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: R&D and Productivity: The Unfinished Business

Chapter Author: Zvi Griliches

Chapter URL: http://www.nber.org/chapters/c8350

Chapter pages in book: (p. 269 - 283)

12 R&D and Productivity: The Unfinished Business

Current work on the role of public and private research in productivity growth has deep roots in the early work of agricultural economists. The first micro production function estimates (Tintner 1944), the first detailed total factor productivity (TFP) calculations (Barton and Cooper 1948), the first estimates of returns to public R&D expenditures (Schultz 1953; Griliches 1958), and the first production function estimates with an added R&D variable (Griliches 1964) all originated in agricultural studies. Other original contributions to applied econometrics by agricultural economists include Waugh (1929) on hedonics, Nerlove (1958) on distributed lags, and Hoch (1955) and Mundlak (1961) on panel data econometrics.[1]

The specific subfield I want to discuss here, the impact of R&D on productivity, has expanded enormously from its modest beginnings. Given the large number of recent surveys of this field, I will not review it again, having just done it in Griliches (1995), except to note that one of the best surveys, Australian Industry Commission (1995, vol. 3, app. QA), lists 27 studies estimating the returns to R&D at the firm level, 28 at the industry level, 10 at the country level, and 20 studies for agriculture alone.[2]

A preliminary version of this paper appears in *Conference Proceedings on Global Agricultural Science Policy for the Twenty-First Century: Invited Papers*, pp. 1–20 (Melbourne, Australia: Conference Secretariat, Department of Natural Resources and Environment, August 1996), and is also forthcoming in the proceedings of the Conference on Economic Growth, Technology, and Human Capital, held 19–20 December 1996 in Tucuman, Argentina. The author is indebted to Jacques Mairesse for helpful comments on an earlier draft, to Steve Bond and Bronwyn Hall for providing him with the panel data update, to Chorching Goh and Aviv Nevo for able research assistance, and to the Mellon Foundation and the National Science Foundation for financial support.

1. See Heady and Dillon (1961); Berndt (1991, chap. 4); Griliches and Mairesse (1998); and Griliches (1996) for historical surveys of some of these topics.

2. Additional surveys can be found in Huffman and Evenson (1993, chap. 7); Mairesse and Sassenou (1991); Mairesse and Mohnen (1995); Nadiri (1993); Alston and Pardey (1996, chap. 6); and Hall (1996).

Major progress was made in the past thirty years in this field: new databases were developed at the firm, business-unit, and project levels, and other measures of innovation were added, especially observations on patents. Still I am not entirely happy. As progress was made, it became clearer how much we still don't know and how thin are our data.

I will divide my remarks into three parts: (1) puzzles about the current results in this field; (2) conceptual problems with the "central" R&D capital model; and (3) econometric problems: simultaneity, heterogeneity, and spillovers.

12.1 Recent Results and Puzzles

The major framework for the analysis of the relationship between R&D and productivity has been the "R&D capital in the production function" model (see Griliches 1973 for an early exposition): $Q = AX^{\beta}K^{\gamma}$, where Q is output, X is an index of conventional inputs including physical capital, K is the "stock of knowledge" (or R&D), A is the level of disembodied technology, and β and γ are the parameters of interest. The focus in such analysis is on estimating γ, the elasticity of output with respect to R&D capital. Recent studies using 1980s data have raised the possibility that γ may have declined over time. The issue is important substantively and needs further investigation. Is it a temporary phenomenon? Has it been reversed recently? Is it a drop in private rather than social returns (Hall 1993)? By the way, a change in γ is not the same as a change in the net rate of return to R&D: $\rho = \gamma\ (Q/K) - \delta$, where δ is the depreciation rate of such capital. The rate of return to R&D may decline if K grows faster than Q and/or if δ rises, without necessarily implying a change in γ.

There are three more bits of unfinished business in the "results" area: (1) Often the productivity growth equation is estimated with the R&D intensity rather than the growth in R&D capital as the relevant variable. Sometimes this version gives "better" results. An argument can be made for it (chap. 9 in this volume), but I have not seen a convincing reconciliation of the results of these two versions of the same model. An encompassing test is in order here. (2) There is a parallel literature on estimating the valuation of R&D capital (or investment) in the framework of market value equations (Griliches 1981; Pakes 1985; Hall and Hall 1993). It should be connected to the production function estimation literature. (3) Work has been done both at the firm and industry levels. One might expect higher estimates of the rate of return to R&D (ρ) at the industry level due to the internalization of spillover effects, but the bulk of the results does not go in this direction and no convincing exploration of the aggregation problem has been done yet, as far as I know, in this context. One possibility is a higher δ at the individual firm (private) level because of obsolescence and the "creative destruction" of rents as against a larger component of more slowly depreciating social returns at the more aggregate levels.

12.2 The "Central" Model and Its Discontents

The "central" model treats R&D as another investment stream, parallel to physical investment, and constructs an analogous "knowledge" capital stock using the perpetual inventory method and an assumed (fixed) depreciation rate δ. But knowledge is not like refrigerators, and each of the steps in the construction of such a "capital" concept is problematic.

The list of problems is long:

The standard approach aggregates R into a K concept linearly, ignoring the possibility that knowledge production depends nonlinearly not only on current R&D efforts but also on previously accumulated results. Moreover, R as a producer of additions to K may be subject to short-run decreasing returns to the intensity of research and to longer-run diminishing returns due to the depletion of technological opportunities, unless they are recharged by science or other sources of new discoveries. This is not a new concern. It is alluded to in chapter 2 in this volume, it was raised in a number of papers by Evenson (e.g., 1984), and it has been revived in a number of recent papers.

Formal properties of models where $\dot{K} = f(R, K)$ have been considered by Bachrach (1990), Hall and Hayashi (1989), Lach (1994), Jones (1995), and Klette (1994), among others. A reasonable version of such a model is:

$$\dot{K}_i = R^{\gamma} K_i^{\phi} K_A^{h}$$

where the ϕ parameter associated with the own stock of knowledge reflects the within-firm spillovers and time interdependencies in the research process, while the h parameter, associated with the aggregate state of knowledge, reflects both positive external spillovers and negative crowding-out effects. Having started with such a model, there is no clear role left for a separate depreciation effect, though some of the authors add a linear depreciation component to such models.

In estimation, such models lead to the solving out of the unobservable K stock and to the estimation of productivity growth as a function of R and lagged levels of output, TFP, or patent stocks. The current results along these lines are interesting but not fully convincing, both because of econometric problems associated with the use of lagged dependent variables, and because of the likely endogeneity of R, a topic to which I shall return below.

Other conceptual problems are associated with the whole notion of depreciation of knowledge and with the question of how knowledge should be incorporated into the production function. Much of what we think of as depreciation is not physical forgetting but rather the dissipation of rents as the result of obsolescence. It is a valid private cost component of innovation but not necessarily a social one. Its implications for measurement depend on the state of price index measurement technology and on the market structure of the relevant industries. In the computer industry, where the incumbents have little market power, prices and revenues fall, but quantities need not. If correctly "deflated," there is little depreciation to knowledge capital in a "true" quality-constant production function. In the pharmaceuticals industry, where incumbents choose to depreciate their patent monopolies optimally and the appearance of new substitutes does not cause incumbent prices to decline, deflated revenues will fall and we would interpret it as the depreciation of private R&D capital and a decline in productivity (since the same set of resources are still used in the industry producing essentially the same quantities as before). All that has happened is that the previously accumulated R&D capital is now available to others in the industry and hence cannot collect much rent. But it is still contributing to the productivity (technology) of the industry. From a social perspective the loss of patent protection does not result in a decline in such capital but rather a rise in its utilization! The fact that in most cases our micro production functions are closer to revenue functions than to true quantities makes the second case more prevalent than the first. But often the data are a mixture of the two, leading to great difficulties in the interpretation of the empirical results.

It is obvious that such capital does not depreciate just due to the efflux of time or to mechanical wear and tear. The obsolescence of privately generated R&D-based knowledge is clearly a function of the activity of others and is unlikely to occur at a constant rate. A major challenge before us is to model this process convincingly. A start has been made by Caballero and Jaffe (1993), but this has yet to be transferred to the work on micro production functions.

The above discussion does not imply that there is no obsolescence in social knowledge. There has surely been loss in the social value of the knowledge stocks associated with making carbon copies of documents and building ships, both in the sense that existing stocks are applied to much smaller industries and hence the implicit social returns, the consumer surpluses attributable to the original invention of these products, become smaller as demand falls, and in the sense that they become much harder to retrieve due to the lack of use, the retirement and death of associated human capital, and just plain forgetting. Such depreciation need not have the usual declining-balance (geometric) form, except possibly in the aggregate, where the population renewal theorem (Jorgenson 1973) comes into play.

The final set of problems is associated with the nonrivalrous nature of knowledge (Arrow 1962; Romer 1990). If $\dot{K}$ is to be measured by the outputs of the knowledge-producing processes, it becomes an index of the level of productivity along the lines of the quality ladders or variety models of Grossman and Helpman (1991) and not a parallel capital input within the list of standard inputs. If $\dot{K}$ is measured by R&D input rather than output, the question is still, should the resulting production function be interpreted as having constant returns including the R&D input? The usual solution to this internal versus external economies of scale question was to treat the own R&D effects as subject to decreasing returns and to include them in the standard list of inputs, while treating the spillovers from the R&D of others as externalities (see chaps.

2 and 11 in this volume), assuring perfect competition within the relevant sectors. But the nonrivalrous nature of R&D results makes perfect competition solutions unlikely, leading to the patent system and other appropriability mechanisms and a divergence between price and marginal costs of production. The recent revival of monopolistic competition theory and its application in this context make it clear that knowledge-producing firms will have nonnegligible markups whose magnitude will depend on the conditions of competition in their industries and the strengths of their appropriability positions. What we have then in our data are revenue functions with nonzero markups and downward sloping demand functions "solving" the increasing returns "problem" (it is only a problem for our models, not necessarily for the real world). In particular, as I will show below, if one assumes that R&D affects only demand, one would interpret estimates of γ as a measure of $-\phi/\eta$, where ϕ is the demand elasticity with respect to R&D and η is the price elasticity of demand. This is equivalent to the Dorfman and Steiner (1954) result for advertising. Even if only partially true this has serious implications for estimation, which is my next topic.

12.3 Econometric Issues

There are a number of sources of misspecification which afflict the "standard" production function estimates of the elasticity of output with respect to R&D capital (γ). The major ones are (1) the simultaneity of the R&D decision, (2) heterogeneity and endogeneity of individual product prices, (3) heterogeneity of the underlying production functions, and (4) the role of spillovers.

The more general topic of the simultaneity of input decisions was discussed recently in Griliches and Mairesse (1998). If R&D is chosen on the basis of economic incentives, it is unlikely to be fully independent of the shocks and errors which affect the production relations we are trying to estimate. This is the simultaneity problem. If all firms face the same production function and the same factor prices, it is not clear why different firms would choose different R&D levels. If they all do the same thing, we may not be able to estimate anything. If they do not, then we need to understand why not. That is the identification problem.

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 might 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.

There are several "solutions" to the simultaneity problem. First, if one has good series on 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.[3] Even if one had the prices, they are likely to be highly collinear over time. There is one possible exception to this pessimistic view. With good data one could construct different "tax prices" of R&D facing different firms, which would provide us with some relevant cross-sectional variation. But, to my knowledge, that has not been implemented yet in this context.

Also, the implicit assumption of certainty about the future underlying such static models makes little sense in the R&D context. What is maximized here is the present value of all future profits, and the relevant output price concept is an expected one and not the current one, especially if current output (and demand) is subject to special and transitory circumstances.

Second, if both time series and cross-sectional data are available and one is willing to assume a simple permanent-transitory model: $u = \alpha + e$, where α is the permanent component which affects input demand choice while e, the transitory component, does not, then consistent parameter estimates can be had from the within-firm covariances. This is equivalent to allowing a separate constant term (dummy variable) for each firm, which would absorb the offending term in it. Unfortunately, such data sets are rare. Moreover, the covariance approach may exacerbate other problems, such as errors in the variables, which also afflict these kinds of data.

Third, one may be able to find other "indicator" variables of interest and they may help to solve the identification problem in such models. I shall discuss one such approach below.

The question whether the R&D stock measure is "contaminated" by simultaneity depends upon what is in the production function disturbance and to what extent it is anticipated by the decision makers. The usual construction of $K_T = \Sigma(1 - \delta)^j R_{t-1-j}$, with j going from zero to infinity, puts only lagged values of R&D into the equation. But to the extent that there are more or less permanent firm effects, reflecting market positions, differences in quality of the labor force, and other misspecifications, they would be correlated also with past R&D decisions. Going "within" or using growth rates eliminates such fixed effects but may still leave other specification errors, such as changing utilization rates and demand conditions. These may still influence current R&D decisions.

An example of current approaches to such problems can be seen in table 12.1, which is adapted from Griliches and Mairesse (1998). The first part, col-

3. First, there are no published R&D deflators at the detailed industry level; second, if they were available, they would still be very highly correlated with the cost of labor and cost of capital indexes, which are likely to be major ingredients of such indexes. What we will not have are changes in "real" R&D costs, in the productivity of such expenditures, in a field or industry, caused by various technological and scientific breakthroughs.

Table 12.1 **Alternative Estimates of Production Function Parameters:[a] U.S. R&D-Performing Firms, 1973, 1978, 1983, 1988 (standard errors in parentheses)**

	Balanced Panel		Full Sample[b]			
			Total OLS		Nonparametric F	
	Total	Within				
Variables[a]	(1)	(2)	(3)	(4)	(5)	(6)
Labor	.496	.685	.578	.551	.591	
	(.022)	(.030)	(.013)	(.013)	(.013)	
Physical capital	.460	.180	.372	.298	.321	.320
	(.014)	(.027)	(.009)	(.012)	(.016)	(.017)
R&D capital	.034	.099	.038	.027	.081	.077
	(.015)	(.027)	(.007)	(.007)	(.016)	(.019)
Investment	—	—	—	.110	—	
				(.011)		
Other variables[c]	—	—	—	—	Powers of h	Polynomial in P and h
N[d]	856		2,971		1,571	

[a]The dependent variable in columns (1)–(4) is the log of sales, while in columns (5) and (6), the dependent variable is the log(value added) $- \hat{\beta} \times$ log(labor).

[b]Consult Griliches and Mairesse (1998) for details of the estimation algorithm leading up to columns (5) and (6).

[c]The other variables in the equations are Year and Year $\times$ Industry 357 (i.e., computers) dummy variables.

[d]The number of observations in the balanced panel for regressions in columns (1) and (2) are the observations for those firms that have continuous data over the period. Similarly, the 2,971 observations in columns (3) and (4) are all the observations in the full sample. (Only six observations had to be discarded because of zero investment.) The number of observations in the last two columns decreases to 1,571 because lagged values of some of the independent variables are needed in the estimation.

umns (1) and (2), presents standard OLS production function estimates for a heavily selected panel of 214 R&D firms in U.S. manufacturing. As usual, the capital coefficient declines as one moves to "within-firm" data, but the estimated R&D coefficient actually increases. Table 12.2, which is new, applies the more general Chamberlain (1984) Π-matrix approach to the estimation of such a model and asks whether the R&D coefficients have declined over time. As can be seen from comparing the estimated γ's in columns (3) and (5) or (4) and (6), they did not. The allowance for correlated effects hits the physical capital coefficient primarily, and the allowance for individual firm heteroskedasticity introduces an additional puzzling instability in the estimated coefficients (compare the estimates in the SUR versus MD columns) but leaves the R&D coefficients largely unchanged. Neither of these estimates, however, takes care of the simultaneity of the employment decision, if the latter is af-

Table 12.2 **Alternative Estimates of Production Function Parameters: U.S. R&D-Performing Firms, 1973, 1978, 1983, 1988, Balanced Panel**

	Uncorrelated Random Effects		Correlated Effects			
Variables	SUR (1)	MD (2)	SUR (3)	MD (4)	SUR (5)	MD (6)
Labor	.594	.686	.664	.805	.671	.818
	(.037)	(.020)	(.020)	(.029)	(.044)	(.030)
Physical capital	.334	.260	.163	.062	.164	.062
	(.031)	(.014)	(.033)	(.022)	(.033)	(.022)
R&D capital	.067	.065	.092	.080		
Combined	(.022)	(.015)	(.035)	(.022)		
1973					.086	.065
					(.036)	(.024)
1978					.087	.072
					(.035)	(.023)
1983					.073	.059
					(.036)	(.024)
1988					.094	.076
					(.035)	(.023)
Chi-square (degrees of freedom)		366 (45)		121 (33)		110 (30)

Notes: Number of observations = 214 firms × 4 = 856. SUR = seemingly unrelated (multivariate) regression estimates; MD = minimum distance (individual heteroskedasticity weighted). Dependent variable is log deflated sales. Other variables in the equation are year dummy variables, computer (357) industry dummy variable, and computer-year interaction variables.

fected by current shocks in production or correlated with unmeasured changes in capacity utilization.

An interesting new approach to the simultaneity problem is presented by Olley and Pakes (1996) in their paper "The Dynamics of Productivity in the Telecommunications Equipment Industry." This paper deals with two topics, selectivity and simultaneity, in an intertwined fashion. The sample selectivity problem may be quite serious for panel data. If observations (and data) are not missing at random, estimates that are based on "clean" and "balanced" subsamples could be badly biased. For example, a bad draw of u may force a firm or plant to exit from the industry. Such a negative correlation between estimated productivity shocks and future probabilities of exit was observed by Griliches and Regev (1995) in their analysis of Israeli industrial firms. They called it "the shadow of death." If the impact of negative u's on exit is stronger for smaller firms (the larger ones having more resources to survive them), then this will induce a negative correlation between u and the stock of capital among the surviving firms and bias the estimated capital coefficient downward in such samples. I will emphasize, however, their suggested solution to the simultaneity problem in this paper. (See Griliches and Mairesse 1998 for a more detailed exposition.)

The major innovation of Olley and Pakes is to bring in a new equation, the investment equation, as a proxy for α, the unobserved transmitted component of u.[4] Trying to proxy for the unobserved α (if it can be done right) has several advantages over the usual within estimators: it does not assume that α reduces to a "fixed" (over time) firm effect; it leaves more identifying variance in the "independent" variables and is therefore a less costly solution to the omitted variable and/or simultaneity problem; and it should also be substantively more informative.

Their argument goes roughly as follows: the investment demand of the firm at time t can be written as a function of the predetermined capital stock variables and that part of the shock in the production function u, the α, that is transmitted to both the employment and the investment decisions. Inverting this relationship and solving for α as a function of investment and capital stock, one can approximate it now semiparametrically and estimate the production function in two steps (three, if one also deals with selectivity at the same time): First one gets a consistent estimate of the coefficient of the labor variable, and then one retrieves the capital coefficient by using the estimated labor coefficient to move the endogenous labor variable to the left-hand side of the equation.

An application of their approach to our data is presented in columns (3)–(6) of table 12.1. Because exit is often a success for our R&D firms (being taken over) rather than a failure, the selection problem is not particularly severe in our data (compare the results in columns [1] and [3]). Once one shifts to the more complete unbalanced samples, the remaining selectivity (mainly attrition) does not appear to be too important (compare columns [5] and [6]).

As far as the simultaneity problem is concerned, either it is of no great import in these data or the introduction of investment and the associated Olley and Pakes procedure does not fully adjust for it. Investment is highly "significant" when added to the production function (see column [4]), but at the end of the procedure (having allowed for selectivity and unbalance), the coefficients change only a little (compare columns [1] and [3] with [6]) except that again we do get a higher R&D coefficient.

The Olley and Pakes solution to the simultaneity problem is a clever way to exploit the fact that the unobserved "productivity shocks" are transmitted to more than just one equation and should be estimated within a system of behavioral equations. It does rest, however, on two very strong assumptions: (1) that there is only *one* single-component unobservable in the system, the α_{it}, which follows a first-order Markov process and is fully transmitted to the investment equation, and (2) that no other variables or errors appear in it. Investment de-

4. In their notation α is ω and they refer to it simply as "productivity."

pends, however, also on other individual factors such as interest rate expectations, tax treatments, and changes in future demand prospects not yet fully captured in the initial state variables (the capital stocks). In principle, there may be additional instrumental variables and other indicators of α, such as R&D, which could help solve the errors in the investment equation problem, except for the extreme nonlinearities introduced by their semiparametric approach.[5]

Other approaches lean more heavily on assumptions about lags in the transmission of the disturbances to the other decision variables and use lagged values as instrumental variables in estimating such models (see Blundell and Bond 1995 and Mairesse and Hall 1996). One can write a simple model of the production function as

$$y_{it} = \beta x_{it} + \gamma k_{it} + \alpha_i + u_{it}, \qquad u_{it} = \rho u_{it-1} + e_{it},$$

where small letters represent the logarithms of the variables, x is a composite of conventional inputs including physical capital, k is a measure of the R&D stock, α_i is an unobserved permanent firm effect, while u is a randomly changing technical disturbance. The innovation in u, the e_{it}, is unpredictable, but whether x and k are independent of it depends on the assumed lag structure of the decisions affecting their evolution. (Of course, u could be modeled as a higher order autoregression.) In such a world, we could solve out u_{it} and rewrite the equation as

$$y = \beta(x - \rho x_{-1}) + \gamma(k - \rho k_{-1}) + \rho y_{-1} + e + (1 - \rho)\alpha_i,$$

and use past differences in x, k, and y, which should be independent of α_i and e_i, as instruments.[6]

In table 12.3, a larger sample (including non-R&D firms) is analyzed in this framework using the Generalized Method of Moments (GMM) approach (see Mairesse and Hall 1996 for a recent exposition in a similar context). Columns (1) and (6) present the OLS estimates for levels and first differences respectively. Column (2) allows for serial correlation and finds it very high (as could be expected). Columns (3) and (4) repeat these level computations allowing for the endogeneity of all the input (and lagged output) variables, using past differences in these variables as instrumental variables. Column (5) is similar, but only instruments the labor and lagged output variables, treating the two capital stocks as predetermined.[7] Columns (7) and (8) present the corresponding estimates of this equation in first differences, instrumented by past levels. The first differences transformation is optimal if $\rho = 1$, or if $\rho = 0$ and the "not-so-fixed effect" is a random walk, that is, $\alpha = \alpha_{-1} + e$. Column (7) uses

5. The current state of estimating nonlinear errors-in-variables models is not completely hopeless, but it is not easy either.

6. This assumption is right for "stationary" α's, where their effect on y is unchanged over time.

7. Using instruments from $t - 3$ rather than $t - 2$ increases the standard errors but has little effect on the reported results.

Table 12.3 **Alternative Estimates of Production Function Parameters: U.S. R&D and Non-R&D Manufacturing Firms, 1982–87, $N = 676$ (standard errors in parentheses)**

	Levels					First Differences		
	OLS		Instrumented by Differences			OLS	Instrumented by Levels	
Variable	(1)	(2)	(3)[a]	(4)[a]	(5)[b]	(6)	(7)[c]	(8)[d]
Labor	.567	.616	.665	.750	.652	.613	.705	.611
	(.008)	(.013)	(.048)	(.027)	(.046)	(.013)	(.024)	(.062)
Capital	.402	.122	.277	.289	.314	.114	.084	.110
	(.007)	(.012)	(.036)	(.027)	(.031)	(.012)	(.019)	(.037)
R&D stock	.016	.041	.033	.025	.030	.030	.046	.059
	(.004)	(.012)	(.017)	(.017)	(.010)	(.013)	(.017)	(.022)
Lagged output	0	.981	0	.573	.654	1	1	1
		(.004)		(.023)	(.031)			

Notes: Estimates in columns (2), (4), and (5) (equation 11) are constrained to the same ρ coefficient in $(x - \rho x_{-1})$ and ρy_{-1}. Additional variables included in the equations: no-R&D dummy variable, year dummies, computer industry dummy, and interaction with year.

[a]Instrument sets: all differences as of $t - 2$ and earlier, for l, c, k, and y_{-1}.

[b]Instrument sets: c and k treated as predetermined. Only l and y_{-1} instrumented.

[c]Instrument sets: levels of l, c, and k as of $t - 2$ and earlier.

[d]Instrument sets: only l instrumented.

past levels as instruments for all three variables, which is appropriate if there are random measurement errors in them or if there is some remaining contemporaneous simultaneity, while column (8) only instruments the labor variable.

The preferred specification, column (5), indicates a substantively and statistically significant R&D coefficient of about .03. It also finds that the individual firm effects are not entirely fixed but include a component which does depreciate, albeit slowly. If one approaches the limit of $\rho = 1$ (first differences), there is hardly any identifying variance left in the annual changes in our measures of physical and R&D capital. Measurement and timing errors now predominate, while the remaining information content in the instruments is too small to allow one to extract whatever signal is still left in these variables. In the end, what is clear is that there seems to be a significant R&D coefficient, but its magnitude is uncertain, varying from about .03 to .08 based on estimates from reasonably robust specifications (table 12.1 col. [6], table 12.2 col. [4], and table 12.3 col. [5]).

The GMM approach uses past values of the inputs and outputs as instruments. What is their identifying content? Inputs today depend on past demand and supply shocks because, presumably, there are lags in adjustment and also erroneous decisions. But without specifying nontrivial real factor demand and supply equations with measurable exogenous shifters of such functions, we have no interesting variables that could be used to interpret (identify) their behavior. There are no measures of shifts in the potential demands for a firm's products, or of changes in technological opportunities, market structure, or individual firm cost of capital. Without such shifters it is hard to tell whether such lagged values represent an interesting experiment which would allow us to identify something.

Another major specification problem revolves around the unlikely assumption that all firms within an industry charge the same price. If product prices are both different and endogenous, then what is estimated is a revenue function, not a production function, with left-out product prices in the residual. This problem is considered by Klette and Griliches (1996) who, reinventing an argument made by Marschak and Andrews (1944), start with a model of firms facing symmetric logarithmic market share (demand) functions

$$y_i - y_I = \eta(p_i - p_I) + e,$$

where y_i and y_I are respectively the real output of the firm and the industry, p_i is the firm's own price (or price index), η is demand elasticity with respect to the relative price of its own products, p_I is the aggregate industry price index (relative to the overall economy price level), and e are all other demand shifters for the products of this firm. If the variable that we observe is not real output y, but deflated revenue (sales)

$$r_i = (y_i + p_i) - p_I,$$

then the "revenue production" function is

$$r = \beta x + \gamma k + u + (p_i - p_I).$$

There would be no problem here if the p_i's were random and exogenous. But if firms have a modicum of market power, at least in the short run, p_i will be set by them and will be correlated with u, x, and k. Setting price equal to marginal revenue and solving out for p_i yields the pseudo–production function

$$r = [\beta x + \gamma k + u]/m - (y_I)/\eta - e/\eta,$$

where the markup coefficient $m = \eta/(1 + \eta)$ is likely to be larger than one. Since y_I and p_I are aggregates, they can be controlled for by the introduction of period dummy variables. It is clear now that the estimates of α and β will be biased *downward* on the order of $1/m$, implying diminishing returns to scale in contexts where there actually may be increasing returns.

This model can be extended by adding R&D capital to the demand function, with ϕ as its elasticity.[8] The coefficient of k in the deflated sales equation is then $(\gamma/m - \phi/\eta)$, a combination of its effects on both productivity and demand, attenuated by the price elasticity of demand. This coefficient can also

8. In this form, R&D capital is a separable demand shifter, leaving the price elasticity of demand unaffected. (See chap. 5 in this volume for an early formulation of this model.) A more complex model might also include an interaction term, making the price elasticity itself a function of K.

be rewritten as $\gamma + (\gamma - \phi)/\eta$, showing that the pure productivity effect of R&D will be underestimated as long as it is smaller than its demand effect (ϕ). Klette and Griliches show that if one has a measure of the demand shifter (they use aggregate industry sales y_I for that) one can identify η and β, but one cannot separate ϕ from γ, unless one assumes $\phi = 0$. Without actual individual firm prices, there may be little that we can do here except be more careful in our interpretation of such results.

All of this discussion has focused on estimating the effects of R&D, but what makes different firms choose to undertake different amounts of R&D? I have already noted the lack of good external causal variables. To the extent that differences in R&D reflect technological opportunities, they could be modeled as differences in γ, firms facing (or possessing) different knowledge-producing technologies (though keeping the conventional input component the same within an industry). But unless one brings in some substantive variables which would explain this heterogeneity, such generality adds very little content. (See Mairesse and Griliches 1990 for a parallel discussion of heterogeneity in the physical capital elasticity.) The open modeling question is how to use the observed differences in R&D intensity to infer something interesting about the underlying sources of the heterogeneity in γ.

The final estimation-specification problem I want to discuss is the estimation of spillover effects. The standard approach (for which I must take some responsibility, cf. chap. 2 in this volume) introduces a distance weighted measure of the research efforts of other firms within the same and/or neighboring industries or technological areas. It is clearly a first step in the right direction, but it is also subject to a serious identification problem: Does it work because a firm benefits from the efforts of others or is it just a reflection of spatially correlated technological opportunities? It could be a response to common differences across fishing grounds or, in more technical terms, the individual firm effects α_i may not be independent of each other but may be subject to some local clustering, which will be picked up by the spillover measures. This issue is discussed in a more general context by Manski (1991), under the title "the reflection problem." It would be nice if someone could come up with an approach that could distinguish between these two interpretations, but that is unlikely since the basic model is not identified without much more explicit parameter restrictions and priors on the possible channels of communication.

I have concentrated today on the unfinished business, not to emphasize the "glass-half-empty" aspect, but rather to indicate the rich research opportunities ahead. "Our song is not finished, it's only beginning!"

References

Alston, J. M., and P. G. Pardey. 1996. *Making science pay: The economics of agricultural R&D policy.* Washington, D.C.: AEI Press.

Arrow, K. J. 1962. Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, 609–25. NBER Special Conference Series, vol. 13. Princeton, N.J.: Princeton University Press.

Australian Industry Commission. 1995. *Research and development*. Report no. 44. Canberra: Government Publishing Service, 15 May.

Bachrach, C. 1990. Essays on research and development and competitiveness. Ph.D. diss., Massachusetts Institute of Technology.

Barton, G. T., and M. R. Cooper. 1948. Relation of agricultural production to inputs. *Review of Economics and Statistics* 30 (2): 117–26.

Berndt, E. 1991. *The practice of econometrics: Classic and contemporary*. Reading, Mass.: Addison-Wesley.

Blundell, R. S., and S. Bond. 1995. Initial conditions and moment restrictions in dynamic panel data models. Working Paper W95/17, Institute of Fiscal Studies, London.

Caballero, R. J., and A. B. Jaffe. 1993. How high are the giants' shoulders? In *Macroeconomics Annual 1993*, ed. Oliver J. Blanchard and Stanley Fischer, 15–74. Cambridge, Mass.: MIT Press.

Chamberlain, G. 1984. Panel data. In *Handbook of econometrics*, ed. Z. Griliches and M. Intriligator, vol. 2, 1247–1318. Amsterdam: North-Holland.

Dorfman, R., and P. O. Steiner. 1954. Optimal advertising and optimal quality. *American Economic Review* 44 (December): 826–36.

Evenson, R. E. 1984. International invention: Implications for technology market analysis. In *R&D, patents, and productivity*, ed. Zvi Griliches, 89–123. Chicago: University of Chicago Press.

Griliches, Z. 1958. Research cost and social returns: Hybrid corn and related innovations. *Journal of Political Economy* 66 (5): 419–31.

———. 1964. Research expenditures, education and the aggregate agricultural production function. *American Economic Review* 54 (6): 961–74.

———. 1973. Research expenditures and growth accounting. In *Science and technology in economic growth*, ed. B. R. Williams, 59–95. London: Macmillan.

———. 1981. Market value, R&D, and patents. *Economics Letters* 7:183–87.

———. 1995. R&D and productivity: Econometric results and measurement issues. In *Handbook of the economics of innovation and technological change*, ed. P. Stoneman, 52–89. Oxford: Basil Blackwell.

———. 1996. The discovery of the residual: A historical note. *Journal of Economic Literature* 34 (September): 1324–30.

Griliches, Z., and J. Mairesse. 1998. Production functions: The search for identification. In *The Ragnar Frisch centennial symposium*, ed. S. Ström. Economic Society Monograph Series. Cambridge: Cambridge University Press, forthcoming.

Griliches, Z., and H. Regev. 1995. Firm productivity in Israeli industry, 1979–1988. *Journal of Econometrics* 65: 175–203.

Grossman, G. M., and E. Helpman. 1991. *Innovation and growth in the global economy*. Cambridge, Mass.: MIT Press.

Hall, B. H. 1993. Industrial research in the 1980s: Did the rate of return fall? *Brookings Papers on Economic Activity: Microeconomics* 2:289–331.

———. 1996. The private and social returns to research and development. In *Technology, R&D, and the economy*, ed. B. Smith and C. Barfield, 140–62. Washington, D.C.: Brookings Institution and AEI.

Hall, B. H., and R. E. Hall. 1993. The value and performance of U.S. corporations. *Brookings Papers on Economic Activity* 1–50.

Hall, B. H., and F. Hayashi. 1989. Research and development as an investment. NBER Working Paper no. 2973. Cambridge, Mass.: National Bureau of Economic Research, May.

Heady, E., and J. Dillon. 1961. *Agricultural production functions.* Ames: Iowa State University Press.
Hoch, I. 1955. Estimation of production function parameters and testing for efficiency. *Econometrica* 23 (3): 325–26.
Huffman, W. E., and R. E. Evenson. 1993. *Science for agriculture.* Ames: Iowa State University Press.
Jones, C. I. 1995. R&D-based models of economic growth. *Journal of Political Economy* 103 (4): 759–84.
Jorgenson, D. W. 1973. The economic theory of replacement and depreciation. In *Econometrics and economic theory,* ed. W. Sellekaerts, 189–221. New York: Macmillan.
Klette, T. J. 1994. R&D, scope economies and company structure: A "not so fixed effect" model of plant performance. Statistics Norway, Oslo. Mimeograph.
Klette, T. J., and Z. Griliches. 1996. The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics* 11:343–61.
Lach, S. 1994. Non-rivalry of knowledge and R&D's contribution to productivity. Working Paper no. 289, Hebrew University, Jerusalem, June.
Mairesse, J., and Z. Griliches. 1990. Heterogeneity in panel data: Are there stable production functions? In *Essays in honor of Edmond Malinvaud,* ed. P. Champsaur et al., vol. 3, 193–231. Cambridge, Mass.: MIT Press.
Mairesse, J., and B. H. Hall. 1996. Estimating the productivity of research and development in French and U.S. manufacturing firms: An exploration of simultaneity issues with GMM methods. In *International productivity differences and their explanations,* ed. K. Wagner and Bart Van Ark, 285–315. Amsterdam: Elsevier Science.
Mairesse, J., and P. Mohnen. 1995. R&D and productivity: A survey of the econometric literature. Institut National de la Statistique et des Études Économiques (INSEE), Paris. Mimeograph.
Mairesse, J., and M. Sassenou. 1991. R&D and productivity: A survey of econometric studies at the firm level. *STI Review* (Paris: OECD) 8:9–43.
Manski, C. 1991. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60 (3): 531–42.
Marschak, J., and W. Andrews. 1944. Random simultaneous equations and the theory of production. *Econometrica* 12:143–205.
Mundlak, Y. 1961. Empirical production function free of management bias. *Journal of Farm Economics* 43:44–56.
Nadiri, M. I. 1993. Innovations and technological spillovers. NBER Working Paper no. 4423. Cambridge, Mass.: National Bureau of Economic Research.
Nerlove, M. 1958. *Distributed lags and demand analysis.* USDA, Agriculture Handbook No. 141. Washington, D.C.: Government Printing Office.
Olley, S., and A. Pakes. 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64 (6): 1263–97.
Pakes, A. 1985. On patents, R&D and the stock market rate of return. *Journal of Political Economy* 93 (2): 390–409.
Romer, P. M. 1990. Endogenous technological change. *Journal of Political Economy* 98 (5): S71–S102.
Schultz, T. W. 1953. *The economic organization of agriculture.* New York: McGraw-Hill.
Tintner, G. 1944. A note on the derivation of production functions from farm records. *Econometrica* 12 (1): 26–34.
Waugh, F. V. 1929. *Quality as a determinant of vegetable prices.* New York: Columbia University Press.