENT-FIRM DATA

ABSTRACT

Studies of firm-level data have shown that a firm's R&D and the R&D of other firms increase conventional factor productivity. We investigate these phenomena further by examining the relationship between plant-level productivity and firm-level R&D. We find that (1) the productivity-enhancing effects of parent firm R&D are diminished by geographic distance from the research lab and "technological" distance between the product-field focus of the R&D and the plants; (2) productivity appears to depend on the intensity of parent firm R&D (R&D per plant), not on the total amount; and (3) spillovers of research effects from technologically related firms are significant but also depend on R&D intensity rather than total industry R&D. These results suggest that, despite the externalities created by spillovers of R&D, the "dilution" of R&D across multiple target plants reduces its potency sufficiently that spillovers may not be a source of industry-wide or economy-wide increasing returns.

James D. Adams
Department of Economics
College of Business Administration
University of Florida
Gainesville, FL 32611-7140
and NBER

Adam B. Jaffe
Department of Economics
Brandeis University
Waltham, MA 02254
and NBER

1. Introduction

It is now well-understood that the non-rival nature of knowledge and information is at the heart of the economics of R&D, technological change, and productivity growth. Numerous studies have shown that "spillovers" of knowledge across firms have important implications for industrial organization (Spence, 1984; Levin and Reiss, 1984 and 1988) and can, under certain circumstances, generate equilibrium growth paths for the economy as a whole in which income per capita grows forever (Romer, 1986 and 1990; Lucas, 1988). Similarly, the ability to "spread" a given amount of R&D over any number of productive units can lead to increasing returns to R&D *within* individual firms (Cohen and Klepper, 1993).

Existing analyses treat spillovers across firms and increasing returns to R&D within firms as distinct phenomena. This has not always been the case. Indeed, Alfred Marshall, who is often credited with being first to write about "knowledge spillovers," viewed such spillovers as allowing small firms to achieve economies associated with large scale operations:

> Many of those economies in the use of specialized skill and machinery which are commonly regarded as within the reach of very large establishments, do not depend on the size of individual factories. Some depend on the aggregate volume of production of the kind in the neighborhood; while others again, especially those connected with the growth of knowledge and the progress of the industrial arts, depend chiefly on the aggregate volume of production in the whole civilized world[1].

In this paper we study the anatomy of the knowledge transfers briefly noted by Marshall. We examine both transfers of knowledge across facilities within a firm, and spillovers across firms. In both cases, the extent of increasing returns is determined by the extent to which the inherent non-rival nature of information itself is tempered by other considerations. First, the extent of increasing returns will be affected by the breadth of technological relevance of knowledge. That is, a given "bit" of information will be extremely useful for some purposes and less useful for others. Whether we look within or across firms, the degree to which knowledge is nonrival will be affected by the extent to which knowledge developed in a specific circumstance is useful in other circumstances. In the spillover literature this has been addressed by recognizing that the magnitude of spillovers between two firms is likely to be a function of the "technological distance" (Jaffe, 1986) between them. In the literature on organizations, this issue is couched as the extent of "know-how complementarities" (Helfat, 1995) among distinct business units within a firm.

Second, for a given bit of knowledge to be widely used it must be effectively transferred across institutional, cultural, and geographic boundaries[2]. The cost of transfers works against increasing returns, so that the extent of increasing returns is limited by the magnitude of the costs. Looked at in this way the boundary of the firm is just one of several sources of transactions costs that may limit increasing returns. It is an empirical question, for example, if costs of learning about research results from another plant are typically higher when the other plant is in another state but owned by the same firm, or next door but owned by a different firm[3].

We focus on manufacturing establishments, and examine the extent to which their productivity is affected by R&D performed in formal research labs[4]. To begin to get a handle on the multiple factors mentioned above, we distinguish the effects of R&D by whether or not it is performed by the firm owning the manufacturing establishment; by the geographic distance between the R&D facility and the manufacturing establishment; and by the extent of match between the "product field" in which the R&D is performed and the product mix of the establishment. To examine these questions we utilize data from several different sources. At the heart of the data set is a panel of manufacturing establishments over time from the Census and Survey of Manufactures (the Longitudinal Research Data or "LRD"), matched by firm and industry to the firm-level R&D survey conducted by the Census for the NSF (the "NSF R&D data"). Because of the laboriousness of this matching process we limit ourselves to plants and firms within the chemicals industry (SIC 28).

There is a growing empirical literature investigating productivity and industrial organization issues using the plant-level data from the LRD. But as yet there has been little work that compares plant-level results from the LRD with firm-level results for the same corporations. We undertake such a comparison and find important differences at the plant level, described below, that ought to be taken into account by anyone using the plant-level data.

The paper is organized as follows. Section 2 sets out an econometric framework for measuring the effects of firm boundaries, geographic distance and technological distance on the effectiveness of transfer of R&D results. Section 3 describes in detail the data on firms and establishments and discusses a number of measurement issues. Section 4 presents the results, and Section 5 provides concluding observations.

2. Modelling framework

We postulate that a plant or manufacturing establishment *i* has an "effective stock of knowledge" K_{it} at time *t*. In general this knowledge may be the result of learning by doing at this and other plants, of informal "research" activities performed at the plant, of formal research of the plant's parent firm, performed at many locations, and of formal research of other firms. By necessity we ignore the unobserved sources of K_{it}: learning by doing and informal research, and focus on formal research of the firm and other firms. We examine the extent to which the impact of R&D on the plant's productivity is affected by the geographic and technological distances involved, by the number of other plants that are sharing the same R&D resources, and by the ownership of the R&D facility as compared to that of the plant.

We model the effect of the stock K_{it} in a total factor productivity framework, assuming a Cobb-Douglas production function for the output of plant *i*:

$$Q_{it} = K_{it} L_{it}^{\alpha_{Li}} C_{it}^{\alpha_{Ci}} M_{it}^{\alpha_{Mi}} \exp(\epsilon_{it}), \qquad (1)$$

where Q_{it} is the output of plant i in year t; L_{it} is labor input, C_{it} is conventional capital inputs, M_{it} is material inputs, and ϵ_{it} is everything else that affects output.[5] Rather than try to estimate (1) directly we increase our ability to identify the effects of knowledge by using factor shares as estimates of the output elasticities α_{Li}, α_{Ci}, and α_{Mi}. That is, we calculate the level of conventional factor productivity:

$$TFP_{it} = \frac{Q_{it}}{L_{it}^{\alpha_{Li}} C_{it}^{\alpha_{Ci}} M_{it}^{\alpha_{Mi}}}. \qquad \textbf{(2)}$$

from the input and output data and the factor shares, thereby allowing for plant level differences in the intensity of factor use. Substituting (1) into (2) suggests that the effect of knowledge on output can then be estimated from a regression of the level of TFP on the effective knowledge stock. Note that this approach assumes constant returns to scale at the plant level in the conventional inputs L, C, and M. This assumption is consistent with our data.

An alternative approach would be to estimate TFP_{it} by regressing the log of Q_{it} on logs of L_{it}, C_{it}, and M_{it}, at the industry level. This would provide industry-wide estimates of the α_j output elasticities in (2). We did not pursue this approach because it concealed genuine differences in the cost structures of our plants. To put the same point differently, use of industry regression estimates forces the output elasticities to be the same over heterogeneous plants and biases the elasticities towards zero.

There were indications that the calculated cost shares or elasticities were close to the true values measured free of errors. For example, our results remained the same when we used calculated cost shares averaged over two years instead of one year, suggesting that errors in the shares were unimportant[6].

Ideally we would construct a proxy for the effective knowledge stock that simultaneously incorporated all of the geographic and product-specific effects of interest. Unfortunately, there are inherent limitations of the R&D data to identify the effects of distance along both geographic and technological dimensions. To estimate both effects one would need data revealing the *joint* distribution of research activity along the two dimensions. Instead we observe just the *marginal* distributions[7]. That is, we know how much of the firm's research is in different states, and how much is in different fields, but we do not know how much in each field is done in each state. This limitation prevents us from using a model that simultaneously captures all effects. We are limited to a series of partial analyses. Each of these analyses takes the general form:

$$K_{it} = \left[\frac{(R_{it}^{c} + \delta R_{it}^{d})^{\beta_f}}{(n^{c}_{it})^{\gamma_c}(n^{d}_{it})^{\gamma_d}}\right]\left[\frac{S_{it}^{\beta_s}}{N_{it}^{\gamma_s}}\right] \tag{3}$$

where R^c_{it} connotes research of i's parent firm that is "close" to plant i, and R^d_{it} connotes research that is "distant," in the sense of either geography or technology. The variables n^c_{it} and n^d_{it} are the total number of plants (including i) that are in the "close" and "far" groups, however defined. S_{it} denotes the potential spillover; it is measured as the research of firms other than i's parent weighted by technological proximity or relevance; N_{it} is likewise the total number of plants in the rest of industry[8]. Notice that the number of plants variables (n^c_{it}, n^d_{it}, and N_{it}) partially serve the function of scaling the different forms of R&D in common terms of R&D per plant, since the effect of total R&D in each form holds the number of plants constant in that form. In addition they pick up any omitted firm or industry characteristics that are associated with numbers of plants.

A number of important assumptions are embedded in this functional form. First, while we treat "close" and "far" knowledge from the parent firm as perfect substitutes (albeit with different productivities), we treat knowledge from the parent firm and other firms as complements. This reflects the view that absorbing spillovers from other firms requires doing research yourself (Jaffe, 1986; Cohen and Levinthal, 1989). Second, we treat both technological and geographic distance as binary rather than continuous variables. This is partly an accommodation to the data, which probably would not support estimation of continuously declining effects with distance. More fundamentally, it is not obvious that the effect of distance is continuous. Our approach reflects the notion that if knowledge sources are nearby, then mechanisms of informal communication that operate among people in the same area can operate; beyond a certain distance, these mechanisms cannot operate and knowledge flows only by more formal means such as publication. We assume that once you are at a distance where informal communication is not available, it does not matter greatly how far away you are. By analogy, we are saying that people at Harvard and M.I.T. communicate more with each other than either do with people at Stanford, but there is not a big difference between the extent of their communication with Stanford and University of Chicago.

Third, the production function that we have specified, while standard in the literature, may leave out important inputs to the plant-level production process. In particular the manufacturing plants of multiplant firms receive management, marketing and other assistance from central office and auxiliary establishments. Conversely, the output of individual plants may be valued at intra-firm transfer prices that implicitly assign part of the value created at an establishment to other establishments within the firm. Thus measured TFP at the plant level may be higher or lower than true TFP, and these biases may be related to the multiplant nature of a firm's operation. Since we find strong empirical effects associated with multiplant operation this partial view of the firm is troubling, but we are limited to the available data, which are for manufacturing plants only.

Fourth, we are assuming that all R&D either reduces the inputs necessary to produce a given amount of output, as in process R&D, or increases the quality and hence the price that the firm charges, leading to increased value of output, as in product R&D. Realistically, increases in the quality of output may not be reflected in higher prices or the construction of price deflators may remove price-related quality increases. For this reason the effect of product R&D on measured TFP is probably understated, and so is the effect of total R&D[9]. The difficulty is compounded by our use of plant-level output measures since plant-level prices reported for the establishments of multi-plant firms may be internal transfer prices that do not correspond to market values. Again there is little that can be done about this problem of a limited view of the overall firm with available data, but it is something to keep in mind in interpreting the results.

Finally, by "normalizing" the knowledge stock in (2) by the number of plants, we allow for the possibility that the transactions costs associated with transferring knowledge may increase with the number of locations across which that knowledge is being shared. Of course these could be zero, suggesting strong increasing returns, at least as long as technological and geographic distances are small. Our original conception was that the magnitude of the parameters would fall between zero and the magnitude of the corresponding parameters. Such a result would suggest that increasing returns were being tempered by knowledge transfer costs. To our surprise, these are often larger than those in plant-level regressions, raising an issue of decreasing returns that we will discuss further below.

For either concept of distance, we obtain an estimable equation by substituting (1) and (3) into (2) and taking logs:

$$\begin{aligned} \ell n(TFP_{it}) = \beta_f \ell n(R_{it}^c + \delta R_{it}^d) + \beta_s \ell n(S_{it}) - \gamma_c \ell n(n_{it}^c) \\ - \gamma_d \ell n(n_{it}^d) - \gamma_s \ell n(N_{it}) + \sum_k \phi_k Z_{it}^k + \mu_{it} \end{aligned} \qquad (4)$$

where the Z^k are additional variables that explain productivity such as time dummies and age effects, and μ_{it} is the residual unexplained effect[10].

3. Description of the data

We choose to study chemicals (SIC 28) as opposed to another industry because production data for this industry tend to be of good quality and there are clear distinctions between technologies in the industry subgroups. These support the construction of meaningful spillover pools, almost by necessity constructed along the lines of the NSF applied product fields. Our data cover the period 1974-1988.

The basic data combine six separate sources: (1) plant level production data from the Annual

Survey of Manufactures and the manufacturing Census, known as the Longitudinal Research data base; (2), firm level data from the R&D survey conducted for NSF by Census; (3), the NBER 4 digit manufacturing data constructed by Wayne Gray, which include deflators for gross investment, value of shipments, and materials; (4), the Bureau of Economic Analysis 2 digit deflators and depreciation rates for capital stocks of equipment and structures; (5), the BLS 2 digit rental rates per constant dollar of equipment and structures; and (6), the Census Picadad file for the calculation of distances between all possible points of latitude and longitude. In addition to these data sources used in the establishment-level regressions, we also merged the data at the firm level with the Standard and Poors Compustat data based on annual reports and 10-K filings. This allows us to compare some of the results that we get in the plant-level data with analogous regressions for the firm as a whole.

Before exclusions the file consists of 1150 chemical firm-years and 21,546 plant-years. Since the sample period is 1974-1988, these statistics translate into roughly 80 chemical firms per year and 1400 chemical plants per year. The mean number of plants per firm is 18, more before 1979 and less afterwards, due to increased selectivity in the Annual Survey of Manufactures (ASM) at this time.

In constructing the data set we attempted to match every observation in the LRD and R&D data that met our criteria for data quality.[11] In the case of the R&D we required that data almost always exist on research expenditures by state and applied product field. Where it did exist we required that it be real and not imputed, and that the state and applied product field components approximately add to totals. In the few cases where the data failed to exist we required that good data exist in adjacent survey years so that we could interpolate[12].

Referring to (2), TFP entails the deflation of nominal values of materials, labor, and output to obtain real values. Also it requires deflation of gross investment in equipment and structures and the construction of real stocks for each form of capital. Finally it requires the construction of factor cost shares.

In terms of the LRD production data, materials input is defined as current expenditure minus the change in materials inventory. Gross investments are defined as expenditures on new equipment and structures. Output in the LRD is the value of shipments plus the increase in work-in-progress and final goods inventories.

Real labor input is simply total employment. Real materials input, gross investment, and output are obtained by dividing nominal values by the NBER 4 digit deflators indexed to 1987.

In order to construct real capital stock we followed the methodology of Lichtenberg (1992). In the initial year for the time series for any plant we deflated gross book values of equipment and structures separately using 2 digit deflators for each type of capital from the Bureau of Economic Analysis[13]. Deflators were given by the ratio of industry net capital stock in 1987 dollars to industry gross capital in historical dollars. Initial real capital stock therefore is

$$C_{ijt} = GBV_{ijt} \, x \, \frac{NCC_{jt}}{GHC_{jt}} \qquad (5)$$

where C_{ijt} is real capital stock of plant i in industry j, GBV_{ijt} is gross book value in historical dollars of the plant, NCC_{jt} is net capital stock of the industry in constant 1987 dollars, and GHC_{ijt} is gross capital stock of the industry in historical dollars. For succeeding years in the time series of each plant we applied the perpetual inventory formula for equipment and structures separately,

$$C_{ijt} = C_{ijt-1}(1-\delta_{jt}) + I_{ijt}, \qquad (6)$$

where C_{ijt-1} is real capital stock from year t-1, δ_{jt} is the BEA depreciation rate by 2 digit industry and each form of capital, and I_{ijt} is gross investment in the plant in constant 1987 dollars. Bailey, Campbell, and Hulten (1992) compare this method of deflation with a more elaborate method. The more detailed method followed each plant from its first appearance in the LRD, and deflated the entire investment stream using the NBER 4 digit deflators, and found that the more careful method of calculation made very little difference in results, largely because of the small share of capital in cost which minimizes the impact of errors in the calculation of capital stock.

The Census-NSF R&D survey reports the flow of R&D expenditure in a given year. An accumulated stock of R&D over time would be a theoretically superior determinant of plant productivity. As a practical matter, however, the stock and flow approaches to R&D will differ in their estimated effects only to the extent that firms vary their real R&D substantially over time. In general, such variation is relatively small, making estimation based on flows econometrically similar to estimation based on stocks. Still, we explore a version of a stock model in which the R&D variable is a partial accumulation of past R&D:

$$RDK_t = \sum_{l=0}^{5} (1-\delta)^l RD_{t-l} \qquad (7)$$

where the depreciation rate is taken to be 15 percent per year (Griliches and Lichtenberg, 1984).

In order to explore spillovers from outside the firm, we included S_{it}, the relevance-weighted R&D of other firms, in Equation (3) determining the plant's stock of knowledge. To construct this variable, we adapted the "technological proximity" approach of Jaffe (1986):

$$S_{it} = \sum_{j \neq i} P_{ij} R_{jt} \qquad (8)$$

where P_{ij} is the proximity between firm i and firm j in research space, calculated as the uncentered correlation between the firms' R&D distribution vectors over the 32 NSF product fields[14]. Hence P_{ij} is unity for two firms that have exactly the same research profile, is zero for two firms who have no research areas in common, and between zero and unity for all other firms.

Since we follow a computational approach to TFP, (2) requires estimates of factor cost shares in order to compute estimates of the α_{Zi} elasticities. We begin with expenditures. Labor expenditures equal wages of production and non-production workers plus supplementary labor costs. Materials expenditures are expenditures on materials and purchased services net of growth in materials inventories. We followed a different procedure for the estimation of capital expenditures. Reported capital spending moves erratically due to lumpiness of investments and nonreporting of the implicit value of rentals on the firm's capital stock. We multiply real capital stock by the Bureau of Labor Statistics' two digit industry rental rates per dollar of capital to obtain an estimate of spending on capital. We perform this procedure separately by equipment and structures and sum the results to obtain capital expenditures. Each of the three expenditures: labor, materials, and capital, are divided by all the expenditures to obtain estimated cost shares. The bulk of costs at the plant level is on materials, with labor second and capital last. While some might object that this procedure imposes constant returns on the data, the alternative regression procedure, which does not impose this restriction, finds the sum of the elasticities close to one[15].

Descriptive statistics are reported in Tables 1 and 2. Table 1 reports the industrial distribution of the plants[16]. About two thirds are in chemicals, petroleum, and rubber. Most of the remainder are clustered in the other high technology industries-- machinery, electrical equipment, and instruments-- with the exception of a final cluster, found in food processing. For most of our analyses we replicate the results using both the complete sample and a sample limited to chemical industry plants. All plant-level regressions include dummies for the industry group of the plant[17]. We also undertake analyses for finer industry subsectors at the end of the paper.

Table 2 reports means and variances by industry group for total factor productivity of the plant, R&D of the parent firm in the same applied product field as the plant's industry, and R&D of the rest of the chemicals industry in the same applied product field as the plant's industry.

The calculations reveal the immense range of plant TFP. These calculations are performed before the exclusion of most outliers. The only restrictions are that output and inputs be positive and not missing, and that expenditures on inputs divided by sales not exceed 10.0. The large standard deviations of TFP suggest the importance of industry differences, births, deaths, and plant idling, and measurement error. Clearly differences in TFP reflect a good deal else than technology.

The statistics on parent firm applied product field R&D listed in column 2 are as expected. They are quite large in the core chemical fields, especially pharmaceuticals, and in some of the affiliated industries. We also see similar concentrations of industry R&D by applied product field, though industry R&D is of course much larger. Table 2 makes it clear that between-industry correlations of productivity and R&D are unlikely to be very high given that productivity is driven by many other factors besides technology.

4. Findings

Table 3 presents the results of our simplest estimation, in which we ignore spillovers from other firms and the effects of geographic and technological distance. We simply look at the effect of firm level R&D on plant productivity, controlling for the number of plants over which the firm's total R&D must be "spread." We also include total employees of the parent firm, and the relative age (plant age in years minus average age of the firm's plants), dummies for year, sub-industries, regions, new plants and plants with large output reductions. We perform the estimation for all plants owned by the chemical firms, and for a subset limited to chemical establishments.

The results are broadly similar whether we look at all plants or the chemical industry subset. The life-cycle effects are quite important, with measured productivity being dramatically lower in both new plants and those that are cutting back. There is also a significant effect of relative plant age, with each year of age relative to the firm mean being associated with about 1% higher productivity. This could be due to "learning by doing," and it runs counter to any bias that might be present as a result of newer plants having unmeasured advantage in the quality of the capital stock. Regional effects also matter, with productivity highest in the North and lowest in the South.

Turning to R&D, the most striking finding is that R&D does not have a measurable impact on productivity unless we control for number of plants. Once we control for the number of plants (eq. 3.2 and 3.6 in the flow version, and 3.3 and 3.7 in the stock version), we obtain estimates of the elasticity of productivity of R&D in the range of .06 to .08, which are only slightly lower than the results from firm-level data (Lichtenberg, 1992; Griliches and Mairesse, 1984). The number of plants is itself extremely significant, and larger in magnitude than the R&D coefficient; this difference is statistically significant[18]. This says that the parameter of equation (3) is actually greater than the parameter; R&D is so rapidly diluted by spreading R&D over multiple plants that R&D must be increased *faster* than proportional to the number of plants in order to maintain its effectiveness at each plant.

This is a perplexing result, implying if it were true that these firms would be better off breaking themselves into pieces, something that they have clearly resisted doing. And yet it is a robust finding in our data. Note that the regressions imply that the number-of-plants effect cannot be interpreted as a "firm size" effect, as the equation also includes the total number of firm employees, which has a very small positive effect except when number of plants is excluded.[19] Nevertheless, equation 3.4, which constrains the specification to the log of R&D per plant, fits the data nearly as well, and the coefficient of R&D per plant is similar to the R&D coefficient in specifications that introduce the log of R&D and the log of number of plants separately.

Most previous studies of the effect of R&D on productivity have used firm-level data, and hence did not explore any effect due to number of plants. Particularly given the surprising magnitude of the effect, we would like to know the extent to which it is real, and the extent to which it might be an artifact of difficulties in measuring inputs and outputs at the plant level. If, for example, multi-plant firms tend to use transfer prices that assign more of value-added to central-office establishments, then they would appear to have lower productivity at the plant level, but the firm as a whole need not show lower productivity. With the data we have we cannot resolve these issues conclusively, but we can narrow the range of explanations. Table 4 simply shows the average TFP for plants belonging to firms with various numbers of plants. This table shows that the simple correlation between productivity and number of plants is very strong, and that it persists, despite an upturn in the second bracket, throughout the range of numbers of plants, rather than being dominated by firms with very few or very many plants.

One piece of information that bears on centralization of the firm and its relation to measured plant productivity in Table 4 is the extent to which R&D is separately conducted in central office and research establishments. In each year of the manufacturing census a survey is conducted of auxiliary establishments, including separate R&D facilities. We matched these R&D auxiliaries to our sample of firms for the census years 1977, 1982, and 1987. We then constructed the fraction of *firm* R&D that is conducted in the separate R&D facilities. Since the auxiliary samples are not full enumerations of all such facilities for our sample of firms, and since the measure of centralization that we have is indirect, the results are suggestive rather than conclusive. They are consistent, however with a transfer pricing explanation for the decline in measured productivity as firm size increases in Table 4.

We find that the share of separately conducted R&D *falls* from 0.46 in firms with 1-10 manufacturing plants to 0.34 in firms with 11-20 plants. The separately conducted share *rises* to 0.59 for firms with 21-40 plants, and then it rises further, to 0.65 for firms with more than 40 plants. These patterns are the mirror image of those in Table 4, where measured productivity rises over the first two ranges of plants, declining thereafter. The pattern is consistent with transfer pricing behavior, in which more of the value of plant output is transferred to central office establishments in more centralized firms -- those with more separately conducted R&D-- thereby lowering measured productivity in the manufacturing facilities of larger firms. Thus it appears that the share of separately conducted R&D captures transfers as well as simple centralization.

As an additional window on this issue, we identified as many of the LRD firms as possible in the Standard and Poors Compustat database of firm-level accounting data. The Compustat or similar data have been the source of the existing literature on firm-level effects of productivity on R&D. By merging the two datasets, we can mimic the typical firm-level regression, but include the number of plants variable as an additional regressor. However, the Compustat data do not contain the information necessary to calculate cost shares, so instead of duplicating our TFP regression we estimated a production function for sales at the firm level as a function of firm-level employment, capital, materials, R&D, and number of plants. Since about one half the LRD chemical firms appear in Compustat, it was necessary to control for changes in sample composition[20]. In order to control for this and for use of the production function rather than the TFP formulation, we also estimated the production function at the plant level for those plants whose parent firms do appear in Compustat. The results are reported in Table 5.

We have already discussed the plant-level data in Table 5. The firm-level data derive from Compustat except for Census-NSF R&D spending and the number of plants, which are the same variables that were used in Table 3. Real output is the log of Compustat deflated value of sales in 1987 dollars. The Bureau of Labor Statistics deflator is for chemicals as a whole since we have no evidence on the distribution of sales by industry subgroup for firms[21]. We also use Compustat data on firm employment, gross book value of capital, R&D, and materials (inclusive of some depreciation expenditures). We express capital in 1987 dollars using the Lichtenberg methodology described above, we use the implicit GDP deflator to express R&D in 1987 dollars, and we deflate materials using the NBER materials deflator indexed to the year 1987. Note that the materials item is not a perfect indicator of current expenditures on materials and services, since it includes expenditures on replacement of capital. But there is little to be done about this, since materials are needed for comparison with the plant data, where they are quite important.

Starting with the plant-level regressions in Columns 1 and 2, the production function formulation exhibits a much smaller R&D elasticity than the analogous TFP regressions in Table 3 (equations 3.2 and 3.3)[22]. It is striking, however, that the *ratio* of the number of plants coefficient to the R&D coefficient in the plant-level production function regressions in Table 5 is very similar to the ratio of those coefficients in the plant-level TFP regressions in Table 3. Thus, the phenomenon of apparent decreasing returns is associated with analyzing the data at the plant level, independent of whether one adopts the TFP or production function approaches.

In the firm level regressions, however, the coefficient on number of plants is essentially *equal* in magnitude to the coefficient on R&D, whether in the flow formulation (columns 5.3-5.5) or the stock formulation (column 5.7). Indeed, the restriction imposed by entering R&D on a per-plant basis as in columns 5.6 and 5.8 is not rejected by the data. Another important difference between the firm and plant regressions is that the coefficient of R&D in the firm regressions is largely independent of whether number of plants is included, as 5.3 and 5.4 show, whereas in the plant regressions R&D fails to matter unless the number of plants is accounted for.

We draw two significant conclusions regarding the number-of-plants effect from Tables 3, 4 and 5. First, the fact that this effect weakens relative to the R&D coefficient in the firm-level regressions suggests that it is likely to be an artifact of output mismeasurement at the plant level. Combined with the strength of the correlation shown in Table 3, this suggests that any productivity-related studies at the plant level need to at least take number of plants into account. Second, while we cannot resolve how much of the effect is spurious and due to data problems in the plant data, its persistence in the firm-level regressions suggests that R&D "dilution" is a very important counterweight to the tendency of R&D to generate increasing returns. Indeed, the *best* we can do for R&D is to get it to the constant-returns-to-scale formulation of R&D per plant. We discuss the implications of this finding further in Section V below.

Table 6 returns to the plant-level TFP approach, and examines the issue of geographic localization. We decompose the firms' R&D into that portion that is "close" to the plant in question and the portion that is "far," and estimate the relative contribution of each using the formulation of Equation

(3). We consider two different conceptions of "close:" being within the same state, and being within 100 miles of a state's R&D center[23]. The results are quite similar to those of Table 3, except that now we find the expected diminution of effectiveness for more distant R&D.[24] R&D performed outside the state or, alternatively, more than 100 miles away, is roughly 10 to 30 percent as effective as R&D performed in the same state. The fact that the discount is generally greater for being beyond 100 miles than it is for being in a different state presumably reflects the fact that, particularly in the Northeast, plants can be within 100 miles but be in different states. The overall R&D elasticity for this composite R&D total is slightly lower than in Table 3, approximately .05 to .07. The "dilution" effect from other plants remains significant, and is generally larger than the R&D elasticity, particularly for plants outside the state.

Table 7 explores the effect of technological rather than geographic distance. In all cases we find that R&D outside the product field is less effective than R&D within the product field, but the magnitude of the "discount" varies greatly across specifications and data samples[25]. In the flow formulation, technologically distant R&D is one-fourth to one-half as effective as same product field R&D. In the stock formulation, the outside-of-product-field stock has essentially no effect on productivity. The overall R&D elasticities fall further from those in Table 3. Technological effects are not estimated as precisely as geographic ones, reflecting greater measurement error in the allocation of firms' R&D across product fields compared with the allocation across states. This measurement error very likely explains the decline in the composite R&D elasticity.

One contrast between Tables 6 and 7 lies in the division of the effect of number of plants. While the overall effects of plant numbers are quite similar between the two tables, in Table 6 plants out of state amortize the effect of R&D more strongly than number of plants in-state, while in Table 7 number of plants in the same applied product more strongly amortizes the effect of R&D than do plant numbers in other applied product areas. While we lack the requisite information to draw a definitive conclusion, these patterns strike us as related to patterns of transfer pricing within the company. The presence of a lot of plants outside the state proxies for the presence of a central office located elsewhere to which value is transferred. The presence of many plants in the same applied product area likewise reflects the existence of a separate facility, perhaps a laboratory that performs R&D targeted on many plants in the same product area, which is then assigned the value of plant output above cost.

Table 8 incorporates spillover effects calculated as in equation (8), that is, using weights based upon similarity of any two firms in the space of applied product R&D. We find that the R&D of other firms does affect a plant's productivity, holding industry constant. For the industry spillovers, geographic proximity does not seem to matter, although this may simply represent the fact that spillovers are weighted by technological closeness of firms, not by their geographic coincidence. In any event, the coefficient on the portion outside 100 miles actually exceeds unity in column 8.3, although this effect is not estimated very precisely and one could not reject its being 1.0 at conventional significance levels. We also find that the elasticity of plant productivity with respect to the relevance-weighted sum of other firms' R&D is actually greater than the elasticity with respect to the parent firm's R&D. Note that industry R&D is a much bigger number, so that the greater elasticity implies that the *marginal product* of industry R&D is approximately 40% as large at the mean of the data as the marginal product of parent firm research[26]. But to get this relatively large spillover effect other firms must increase their R&D *intensity*: There is a "dilution" effect for spillovers as there was for within-firm R&D. Indeed, omitting the number of industry plants from the equation destroys the spillover effect just as omitting the number of firm plants from the equation destroyed the firm R&D effect in Table 3. Overall, the spillover and number-of-plants effects are very similar in magnitude, suggesting that plant level productivity is determined by the average R&D per plant ratio of technologically similar firms.

In view of the discussion of localization in the Tables 6 to 8, we decided to study the patterns of plant location in our sample. Table 9 displays fractions of plants for several industry groups located in ten states that are the most densely populated by plants in our full sample. In addition to the "All

Industries" comparison group, we present five industry groups: food and kindred products; industrial organic and inorganic chemicals; drugs and agricultural chemicals; plastics, soaps, paints, and miscellaneous chemicals; and rubber and plastics. These groups comprise three-fourths of the sample and thus qualify as leading groups. In the "All Industries" column, and indeed in the others, we list the five most populous states in bold face type with their rank below in parentheses[27]. It is clear from this column that plants congregate in four geographic centers: California; the Middle Atlantic region comprised of New York, New Jersey, and Pennsylvania; the East North Central Region centered on Michigan, Illinois, and Ohio; and Texas. To an extent this pattern repeats for the sub-industries but with many differences. We see that the fractions exhibit at least a two to one range, and usually larger, for a given state in any row. In addition the ranks of the top five states change markedly when industry changes. For example, California is ranked first in rubber and plastics but fifth in industrial organic and inorganic chemicals. A fourth of all drug and agricultural chemicals plants are located in the Middle Atlantic, but the share of the Middle Atlantic is only one-eighth of all food and kindred plants. These results suggest patterns of geographic specialization for sub-industries within the four main centers characteristic of the industry as a whole. Drugs and agricultural chemicals are centered on Middle Atlantic, but rubber and plastics are centered on East North Central.

Our final piece of empirical work is to estimate equations like those in Table 8 separately for 3 industry subgroups: Industrial Organic and Inorganic Chemicals; Drugs and Agricultural Chemicals;[28] and Plastics, Soaps and Miscellaneous Chemicals. The results are presented in Table 10. Not surprisingly, many of the effects of interest vary in magnitude across the subgroups. The drugs group exhibits a higher R&D elasticity, less diminution with distance, and a greater spillover elasticity than the other two groups. Geographically distant R&D is almost worthless in the industrial chemicals group. All groups exhibit spillover effects that are close to being pure intensity effects (R&D per plant of other firms), as well as own-plants effects that are larger than own-R&D effects, just as before.

5. Discussion and conclusions

Our analysis reveals several important empirical regularities. Although based on a single industry sector, they suggest new ways of thinking about plant-level production data and about the returns to R&D. First, despite the increasing use of the LRD data, the strong correlation between plant-level productivity and the number of plants of the parent firm has not, to our knowledge, been reported previously.[29] While we have a less than complete understanding of the phenomenon, its magnitude and robustness suggest that this is an important area for future research. Further, the fact that it is attenuated when regressions are run at the firm level suggests that it is at least partly an artifact of plant-level measurement. Thus an element of caution is in order when comparing any plant-level productivity results to results that have been previously obtained in firm-level data.

Turning to R&D, the broad picture painted by our results is one in which spillovers of R&D are important, both within and across firms. We find, however, that these spillovers are significantly diluted as the number of recipients grows, calling into question the theoretical argument that R&D spillovers imply industry-wide or economy-wide increasing returns to scale. Indeed, our overall results fit a constant-returns model in which R&D *intensity* is what matters, rather than the overall level of research activity, either within the firm or within the industry.

We are less successful at pin-pointing exactly the forces that prevent knowledge from being a broadly usable public good. Our own view is that growth almost never occurs by replication of results for identical production processes, so that spreading R&D over more units inevitably means spreading over more disparate units. Our results for the relative value of R&D in other product fields are consistent with this conjecture, but we believe that the product-field breakdown is far too coarse and far too crudely applied to capture much of the technological differences across plants within a firm or industry. To put it another way, a firm that is twice as large is, to first approximation, doing twice as many different things. Hence to achieve the same overall R&D effect, it must do twice as much R&D.

This does not imply that spillovers are unimportant; nor is it inconsistent with a significant gap between the private and social rates of return to R&D. On the margin, a firm that increases its R&D *does* increase R&D intensity, thereby creating benefits for other units of the firm and for other firms. But our results do suggest that endogenous growth models in which labor productivity is modeled as a function of the economy's stock of R&D may have to be re-cast. Such a re-formulation would be consistent with the literature on organizations and hierarchies, which suggests that problems of information associated with increasing complexity, and thus increasing moral hazard problems within organizations, are likely to eventually create diseconomies of scale.[30] It would also be consistent with the fact that the U.S. does not apparently grow much faster than say, the Netherlands despite our much larger stock of R&D.

The results also confirm significant geographic localization of the effects of R&D within firms. We interpret this effect as reflecting increased communication costs with distance.[31] As Table 9 shows, however, more distant plants are also less likely to be in the same product field. Hence it is likely that what appears to be geographic localization is partially a manifestation of technological localization. We do not find evidence of significant geographic localization of spillovers, but these parameters are very imprecisely estimated, and we are unable to focus the spillover on geographic coincidence at the same time that we focus on technological similarity[32].

The last decade has seen steady progress in empirical studies of the returns to R&D, particularly with respect to measurement of spillover effects. This empirical literature has grown alongside an increasing theoretical literature exploring the consequences for growth of the non-rival nature of knowledge. This paper adds to that literature, confirming the empirical significance of spillovers, but adding empirical estimates of important forces that limit the extent to which spillovers appear to be a source of industry-wide increasing returns to scale.

Bibliography

Adams, J.D. and Peck, S. "A Guide to R&D Data at the Center for Economic Studies U.S. Bureau of the Census," Working Paper CES 94-9, Center for Economic Studies, August 1994.

Bailey, M., Campbell, D., and Hulten, C. "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity: Microeconomics* (1992), pp. 187-249.

Calvo, G., and Wellisz, S. "Supervision, Loss of Control, and the Optimum Size of the Firm," *Journal of Political Economy* Vol. 86 (September/October 1978), pp. 943-952.

Cohen, W.M. and Klepper, S. "A Reprise of Size and R&D," Mimeo, Carnegie-Mellon University, 1994.

_______________, and Levinthal, D.A. "Innovation and Learning: the Two Faces of R&D," *Economic Journal*, Vol. 99 (September 1989), pp. 569-597.

Dearden, J., Ickes, B.W., and Samuelson, L. "To Innovate or Not to Innovate: Incentives and Innovation in Heirarchies," *American Economic Review*, Vol. 80 (December 1990), pp. 1105-1124.

Geanokoplos, J. and Milgrom, P. "A Theory of Heirarchies Based on Limited Managerial Attention," *Journal of the Japanese and International Economies*, Vol. 5 (September 1991), pp. 205-225.

Griliches, Zvi. "Issues in Assessing the Contribution of Research and Development to Productivity Growth," *Bell Journal of Economics*, Vol. 10 (Winter 1979), pp. 92-116.

_________ "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, Vol. 94 (Supplement 1991), pp. 29-47.

__________ and Mairesse, J. "Productivity and R&D at the Firm Level," in Zvi Griliches, ed., *R&D, Patents, and Productivity*. Chicago: University of Chicago Press for NBER, 1984.

__________ and Hausman, J. "Errors in Variables and Panel Data," *Journal of Econometrics*, Vol. 31 (February 1986), pp. 93-118.

__________ and Klette, Tor. "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous," NBER Working Paper no. 4026, 1992.

Helfat, Constance E. "Know-how complementarities and Knowledge Transfer Within Firms: The Case of R&D," Mimeo, University of Pennsylvania, March 1994.

Holmstrom, B.R. and Tirole, J. "The Theory of the Firm," Chapter 2 of *Handbook of*

Industrial Organization, Vol. 1, Amsterdam: North-Holland, 1989.

Hsiao, Cheng. *Analysis of Panel Data*, Cambridge, U.K.: Cambridge University Press, 1986.

Jaffe, Adam B. "Technological Opportunity and Spillovers of R&D," *American Economic Review*, Vol. 76 (December 1986), pp. 984-1001.

_____________ "Real Effects of Academic Research," *American Economic Review*, Vol. 79 (December 1989), pp. 957-970.

_____________, Henderson, R., and Trajtenberg, M. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, Vol. 108 (August 1993), pp. 577-598.

Jarmin, Ronald S. "Learning by Doing and Competition in the Early Rayon Industry," *Rand Journal of Economics*, Vol. 25 (Fall 1994), pp. 441-454.

Jovanovic, Boyan. "The Diversification of Production," *Brookings Papers on Economic Activit:, Microeconomics*, 1993, pp.1.

Keren, M. and Levhari, D. "The Internal Organization of the Firm and the Shape of Average Costs," *Bell Journal of Economics*, Vol. 14 (Fall 1983), pp. 474-486.

Klette, Tor. "R&D, Scope Economies, and Company Structure: A 'Not-so-Fixed Effect' Model of Plant Performance," Discussion Paper no. 120, Statistics Norway.

Krugman, Paul. *Geography and Trade*, Cambridge: M.I.T. Press, 1991.

Lichtenberg, Frank. *Corporate Takeovers and Productivity*, Cambridge, Massachusetts, M.I.T. Press, 1992.

Levin, R.C. and Reiss, P. "Tests of a Schumpeterian Model of R&D and Market Structure," in Zvi Griliches, ed., *R&D, Patents, and Productivity*, Chicago: University of Chicago Press for NBER, 1984.

__________________________ "Cost-reducing and demand-creating R&D with spillovers," *Rand Journal of Economics*, Vol. 19 (Winter 1988), pp. 538-556.

Lucas, Robert E., Jr., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, Vol. 22 (July 1988), pp. 3-42.

McAfee, R.P. and McMillan, J. "Organizational Diseconomies of Scale," Manuscript, May 1992.

Marshall, Alfred. *Principles of Economics*, 8th ed., London: Macmillan, 1920.

Romer, Paul M., "Increasing Returns and Long Run Growth," *Journal of Political Economy*, Vol. 94

(October 1986), pp. 1002-1037.

_____________ "Endogenous Technological Change," *Journal of Political Economy*, Vol. 98 (Supplement, October 1990), pp. S71-S102.

Spence, A. Michael. "Cost Reduction, Competition, and Industry Performance," *Econometrica*, Vol. 52 (1984), pp. 101-121.

Williamson, Oliver E. "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy*, Vol. 75 (April 1967), pp. 123-138.

Endnotes

1. Alfred Marshall, *Principles of Economics*, MacMillan (1920), Book IV, Chapter VIII, pp. 220.

2. The last line of the passage from Marshall suggests that he sometimes thought geography unimportant for the transfer of "knowledge and the progress of the industrial arts." For a contrary view see Krugman (1991); for evidence on geographic localization of knowledge spillovers see Jaffe, Henderson and Trajtenberg (1993).

3. There is a big difference between the firm boundary and geographic or other boundaries from a *strategic* point of view. The firm presumably tries to minimize the costs of internal transfer and maximize the costs of external transfer. In this paper we abstract from such strategic considerations, and simply estimate how important in practice these different costs seem to be.

4. As such we examine "learning by studying" as opposed to "learning by doing." Jarmin (1994) examines the extent to which learning by doing is a non-rival good.

5. Note that we constrain the elasticity of output with respect to the knowledge stock to unity. Since knowledge is unobserved, this has no empirical implications, so long as we permit the elasticity of the knowledge stock with respect to observables (such as R&D) to be estimated.

6. To check for errors associated with the cost shares we took averages of the α_j between adjacent periods and used the averaged cost shares to recalculate (2). However, the averaging had little effect on the regression findings noted below, suggesting that the cost shares are measured reasonably well. For more on the downward bias caused by using industry shares, see footnote 22.

7. The reason for this limitation is readily apparent: with 32 product areas and 50 states, the reporting burden imposed by the estimation of the joint distribution would exceed the patience and perhaps capacity of firms to answer. Availability of the marginal distributions is already unusual among most data on R&D.

8. Alternatively, let the effect K_{it} be

$$K_{it} = \left[\frac{(R_{it}^{c} + \delta R_{it}^{d})^{\beta_f}}{(n_{it})^{\gamma}}\right]\left[\frac{S_{it}^{\beta_s}}{N_{it}^{\gamma_s}}\right]$$

where n_{it} is the number of the firm's plants whether close by or distant. This difference in specifications generally matters little to the estimated effect of firm R&D.

9. See Griliches (1979) for a discussion, and Griliches and Klette (1992) for an attempted solution of the problem.

10. In firm level data, the preferred approach to measuring R&D/productivity effects in panel data is to use fixed effects or differences, in order to allow for unobserved permanent differences among the observation units. This greatly decreases the signal to noise ratio. In the plant-level data we were unable to get meaningful results with any estimation method that allows for unobserved effects, including the long-difference estimator proposed by Griliches and Hausman (1986). We therefore rely on the included control variables to capture the important effects.

11. We say attempted, because firm id numbers in the R&D survey are not updated with ownership changes as they are in the LRD. We achieved a 95% match rate for R&D firms in census years and a 74% match rate in Annual Survey of Manufactures (ASM) years, when the majority of small plants and hence small firms are not surveyed.

12. This criterion, combined with the appearance and disappearance of firms from the ASM, has the effect of introducing perforations --frequent starts and stops-- in the merged data. Hsiao (1986), Ch.8 contains a discussion of econometric methods for dealing with perforated data.

13. We thank John Musgrave of BEA for the industry deflators.

14. See Adams and Peck (1994) for definitions of the product fields in the R&D survey.

15. The regression approach to TFP performs regressions of the log output on a vector of real inputs in logarithmic form. The regression coefficients are average output elasticities, and need not sum to 1.0, that is impose constant returns to scale. However the sum is usually close to 1.0 because the average plant operates at minimum average cost.

16. As one would expect of this industry, over half the plants are concentrated in seven localities: California, Illinois, New Jersey, New York, Pennsylvania, Ohio, and Texas.

17. The industry groups are at the 3 digit industry level within chemicals, and the 2 digit level outside chemicals. The groups correspond to those in Table 1.

18. For the full sample the F statistic is 148.6. For the sample of chemical plants the F statistic is 317.7.

19. A related interpretation might be that there is no dilution, so firms with more plants get a greater total return to R&D; this causes them to spend more on R&D, driving down its marginal product and thereby making it appear that R&D has less effect for firms with more plants (Cohen and Klepper, 1993). In addition to including number of employees to control for size effects, we also estimated the model with higher order terms in R&D, and this does not cause the number-of-plants effect to diminish.

20. Not all parent firms owning the manufacturing plants are listed on the New York Stock Exchange, which is a requirement to be in Compustat. Many are foreign firms listed on foreign exchanges. Thus we lose about half the firms in the Census LRD.

21. It is for this reason that we omit industry controls from Table 5, so that the plant and firm level results are comparable: industry shares are not available at the firm level.

22. One reason for the smaller elasticity is that we are imposing common output elasticities on the plants, which are in different industries and use different technologies. This is a specification error forced by the need to compare plants with firms.

23. By a state's R&D center, we mean its major center of urban population. We used the Census Picadad file to calculate distances on the surface of the earth between all of the latitude and longitude points represented by these centers. The implicit logic is that plants are situated around these same locations, so that our distance calculation is correct on average. This approximation is imposed on us by the aggregation of R&D at the state rather than the laboratory level.

24. In this and all subsequent Tables, we suppress the estimated effects for regions, life-cycle status, number of firm employees and relative plant age; their general nature does not change in the different specifications.

25. See Klette (1994) for an analysis of economies of scope of R&D within Norwegian manufacturing.

26. The marginal product of each R&D stock is given by the product of the estimated elasticity η_i and the ratio of output Q to the respective amount of R&D, R_i. Thus, the marginal product is $\partial Q/\partial R_i = \eta_i Q/R_i$. Therefore the ratio of the marginal products of spillover R&D (S) to firm R&D (F) is given by $\eta_S Q/R_S \div \eta_F Q/R_F = (\eta_S/\eta_F)(R_F/R_S)$. This calculation yields (.24/.06)*(23/350)=.39. The first ratio derives from Table 8, while the second ratio is mean firm R&D divided by mean rest of industry R&D as given by Table 2.

27. The fifth-ranked state in the food industry, Missouri, is not among the ten states most populated by plants.

28. Agricultural chemicals are dominated in these data by pesticide manufacturers, who have R&D and manufacturing processes similar to drugs, and thus a similar set of manufacturers. We are indebted to Michael Ollinger of the USDA for this piece of information.

29. We explored several additional variations to determine if the number of plants was proxying for something else. In particular, the number-of-plants effect is not significantly diminished by controlling for firm diversification or industry concentration (both measured at the 4-digit SIC level). Interestingly, in the presence of the plants variable, diversification was positively associated with productivity.

30. See, for example, Williamson (1967), Calvo and Wellisz (1978), Holmstrom and Tirole (1989), Dearden, Ickes and Samuelson (1990), Geanokoplos and Milgrom (1991), and McAfee and McMillan (1992).

31. In one of our experiments we split our sample into earlier and later periods, the idea being one of capturing increasing flows of information across different localities with the improvement of telecommunications. We do in fact find that R&D is less localized in the later period. However, the later period is limited to larger plants in the Census data owing to a shift in the sample design of the LRD, so that we cannot disentangle this effect from improvements in information flows within firms.

32. This is the problem already alluded to, of having marginal distributions of R&D by applied product and locality but not the joint distribution along both dimensions of R&D.

Table 1
The Distribution of Plants by Industry Group

Industry Group (SIC in parentheses)	Number of Plant Years (% of total in parentheses)
Industrial Inorganic and Organic Chemicals (281, 286)	5572 (25.9)
Plastics, Resins, and Fibers (282)	1287 (6.0)
Drugs (283)	1598 (7.4)
Agricultural Chemicals (287)	711 (3.3)
Soaps, Paints, Other Chemicals (284, 285, 289)	3530 (16.4)
Total for Chemicals (28)	**12698 (58.9)**
Food (20)	1141 (5.3)
Petroleum Refining (29)	581 (2.7)
Rubber and Miscellaneous Plastics Products (30)	1370 (6.4)
Machinery (35)	635 (2.9)
Electrical Equipment (36)	730 (3.4)
Instruments (38)	1247 (5.8)
Other Manufactures	3144 (14.6)

Notes: Period is 1974-1988. Plants are restricted to those owned by chemical concerns. The definition of chemical firms follows the research and development survey.

Table 2
Means of Total Factor Productivity and Applied Product Field R&D by Industry Group
(Standard Deviations in Parentheses)

Industry Group	TFP	Product Field R&D of Parent Firm	Product Field R&D of Rest of Industry
All Industries	7.3	23,323	350,243
	(9.0)	(45,033)	(336,038)
Food	4.6	16,948	85,160
	(6.2)	(24,718)	(29,226)
Textiles and Apparel	6.0	2,030	20,565
	(4.3)	(4,525)	(32,095)
Lumber, Furniture, and Paper	6.2	0	0
	(7.5)	(0)	(0)
Chemicals Industry			
Industrial Organic and Inorganic Chemicals	4.4	37,117	463,501
	(6.5)	(44,721)	(93,241)
Plastics, Resins, and Fibers	3.9	50,564	467,111
	(3.4)	(91,527)	(142,808)
Drugs	15.1	57,019	1,239,765
	(12.9)	(63,176)	(286,646)
Agricultural Chemicals	4.9	11,584	250,055
	(10.6)	(17,905)	(62,386)
Paints, Soaps, and Other	6.6	16,528	473,199
	(6.0)	(35,986)	(98,029)
Petroleum Refining	2.9	2,591	36,146
	(2.7)	(9,625)	(16,683)
Rubber and Plastics	7.4	7,276	92,665
	(6.5)	(30,940)	(92,362)
Stone, Clay, and Glass	14.9	2,088	16,057
	(23.6)	(3,739)	(3,904)
Primary and Fabricated Metals	8.5	7,230	67,163
	(7.5)	(18,840)	(44,470)
Machinery and Transportation Equipment	12.5	4,971	29,405
	(9.4)	(15,367)	(20,536)
Electrical Equipment	10.5	22,900	67,604
	(9.5)	(39,595)	(51,058)
Instruments and Miscellaneous	13.2	9,265	91,581
	(9.7)	(16,680)	(62,694)

Notes: See Equation (2) and the accompanying text for the definition of TFP. R&D variables are in thousands of 1987 dollars.

Table 3
Baseline Model of Firm R&D Effects on Plant Productivity
(t-Statistics in parentheses)

Variable or Statistic	All Plants				Chemical Plants		
	Eq. 3.1	Eq. 3.2	Eq. 3.3	Eq. 3.4	Eq. 3.5	Eq. 3.6	Eq. 3.7
log of firm R&D (flow)	0.01 (2.2)	0.06 (14.6)			-0.01 (-1.2)	0.06 (11.7)	
log of firm R&D (stock)			0.08 (12.0)				0.08 (9.6)
log of firm R&D per plant (flow)				0.07 (17.4)			
log (number of plants)		-0.17 (-25.3)	-0.20 (-19.8)			-0.19 (-22.8)	-0.21 (-16.7)
log (number of firm employees)	-0.01 (-2.0)	0.01 (3.0)	0.01 (1.2)	-0.02 (-5.8)	0.00 (0.4)	0.03 (4.8)	0.02 (2.6)
Relative Plant Age[a]	0.01 (14.0)	0.01 (14.8)	0.01 (11.2)	0.01 (14.4)	0.02 (14.6)	0.02 (14.8)	0.02 (11.9)
Birth dummy	-0.29 (-5.3)	-0.27 (-5.1)	-0.31 (-3.7)	-0.29 (-5.3)	-0.47 (-6.6)	-0.47 (-6.8)	-0.79 (-6.6)
Slowdown or Death dummy	-0.58 (-13.9)	-0.58 (-14.0)	-0.70 (-12.8)	-0.58 (-13.9)	-0.49 (-8.7)	-0.49 (-8.9)	-0.57 (-8.0)
Regional Dummies							
South	-0.06 (-3.6)	-0.06 (-3.5)	-0.05 (-2.1)	-0.06 (-3.9)	-0.07 (-3.4)	-0.07 (-3.4)	-0.04 (-1.4)
North	0.16 (9.5)	0.16 (9.3)	0.21 (8.9)	0.16 (9.1)	0.17 (7.4)	0.16 (7.1)	0.21 (6.6)
West	0.00 (0.2)	0.00 (0.1)	0.03 (1.5)	-0.00 (-0.3)	-0.01 (-0.5)	-0.01 (-0.3)	0.04 (1.5)
Adjusted R^2	0.37	0.39	0.40	0.38	0.35	0.38	0.41
N	19345	19345	9892	19345	11399	11399	5923

Notes: Dependent variable is log of total factor productivity of the plant. Estimation method is OLS. All equations include industry and year dummies. See text for construction of stock of R&D.

[a]Age of plant in years minus average age of all of firm's plants.

Table 4
Means of Plant Level Total Factor Productivity
Classified by the Number of Plants in a Firm

Range of Plants in a Firm	Number of Plant Years	Mean of Plant Level TFP
0-10	2458	8.43
11-20	3269	8.60
21-40	4834	7.58
41-70	3886	5.57
71-100	2922	5.95
101+	5348	5.55

Notes: See Equation (2) and the accompanying text for the definition of TFP.

Table 5
Firm and Plant Level Production Functions
Matching Samples from the Compustat and LRD Data
(t-Statistics in parentheses)

Variable or Statistic	Plant Level Regressions		Firm Level Regressions					
	Eq. 5.1	**Eq. 5.2**	**Eq. 5.3**	**Eq. 5.4**	**Eq. 5.5**	**Eq. 5.6**	**Eq. 5.7**	**Eq. 5.8**
log of Compustat R&D (flow)					0.10 (13.6)			
log of Census-NSF R&D (flow)	0.01 (3.5)		0.06 (8.2)	0.06 (7.9)				
log of Census-NSF R&D per plant (flow)						0.07 (12.3)		
log of Census-NSF R&D (stock)		0.03 (5.3)					0.10 (7.8)	
log of Census-NSF R&D per plant (stock)								0.10 (11.9)
log (number of plants)	-0.02 (-3.5)	-0.06 (-6.2)		-0.09 (8.8)	-0.07 (-7.2)		-0.10 (-7.9)	
log (employees)	0.26 (47.3)	0.23 (31.3)	0.41 (32.3)	0.41 (35.1)	0.39 (33.8)	0.41 (35.2)	0.42 (27.1)	0.42 (28.1)
log (stock of capital)	0.13 (28.5)	0.14 (21.2)	0.07 (4.6)	0.08 (5.8)	0.02 (1.1)	0.07 (5.4)	0.06 (3.1)	0.06 (3.6)
log (materials)	0.59 (122.4)	0.61 (93.3)	0.43 (26.5)	0.48 (29.7)	0.52 (32.8)	0.47 (29.6)	0.46 (21.9)	0.46 (22.0)
Adjusted R^2	0.94	0.95	0.99	0.99	0.99	0.99	0.99	0.99
N	13286	7024	534	534	524	534	292	292

Notes: Dependent variable is log of sales. All regressions include time dummies. Plant-level regressions include industry dummies.

Table 6
Geographic Localization of R&D Within Firms
(Asymptotic t-Statistics in Parentheses)

Variable or Statistic	All Plants			Chemical Plants		
	Eq. 6.1	Eq. 6.2	Eq. 6.3	Eq. 6.4	Eq. 6.5	Eq. 6.6
	Proximity concept			Proximity concept		
	same state	same state	circle of 100 miles	same state	same state	circle of 100 miles
log of firm R&D (flow)	0.06 (13.6)			0.05 (10.8)		
log of firm R&D (stock)		0.07 (11.6)	0.07 (10.9)		0.07 (8.9)	0.06 (8.6)
Fractional effect of "far away" R&D[a]	0.19 (3.1)	0.28 (2.8)	0.16 (2.8)	0.15 (2.5)	0.13 (2.1)	0.11 (2.2)
log (number of plants, nearby)	-0.06 (-6.4)	-0.06 (-4.0)	-0.03 (-3.3)	-0.12 (-9.3)	-0.11 (-5.9)	-0.05 (-4.6)
log (number of plants, far away)	-0.12 (-17.8)	-0.15 (-14.8)	-0.16 (-16.5)	-0.12 (-14.6)	-0.14 (-11.1)	-0.16 (-13.0)
Adjusted R^2	0.39	0.4	0.4	0.39	0.41	0.41
N	19348	9893	9877	11402	5924	5912

Notes: Dependent variable is log of total factor productivity. Estimation method is NLLS. Other variables in the regressions include dummies for year, industry, plant operating status (birth, slowdown, and death), region, relative plant age, and log(firm employees), as in Table 3. See text for construction of the stocks of R&D.

[a] Relative value of R&D outside of the state or outside 100 miles, if "nearby" R&D has value of unity. See text for more detail.

Table 7
Technological Localization of R&D Within Firms:
Applied Product Field Decomposition
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	All Plants		Chemical Plants	
	Eq. 7.1	Eq. 7.2	Eq. 7.3	Eq. 7.4
log of firm R&D (flow)	0.04 (10.7)		0.04 (8.3)	
log of firm R&D (stock)		0.04 (6.4)		0.05 (5.8)
differential effect of firm R&D in other product fields[a]	0.54 (2.5)	0.02 (1.3)	0.27 (2.3)	0.06 (1.4)
log (number of plants, same product field)	-0.15 (-26.8)	-0.18 (-22.6)	-0.21 (-28.6)	-0.22 (-21.8)
log (number of plants, other product fields)	-0.03 (-5.0)	-0.02 (-2.3)	-0.01 (-1.7)	-0.00 (-0.3)
Adjusted R^2	0.40	0.41	0.40	0.419
N	19348	9893	11402	6147

Notes: Dependent variable is log of total factor productivity. Estimation method is NLLS. Other variables in the regressions include dummies for year, industry, plant operating status (birth, slowdown, and death), region, relative plant age, and log(firm employees), as in Table 3. See text for construction of the stocks of R&D.

[a] Relative value of R&D outside of the product field, if product-field R&D has value of unity. See text for more detail.

Table 8
Within-Firm and Industry-Spillover R&D Effects Within and Beyond a Circle of 100 Mile Radius
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	All Plants			Chemical Plants		
	Eq. 8.1	Eq. 8.2	Eq. 8.3	Eq. 8.4	Eq. 8.5	Eq.8.6
log of firm R&D (flow)	0.05 (11.1)	0.04 (9.6)	0.04 (8.5)	0.05 (9.0)	0.04 (8.0)	0.04 (7.0)
fractional effect of firm R&D > 100 miles away[a]	0.07 (2.7)	0.04 (2.2)	0.02 (1.8)	0.07 (2.2)	0.03 (1.8)	0.01 (1.5)
log (relevant industry R&D[b])	0.07 (6.2)	0.24 (14.0)	0.24 (13.1)	0.07 (4.7)	0.21 (9.1)	0.22 (9.2)
fractional effect of relevant industry R&D > 100 miles away[a]			1.92 (2.5)			0.98 (2.8)
log (number of firm plants within 100 miles)	-0.04 (-6.6)	-0.04 (-2.2)	-0.04 (-4.6)	-0.08 (-9.4)	-0.08 (-4.7)	-0.07 (-6.8)
log (number of firm plants outside 100 miles)	-0.13 (-19.1)	-0.11 (-16.9)	-0.11 (-15.5)	-0.14 (-16.1)	-0.13 (-15.0)	-0.12 (-13.7)
log (number of industry plants)		-0.27 (-11.9)			-0.24 (-7.8)	
log (number of industry plants within 100 miles)			-0.01 (-3.1)			-0.01 (-2.6)
log (number of industry plants outside 100 miles)			-0.26 (-11.0)			-0.22 (-7.1)
Adjusted R^2	0.39	0.41	0.39	0.39	0.39	0.38
N	18581	18581	17517	10903	10903	10157

Notes: Dependent variable is log of total factor productivity. Estimation method is NLLS. Other variables in the regressions include dummies for year, industry, plant operating status (birth, slowdown, and death), region, relative plant age, and log(firm employees), as in Table 3. See text for construction of the stocks of R&D.

[a] Relative value of R&D outside 100 miles, if "nearby" R&D has value of unity. See text for more detail.

[b] Following Jaffe (1986), relevant industry R&D is the weighted average of all other firm's R&D, $\sum_j \cos \theta_{ij} R_j$, where $\cos \theta_{ij}$ is the uncentered correlation between the vectors of applied product R&D in the particular firm i and all others j, and R_j is the R&D of firm j. These correlations are based on the 32 field breakdown in the Census-NSF R&D Survey. See text for more detail.

Table 9
Fractions of Plants in Leading States, by Industry
(State Rank in an Industry in Parentheses)[a]

State	All Industries	Food	Industrial Organic & Inorganic Chem.	Drugs & Agricultural Chem.	Plastics, Soaps, Paints, & Misc Chem.	Rubber and Plastics
California	**0.098** (1)	**0.128** (1)	**0.059** (5)	**0.087** (2)	**0.105** (2)	**0.122** (1)
Illinois	0.060	**0.066** (3)	0.051	0.036	**0.089** (3)	**0.074** (4)
Michigan	0.031	0.025	0.036	0.026	0.040	0.011
New Jersey	**0.081** (2)	0.035	**0.083** (2)	**0.128** (1)	**0.116** (1)	**0.083** (3)
New York	0.054	**0.061** (4)	0.035	**0.076** (3)	0.035	0.052
North Carolina	0.031	0.018	0.025	0.041	0.040	0.031
Ohio	**0.069** (3)	0.041	**0.078** (3)	0.042	**0.069** (4)	**0.096** (2)
Pennsylvania	**0.063** (5)	0.039	**0.061** (4)	**0.076** (4)	0.037	**0.070** (5)
Tennessee	0.030	0.045	0.042	0.028	0.028	0.018
Texas	**0.068** (4)	**0.086** (2)	**0.108** (1)	**0.060** (5)	**0.066** (5)	0.016

Notes: Selection of states is as follows. The state must be among the top ten ranked by number of manufacturing plants in the sample. Selection of industries coincides with five of the most important for the sample, as measured by number of plants. [A] Rank is determined by the share of plants in a state, for the leading 5 states, in a particular industry. If states listed as among the top ten for "All Industries" do not include one or more of the leading 5 states for a particular industry listed to the right of "All Industries," then some ranks will be missing.

Table 10
Within-Firm and Industry-Spillover Effects by Chemical Sub-Industry
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	Industrial Organic and Inorganic Chemicals		Drugs and Agricultural Chemicals		Plastics, Soaps, Paints, and Miscellaneous	
	Eq. 10.1	Eq. 10.2	Eq. 10.3	Eq. 10.4	Eq. 10.5	Eq. 10.6
log of firm R&D (flow)	0.05 (6.2)	0.04 (5.2)	0.13 (7.3)	0.11 (6.0)	0.05 (8.0)	0.04 (5.1)
fractional effect of firm R&D > 100 miles away[a]	0.02 (1.4)	0.01 (1.1)	0.65 (2.0)	0.60 (1.7)	0.23 (1.6)	0.16 (1.3)
log of relevant industry R&D (flow)[b]		0.22 (5.6)		0.63 (8.2)		0.18 (5.9)
log (number of firm plants inside 100 miles)	-0.09 (-8.5)	-0.09 (-8.1)	-0.12 (-4.2)	-0.13 (-4.7)	-0.09 (-7.1)	-0.09 (-6.9)
log (number of firm plants outside 100 miles)	-0.17 (-13.2)	-0.16 (-12.8)	-0.35 (-13.1)	-0.21 (-7.2)	-0.13 (-10.9)	-0.13 (-10.4)
log (number of industry plants)		-0.24 (-4.4)		-0.84 (-9.2)		-0.28 (-7.1)
Adjusted R^2	0.23	0.24	0.29	0.32	0.12	0.13
N	4843	4843	1992	1992	4068	4068

Notes: Dependent variable is log of total factor productivity. Estimation method is NLLS. Other variables in the regressions include dummies for year, industry, plant operating status (birth, slowdown, and death), region, relative plant age, and log(firm employees), as in Table 3. See text for construction of the stocks of R&D.

[a] Relative value of R&D outside 100 miles, if "nearby" R&D has value of unity. See text for more detail.

[b] Following Jaffe (1986), relevant industry R&D is the weighted average of all other firm's R&D, $\sum_j \cos \theta_{ij} R_j$, where $\cos \theta_{ij}$ is the uncentered correlation between the vectors of applied product R&D in the particular firm i and all others j, and R_j is the R&D of firm j. These correlations are based on the 32 field breakdown in the Census-NSF R&D Survey. See text for more detail.