The Span of the Effect of R&D in the Firm and Industry*

by

James D. Adams
Center for Economic Studies, U.S. Bureau of the Census
and University of Florida

and

Adam B. Jaffe
Harvard University and NBER

CES 94-7 June 1994

information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Room 1587, FB 3, U.S. Bureau of the Census, Washington, D.C. 20233-6300, (301) 763-2065 or INTERNET address snguyen@info.census.gov.

Abstract

Previous studies have found that the firm's own research and spillovers of research by related firms increase firm productivity. In contrast, in this paper we explore the impact of firm R&D on the productivity of its individual plants. We carry out this investigation of within firm R&D effects using a unique set of Census data. The data, which are from the chemicals industry, are a match of plant level productivity and other characteristics with firm level data on R&D of the parent company, cross-classified by location and applied product field.

We explore three aspects of the span of effect of the firm's R&D: (i), the degree to which its R&D is "public" across plants; (ii), the extent of its localization in geographic space, and (iii), the breadth of its relevance outside the applied product area in which it is classified. We find that (i), firm R&D acts more like a private input which is strongly amortized by the number of plants in the firm; (ii), firm R&D is geographically localized, and exerts greater influence on productivity when it is conducted nearer to the plant; and (iii), firm R&D in a given applied product area is of limited relevance to plants producing outside that product area. Moreover, we find that while geographic localization remains significant, it diminishes over time. This trend is consistent with the effect of improved telecommunications on increased information flows within organizations.

Finally, we consider spillovers of R&D from the rest of industry, finding that the marginal product of industry R&D on plant productivity, though positive and significant, is far smaller than the marginal product of parent firm's R&D.

KEYWORDS: R&D, technical change, productivity.

* We thank Mark Doms and Su Peck of the Center for Economic Studies, Bureau of the Census, for advice concerning the data on plants and firms used in this paper. Lawrence Kenny, William Long, and David Sappington offered helpful suggestions, along with seminar participants at Bureau of the Census, Florida, and Maryland. The paper also benefited from presentations at the <u>Fourth International Symposium on the Management of Technology</u>, Sainte Adele, Quebec, October 1993, and Meetings of the American Economic Association, Boston, January 1994.

I. 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 generate equilibrium growth paths for the economy as a whole in which income per capita can grow forever (Romer (1986, 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 quite distinct phenomena. This has not always been the case. Indeed, Alfred Marshall, who is often credited with being the first to write about the phenomena that we now call "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 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, less useful but still relevant for some others, and useless for others. Whether we look within or across firms, the "effective non-rivalnous" of knowledge will be affected by the extent to which knowledge developed in a given specific circumstance is, in fact, 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, 1994) among distinct business units within a firm.

---

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

Second, for a given bit of knowledge to be widely used it must be effectively transferred across institutional, cultural and/or geographic boundaries.[2] The cost of this transfer process works against increasing returns, so that the extent of increasing returns is powerfully affected by the magnitude of these costs. Looked at this way, the boundary of the firm is just one of several important sources of transactions costs that may limit increasing returns. It is an empirical question, for example, if the cost of learning about and absorbing research results from another plant is typically higher if 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 established research labs.[4] To begin to get a handle on the

---

2. The last line of the passage from Marshall suggests that he 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 the 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 (outbound) 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 (1993) examines the extent to which learning by doing is a non-rival good.

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 dataset 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 ("NSF R&D data"). Because of the laboriousness of this matching process, we limit ourselves to establishments and firms within the chemical industry (SIC 28).

The paper is organized as follows. Section II 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 III describes in detail the data on firms and establishments, and discusses a number of measurement issues. Section IV presents the results. Section V discusses those results, focussing particularly on what we find to be apparently strong *decreasing* returns within firms. Section VI contains concluding remarks.

II. Modelling Framework

We postulate that a plant (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. In this paper, we ignore learning by doing and informal research as knowledge sources, and focus on the 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*:

$$Q_{it} = K_{it} L_{it}^{\alpha_{Lt}} C_{it}^{\alpha_{Ct}} M_{it}^{\alpha_{Mt}} \exp(\epsilon_{it}), \qquad (1)$$

where Q_{at} 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 $,_{it}$ is everything else that affects output.[5] Rather than try to

---

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.

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:

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

from the input and output data and the factor shares. Substituting (2) into (1) 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.

Ideally, we would construct a proxy for the effective knowledge stock that simultaneously incorporated all of the effects of interest. Unfortunately, there are inherent limitations on the ability of the data to simultaneously identify the effects of distance along 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 only the *marginal* distributions. 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:

$$K_{it} = \left[\frac{(R_{it}^{c} + \delta R_{it}^{d})^{\beta_f}}{(n^{c}_{it})^{\gamma_c}(n^{d}_{it})^{\gamma_d}}\right]\left[\frac{R_{It}^{\beta_I}}{n_{It}^{\gamma_I}}\right] \tag{3}$$

where R^{c}_{it} connotes research of i's parent firm that is in some sense "close" to plant i, and R^{d}_{it} connotes research that is in some sense "distant." As indicated above, the meaning of close and distant could be either geographic or technological. The variables n^{c} and n^{d} are the total number of plants (including i) that are in the "close" and "far" groups, however defined. R_{It} denotes the research of firms other than i's parent; n_{It} is the total number of plants in the industry[6].

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 potentially 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

---

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

$$K_{it} = \left[\frac{(R_{it}^{c} + \delta R_{it}^{d})^{\beta_f}}{(n_{it})^{\gamma}}\right]\left[\frac{R_{It}^{\beta_I}}{n_{It}^{\gamma_I}}\right]$$

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

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 intrinsically continuous. Our approach is 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 with University of Chicago.

Finally, by "normalizing" the knowledge stock 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 the (s could be zero, suggesting strong increasing returns, at least as long as technological and geographic distances are kept small. Our original conception was that the magnitude of the (parameters would fall between 0 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, the (s are often larger than the $s, suggesting a form of decreasing returns that we will discuss further below.

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

$$\begin{aligned} \ln(TFP_{it}) = \beta_f \ln(R_{it}^c + \delta R_{it}^d) + \beta_I \ln(R_{It}) - \gamma_c \ln(n_{it}^c) \\ - \gamma_d \ln(n_{it}^d) - \gamma_I \ln(n_{It}) + \sum_k \phi_k Z_{it}^k + \mu_{it} \end{aligned} \qquad (4)$$

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.[7]

III. Description of the Data

We study chemicals (SIC 28) in this paper 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 span the period 1974-1988.

---

7. 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 must therefore rely on the hope that the included control variables capture most of the important effects.

The 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.

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 Survey of Manufactures 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[8]. In the case of the R&D we required that data

8. 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 ASM years.

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[9].

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.

---

9. 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.

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[10]. 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, $*_{jt}$ 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

---

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

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.

Since we follow a computational approach to TFP, then (2) requires estimates of factor cost shares in order to compute estimates of the "$_{Zi}$ elasticities[11]. We begin with expenditures. Labor expenditures equal wages of production and non-production workers plus supplementary labor costs. Materials expenditures are expenditures 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 shadow value of rentals on the firm's capital stock. We multiply real capital stock by 2 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

---

11. The regression approach to TFP performs regressions of the log of real 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.

expenditures, on 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 materials last. While some might object that this procedure imposes constant returns on the data, the alternative regression procedure, which does not impose this restriction, generally finds the sum of the elasticities close to one.

Descriptive statistics are reported in Tables 1 and 2. Table 1 reports the industrial distribution of the plants[12]. 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-- and food processing. This pattern of concentration of plants in industries that are strongly affiliated with chemicals naturally conditions our analysis of industry groups, since the study of outliers requires reliable indicators of central tendency.

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.

---

12. 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.

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 value of sales not exceed 10.0.

The low end of the range of TFP is populated by plant births, for which output has not as yet caught up with input, and it is populated by plants that are idled. Paradoxically, a rather high productivity can be implied by plant death, since inputs can be set at a low level as the plant subsists off the sale of final goods inventories. The rather high standard deviations of TFP suggest the importance of industry differences, births, deaths, and plant idling, as well as measurement error. Clearly differences in TFP are influenced by a good deal besides technology. In particular they are influenced by industry variations in overhead costs, such as marketing and other central office expenses.

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.

As a final data issue, we confront the theoretical expectation that the effective stock of knowledge should depend on the *history* of research investments on which the plant draws, not just current R&D. As a practical matter, 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:

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

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

V. 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 dummies for year, sub-industries, regions, new plants and plants with large output reductions. We perform the estimation for all plants owned by the identified 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. 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, 3.4, 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 .07, which are 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[13]. 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

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

maintain its effectiveness at each plant. This is a disturbing result, implying that firms would be better off breaking themselves into pieces. It is a robust finding in these data. Nevertheless, equation 3.3, 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 scarcely larger than the specifications that introduce the log of R&D and the log of number of plants separately.

In Table 4 we introduce the first distance distinction into the regressions. We decompose the firms' R&D into that portion that is in the same state as the plant, and all other, and estimate the relative contribution of each using the formulation of Equation (3). The results are quite similar to those of Table 3, except that we find the expected diminution of effectiveness for more distant R&D.[14] We find that R&D performed outside the state is roughly 10 to 20 percent as effective as R&D performed in the same state. 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.

---

14. In this and all subsequent Tables, we suppress the estimated effects for regions and life-cycle status; their general nature does not change in the different specifications.

Table 5 explores the effect of technological rather than geographic distance. We find that R&D outside the plant's product field is roughly one-third as effective as R&D in the plant's product field. The overall R&D elasticities fall further from those in Table 3. Technological effects are not estimated as precisely as geographic ones, probably reflecting greater measurement error in the allocation of firms' R&D across fields relative to the allocation across states.

Table 6 is analogous to Table 4, but broadens the notion of "close" to include all states within 100 miles of the plant. This is intended to allow for the reality that, particularly in small states in the Northeast, research could be close while being in another state. The results are qualitatively similar. As expected, the implied discount for being "far" is now even greater; research in states beyond 100 miles is only 7 to 11 percent as useful as research done inside that radius.

Table 7 incorporates spillover effects. We find that the R&D of other firms does affect a plant's productivity.[15] We also find that the elasticity of plant productivity with respect to other firms' R&D is approximately the same as the elasticity with respect to the parent firm's R&D. Note that industry R&D is a

---

15. Note that, unlike the previous, these results do not include industry dummies in the regression. There is simply too little within-industry variation in the spillover variables, even with geographic effects, to identify the spillover effects in the presence of industry dummies.

much bigger number, so that the similar elasticities imply that the *marginal product* of industry R&D is approximately one-fifteenth as large as the marginal product of parent firm research. (See Table 2 for means of R&D variables.) In other words, each dollar spent by another firm is much less useful than a dollar spent by the parent, but because there are so many more of them their collective effect is of the same order of magnitude. It is interesting to note that the number of plants in the industry does *not* reduce productivity. In our model, this is interpreted to mean that once knowledge makes it past the boundary of the firm (which significantly reduces its potency), there is no further dilution connected with the number of spillover beneficiaries.

Table 8 concludes the presentation of results by breaking up the data between the first and second halves of the time period. To test robustness to the choice of breakpoint, we compare 1974-78 with 1979-88 (Columns 8.1 and 8.4) and also 1974-1980 and 1981-1988 (Columns 8.2-8.3 versus 8.5-8.6). The samples include all plants, not merely chemical plants. The basic finding, which is insensitive to the breakpoint, is that the return to R&D has increased in the more recent period, and that geographic localization has decreased[16]. These results are consistent with the idea that the pace of technical change has quickened in the

16. We are indebted to David Sappington for suggesting that we stratify the regressions by time period.

most recent period, and with the notion that improvements in communications and information technology have lessened the importance of distance.

V. Discussion and Conclusions

The biggest puzzle in the results is the persistent, strong, large "dilution" effect whereby plant-level productivity falls with the number of plants owned by a single firm. To emphasize the significance of the number-of-plants effect, consider the following stylized summary of our results. We assume constant returns to scale in conventional inputs, and then find elasticities with respect to parent firm R&D of 4-8% and industry R&D of about 6%. Hence, *holding the number of plants constant* we find private returns to scale would fall in the range of 1.04-1.08, while industry returns to scale would be about 1.10-1.14. We find, however, that, holding all else constant, the elasticity of output with respect to the number of plants is about -0.16, suggesting overall *decreasing* returns to scale.

Since this suggests non-optimizing behavior on the part of multi-plant firms, we are naturally inclined to search for other explanations. We are dependent on the plant-level reported sales data, which are to some extent an artifact of transfer prices used by the firms. If firms with more plants tend to use transfer prices that impute more value to headquarters or marketing, this would make the plants of such firms "look" less productive. We are skeptical that this story could account for all of the large number-of-plants effect, since the plant effect could be an additional manifestation of technological distance. All else equal, a firm with many plants will tend to make more different kinds of products. This means that the fraction of the firm's research that is devoted to problems of interest to any particular plant will fall as the number of plants increases. To the extent that product fields are a crude technological classification, and/or firms have difficulty classifying their research by product fields, this effect would not all be captured by the product field distinction in our model. Still, this story would explain why (might approach $; it does not explain why it would exceed it.

Although this result may be an artifact of measurement problems, its size and robustness[17] suggests some consideration of whether it could be real. Williamson (1967) and Calvo and Wellisz (1978) explore the idea that layers of hierarchy create costs in the form of information and directives being inaccurately or inadequately passed down to subordinates. Keren and Levhari (1983) develop a model in which this cost is

17. We explored several 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.

optimally traded off against the benefits of hierarchical organization. It is hard to see how our results are consistent with _optimal_ hierarchy size.

Dearden, Ickes, and Samuelson (1990) model the problem of innovation in hierarchies as a two-period game with managers as principals and subordinates as agents. In this game the principal is only able to observe output, and is unable either to separate job productivity from worker quality, or to ascertain whether an innovation has or has not been adopted. The asymmetry of information allows high productivity agents to shirk work effort and innovation; the optimal compensation structure therefore results in too little innovation that diffuses too slowly, relative to the first best. McAfee and McMillan (1992) also study the additional costs of hierarchy due to the fact that information is one-sided with agents. They point out that hierarchies provide benefits, as well as imposing informational costs, in the form of output coordination and extraction of monopoly rents. Geanokoplos and Milgrom (1991) pursue the possibility of cost savings in detail. Using a quadratic cost objective they demonstrate the possibility of advantages to output coordination that complement the necessity of specialization inside the enterprise which they demonstrate more generally.

Thus, a variety of theoretical approaches assume imperfect information on the part of managers or principals. All entail moral hazard on the part of subordinates or agents or another informational failure that ultimately brings about organizational diseconomies. These explain why organizations will not grow infinitely large, even in the presence of strong economies of scale. If, however, size is anything like optimal, it is hard to reconcile our results with the presence of large multi-plant firms. There would have to be something associated with multiplant operation that powerfully affects profits but not productivity as we measure it.

Putting aside the number of-plants-effect, the results present a plausible picture of the productivity effects of the flows of knowledge emanating from formal research programs. Distance does matter; research labs that are farther away or focussed on other product fields do not have as large effects on productivity at the plant level. There is evidence of research spillovers, suggesting the existence of significant technological externalities associated with chemical research programs.

One important caveat is that much R&D is devoted to product improvement rather than process improvement. In principle, increases in product quality that yield greater sales revenues can be incorporated in the TFP framework. It is unlikely, however, given the way real output is typically measured, that very much quality improvement does show up in TFP as we measure it (Griliches, 1979). This difficulty is confounded by our use of plant-level output measures, since the plant-level prices reported for the establishments of multi-plant firms may be internal transfer prices that do not correspond to market values. These considerations suggest that we would underestimate the effect of R&D on productivity, both within firms and from spillovers.

Bibliography

Bailey, Martin, Campbell, David, and Hulten, Charles, "Productivity Dynamics in Manufacturing Plants," **Brookings Papers on Economic Activity: Microeconomics** (1992): 187-249.

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

Dearden, James, Ickes, Barry W. Ickes, and Larry Samuelson, "To Innovate or Not to Innovate: Incentives and Innovation in Heirarchies," **American Economic Review** 80 (December 1990): 1105-1124.

Cohen, Wesley M., and Stephen Klepper, "A Reprise of Size and R&D," mimeo, Carnegie-Mellon University, 1993.

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

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

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

__________, "The Search for R&D Spillovers," Cambridge, Mass., NBER Working Paper 3768, July, 1991.

__________, and Mairesse, Jacques, "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, Jerry, "Errors in Variables and Panel Data," **Journal of Econometrics** 31:1 (February 1986), 93-118.

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

Holmstrom, Bengt R., and Tirole, Jean, "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** 76 (December 1986): 984-1001.

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

______________, Henderson, Rebecca, and Trajtenberg, Manuel, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," **Quarterly Journal of Economics** 108 (August 1993): 577-598.

Jarmin, Ronald S., "Learning by Doing and Competition in the Early Rayon Industry," Center for Economic Studies Working Paper, April 1993.

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

Keren, Michael, and David Levhari, "The Internal Organization of the Firm and the Shape of Average Costs," **Bell Journal of Economics** 14 (Fall 1983): 474-486.

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

Lichtenberg, Frank, **Corporate Takeovers and Productivity**, Cambridge, Massachusetts, MIT Press, 1992.

Levin, Richard C., and Peter Reiss, "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** 19 (Winter 1988): 538-556.

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

McAfee, R. Preston, and John McMillan, "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** 94 (October 1986): 1002-1037.

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

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

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

Table 1
The Distribution of Plants by Industry Group

Industry Group (SIC in parentheses)	**Number of Plant Years (% of total in parentheses)**
Food (20)	1141 (5.3)
Chemicals (28)	12698 (58.9)
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)
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	R&D of Parent Firm in the Product Group	R&D of Rest of Industry in the Product Group
All Industries	7.3 (9.0)	23,323 (45,033)	350,243 (336,038)
Food	4.6 (6.2)	16,948 (24,718)	85,160 (29,226)
Textiles and Apparel	6.0 (4.3)	2,030 (4,525)	20,565 (32,095)
Lumber, Furniture, and Paper	6.2 (7.5)	0 (0)	0 (0)
Chemicals Industry			
Industrial Organic and Inorganic Chemicals	4.4 (6.5)	37,117 (44,721)	463,501 (93,241)
Plastics, Resins, and Fibers	3.9 (3.4)	50,564 (91,527)	467,111 (142,808)
Drugs	15.1 (12.9)	57,019 (63,176)	1,239,765 (286,646)
Agricultural Chemicals	4.9 (10.6)	11,584 (17,905)	250,055 (62,386)
Paints, Soaps, and Other	6.6 (6.0)	16,528 (35,986)	473,199 (98,029)
Petroleum Refining	2.9 (2.7)	2,591 (9,625)	36,146 (16,683)
Rubber and Plastics	7.4 (6.5)	7,276 (30,940)	92,665 (92,362)
Stone, Clay, and Glass	14.9 (23.6)	2,088 (3,739)	16,057 (3,904)
Primary and Fabricated Metals	8.5 (7.5)	7,230 (18,840)	67,163 (44,470)
Machinery and Transportation Equipment	12.5 (9.4)	4,971 (15,367)	29,405 (20,536)

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

Industry Group	TFP	R&D of Parent Firm in the Product Group	R&D of Rest of Industry in the Product Group
Electrical Equipment	10.5 (9.5)	22,900 (39,595)	67,604 (51,058)
Instruments and Miscellaneous	13.2 (9.7)	9,265 (16,680)	91,581 (62,694)

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

Table 3
Firm R&D Effects on Plant Productivity:
Chemicals Industry
(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**
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant Operating Dummies							
Birth	-0.37 (-6.9)	-0.35 (-6.6)	-0.38 (-4.6)	-0.36 (-6.8)	-0.59 (-8.5)	-0.59 (-8.7)	-0.96 (-8.0)
Slowdown or Death	-0.60 (-14.7)	-0.60 (-14.8)	-0.71 (-13.2)	-0.60 (-14.7)	-0.51 (-9.3)	-0.51 (-9.4)	-0.57 (-8.1)
Regional Dummies							
South	-0.07 (-4.2)	-0.07 (-4.3)	-0.06 (-2.7)	-0.07 (-4.4)	-0.09 (-4.4)	-0.09 (-4.5)	-0.07 (-2.3)
North	0.17 (10.0)	0.16 (9.8)	0.22 (9.5)	0.16 (9.7)	0.18 (7.7)	0.16 (7.3)	0.21 (6.9)
West	0.01 (0.5)	0.01 (0.3)	0.04 (1.8)	0.00 (0.1)	-0.02 (-0.8)	-0.02 (-0.7)	0.04 (1.2)
Measures of Firm R&D							
log (flow of total R&D)	0.006 (2.0)	0.059 (15.8)			-0.004 (-1.2)	0.064 (13.4)	
log (flow of total R&D per plant)				0.061 (15.9)			
log (stock of total R&D)[a]			0.076 (12.8)				0.079 (10.6)
log (number of plants)		-0.15 (-24.4)	-0.19 (-19.9)			-0.17 (-21.8)	-0.20 (-16.5)
Adjusted R^2	0.361	0.379	0.398	0.368	0.342	0.366	0.398
N	20022	20022	10294	20022	11845	11845	6147

Notes. Dependent variable is log (total factor productivity). Estimation method is OLS.
[a] The stock of total R&D is given by

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

where *=0.15. The lag on R&D investments is limted to 5 periods so RDK_t is a partial stock of R&D capital.

Table 4
Geographic Localization of R&D Effects
Within Firms[a]
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	**All plants**		**Chemical Plants**	
	Eq. 4.1	**Eq. 4.2**	**Eq. 4.3**	**Eq. 4.4**
Dummies[b]	Yes	Yes	Yes	Yes
Flow of R&D				
total firm R&D	0.054 (14.5)		0.058 (12.3)	
differential effect of firm R&D in other states	0.115 (3.1)		0.081 (2.5)	
Stock of R&D[c]				
total firm R&D		0.072 (12.0)		0.053 (8.2)
differential effect of firm R&D in other states		0.169 (2.8)		0.008 (1.7)
log (number of plants, same state)	-0.066 (-6.8)	-0.063 (-4.5)	-0.132 (-10.2)	-0.145 (-7.9)
log (number of plants, other states)	-0.110 (-16.7)	-0.140 (-14.2)	-0.106 (-13.1)	-0.094 (-8.9)
Adjusted R^2	0.381	0.399	0.372	0.401
N	20123	10294	11845	6147

Notes. Dependent variable is total factor productivity. Estimation method is NLLS. [a] Specification of firm R&D effects is $b \cdot \log(rd_s + c \cdot rd_o)$, where b is the effect of total firm R&D, rd_s is firm R&D in the same state as the plant, c is the subsidiary effect of firm R&D conducted in other states, and rd_o is firm R&D in other states.
[b]Other variables in the regressions include dummies for year, industry, plant operating status (birth, slowdown, and death), and region, all as noted in Table 3. [c]See notes to Table 3 for the stock of total R&D.

Table 5
Localization of R&D Effects within Firms
In Technology Space[a]
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	All plants		Chemical Plants	
	Eq. 5.1	Eq. 5.2	Eq. 5.3	Eq. 5.4
Dummies[b]	Yes	Yes	Yes	Yes
Flow of (R&D)				
log (firm R&D)	0.044 (11.3)		0.049 (9.9)	
differential effect of firm R&D in other product fields	0.326 (2.6)		0.201 (2.5)	
Stock of (R&D)[c]				
log (firm R&D)		0.039 (6.4)		0.044 (5.6)
differential effect of firm R&D in other product fields		0.010 (1.3)		0.010 (1.0)
log (number of plants, same product field)	-0.150 (-26.8)	-0.178 (-22.9)	-0.213 (-29.2)	-0.224 (-22.1)
log (number of plants, other product fields)	-0.024 (-3.7)	-0.013 (-1.9)	-0.001 (-0.2)	0.013 (1.5)
Adjusted R^2	0.387	0.407	0.372	0.419
N	20123	10294	11845	6147

Notes. Dependent variable is log (total factor productivity). Estimation method is NLLS. [a] Specification of firm R&D effects is $b \cdot \log(rd_a + c \cdot rd_o)$, where b is the effect of firm R&D, rd_a is R&D in the same product field as the plant's, c is the differential effect of firm R&D conducted in other product fields, and rd_o is firm R&D in other product fields. [b] Other variables in the regressions include dummies for year, industry, plant operating status (birth, slowdown, death), and region, all as noted in Table 3. [c] See notes to Table 3 for the stock of total R&D.

Table 6
Geographic Localization of R&D Effects
In a Circle of Given Radius[a]
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	**All plants**		**Chemical Plants**	
	Eq. 6.1	**Eq. 6.2**	**Eq. 6.3**	**Eq. 6.4**
Dummies[b]	Yes	Yes	Yes	Yes
Flow of R&D, Radius=100 miles				
log (firm R&D)	0.056 (14.5)		0.063 (13.1)	
differential effect of firm R&D>100 miles away	0.071 (3.3)		0.091 (3.0)	
Stock of R&D, Radius=100 miles[c]				
log (firm R&D)		0.065 (11.4)		0.066 (9.2)
differential effect of firm R&D>100 miles away		0.107 (2.8)		0.067 (2.2)
log (number of firm plants within 100 miles)	-0.040 (-6.7)	-0.025 (-3.1)	-0.073 (-9.2)	-0.046 (-4.2)
log (number of firm plants outside 100 miles)	-0.120 (-19.0)	-0.152 (-16.4)	-0.133 (-16.0)	-0.146 (-12.5)
Adjusted R^2	0.385	0.399	0.373	0.401
N	19567	10314	11532	6162

Notes. Dependent variable is log (TFP). Estimation method is NLLS. [a] Specification of firm R&D effects is bClog (rd_r +cCrd_o), where b is the effect of total firm R&D, rd_s is total R&D within a radius of 100 miles, c is the differential effect of R&D conducted outside 100 miles, and rd_o is R&D outside the 100 mile radius. [b] Other variables in the regressions include dummies for year, industry, plant status (birth, slowdown, death), and region, all as noted in Table 3. [c] See notes toTable 3 for the stock of total R&D.

Table 7
Firm and Industry R&D Effects
In a Circle of Given Radius[a]
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	All plants		Chemical Plants	
	Eq. 7.1	Eq. 7.2	Eq. 7.3	Eq. 7.4
Industry Dummies	No	No	No	No
Other Dummies[b]	Yes	Yes	Yes	Yes
Flow of R&D (Radius=100,200, 400 miles)				
log (firm R&D)	0.085 (19.4)	0.088 (20.4)	0.108 (20.6)	0.110 (21.8)
differential effect of firm R&D>100 miles away	0.149 (4.3)		0.226 (4.7)	
differential effect of firm R&D>200 miles away		0.353 (4.6)		0.256 (4.7)
log (industry R&D within 400 miles)	0.067 (9.8)	0.084 (10.6)	0.051 (6.7)	0.067 (7.7)
log (number of firm plants within 100 miles)	-0.037 (-5.3)	-0.074 (-8.7)	-0.126 (-14.5)	-0.130 (-12.3)
log (number of firm plants outside 100 miles)	-0.221 (-30.7)	-0.199 (-24.5)	-0.276 (-31.5)	-0.250 (-25.7)
log (number of industry plants within 400 miles		-0.011 (-1.5)		-0.015 (-1.8)
Adjusted R^2	0.136	0.137	0.236	0.236
N	19561	19561	11529	11529

Notes. Dependent variable is log (total factor productivity). Estimation method is NLLS. [a] Specification of firm R&D effects is b$\cdot$log ($rd_r + c \cdot rd_o$), where b is the effect of the log of firm R&D, rd_s is firm R&D within a radius of R miles, c is the differential effect of R&D conducted outside R miles, and rd_o is R&D outside the R mile radius. [b] These are dummies for year, plant status (birth, slowdown, death), and region. [c] See notes toTable 3 for the stock of total R&D.

Table 8
Changes in Localization Over Time:
Firm and Industry R&D Effects
In a Circle of Given Radius[a]
(Asymptotic t-Statistics in parentheses)

Variable or Statistic	**Initial Period**			**Concluding Period**		
	Eq. 8.1	**Eq. 8.2**	**Eq. 8.3**	**Eq. 8.4**	**Eq. 8.5**	**Eq. 8.6**
Time Period	1974-1978	1974-1980	1974-1980	1979-1988	1981-1988	1981-1988
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Other Dummies[b]	Yes	Yes	Yes	Yes	Yes	Yes
Flow of R&D, (Radius= 100 & 400 miles)						
log (firm R&D)	0.029 (6.8)	0.036 (6.8)	0.041 (6.7)	0.074 (15.6)	0.074 (14.1)	0.135 (24.2)
differential effect of firm R&D>100 miles away	0.013 (1.1)	0.022 (1.6)	0.022 (1.5)	0.191 (3.6)	0.215 (3.1)	0.371 (4.8)
log (industry R&D within 400 miles)			0.077 (7.6)			0.071 (5.8)
log (number of firm plants)	-0.113 (-12.4)	-0.125 (14.8)	-0.209 (-20.6)	-0.164 (-18.5)	-0.148 (-14.9)	-0.277 (-24.4)
log (number of industry plants within 400 miles)			-0.030 (-3.1)			0.010 (0.9)
Adjusted R^2	0.390	0.395	0.113	0.398	0.396	0.178
N	9283	11636	11636	10221	7868	7868

Notes. Dependent variable is log (TFP). Estimation method is NLLS. [a] Specification of firm R&D is b·log (rd_r +c·rd_o), where b is the effect of log (firm R&D), rd_s is R&D within a radius of R miles, c is the differential effect of R&D outside R miles, and rd_o is the R&D outside the radius. [b] Dummies stand for year, plant status (birth, slowdown, death), and region. [c] See notes to Table 3 for the stock of total R&D.