This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: R&D and Productivity: The Econometric Evidence

Volume Author/Editor: Zvi Griliches

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-30886-3

Volume URL: http://www.nber.org/books/gril98-1

Publication Date: January 1998

Chapter Title: Issues in Assessing the Contribution of Research and Development to Productivity Growth

Chapter Author: Zvi Griliches

Chapter URL: http://www.nber.org/chapters/c8340

Chapter pages in book: (p. 17 - 45)

2 Issues in Assessing the Contribution of Research and Development to Productivity Growth

2.1 Introduction

Economists have used two rather different styles of research in their attempts to assess the contribution of research and development (R&D) expenditures to economic growth: historical case studies and econometric estimates of production functions containing an R&D variable. There have been a number of detailed case studies of particular innovations tracing out their subsequent consequences (see Griliches (1958) and Mansfield et al. (1977) for examples and Griliches (1973) for a survey). Much can be and has been learned through such studies. They are, however, very data- and time-expensive and are always subject to attack as not being representative, since they tend to concentrate on prominent and successful innovations and fields. Thus, it is never quite clear what general conclusions one can draw on the basis of such studies.

The econometric production function approach tries to meet these objections by abandoning the interesting detail of specific events and concentrating instead on total output or total factor productivity as a function of past R&D investments (and other variables). Here *all* productivity growth (to the extent that it is measured correctly) is related to *all* expenditures on R&D and an attempt is made to estimate statistically the part of productivity growth that can be attributed to R&D (and sometimes, also, to its components). While the production function approach is more general than the case study approach, it is also coarser and suffers from all the problems that beset attempts to infer causality from behavioral data on the basis of correlational techniques. Never-

Reprinted from the *Bell Journal of Economics* 10, no. 1 (Spring 1979): 92–116. Copyright © 1979 American Telephone and Telegraph Company.

This paper was written during my stay at the Institute for Advanced Studies, The Hebrew University, Mount Scopus, Jerusalem. The work on this paper was supported by NASA and NSF. I am indebted to Henry Herzfeld, Carol Kitti, Ariel Pakes, and Mark Schankerman for helpful comments.

theless, currently it is the only available general way of trying to answer questions about the contribution of R&D to growth. It is the purpose of this paper then to explore both the promise and the problems raised by this approach. Some of these problems are conceptual and semantic: What do we want to measure? Others are substantive: What do we measure? And others are statistical-methodological: How can we tell from the data what happened?

Productivity and its growth are best discussed in the context of a "production function," $Y = F(X, \ldots)$, which describes the relationship between various inputs X and final output Y. Productivity ($A = Y/X$) is then defined as the ratio of output (Y) to some index of the total input X and its determinants are then discussable in terms of the list of variables included in X, the mathematical form assumed for the production function $F(\ )$, the particular empirical observations chosen to represent Y and X and the statistical methods used to infer the properties of $F(\ )$ from the data.

A major conceptual issue in any such study is the definition and scope of Y. Are we talking about GNP, the change in some measure of national wealth, or an even broader concept of economic welfare? In the health sector, for example, a major product of research and development is the reduced morbidity of the population. To the extent that R&D affects the workforce and increases hours of work, it will affect X and Y in a parallel fashion and have no effect on A as it is conventionally measured. Moreover, to the extent that it affects the morbidity of children and nonworking wives, it may not show up in our measures at all or show up perversely (e.g., an improvement in the health of children would be associated with less hospitalization, a smaller demand for drugs, and a resultant decline in GNP as it is conventionally measured).[1]

A semantic issue is what do we mean by the "contribution of research and development to growth"? Is it the partial derivative of Y with respect to some measure of research and development flow (R) or stock (K) or is it the total derivative, including the indirect effects of a change in research and development on Y through the induced changes in X? That is, are we asking about the marginal effect of an additional dollar spent on R&D holding all other investments constant, or are we asking for the total effect of a particular R&D investment, including the contribution of all the other investments (in equipment and training) induced by it? To some extent this is a distinction between economic-accounting approaches and causal-historical ones. Most of the economic literature has interpreted the question in the first, partial sense, though occasionally one can find some dissenting voices (Gordon, 1968 and Rymes, 1971). It is a semantic issue, since a complete model of this process can answer both questions and derive the answer to one of them from the other. But it does

1. Of course, some of the freed resources will be spent on something else. The point of the illustration is, however, the fact that the "productivity" in the health sector as conventionally measured may in fact decline as the result of R&D.

reflect differences in emphasis and should be kept in mind in interpreting and framing statements intended to affect research and development policy.

Measurement issues arise both in the case of output and in the case of inputs. Difficult problems exist in the measurement of output in the government and service sectors. Problems arise also in the measurement of output in the private goods sector for complex and changing goods. They are essentially the dual of the familiar "quality change" problem in the construction of price indices. Unfortunately, the more research and development intensive is an industry, the more likely is its output to be subject to such measurement problems. Problems arise also in defining the "stock" of research and development and in developing appropriate deflators for it.

Serious difficulties in econometric inference result from the fact that most of the variables of interest tend to move together over time and space, making it hard to untangle their separate effects. Moreover, it is not easy to establish causality. Research and development investments are themselves affected by the level of output and by past profits and productivity, forcing one to formulate simultaneous equations models and to turn towards much more complex estimation techniques.

Clearly the issues alluded to above exceed by far the scope of one modest survey paper. In what follows I shall, by and large, accept the existing national income accounting framework and interpret the question of the contribution of research and development to productivity as referring to aggregate or industrial productivity measures as they emerge from these accounts.[2] Also, I shall join the majority of the writers on this subject by trying to estimate the *partial* contribution of R&D to growth, holding the contribution of other inputs (labor and capital) constant. Since I have discussed the conceptual and measurement issues at some lengths earlier (Griliches, 1973, 1977), here I shall only briefly recapitulate some of the major points.

The paper proper starts with a brief outline of the production function model used in analyzing returns to R&D and then proceeds to discuss in turn two very difficult problems: the measurement of output in R&D intensive industries and the definition and measurement of the stock of R&D "capital." The latter concept leads us to a discussion and modeling of the spillover effects of R&D (Section 2.2) and to suggestions for possible measurement of such effects via the concept of technological distance between firms and industries (Sections 2.3 and 2.4). Somewhat more familiar econometric problems (multicollinearity and simultaneity) are taken up in Section 2.5, and Section 2.6 is devoted to problems arising more specifically in the R&D context. Conclusions and

2. There is a whole literature criticizing the aggregate production function approach. Within its own frame of reference there are serious problems of aggregation (Fisher, 1969) and one can also raise questions about the profit maximizing or cost minimizing framework that underlies much of this work (Nelson and Winter, 1974 and elsewhere).

recommendations for additional data collection and new research are to be found in the last section of the paper.

2.2 A Model

Let $Y = F(X, K, u)$ be the "production" function connecting some measure of output, Y, at the micro or macro level, to the "inputs" X, K, and u, where X stands for an index of conventional inputs such as labor and capital, K is a measure of the current state of technical knowledge, determined in part by current and past research and development expenditures, and u stands for all other unmeasured determinants of output and productivity. Define also $A = Y/X$ as the level of (total factor) productivity, and $\tau = y - x = (dY/dt)/Y - (dX/dt)/X$ as its rate of growth. Let us also assume that there exists a relationship between K, the current level of technological knowledge, and $W(B)R$, an index of current and past levels of research and development expenditures, where $W(B)$ is a lag polynomial, describing the relative contribution of past and current research and development levels to K, and B is the lag (backward shift) operator. Thus,

$$K = G[W(B)R, v],$$

where v is another set of unmeasured influences on the accumulated level of knowledge and

$$W(B)R_t = (w_0 + w_1B + w_2B^2 + \ldots)R_t = w_0R_t + w_1R_{t-1} + w_2R_{t-2} + \ldots$$

Now the various issues mentioned above can be restated in terms of what we would like conceptually Y, X, K, and R to measure, what the problems with actual measures available to us are, what we need to assume about the world for the relationships $F(\ )$, $G(\ )$, and $W(B)$ to exist and be of interest to us, what we want to know about them, and how we are going to find it out in the face of imperfect data and with the presence of unmeasured forces u and v.

For example, just to write the function $F(X, K, u)$ implies the assumption of separability of the conventional inputs X (labor and capital) from the series of past and current research and development investments R. Similarly, writing $W(B)R_t$ as a linear function of all past R&D investments implies that there are no diminishing returns or rising costs at the annual R&D level.[3] One can look at such assumptions either as a statement about the properties of the real world or as a statement about the conditions for the approximations implied in the construction of aggregate indices of total input (X) or the stock of knowledge capital (K), not to mislead us too much. Given the limited quantity and quality of the data available to us, such assumptions are not really testable (for attempts at such tests cf. Berndt and Christensen (1974)). In what follows I shall

3. An alternative approach would complicate this model further by adding an annual knowledge production function of the form $\dot{K} = H(R,K)$ and defining K accordingly.

ignore these "garden variety" index number problems and concentrate instead on the problems peculiar to research and development.[4]

Nor will I worry much here about the functional form of $F(\ )$. For expositional simplicity I shall assume it to be Cobb-Douglas and assume that the unmeasured factors u can be considered as random after the introduction of a time trend into the equation to represent the systematic component of the unmeasured factors. Then we can rewrite F as:

$$Y = DC^{\alpha}L^{\beta}K^{\gamma}e^{\lambda t+u},$$

where D is a constant, t is a time index, e is the base of natural logarithms, and α, β, γ, and λ are some of the parameters we are interested in estimating. If we had more and better data, we could try for a more complex description of the production process, using more general functional forms such as the CES or the translog, and introducing more parameters to be estimated.[5] But for the purposes of this paper, this simple characterization will suffice.

Let us define a conventional total input index X as:

$$X = C^{s}L^{1-s},$$

where s is the observed factor share of physical capital. Let us assume, for a while, that s is observed correctly and that it is proportional to the true coefficient of capital, i.e., $s = \alpha/(\alpha + \beta)$, and there is no error in computing the true *relative* shares of labor and capital. Then measured total factor productivity,

$$A = Y/X = DX^{\alpha+\beta-1}K^{\gamma}e^{\lambda t+u},$$

depends not only on the contribution of research capital K, the contribution of the trend t in the other unmeasured factors, and the random factor u, but also on the level of other inputs X, as long as there are nonconstant returns to scale ($\alpha + \beta \neq 1$). Moreover, any error of measurement in one of the inputs will transmit itself directly to the productivity measure. For example, let the "true" relevant measure of L be given by $L = Q_L N$, where Q_L is the average "quality" per worker and N is the total number of workers. Now if N rather than L is used in the construction of X, then measured productivity will be given by:

$$\tilde{A} = Y/\tilde{X} = D\tilde{X}^{\alpha+\beta-1}Q_L^{\beta}K^{\gamma}e^{\lambda t+u},$$

where $\tilde{X} = C^{s}N^{1-s}$, and similarly for errors in the measurement of C. Since the issue of errors of measurement in C and L has been discussed at length else-

4. See Diewert (1977) for a recent discussion of the necessary assumptions for the existence of such composite commodities as "labor," "capital," and "research capital."

5. To estimate a more general functional form we would need to observe firms utilizing very different combinations of factors of production. But since most firms face rather similar factor prices, why should they have very dissimilar factor ratios? In our context, the issue of the functional form $F(\ )$ is not very interesting or crucial unless we are interested in the specific interaction of K with a particular input (e.g., if we suspect a particular complementarity between physical and research capital). As far as the role of K itself is concerned, the functional form issue can be investigated by adding the square or other nonlinear functions of K to the equation.

where (Griliches, 1963; Jorgenson and Griliches, 1967), I only want to remind the reader about their potential importance in any practical context.

The issues we want to concentrate on here are the definition of Y, the measurement of K, and the estimation of γ.

2.3 The Measurement of Output in Research and Development Intensive Industries

One of the major difficulties in measuring the contribution of research and development to economic growth is the fact that much of it is performed in industries whose product is itself badly measured. In three major areas of public research and development investments, defense, space, and health, the output measures are based on inputs and hence cannot and do not reflect the improvements in productivity which have been achieved with the help of research and development investments. In space exploration output is measured by man-days and expenditures on equipment and does not rise or fall with the success of the venture. In defense, companies sell equipment to and build battleships for the government. Again, there are no price indices of military aircraft that take into account their improved performance in terms of speed or maneuverability nor are there price indices that reflect the improved resolution and range of radar equipment. Defense purchases are almost always deflated by cost indices and, except for the fact that such indices are often not exactly consistent with the industry data, the resulting productivity measures should show no growth, even though large research and development expenditures have been made successfully to improve the performance of such items.

A similar problem arises in the measurement of output in the service sector, especially in health. The output of physicians is measured basically by the number of patient visits and the output of hospitals is measured by patient days. Any improvement in the performance of these activities in the sense of a higher frequency of cure of specific ailments not only will not show up as an increase in the product of this sector, but might actually lead it to register a decline.

The problem is more complex with new or improved products which are sold directly to consumers, such as pocket calculators or drugs. Here the measurement of output will depend on the market structure of the industry and the procedures used by the statistical agencies (e.g., the Bureau of Labor Statistics in the United States) in constructing the relevant price indices. If the producer of the new item were a perfectly discriminating monopolist, he would capture all the social returns to the innovation. Since most monopolists cannot discriminate perfectly and since their market position is far from secure, the actual revenue received by them will fall short of what could have been realized by a discriminating and secure monopolist. What happens to price indices will depend on whether they allow for the "quality" improvements embedded in the

new item or not. By and large they do not make such quality adjustments.[6] Instead, the new product is "linked in" at its introductory (or subsequent) price with the price indices left unchanged. Hence measured output goes up by the revenue received, costs go up by the increased utilization of inputs in production, and productivity goes up by the amount of the total return to the innovation that the innovator succeeds in appropriating for himself. Subsequently, the entry of other competitors with similar products may force him to lower the price, but that will not increase measured output, since his revenues will decrease proportionately except as the reduction in price results in growth of the overall market for this product. The social return from this erosion of the original innovator's monopoly position will be recorded in the conventional measures. More generally, the amount of social returns reflected in the productivity measures will depend, in part, on how early the price of a new item is included in the official price index. The tendency of statistical agencies is to include it rather late, after the new product has "matured" a bit, and after its price has declined to "normal" levels, thereby missing much of the social contribution of such an invention. Actual scenarios are, of course, even more complicated than outlined above, but the long and the short of it is that official productivity measures in consumer goods industries such as drugs, where no "quality" adjustments are made, reflect only the original private returns appropriated by the inventors and the consumer surplus arising from the subsequent research of competitors which erodes their original market positions.

A simplified example may be of some help here. Imagine an invention of a more sensitive (faster) photographic color film. Let the demand curve for this film be as follows: at $100 per roll only 100 rolls would be sold, at $1 per roll 10,000 rolls will be sold, with all other prices and quantities given by the straight line passing through these points ($q = 10{,}100 - 100p$). Let the average and marginal cost of producing a roll of this film (after the invention) equal $1, including normal markup for overhead. This is also the price of the old film on the market of which 8,000 units were being sold originally. We are assuming then that the new film substitutes for the old film, that at $2 everyone will switch over to the new film, and that at the $1 price the market will expand by an additional 2,000 units.

Now there are several possibilities:

(1) Imagine that this new film was developed by NASA and is licensed freely to everybody. Then the price will be $1, and 10,000 units of the new film (and none of the old) will be sold. Total cost in the film industry will go up from 8,000 to 10,000, as will total revenue. If, as is likely, the statistical

6. Methods for incorporating more extensive "quality" adjustments into the official indices are available but require a much more detailed data base. See Griliches (1971) for a review of the "hedonic" indices literature.

agencies do not interpret this as a fall in the real price of color film, both input and output will have been measured as having increased proportionately and *no increase* in productivity will be recorded.

(2) If this film instead has been produced by a private inventor with a foolproof patent on it, then it would pay him as a monopolist to produce 5,000 units at a price of $5.01 per unit and leave 3,000 units of the old film to be produced at $1 per unit. Now industry revenue will increase to 25,050 + 3,000 = 28,050, while the total cost remains at 8,000, implying an increase in measured productivity of 250 percent (as against zero in the previous competitive, freely licensed case).

(3) Now imagine the possibility that this monopolist could discriminate in the sales of this film individually and get each individual to pay him the maximum price he would be willing to pay for the privilege of using this film. Such a monopolist could collect $510,000 from his sales of 10,000 units at an average price of $51 per unit. In this case revenue will have increased by 511/8 = 64-fold, while costs will have gone up by only 25 percent, which implies a 52-fold increase in productivity. This is exactly the "true" increase in productivity which occurred in case 1 but could not be measured there, since the whole benefit was passed on to the consumer without its being measured anywhere.

(4) It is possible that in situation 1 or 3 the statistical agency notices that all of the previous users of the old film (at the $1 price) are willing to pay at least $2 per roll for the new film and declares that the "quality" of the new film is at least double that of the old one or, in other words, that the real price of film has declined by 50 percent. Then real output will have been measured to increase from 8,000 to 20,000 (in old film units) or by 150 percent while costs (inputs) will have gone up only by 25 percent, which implies a (approximately) 100 percent growth in measured productivity. This is clearly not the right number (case 3 has the correct computation), but it is better than nothing. Unfortunately, such computations are made only in a few industries. Among the major CPI components only automobiles and, recently, housing prices are subject to quality adjustments. Lack of such adjustments can lead to very serious biases in some industries. In computers, for example, the national income accounting convention has been to show no price change whatsoever. There is independent evidence, however, that the "real" price of computers has been falling by about 20 to 30 percent per year in the last decade, leading to a very serious downward bias in the estimate of output and productivity growth in this industry. (See Chow (1967) and Sharpe (1967), among others.)

(5) If the innovation is not a consumer product but is used instead as an input in another industry, then the missing productivity is only misplaced (provided the output of the using industry is correctly measured). If the film discussed above is used only commercially to replace elaborate lighting arrangements, then its contribution would show up in an increase in the

productivity of the photographic services industry (rather than the film industry). Similarly, computer-using industries have really bought more input than has been measured and will therefore show a rise in their productivity, even though the true productivity increase occurred in the computer-producing industry. Thus, to the extent that the output of the invention-using industries is well measured, the returns to such research and development can be found in the aggregate data (though not correctly attributed). But as we have already noted, in many important invention-using sectors such as government, health, and education, output is not well measured, and hence the productivity of the purchased and mismeasured inputs will also not be reflected there.

Up to this point we have been discussing primarily product innovations. Cost-reducing innovations will, of course, be reflected in the productivity measures as long as they occur in industries where output is measured independently of input. Conventional productivity measures reflect, therefore, the cost-reducing inventions made in the industry itself, the privately appropriated part of product innovations within the industry, and the social product of inventions in the input-producing industries which have not already been reflected in the price of purchased inputs.

A word also should be said here about simple cross sectional comparisons. The revenues of a firm reflect its current and past research and development activities. These private returns, however, can exceed the social ones if they occur at the expense of another firm. If, for example, the research and development of a particular firm doubled its revenue by taking it away from a preexisting rival, then, this is not a social return. It is just a transfer. It is more likely, however, that social returns exceed private ones, since only a fraction of them is appropriated by the original inventors.

To summarize, much of reported research and development is expended in areas where its direct contribution cannot be measured. In an earlier paper (Griliches, 1973), I estimated these areas to account for about half of all reported research and development. An additional large component of research and development is aimed at final consumer product rather than process innovations and is reflected in productivity measures only to the extent that producers succeed in appropriating its fruits. Since much of the product of research and development is entirely unmeasured and much of the rest is mismeasured, it is not surprising that it has been rather difficult to find its traces in the data.

2.4 The Measurement of Research and Development Capital

In Section 2.2 we talked about knowledge capital (K). Empirically it is too broad a concept; it aspires to and contains too much. We can, however, focus on the contribution of identified investments in advancing the state of knowledge in a particular (or related) area(s). The contribution of "science" in gen-

eral to a particular industry is probably not measurable, since there is no way of knowing how much "science" is actually used in one industry versus another.

Focusing on the contribution of industrial R&D, we can rewrite the definition of K as:

$$K_t = a_0[W(B)R_t]^{\eta}e^{\mu t+v},$$

where $W(B)R$ is some lag function of past R&D investments, μt is the trend component of all other influences on the state of knowledge, and v is the random transitory component of it. Substituting this into our production function formula, one can absorb η into the γ parameter of the production function, the trend component μt into the general efficiency trend term λt, and the v term into the overall disturbance term u. This leaves us with $W(B)R$, a measure of R&D "capital," as the topic to be discussed in this section.

There are three major issues in the measurement of such "capital":

(1) The fact that the research and development process takes time and that current research and development may not have an effect on measured productivity until several years have elapsed forces one to make assumptions about the relevant lag structure $W(B)$.
(2) Past research and development investments depreciate and become obsolete. Thus the growth in the net "stock" of research and development capital is not equal to the gross level of current or recent resources invested in expanding it.
(3) The level of knowledge in any one sector or industry not only is derived from "own" research and development investments but also is affected by the knowledge borrowed or stolen from other sectors or industries. Thus, the productivity of industry i will depend also on the research and development investments of industries j and h, among others.

Before we proceed to discuss these issues in turn, two other remarks are in order. First, we are obviously aware that there are very different levels and types of knowledge and that research and development results are embodied in people, blueprints, patents, books, and oral tradition. To try to aggregate such diverse items into one notion (index) of research and development "capital" is quite presumptuous. The "sin" of aggregation, however, is not so different here from that committed when constructing measures of national output where bushels of wheat, haircuts, and striptease hours are aggregated into one GNP figure, or "physical" capital which aggregates buildings, planes, computers, and shovels. The difference is, and this is the import of the second remark, that in most of the conventional cases the components to be aggregated have an observable market counterpart to which a value can be attached. A piece of equipment is sold and can be resold at a market price. The results of research and development investments are by and large not directly observable. Re-

search and development capital is fundamentally an "input" rather than an "output" measure. In this, however, it is not so different from, say, the plant construction component of physical capital measures, which are also largely based on cost cumulations rather than on a market valuation of the final results.[7] Nevertheless, the lack of direct measures of research and development output introduces an inescapable layer of inexactitude and randomness into our formulation.[8]

There are several lags involved in the research and development–productivity nexus. First, a particular research and development project may take more than a year to complete. Second, when complete and if successful, it may still take some time before a decision is made to use it or produce it. Once an innovation decision is made, it may show up in the firm's revenue stream only with another lag. If a process innovation, it may be introduced gradually, affecting only parts of the firm's cost structure in the beginning. If a product innovation, it may take time for consumers to find out about it and to accept it. At the aggregate firm level there are many such projects that have started at different dates and are in different stages of fruition. More generally, the convolution and aggregation of many such lag structures should lead to a rather flat but somewhat bell-shaped lag structure connecting total firm research and development to its subsequent productivity (Griliches, 1967).

There is also the issue of depreciation or obsolescence of this capital. If one distinguishes between the firm-specific knowledge capital and the general state of knowledge in the industry as a whole, then at least as far as the first is concerned, it is quite clear that its earning capacity erodes over time, both because better products and processes become available and because its own knowledge begins to lose its specificity (it leaks to other firms in the industry). Thus, from the private point of view there is depreciation of this capital, probably at a rather high rate (see Pakes and Schankerman (1978) for some scattered evidence on this point). This fact, together with the rather short-term nature of much commercial research and development (see the discussion by Mansfield and others in Williams (1973, pp. 87–90)), would imply a research and development lag structure that peaks somewhere between three to five years earlier and then declines rather rapidly, with little of the original research and development product remaining "private" past ten years or so.

The question of depreciation is much more complicated for social research and development capital measures at the industry or national level. The fact that private knowledge loses its privacy and hence its value is a private loss, not a social one. Nevertheless, there is likely also to be some depreciation in

7. Note that nothing tangible corresponds to this notion of R&D "capital." It is just an alternative to expensing R&D as a current input. Ideally it would equal the value of the firm's "know how" and should be related to the excess of market value over appropriately depreciated and deflated book value. See Ben-Zion (1977) for an attempt at measurement in this vein.

8. I shall ignore here the possibility of using patents or publication counts as indices of research and development output. They are only available for a limited range of sectors and industries.

social knowledge. Some of the new products may make the knowledge about older products and processes redundant. Alternatively, much of what is measured as investment in research and development is not "net" investment from a social point of view, since it replaces already existing knowledge. If 50 percent more investment is needed to produce an alternative product that is 10 percent superior to the older version, it may be pursued from the private point of view, since with about half the investment one can appropriate most of the rent collected by the inventors of the earlier version. But from the social point of view, if research and development capital were measured by "output," it would only go up by 10 percent rather than 50. We can say then, equivalently, that either only a fraction of current research and development flow is to be thought of as a net addition to the social stock of knowledge capital or that some fraction of the preexisting stock of this capital is replaced (depreciated) annually. The real problem here is our lack of information about the possible rates of such depreciation. The only thing one might be willing to say is that one would expect such social rates of depreciation to be lower than the private ones.

The problem is even more difficult as far as the measurement of public R&D "capital" (such as is generated by NASA, the Department of Agriculture, universities, and other similar organizations) is concerned. Components of this type of R&D capital that contribute to productivity growth in the industrial sector are likely to be subject to both lengthier lags and lower depreciation rates. It is hard to see, however, where one could get relevant evidence on this topic.[9]

The last major issue is that of "spillovers," the effect of "outside" knowledge capital—outside the firm or industry in question—on the within-industry productivity. The level of productivity achieved by one firm or industry depends not only on its own research efforts but also on the level of the pool of general knowledge accessible to it. Looking at a cross section of firms within a particular industry, one will not be able to distinguish such effects. If the pools of knowledge differ for different industries or areas, some of it could be deduced from interindustry comparisons over time and space. Moreover, the productivity of own research may be affected by the size of the pool or pools it can draw upon. This would lead to the formulation of models allowing for an interaction between the size of individual and aggregate research and development effort.[10]

A simple model of such within-industry spillover effects is given by

$$Y_i = BX_i^{1-\gamma}K_i^{\gamma}K_a^{\mu}$$

9. Some work has been done on the contribution of publicly supported research in agriculture to productivity. See Griliches (1973) for a review of this literature.

10. See Evenson and Kislev (1975, Chapter 4) for an example of such a "borrowing" function involving scientific publications.

where Y_i is the output of the ith firm which depends on its index of conventional inputs X_i, its specific knowledge capital K_i, and on the state of aggregate knowledge in this industry K_a. Note that I have assumed constant returns in the firm's own inputs, X_i and K_i. This simplifies the example greatly. Now let us assume that: (1) the aggregate level of knowledge capital $K_a = \Sigma_i K_i$ is simply the sum of all specific firm research and development capital levels; and (2) own resources are allocated optimally and all firms in the industry face the same relative factor prices. Then we know that the individual K_i to X_i ratios will be given by

$$\frac{K_i}{X_i} = \frac{\gamma}{1-\gamma}\frac{P_x}{P_k} = r,$$

where P_x and P_k are the prices of X and K, respectively, and r, the K/X ratio, does not depend on i. We can aggregate our individual production functions:

$$\sum_i Y_i = \sum_i BX_i(K_i/X_i)^\gamma K_a^\mu = \sum_i BX_i r^\gamma K_a^\mu = Br^\gamma K_a^\mu \sum_i X.$$

Now since the K_i/X_i ratios are equal to r, so also is $\Sigma K_i/\Sigma X_i$, which we can substitute back into this equation, yielding:

$$\sum_i Y_i = B\left(\sum_i K_i \Big/ \sum_i X_i\right)^\gamma K_a^\mu \sum_i X_i = B\left(\sum_i X\right)^{1-\gamma} K_a^{\mu+\gamma},$$

where, by assumption, $\Sigma_i K_i = K_a$. Thus, we get an aggregate production function with the coefficient of aggregate knowledge capital being higher ($\gamma + \mu$) than at the micro level (γ only), since at the aggregate level it reflects not only the private but also the social returns to research and development.[11] The above formula provides a framework for reconciling micro and macro results in this area.[12]

Of course, this formula is rather simplistic and is based on a whole string of untenable assumptions, the major ones being: the assumption of constant returns to scale with respect to X_i and K_i and the assumption of common factor prices for all firms within an industry. These assumptions could be relaxed. This would add a number of "mix" terms to the equation, indicating how aggregate productivity would shift if the share of, say, the larger firms, were to increase (as in the case of economies of scale). If the mix of firms stays stable, such refinements do not add much. In any case, the above formula was presented to suggest the nature of a class of results that one could get from such assumptions, rather than to set out a final model of such phenomena.

11. The effect of aggregated private "knowledge" is K^γ. The total spillover effect of this knowledge is K^μ. Since we have assumed that all private knowledge spills over to some extent and have measured it as $K_a = \Sigma K_i$, the total effect of all private knowledge at the aggregate level is given by $\gamma + \mu$, rather than just by γ.

12. The first time I saw this type of formulation was in an unpublished note by Grunfeld and Levhari (1962).

The problem is much more complicated when we realize that we do not deal with one closed industry, but with a whole array of firms and industries which "borrow" different amounts of knowledge from different sources according to their economic and technological distance from them. The concept of such a "distance" is very hard to define empirically. If we return to our previous example and now interpret the index i as referring to industries rather than firms, it makes little sense to define K_a as $\sum_i K_i$. Rather

$$K_{a_i} = \sum_j w_{ij} K_j$$

is the amount of aggregate knowledge borrowed by the ith industry from all available sources. K_j measures the levels available in these sources, while w_{ij}, the "weighting" function, can be interpreted as the effective fraction of knowledge in j borrowed by industry i. Presumably w_{ij} becomes smaller as the "distance," in some sense, between i and j increases. Thus we need an additional distributed (lag) over space function to construct a measure of the stock of borrowed knowledge.

On what should such a weighting function be based? There have been earlier suggestions based on "vertical borrowing" concepts: Brown and Conrad (1967) used the input-output table to measure the "closeness" of industries proportional to their purchases from each other, while Terleckyj (1974) used the capital and intermediate inputs purchases matrix weights, assuming that "borrowed" research and development is embodied in purchased inputs. Raines (1968) used the "horizontal" product field classification of NSF to include inputs to an industry's research and development and also the research and development expenditures of other industries which were reported as belonging to its product field.

Actually, there are two distinct notions of research and development "spillovers" here which are often confused in the literature. The first one, research and development intensive inputs purchased from other industries at less than their full "quality" price, is related to issues in the measurement of capital equipment and materials and their prices and is not really a case of pure knowledge spillover. If capital equipment purchase price indices reflected fully the improvements in their quality, i.e., were based on hedonic calculations (Griliches, 1971), there would be no need to deal with this problem. As currently measured, however, total factor productivity in industry i is affected not only by its own research and development but also by productivity improvements in industry j to the extent of its purchases from that industry and to the extent that the improvements in j have not been appropriated by its producers and/or have not been incorporated in the official price indices of that (i) industry by the relevant statistical agencies. The use of purchase flow weighted research and development measures assumes that social returns in industry j are proportional to its research and development investment levels and that the amount

of such returns transferred to industry *i* is proportional to its purchases (or stocks) from industry *j*.

A good example of such productivity transfers would be the computer industry. It has had a tremendous real productivity growth, most of it unmeasured in its official indices, and most of it unappropriated within the industry itself (because of rather intensive competitive pressures). Different industries have benefited differentially from it, depending on their rate of computer purchases. One way of accounting for it would be to adjust upward the relevant capital equipment figures by their computer content. The alternative is to "import" the computer industry's research and development in proportion to the purchases from it.

But these are not real knowledge spillovers. They are just consequences of conventional measurement problems. True spillovers are the ideas borrowed by the research teams of industry *i* from the research results of industry *j*. It is not clear that this kind of borrowing is particularly related to input purchase flows. The photographic equipment industry and the scientific instruments industry may not buy much from each other but may be, in a sense, working on similar things and hence benefiting much from each other's research. One could argue that this is what the SIC classification is for. Presumably, the usefulness of somebody else's research to you is highest if he is in the same four-digit SIC classification as you are; it is still high if he is in the same three-digit industry group as you are; and, while lower than before, the results of the research by a firm in your own two-digit classification (but not three-digit) is likely to be more valuable to you than the average results of research outside of it.[13] The problem arises when we want to extend this scale across the other two-digit industries. Here there is no natural order of closeness (e.g., is "leather" closer to "food" or to "textiles"?).

The following alternatives appear reasonable and worth trying: (1) grouping three-digit SIC categories into clusters based on a priori notions about the extent of commonality in their technological and scientific base; (2) using the NSF's applied research and development product field by industry table to induce a distance metric, on the assumption that if an industry is doing research and development on some other industry's products, it is in some sense closer to it technologically than if it does not; (3) using company industrial diversification data from the Census of Enterprises to compute an alternative technological closeness metric; and (4) using information on rates of cross referencing of patents across product fields to infer the technological distance between them.[14] In each of these cases one will have to assume some simple weighting

13. The situation is further complicated by the fact that the major R&D performers are conglomerates, spanning several SIC four-, three-, and even two-digit classifications. The NSF's applied R&D product field data help here a little but not enough. Complete within-firm product line breakdowns along the lines suggested by the FTC would be very useful in this context.

14. Work along lines (2) and (4) is being pursued by Schankerman (1979) and myself.

functions (e.g., the influence declining exponentially with the particular concept of distance) or group the data into a few categories: immediate neighborhood, related fields, and the rest. The available data will not support very refined approaches. There just are not enough degrees of freedom or independent variations in such productivity and research and development series to allow us to estimate very fancy distributed lag schemes over both time and all other industries.

The alternative to the search for such a concept of technological closeness or distance is to use the research investments of different industries as separate variables. But that is not really feasible. At best we would have about 20 years of data for each of about 20 industries. We cannot include 20 separate R&D variables in each of the industry equations; there simply will not be enough degrees of freedom there. We therefore have to aggregate somehow and that is what the idea of technological distance is for: to tell us how to weight the different research series and collapse them into one or a few variables so that we can estimate and assess the empirical importance of R&D spillovers. With such estimates we would compute not only the return to particular R&D in its "own" industry but also the total returns to R&D including the spillovers beyond its borders.

In trying to construct a research and development capital stock measure at the firm or industry level we face, thus, two major tasks: deciding on the appropriate lag structure and finding the right weights for "outside" research and development to represent borrowed knowledge and spillovers. Unfortunately, we have not enough theoretical or factual knowledge to guide us in these tasks. They will have to be solved empirically. The available data base, however, does not inspire much confidence in our ability to do so.

Before we conclude this section, we have to mention one other garden variety measurement problem: the lack of a research and development deflator. There is no official research and development deflator index available currently. NSF is using the implicit GNP deflator which is not so high-skill labor intensive as would be appropriate for a research and development input price index. Battelle publishes a research and development "cost" index, based on Milton (1972), but that appears to be an index of total cost (including other inputs) per scientist. It does not hold either the composition of the research and development labor force or the quantity of other inputs purchased constant. In short, it is not a price index. S. Jaffe reviewed the problem in NSF (1972) and suggested the use of a weighted index of labor compensation and the implicit price index in the nonfinancial corporate sector. That seems the best that one can do at the moment from secondary sources, but there is no reason why a better research and development input price index could not be constructed on the basis of primary sources. The NSF surveys ask for quite a bit of detail on the composition of research and development expenditures while the Battelle (previously Los Alamos) research and development scientists' compensation surveys have a great deal of data on the composition and compensation of the

research and development labor force. This detail could be used to construct a whole set of separate research and development deflators for each of the major research and development performing industries.

2.5 Econometric Problems

Besides the measurement problems discussed above, there are two other serious econometric problems facing the analyst in this area: multicollinearity and simultaneity. Although both are common, "garden variety" econometric problems, each has serious consequences. The problem of multicollinearity arises from the fact that many of the series we are interested in moved very much together over the period of observation. That being the case, it is then difficult (often impossible) to infer their separate contributions with any precision. There are no cheap solutions to this problem. It requires either less-collinear data, more prior information, or a reduction in the aspiration level of the questions to be asked of the data.

Micro-time-series data at the individual firm or establishment level are probably the best way to reduce the multicollinearity that plagues such series at the industry- or economy-wide level. There is much more variability in the R&D histories of particular firms than in the R&D histories of the corresponding industries. This variability can be used to answer questions about the R&D lag structure, the relative effects of government- versus privately financed R&D, or of basic versus applied research, and so on. Such "gains," however, come only at a serious cost. As noted above, at the micro level one can ask only questions about private returns to R&D. Very little of their social returns is detectable at this level.[15]

The analysis of industry-level data can be facilitated somewhat by the imposition of reasonable a priori restrictions. Thus, the use of total factor productivity measures as dependent variables already imposes on the data a set of implicit assumptions about the functional form of the production function and about specific values of its parameters; for example, it sets the output elasticities equal to their observed factor shares.

A similar problem will arise in trying to determine the exact shape of the R&D lag structure. To do that one has to have many years of R&D data and treat each of these years as a separate variable. But, in fact, R&D expenditures are highly correlated from year to year. Thus, it is unlikely that one can estimate their separate contribution with any precision (see Griliches (1967) for a more detailed discussion of such problems). Here it is probably best *to assume*

15. This does not mean that one cannot study the social returns to R&D at the micro level, but that would require a much more detailed data base with information on the magnitude of the actual technological breakthroughs and estimates of the relevant demand elasticities. See Mansfield et al. (1977) for details of such an approach. It goes far beyond the production function framework considered here. It is based, however, largely on confidential private information and is thus difficult to reproduce, extend, and evaluate.

a functional form for the lag distribution on the basis of prior knowledge and general considerations and not to expect the data to answer such fine questions. That is, a "solution" to the multicollinearity problem is a moderation of our demands on the data—our desires have to be kept within the bounds of our means.

The simultaneity problem refers to the possible confusion in causality: future output and its profitability depend on past R&D, while R&D, in turn, depends on both past output and the expectation about its future. With long time series and detailed lag assumptions one may be able to analyze a recursive equations system with current output depending on past R&D, and past R&D depending on past rather than current output. In cross sectional data with only a few observations per firm, it is much harder to make such distinctions, particularly since current expectations about the future are based on current and past data.

It may be useful to outline the problem in somewhat more detail. In the context of simple static profit maximization a complete production and input choice model would consist of the production function,

$$y = \alpha c + \beta l + \gamma k + u,$$

assuming a Cobb-Douglas functional form and denoting the logarithms of corresponding variables by lower case letters, and marginal revenue product equals marginal cost conditions of the form,

$$y - c = a_1 + \pi_1 + v_1,$$

$$y - l = a_2 + \pi_2 + v_2,$$

$$y - k = a_3 + \pi_3 + v_3,$$

where the π's are the respective real factor prices of the various inputs, the v's are random errors in factor demand conditions, and a's are constants which depend on the coefficients of the production function and the product demand and factor supply elasticities (in the case of imperfect competition). (See Nerlove (1965) for more detail on this type of model.) We can solve this system of equations for k in terms of the π's, v's, and u, and write the resulting "reduced form" equation as

$$k = D\{-\alpha(\pi_1 + v_1) - \beta(\pi_2 + v_2) - (1 - \alpha - \beta)(\pi_3 + v_3) + u\},$$

where

$$D = 1/(1 - \alpha - \beta - \gamma).$$

Under these assumptions, it is clear that k is a function of u and that a simple OLS estimate of the production function would result in a biased estimate of γ, since the "independent" variable of interest (k) is in fact correlated with the disturbance in the same equation.

There are several "solutions" to the simultaneity problem. First, if one has good series on the π's, the real factor costs of the various inputs, one could use them as instrumental variables for the estimation of the production function. Unfortunately, in the R&D context one is unlikely to have good factor price series.[16] Even if one had the prices, they are likely to be highly collinear over time. Second, if one is willing to assume that disturbances in the input choice equations, the v's, are independent of the production function disturbance u, one could use input-output ratios, say $k - y = \pi_3 + v_3$, as instruments, since by hypothesis they are independent of u. This is known in the literature as the "indirect least squares" method. The implicit assumption of certainty about the future underlying such a static model makes little sense, however, in the R&D context. What is maximized here is the present value of all future profits, and the relevant output concept is an expected one and not the current one, especially if current output (and demand) is subject to special and transitory circumstances. In such a context not all of the "u" is transmitted to the factor demand equations (see Mundlak and Hoch (1965) and Mundlak (1963) for a more detailed exposition of such "partial transmission" models) and the indirect least squares method alluded to above is not consistent anymore. One needs a way of estimating the ratio of the permanent (transmitted) variance in u before one can use such methods again. Third, if both time series and cross sectional data are available and one is willing to assume a simple permanent-transitory transmission model: $u = \mu + t$, where μ is the permanent component which affects input demand choice while t, the transitory component, does not, then consistent estimates can be had from the within-firm covariance estimates of the parameter. This is equivalent to allowing a separate constant term (dummy variable) for each firm, which would absorb the μ term in it. Unfortunately, such data sets are rare. Moreover, the covariance approach may exacerbate other problems, such as errors in the variables, which may also afflict these kinds of data. Fourth, one may be able to find other "indicator" variables which may be related to the permanent components of the variables of interest and hence may help to solve the identification problem in such models.

For example, data may be available on the stock market value of the firm S. If this measure is proportional to the permanent output of the firm, then taking logarithms and ignoring constants, we can write

$$s = y - t + \varepsilon,$$

where ε is a random term uncorrelated with either μ or t. Now one can use input to value of the firm ratios ($k - s$, etc.) as components of the indirect

16. First, there are no published R&D deflators at the two-digit SIC level; second, if they were available, they would still be very highly correlated with the cost of labor and cost of capital indices, which are likely to be major ingredients of such indices. What we will not have are changes in "real" R&D costs in a field caused by various technological breakthroughs.

least squares procedure alluded to above.[17] More generally, a more explicit formulation of the expectation formation mechanism, together with the use of additional indicator variables such as the value of the firm, its physical capital investment policy, and the number of patents it has received, may allow us to formulate and estimate a more general "unobservables" type model (Griliches, 1974).

To recapitulate, the possible mutual dependency of R&D investments on past and future expected output requires careful attention to model formulation and specification and better and more detailed data to support the application of more sophisticated estimation techniques. Without a careful consideration of such issues, one may wind up reporting something as an estimate of the effect of R&D on output which may be mostly a reflection of the effect of output on R&D rather than vice versa.

2.6 Special Problems

There are several special questions that one would like to have answered in this context:

What are the relative returns to basic versus applied research?

Are the returns to government-financed R&D similar to those of company-financed R&D?

What are the "spillover" effects of government-financed R&D in sectors where the direct effects are almost unmeasurable (e.g., NASA R&D spending)?

How can one distinguish between economies of scale and R&D-induced productivity growth?

It would be interesting to distinguish between returns to basic and applied research, especially at the aggregate level. At the micro level, if firms allocate their resources rationally, the *private* ex post rates of return on different types of research should be about the same (except, possibly for some differences due to their differential riskiness) and we should not be able to distinguish among them. At the industry level one could conceivably find significant differences if, say, the appropriability of the results of basic research is more difficult than the appropriability of applied research. Also, the definition of "basic" implies a potentially wider range of applicability and an a priori higher externality component in its results. Thus, one might expect that the discrepancy between social and private returns would be higher, other things equal, in industries with a higher basic to total R&D ratio. Unfortunately, it may prove very difficult to isolate such effects. Basic research accounts for only about 3

17. The actual model will have to be more complicated, since the value of the firm depends not on expected output but on expected profits, hence also indirectly on planned input expenditures. See Ben-Zion (1977) for an attempt to use the market value of the firm to infer the depreciation rate of R&D expenditures.

percent of the total R&D expenditures in industry. Also, the lag structure of its effects is likely to be flatter and more variable, and hence more difficult to estimate. Therefore, one should not be too surprised if attempts to estimate the separate contribution of basic research fail. Nor should this be interpreted as implying the unimportance of basic research. It may only reflect the inability of our data to reveal such effects.[18]

There is much more basic research done in universities and government institutes than in industry, but it is almost impossible to assess its independent effect on productivity. At the industry level all industries are to some extent beneficiaries of the same research effort. There are no data, however, (except for agriculture) which could connect different expenditures on basic research to specific industries. At best there is a breakdown by fields of science but there is little quantitative information on the differential importance of "scientific fields" for different industries. One could try to construct differentially weighted "relevant" basic research series for different industries, but the results are likely to be quite collinear, and this approach does not appear to be too promising. At the aggregate level one has only one time series to explain and the basic research time series is likely to be no different from many other trending aggregate series. Without much more detail on the structure of basic research and on how it is used (if at all) in industry, it is unrealistic to expect to uncover its effects by econometric methods.[19]

Similar issues arise in the attempt to distinguish between the effects of publicly and privately financed R&D. Within any firm one would not expect to find much of a difference. A dollar is a dollar irrespective of source (unless there are explicit expenditure and accounting rules connected with the use of federal R&D money which lead to inefficiencies). But a concentration of federally supported R&D expenditures in one area may lead to an overall decline in the rate of return to all R&D there. This may explain the difference between the results of Griliches (1975a), who found no differential effect of federal versus private company R&D dollars on the levels and rates of growth of total factor productivity at the *firm* level, and Terleckyj (1975), who found a significant discount of federal R&D dollars as far as their effect on productivity at the two-digit aggregated industry level is concerned.[20] Within an industry, at the firm level there should not be much of a difference between the effectiveness of different dollars, while at the aggregate level different rates of federal

18. In a recent still unpublished paper Mansfield (1980) does find a significant separate impact of basic R&D expenditures on industry productivity growth rates. This indicates that the paragraph above may be too pessimistic.

19. There is always the route of detailed historical case studies but they are subject to problems of selection bias and incomplete coverage.

20. Both studies are based on cross sections of growth rates. Terleckyj covers the years 1948–1966 while Griliches uses only 1957–1965. Terleckyj uses 1958 R&D intensity data while Griliches uses 1957–1965 R&D levels and growth rates. Griliches also finds that the estimated rate of return is lower in industries with high levels of federal R&D support (e.g., electrical equipment, aircraft, and missiles).

R&D investment must imply differences in ex post private and social returns, unless these expenditures are directed primarily at areas where there is private underinvestment in R&D (a large ex ante gap between social and private returns) and concentrated on the less appropriable portions of it.

It might be interesting to look at this more from the "R&D investment function" side. Does federal R&D substitute for or complement private R&D investments? This would depend, in part, on the specific conditions and rules which accompany such federally financed expenditures. Holding federal funds constant, a firm that invests more of its own money than another presumably faces better investment opportunities or is more certain of appropriating a larger fraction of the total return.

Both in this case and in the previously discussed basic versus applied research context, questions are raised as to whether the different types of research are "substitutes" or "complements." Technically, this is a question about the functional form of the production function. In the Cobb-Douglas case all inputs are "complements." This is true of inputs that are considered separately (e.g., labor and capital, and basic and applied research if they have been entered separately with different exponents). When inputs are just added together into one aggregate total R&D measure, the implicit assumption is made that they are infinitely substitutable at the dollar-per-dollar rate. It is rather difficult to investigate questions of functional form with the usual kinds of data (see Griliches and Ringstad (1971) for a more detailed discussion of such difficulties). If one knew more about the differential price (real capital costs) of federal versus company R&D investments, one could investigate this problem through an analysis of input demand equations. Functional form differences are difficult to detect at the production function level, since the different curvature parameters are of second-order importance there. But at the input demand structure level these same parameters (such as the elasticity of substitution) have a first-order effect and thus may be easier to estimate. This, unfortunately, requires us to have the "right" prices for the different types of input—a feasible but very difficult task.

The third question, the detection of the "spillovers" from governmental R&D, was discussed earlier in the context of defining a notion of "borrowed" R&D. A model was outlined which "weighed" the contribution of the jth industry's R&D to the stock of borrowed knowledge in industry i in proportion to some notion of the relevant technological distance between them. Besides the "borrowability" of a particular piece of knowledge of industry i from industry j, there is also the question of differential social productivity of R&D in different industries. There may be effectively more to borrow from some industries than others, per dollar of their R&D expenditures. We would like to be able to test hypotheses of the form "NASA R&D was more productive in the sense of having higher spillovers into the civilian economy than, say, DOD-financed R&D." Statistically this would consist of separating our measures of the total stock of borrowed R&D into components (NASA R&D, DOD R&D,

other federal R&D, industry R&D, etc.) and asking whether they have consistently different effects on the productivity growth of other industries. It is feasible to try such an approach. Whether the available data can sustain such an effort is unclear, however.

The last question, the role of economies of scale in all of this, refers really to two somewhat distinct phenomena. The usual measures of total factor productivity are very much affected by short-term fluctuations in capacity utilization. To get a correct measure of the shift in technological opportunities of an economy (or industry), some adjustments or allowances have to be made for it. Also, there are industries where the longer-run trends in productivity are affected by changes in the scale of consumption. For example, part of the productivity growth in the telephone communications industry is attributable to the rise in calls per subscriber. Similarly, in the electric utility industry, spreading the peak load may lead to significant gains in measured productivity. Some of these changes may be related to earlier R&D expenditures but most are not. The data should be adjusted, if possible, so as not to confuse the issues.

A second type of interaction of R&D with scale occurs in industries where the technological results have been biased towards larger-scale enterprises. Here the payoff to R&D may be delayed if the market does not grow so quickly as anticipated. Moreover, the growth in the market and the growth in the number of "large" customers may, in turn, affect the direction of the R&D effort itself.

There is no satisfactory theoretical framework yet for discussing both the R&D decision process and the returns to it in a world of increasing returns to scale. These few cryptic remarks can at best serve as a placemark for future research. To pursue this topic further here would take us, however, too far afield.

2.7 Previous Work

In this section I shall review briefly several recent studies of R&D returns which fall within the framework of this paper in the sense that they use both the production function framework and econometric methodology to derive their results.

The study which deals with the most aggregate level is the Chase Econometrics Associates' (1975) report, "The Economic Impact of NASA R&D Spending." It is a time-series study (14 to 18 years) of aggregate total factor productivity growth as it relates to distributed lag terms of past NASA R&D expenditures and "other" R&D expenditures. It finds extremely high returns to NASA R&D (on the order of 40 percent per year in perpetuity and more than double the returns to all other R&D). There are several difficulties with this study (see Griliches (1975b) for detailed comments) and it is difficult to accept its results at face value. First, it looks for returns to NASA R&D at the total GNP level, rather than in the sectors where one could have hoped to measure

the effect of the spillover of such R&D nontautologically. Second, there is a relatively short time series available to the authors for this purpose. Finally, the results are brittle with respect to the various assumptions and data adjustments made. Minor changes in procedure and variable definitions lead to substantial changes in the results. The study is interesting in that it disaggregates total R&D into two components (NASA vs. other), uses a distributed lag formulation to capture the delays in the effects of R&D over time and makes various adjustments for changes in capacity utilization over time. Unfortunately, the shortness of the time series and the use of a problematic dependent variable make their results both imprecise and difficult to interpret.[21] In addition, they impose the same lag structure on both NASA R&D and all other R&D. But as I have argued above, only the spillovers of NASA R&D should show up in such aggregate productivity figures, and surely they must have a different lag structure from most of the other R&D which is short term and has relatively quick direct returns.

At the aggregate two-digit industry level, there is a series of studies by Terleckyj (1974, 1975, and 1977) which relate a cross section of estimated rates of total factor productivity growth (1948–1966) for 20 to 33 two-digit industries to a number of R&D intensity ratios and other variables. Two findings stand out in these studies: (1) variables that "work" (at this level) are functions of private R&D investments and not of federally financed ones (as contrasted with the CEA study of the effects of NASA R&D); and (2) an average of R&D in "other" industries, weighted in proportion to the industry's purchases from them, has a significant influence on productivity growth. The significance of the "own" R&D intensity ratio in such studies is not surprising, and was observed in earlier studies by others (e.g., see Griliches (1973)). The major new finding is the large and statistically significant coefficient of his "borrowed" R&D measure. Terleckyj does not distinguish, however, between whether this variable works because of the mismeasurement of the inputs bought by the respective industries and whether the variable works because of the use of the input-output matrix weights as measures of technological "distance" between industries. If the first reason is true, the results are less interesting than they appear at first sight. For example, it is quite likely that his productivity measures (taken from Kendrick (1973)) underestimate the contribution of equipment capital and hence overestimate productivity growth in industries which had a relatively high investment in equipment. This occurs because the Kendrick measures weight capital by stock rather than flow weights and overdeflate them (cf. Jorgenson and Griliches (1972) for additional discussion of these and related issues). Since equipment industries are relatively R&D intensive, the resultant variable may be nothing more than a correction to the originally er-

21. They use total GNP instead of the more relevant private civilian economy concept. Also their use of the Council of Economic Advisers' "potential" output series is debatable. The results do not survive the switch to Denison's potential output or Jorgenson's real output series.

roneously constructed productivity series. If the measure is supposed to represent technological distance and capture the pure spillover effects of R&D elsewhere, its logic would require the use not of the direct input-output matrix weights but of the corresponding "total" weights, which would take into account the flows of information both directly from i to j and indirectly from i through h and g to j, and so forth. Here, too, the sample is small and the evidence is brittle.[22]

At the micro (firm) level Griliches (1975a) related total factor productivity growth measures for 883 large U.S. companies during the 1957–1965 period to various measures of the growth in R&D capital and found a sizeable and significant effect of R&D on productivity growth. This study is also subject to serious limitations. First and most importantly, because of the nature of the data base (individual company reports), only questions about *private* returns to R&D could be asked. Second, because of the shortness of the available time series, no effort was made to investigate the lag structure of R&D effects. And third, only rudimentary attention was paid to the simultaneity problem.

2.8 Some Conclusions and Recommendations

One of the major points of this paper is the plea for realism as to what the production function approach can and cannot accomplish. Given good data, it can tell us something about average returns to R&D investments in the past and whether they appear to be changing over time. It may be able to indicate industries where returns have been especially high or low, but it will not be able to tell us whether a particular proposed R&D project is a good bet or not.

Given the kind of data we have and are likely to have in the near future, there are questions that one is unlikely to be able to answer from such data: e.g., questions about the exact time structure of the effects of R&D on productivity or the role of "science" and "basic" research in all of this. The level of "science" in the aggregate changes only very slowly, and we have no good way of assessing its differential impact on different industries. Unless a cause (stimulus) has varied much in its intensity over time or in its effects over industries, one is unlikely to be able to isolate it reliably by standard econometric techniques. In studying the contribution of NASA's R&D to the growth of the U.S. economy, we are faced with the fact that at the aggregate level we have only one cycle to work with: a rapid and continuous rise in NASA's R&D expenditures to the mid-1960s and then a more or less continuous decline to the mid-1970s. Given reasonably long and variable lags in the effects of such investments, we have, at best, only two observations, one up and one down. Given all else that is happening at the aggregate level at the same time, we are unlikely to

22. A recent study by Mathtech (Agnew and Wise, 1978) using similar data but annual observations (instead of growth rates) is unable to detect a significant effect of such input-output table weighted borrowed R&D measures.

be able to estimate reliably the contribution of a cyclical factor whose one cycle roughly corresponds to the time range of the available data.

This does not mean that the topic cannot be studied at all, only that the production function framework is not the suitable avenue for it. One could presumably identify a list of specific *products* of NASA research and trace out their subsequent impact on productivity in other industries. This, however, would require a much more micro-oriented approach.

At the moment, the lack of relevant data and the conceptual poverty of our models are the major impediments to progress in this area. As far as data are concerned, we have needs at both the macro (industry) and micro levels. At the macro (industry) level a consistent set of total factor productivity measures corresponding to the 24 two- and three-digit industry detail given in the NSF's R&D publications (or the 32 industries corresponding to the applied R&D detail) is needed. These productivity measures should use consistent output and capital deflation methods, be clear about the treatment of sales to government, if possible exclude the R&D expenses and labor force from its measures of current inputs, and be up to date (i.e., go through 1976 or 1977 rather than end in the late 1960s as most of the studies have done). This will require a review and construction of industry price indices which should be consistent, as much as possible, in their treatment of quality change.

A price deflator is needed for the conversion of the R&D expenditures to "real" magnitudes. The earlier work of Jaffe (NSF, 1972) can and should be updated and extended.

The data on R&D collected by the NSF could also be improved. Industry codes use 1967 definitions which lack detail in some high technology areas and may be also out of date by now. Conversion to 1972 codes will have to be made. Coverage of subcontractors of R&D is almost nonexistent. Also, the basic unit surveyed is a firm, not an establishment. Many of the large R&D performing companies are conglomerates. An effort should be initiated to collect R&D data either at the establishment level (which has a more clearly defined industrial identity) or at the "product line" level, as has been recently suggested by the FTC. More effort should also be invested in improving the "Applied R&D by Product Field" estimates.[23]

At the micro (company) level the Census-Griliches-NSF Large-Company panel (1957–1965) should be reconstructed and updated to 1975. Efforts in this direction are now in progress.

Because of confidentiality problems and difficulties in access to census-based data, another "open" firm panel should be created from the published (mainly SEC) record. See, for example, the recent compilations in *Business Week* on this topic. This data set could be matched to other data for the same

23. NSF-Census should also write up and publish the results of their "Response Analysis" efforts of the Industrial R&D Survey rather than just allude to them (see p. 20 of NSF 1977).

firms available in such sources as the Compustat tape, and the Patent Office records on patents applied for and granted by product field.

At the conceptual level we need more research on two very difficult topics: (1) how to measure the public product (and hence also the returns to R&D) in such important sectors as space, defense, and health; and (2) how to conceptualize and estimate technological distance between firms and industries and the associated notions of externalities and spillovers in research.

The first really corresponds to efforts to improve and expand the national income accounts towards more welfare-oriented measures. There do appear to be data and there is no insuperable conceptual problem in constructing a more relevant health sector accounting scheme. Whether something comparable could be done in space and defense is not clear at the moment. Suggestions for possible attacks on the second question, measuring the technological distance between firms and industries, have been advanced in the body of this paper and will not be repeated here. They too appear to be feasible.

At the empirical-econometric level we need studies that compare and carefully contrast results at different levels of aggregation in an attempt to detect and measure the importance of social returns to R&D. We also need studies that pay much more attention to the estimation of the various lag structures between R&D expenditures and productivity growth, and at the same time recognize the problem of simultaneity and tackle it in a nontrivial manner. In addition, with patent data now available for 55 SIC product fields, a serious effort should be mounted to determine whether they actually measure something of interest and exactly what it may be. Especially, one should be able to tell whether there is any connection between the rate of patent activity and subsequent productivity growth in the various industries.

This is a large order, but we are talking about a major source of economic growth and about one of the few variables (R&D) which public policy can affect in the future (and has affected in the past). Hence it is of the utmost importance that we improve our knowledge in this area. It will not be easy, but it can be done, and it should not take a very large increase in research resources to accomplish at least some of it.

References

Agnew, C. E. and Wise, D. E. "The Impact of R&D on Productivity: A Preliminary Report." Paper presented at the Southern Economic Association Meetings. Princeton: Mathtech, Inc., 1978.

Ben-Zion, U. "The Investment Aspect of Nonproduction Expenditures: An Empirical Test." Unpublished manuscript, 1977.

Berndt, E. and Christensen, L. "Testing for a Consistent Aggregate Index of Labor Input." *The American Economic Review,* Vol. 44 (1974), pp. 391–404.

Brown, M. and Conrad, A. "The Influence of Research on CES Production Relations" in M. Brown, ed., *The Theory and Empirical Analysis of Production*, Studies in Income and Wealth, Vol. 3, New York: Columbia University Press for NBER, 1967, pp. 275–340.

Chase Econometrics Associates Inc. "The Economic Impact of NASA R&D Spending." Bala Cynwyd, Pa.: 1975.

Chow, G. "Technological Change and the Demand for Computers." *The American Economic Review*, Vol. 57 (1967), pp. 1117–1130.

Diewert, E. "Aggregation Problems in the Measurement of Capital." University of British Columbia Discussion Paper No. 77–09, 1977.

Evenson, R. E. and Kislev, Y. *Agricultural Research and Productivity*. New Haven: Yale University Press, 1975.

Fisher, F. M. "The Existence of Aggregate Production Functions." *Econometrica*, Vol. 37, No. 4 (1969), pp. 553–577.

Gordon, R. J. "The Disappearance of Productivity Change." Harvard ED Report No. 105. Mimeographed, 1968.

Griliches, Z. "Research Cost and Social Returns: Hybrid Corn and Related Innovations." *Journal of Political Economy* Vol. 66, No. 5 (1958), pp. 419–431.

———. "The Sources of Measured Productivity Growth, U.S. Agriculture, 1940–1960." *Journal of Political Economy* (August 1963).

———. "Distributed Lags: A Survey." *Econometrica* (January 1967).

———. "Research Expenditures and Growth Accounting" in B. R. Williams, ed., *Science and Technology in Economic Growth*, London: MacMillan, 1973, pp. 59–95.

———. "Errors in Variables and Other Unobservables." *Econometrica*, Vol. 42, No. 6 (1974), pp. 971–998.

———. "Returns to Research and Development Expenditures in the Private Sector" (1975a) in J. W. Kendrick and B. Vaccara, eds., *New Developments in Productivity Measurement*, Studies in Income and Wealth, Vol. 41, forthcoming. [Reprinted as chap. 3 in this volume.]

———. "Comments on CEA: The Economic Impact of NASA R&D Spending." Mimeographed, 1975b.

———. "Economic Problems of Measuring Returns to Research" in Y. Elkana et al., eds., *Toward a Metric of Science: The Advent of Science Indicators*, New York: John Wiley & Sons, Inc., 1977.

———, ed. *Price Indices and Quality Change*. Cambridge: Harvard University Press, 1971.

Griliches, Z. and Ringstad, V. *Economies of Scale and the Form of the Production Function*. Amsterdam: North-Holland Publishing Company, 1971.

Grunfeld, Y. and Levhari, D. "A Note on External Economies." Mimeographed, 1962.

Jorgenson, D. W. and Griliches, Z. "The Explanation of Productivity Change." *Review of Economic Studies*, Vol. 34, No. 3 (1967), pp. 249–283.

——— and ———. "Issues in Growth Accounting: A Reply to E. F. Denison." *Survey of Current Business* (May, Part II, 1972).

Kendrick, J. W. *Postwar Productivity Trends in the U.S., 1948–1969*. New York: Columbia University Press for NBER, 1973.

Mansfield, E. "Basic Research and Productivity Increase in Manufacturing." *The American Economic Review*, Vol. 70 (1980), pp. 863–873.

Mansfield, E., Rapaport, J., Romero, A., Wagner, S., and Beardsley, G. "Social and Private Rates of Returns from Industrial Innovations." *Quarterly Journal of Economics*, Vol. 91, No. 2 (1977), pp. 221–240.

Milton, H. S. "Cost-of-Research Index, 1920–1970." *Operations Research* (1972), pp. 1–17.

Mundlak, Y. "Estimation of Production and Behavioral Functions from a Combination of Cross Section and Time Series Data" in C. Christ et al., eds., *Measurement in Economics*, Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld, Stanford: Stanford University Press, 1963.

——— and Hoch, I. "Consequences of Alternative Specifications of Cobb-Douglas Production Functions." *Econometrica*, Vol. 33 (1965), pp. 814–828.

National Science Foundation. "A Price Index for Deflation of Academic R&D Expenditures." NSF 72–310. Washington, D.C., 1972.

———. *Research and Development in Industry, 1975.* NSF 77–324. Washington, D.C.: 1977.

Nelson, R. and Winter, S. "Neoclassical vs. Evolutionary Theory of Economic Growth: Critique and Prospectus." *Economic Journal*, Vol. 84, No. 336 (1974), pp. 886–905.

Nerlove, M. *Estimation and Identification of Cobb-Douglas Production Functions.* Chicago: Rand McNally & Company, 1965.

Pakes, A. and Schankerman, M. A. "The Rate of Obsolescence of Knowledge, Research Gestation Lags, and the Private Rate of Return to Research Resources." Harvard Institute of Economic Research, Discussion Paper 659, October 1978. [Published as "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources" in Z. Griliches, ed., *R&D, Patents, and Productivity*, Chicago: University of Chicago Press, 1984, pp. 73–88.]

Raines, F. "The Impact of Applied Research and Development on Productivity." Washington University Working Paper No. 6814, 1968.

Rymes, T. K. *On the Concept of Capital and Technical Change.* Cambridge: Cambridge University Press, 1971.

Schankerman, M. "The Determinants, Rate of Return, and Productivity Impact of Research and Development." Unpublished Ph.D. dissertation, Harvard University, 1979.

Sharpe, W. F. *The Economics of Computers.* New York: Columbia University Press, 1967.

Terleckyj, N. E. *Effects of R&D on the Productivity Growth of Industries: An Exploratory Study.* Washington, D.C.: National Planning Association, 1974.

———. "Direct and Indirect Effects of Industrial Research and Development on the Productivity Growth of Industries" (1975) in J. W. Kendrick and B. Vaccara, eds., *New Developments in Productivity Measurement*, Studies in Income and Wealth, Vol. 41, forthcoming.

———. "Output of Industrial Research and Development Measured as Increments to Production of Economic Sectors." Paper given at the 15th Conference of the International Association for Research in Income and Wealth, York, England, 1977.

Williams, B. R. *Science and Technology in Economic Growth.* London: MacMillan, 1973.