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

Volume Title: R & D, Patents, and Productivity

Volume Author/Editor: Zvi Griliches, ed.

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-30884-7

Volume URL: http://www.nber.org/books/gril84-1

Publication Date: 1984

Chapter Title: An Exploration into the Determinants of Research Intensity

Chapter Author: Ariel Pakes, Mark Schankerman

Chapter URL: http://www.nber.org/chapters/c10050

Chapter pages in book: (p. 209 - 232)

9 An Exploration into the Determinants of Research Intensity

Ariel Pakes and Mark Schankerman

This paper explores the economic factors that determine the distribution of research effort across firms. Our main objectives are to provide a general framework for analyzing the demand for research by private firms and to document empirically certain stylized facts about R & D intensity and its determinants at different levels of aggregation.

Three competing explanations of the distribution of research are in the literature, each emphasizing a different aspect of the problem. Schmookler (1966) and Griliches and Schmookler (1963) emphasize the importance of expected market size as an inducement to research effort. They recognize that the cost of reproducing the knowledge generated by research is low relative to the original cost of producing it, and therefore, that the private return to research varies directly with the number of units of output embodying the knowledge or with the size of the market. Differences across industries in the cost of producing knowledge are downplayed, based on the argument that scientific knowledge is sufficiently well developed to make the supply of new industrial knowledge highly elastic at the same level of costs for all industries. Rosenberg (1963, 1969, 1974) and Scherer (1965), while granting the importance of market size, argue that the body of scientific and engineering knowledge

Ariel Pakes is a lecturer in the Department of Economics in the Hebrew University of Jerusalem, and a faculty research fellow of the National Bureau of Economic Research. Mark Schankerman is an assistant professor in the Department of Economics at New York University, and a faculty research fellow of the National Bureau of Economic Research.

This paper is a consolidated and extended version of work originally contained in Pakes and Schankerman (1977) and Pakes (1979). The authors would like to thank Zvi Griliches for many constructive discussions of the issues in this paper, and Gary Chamberlain and Bronwyn Hall for econometric and computing suggestions. The authors gratefully acknowledge the financial assistance of the National Science Foundation, grants GS-2726X and PRA-8108635, and the C. V. Starr Center for Applied Economics at New York University.

grows at different rates in different areas and suggest that these differences in the cost of producing industrial knowledge, or technological opportunity, are a major determinant of the observed distribution of research effort. Schumpeter (1950), on the other hand, argues that research effort generates temporary monopoly power for the innovating firm and that the private benefits from the production of knowledge must be a result of quasi-rents appropriated by the producer of the innovation. Schumpeter therefore emphasizes the determinants of the degree of appropriability, such as entrepreneurial ability, industrial market structure, and the general institutional framework (including patent rights) in which the firm operates.

We develop a simple model consistent with the theoretical argument that the output of research activities (industrial knowledge) possesses unique economic characteristics. Our model implies that research intensity depends on three factors: appropriability, technological opportunities, and expected market size or demand inducement. This is a richer set of determinants than those underlying the demand for traditional inputs and is therefore consistent with the empirical observation that the coefficient of variation of research intensity is an order of magnitude larger than those of traditional inputs. We specify an explicit set of stochastic disturbances in a set of factor demand equations and estimate the model both at the intraindustry and interindustry levels of aggregation. The empirical results of the intraindustry analysis imply that, though part of the variance in R & D intensity is attributable to measurement and decision errors, the bulk of the variance in observed research intensity is structural in the sense that it is consistent across factor demand equations. Growth rates of output account for very little of this structural variance. We then explicitly aggregate the micro relations to the interindustry level of observation, and the empirical results at this level of aggregation are strikingly different. In particular, differences in industry growth rates account for well over 50 percent of the interindustry variance in research intensity. We explain how these differences can arise and demonstrate their empirical importance.

Section 9.1 specifies the production relationships in the model. In section 9.2 a model of the private returns to R & D is presented. Section 9.3 specifies the stochastic structure and discusses identification. Section 9.4 applies the model to the intraindustry variance in research intensity. In section 9.5 we explicitly aggregate and estimate the model at the industry level of observation. Brief concluding remarks follow in section 9.6.

9.1 Production Relationships

This paper is based on an extended Cobb-Douglas production function in which research resources enter the production process by raising the productivity of traditional factors of production in a disembodied manner. Problems involved in the construction of R & D variables have made this specification the most widely used framework in the empirical analysis of the role of research resources in production (see Griliches 1973, 1980). Our specification differs from the conventional one in two respects: (1) we decompose R & D resources into research capital and research labor, and (2) we explicitly incorporate an R & D gestation lag and a rate of obsolescence of produced knowledge, both of which influence the optimal R & D intensity of the firm.

We begin with the traditional production function

$$Q = \gamma_0 K N^{\gamma_1} H^{\gamma_2}, \tag{1}$$

where Q is output (value added), γ_0 is a constant (which may be both firm and time specific), K is the stock of accumulated and still productive knowledge produced by the firm, N and H are traditional labor and capital services, and all firm and time subscripts have been omitted for convenience. Since K is not observable, its units are arbitrary, and we normalize it so that a 1 percent increase in K raises output by 1 percent.

The generation of knowledge is summarized by its production function

$$\dot{K}_t^G = A_1 L_{t-\theta}^a C_{t-\theta}^b, \tag{2}$$

where $\dot{K}_t^G$ is the gross increment in produced knowledge in period t, A_1 is a constant (which may be both firm and time specific), $L_{t-\theta}$ and $C_{t-\theta}$ are research labor and research capital services in $t - \theta$, and θ is the mean lag between the time research is undertaken and its embodiment in the traditional production processes of the firm.[1] The parameters a and b are the elasticities of research labor and research capital in the production of increments to K, and they will be assumed not to differ among firms in a given industry (though differences among industries are permitted). These parameters are indices of the technological opportunities of the industry, that is, they reflect the ease with which the underlying scientific and engineering knowledge permits firms in a given industry to transfer their research inputs into cost-reducing innovations (see Scherer 1965 and Rosenberg 1969).

Assuming geometric decay of knowledge at the rate of δ_1 and taking the growth rates of research capital and research labor to be constant both during and prior to the period of analysis (as required by our data), the net increment to knowledge $\dot{K}_t^N$ is

1. Since our data cannot sustain an investigation of the distributed lag between the expenditure of research resources and the resultant increases in a firm's productivity, we use the simplification of a mean lag which applies to all units of research resources equally. Note also that research capital refers to an aggregate of all research resources other than research labor, and that the constant term A_1 captures both the effects of "learning by doing" and of other firms' research as inputs in the production process of the firm in question.

(3) $$\dot{K}_t^N = Ae^{(a\hat{L}+b\hat{C})t} - \delta_1 K_t,$$

where $A = A_1 L_0 C_0 e^{-(a\hat{L}+b\hat{C})\theta}$, and a caret denotes a rate of growth. Solving this differential equation and assuming that $\lim_{t \to -\infty} K_t = 0$, the stock of productive knowledge becomes:

(4) $$K_t = \frac{A_1 L_{t-\theta}^a C_{t-\theta}^b}{\delta_1 + a\hat{L} + b\hat{C}}.$$

This concludes the specification of the production relationships. However, we will require expressions for the reduction in unit costs attributable to an increase in research labor and research capital. Assuming that the firm is a cost minimizer facing fixed input prices, the unit cost function associated with (1) can be expressed as

(5) $$Z = \frac{h(w, p_H)}{K},$$

where Z represents unit costs, and w and p_H denote the (fixed) wage and rental rates for traditional labor and capital services, respectively. Substituting (4) into (5) and differentiating the cost function at time $t + \theta$ with respect to research labor and research capital services at time t, we obtain

(6) $$-\frac{\partial Z_{t+\theta}}{\partial L_t} = \frac{aZ_{t+\theta}}{L_t}, \text{ and } -\frac{\partial Z_{t+\theta}}{\partial C_t} = \frac{bZ_{t+\theta}}{C_t}.$$

9.2 Optimal Factor Intensities for Research Inputs

If private firms are motivated by potential profits, the level of their research effort will be determined by the expected net income generated by investment in research resources. The large observed variance across firms in research intensity should be attributable to the variance in the expected private returns to research. The objective factors that could cause differences in the expected net income generated by the use of research resources are: (1) variation in the costs of research inputs; (2) differences in the productivity of research resources in generating usable industrial knowledge; and (3) differences in the ability to derive monetary benefits from a given unit of produced knowledge. Variation in costs of research inputs will be incorporated into the model and discussed in section 9.3. In connection with the productivity of research resources, the basic model assumes that all firms in a given industry produce a single homogeneous output subject to the same production conditions (as specified in section 9.1), and the model is tested separately for each industry in our data set. Consequently, differences in the expected returns from research beyond those caused by differences in the cost of research inputs will be associated with differences in the ability to derive monetary benefits from a given unit of produced knowledge. However, the industries in the data set are defined quite broadly, so there could be some intraindustry differences in the output elasticity of research resources. At this stage, we do not separate these supply-side differences from those differences in ability to capture the monetary benefits from knowledge. We return to this problem later in this paper.

The difficulty in specifying a mechanism that determines the stream of private benefits accruing to new industrial knowledge is a result of the fact (stressed by Arrow 1962) that knowledge has no, or a very small, cost of reproduction. Since any economic agent aware of the information embodied in the innovation can exploit it, the private benefits from the production of industrial knowledge must be a result of quasi-rents or temporary monopolies accruing to the producer of the innovation (Arrow 1962; Machlup 1962; Nordhaus 1969a, 1969b). The strength of these monopolies, that is, the abilities of firms to appropriate the benefits from the knowledge which they have developed, will determine the private return to research resources and therefore the research intensity of firms. The private return to the development of a new cost-reducing technique will depend on the number of units of output embodying this new knowledge and the fraction of the cost reduction attributable to the innovation apropriated by the innovating firm.

We begin by reviewing the "maximum appropriability environment," first described by Arrow (1962) and later adapted to determine the rate of return to research resources by Nordhaus (1969b). Consider a constant cost industry in competitive equilibrium and an innovation which reduces the cost of production for the firms in the industry and only for such firms. The maximum appropriability environment is based on the assumption that the innovator patents the innovation costlessly and leases the cost-reducing technique to all firms in the industry (including itself), subject to the condition that the final product must sell at a uniform price to consuming units. The lease can be defined in terms of a royalty per unit of output produced with the innovation, ρ_0. The lessor acts as a monopolist and sets the initial royalty to maximize profits subject to the constraint that the royalty plus the new cost of production ($\rho_0 + Z_1$) is less than or equal to the preinnovation cost of production ($Z_0 = P_0$). In virtually all cases the profit-maximizing royalty at the date of introduction will be $\rho_0 = Z_0 - Z_1 = \Delta Z$.[2] The revenue collected in the first year of the

2. If P_1 is the profit-maximizing price for a monopolist with constant unit cost Z_1, the Arrow royalty described in the text will yield maximum profits if and only if $P_1 > P_0$. If the industry demand is price inelastic over the relevant range, the Arrow royalty will be optimal regardless of the magnitude of the cost reduction from the innovation. If the industry demand is price elastic, the condition $P_1 > P_0$ can be written $\Delta Z / Z_0 < |\eta(P_1)^{-1}|$, where η denotes the price elasticity of industry demand. It is apparent from this inequality that the Arrow royalty will be optimal for all but the most major innovations and will certainly be optimal for the cost reduction resulting from the employment of the marginal research resource.

innovation in that appropriability environment is $\rho_0 Q_0^I = \Delta Z Q_0^I$, where Q_0^I denotes the industry output in the year the innovation is introduced. We now extend the analysis to firm-specific, nonmaximal appropriability environments and to the calculation of the entire revenue stream accruing to the innovation. Let $k_{i\tau}$ be the fraction of industry output from which firm i receives royalties on its innovation of age τ, $\rho_{i\tau}$ the royalty per unit of output, and $B_{i\tau}$ the total revenues accruing to the innovation in year τ. Then the discounted value of the stream of revenues generated by the innovation is

$$\text{(7)} \qquad \Pi = \int_0^\infty B_{i\tau} e^{-r\tau} d\tau = \int_0^\infty \rho_{i\tau} k_{i\tau} Q^I e^{-r\tau} d\tau,$$

where r is the discount rate.

The specification of the appropriability environment is based on the following two assumptions:

1. It is easier, or less costly, for a firm to capture the benefits of the knowledge it produces through embodiment in its own output (internal appropriation) than through embodiment in the output of other firms. Internal appropriation is less costly because of the difficulties involved in establishing an effective market for information (for more discussion see Arrow 1962).

2. The revenues accruing to an innovation decline with the age of the innovation. This occurs because new techniques are developed by the firm and its competitors which substitute for the original innovation and because the use of the information in any productive way reveals and spreads it. This tends to erode both the unit royalty that can be charged and the part of industry output from which royalties accrue.

To maintain a specification that is both as general as possible and consistent with the preceding two assumptions, we let

$$\text{(8)} \qquad \rho_{i\tau} k_{i\tau} = (\rho_{i0} e^{-\delta\tau}) e^{\bar{k} + k_i} \frac{Q_{i\tau}}{Q_\tau^I} = \Delta Z e^{\bar{k} + k_i - \delta\tau} \frac{Q_{i\tau}}{Q_\tau^I},$$

where $\Sigma_{i=1}^n k_i = 0$ by construction (n is the number of firms in the industry), and $Q_{i\tau}$ denotes the expected output of firm i at time τ. Revenues in period τ become

$$\text{(9)} \qquad B_{i\tau} = \rho_{i\tau} k_{i\tau} Q_\tau^I = (\Delta Z e^{-\delta\tau}) e^{\bar{k} + k_i} Q_{i\tau}.$$

We interpret the parameters in the following manner: δ is the rate of decay in the unit royalty, exp $(\bar{k} + k_i)$ is the proportion of firm i's share of industry output from which the firm receives this royalty, and exp $(\bar{k})$ is the (geometric) mean of this proportion over all firms in the industry. However, it is impossible to distinguish empirically between a rate of decay in the proportion exp $(\bar{k} + k_i)$ and δ, or between a firm-specific component in the rate of decay and exp(k_i). Since appropriable revenues alone suffice to determine the private benefits from an innovation, it is immaterial whether the firm-specific component applies to the royalty (the price side of revenues) or to the number of units from which the firm receives these royalties (the quantity side of revenues). Hence, these relationships may be interpreted as saying that the revenues generated by a given innovation of age τ depend on: (1) the importance of the innovation, ΔZ; (2) the age of the innovation, τ (through the rate of obsolescence of the *private* returns from knowledge, δ); (3) a firm-specific structural parameter, $\exp(k_i)$, which determines the extent to which the firm can monopolize the information produced by its research resources; and (4) the expected output of the innovating firm, $Q_{i\tau}$, because of the relative ease of internal appropriation.

To obtain the present value of revenues generated by the employment of the marginal unit of research labor (Π_ℓ), substitute (6) and (8) into (7). Setting the price of output equal to one (as it implicitly is in our data) and recalling that the cost reduction does not occur until θ years after the employment of the unit of research labor, we have

$$\Pi\ell = \int_\theta^\infty \frac{a_0}{L_0} e^{\bar{k} + k_i - r\tau - \delta(\tau - \theta)} Q_{0i} e^{g_i^* \tau} d\tau,$$

where g_i^* is the expected rate of growth of output of firm i. Equating Π_ℓ to the wage rate for research labor (w_r), taking a first-order expansion of $\log(\delta + r - g^*)$ around $\log(\delta + r)$, and rearranging terms, we can express the optimal research labor intensity (and following an analogous procedure, the optimal research capital intensity) as:[3]

$$\begin{aligned} (10) \qquad \log(w_r L/Q) &= \log \beta_0 + \alpha g^* + k_i, \\ \log(p_c C/Q) &= \log\beta_1 + \alpha g^* + k_i, \end{aligned}$$

where $\log \beta_0 = \log[a/(r + \delta)] - r\theta + \bar{k}, \log \beta_1 = \log[b/(r + \delta)] - r\theta + \bar{k}$, p_c is the price of research capital services, and $\alpha = (r + \delta)^{-1} + \theta$.

Several features of equation (10) are worth noting. First, since the returns from both research labor and research capital are derived from the returns to industrial knowledge, any factor that affects the returns to knowledge will influence the optimal intensities of both research variables. This fact permits econometric identification of the relative importance of the unobserved structural parameter (k_i) in determining the research intensities of firms. Also noted that the indices of technological opportunity at the industry level and the average degree of appropriability (a, b, and $\bar{k}$) affect the research intensities of firms since they appear in

3. Since only the moment matrix of the variables was available, we were limited to linear combinations of the original variables and forced to use Taylor approximations. The approximation error evaluated at the means is about 2 percent, and if g is distributed symmetrically this will not affect the estimate of α.

the constant terms in (10). These parameters are assumed not to vary within a given industry, but they may vary across industries and, in fact, could be endogenously determined in a more complete model. Second, equation (10) indicates that the firm's employment of research resources will vary directly with its expected market size (its appropriability base) and inversely with the rate of obsolescence and the rate of discount. The importance of expected market size in determining the optimal level of research resources follows directly from the fact that knowledge has a low cost of reproduction.[4] For a given value of initial revenues accruing to an innovation, the higher rate of obsolescence, the smaller the total value of private benefits from the innovation, and therefore, the less intense the research effort will be. Moreover, since research produces a stock (knowledge) whose benefits accrue over the future, the optimal research intensity will vary inversely with the rate of discount (see Lucas 1967 for an empirical test on aggregate data). Finally, the longer the gestation lag, the larger the influence of the future is in determining the returns to R & D, and hence, the more important the effect of expected growth on the optimal R & D intensity.[5]

The model presented here posits a set of firms that produce knowledge from research resources and produce output by combining this knowledge with traditional factors of production. The price of output is determined by the cost of traditional factors plus quasi-rents generated by temporary monopoly power over the information produced by the research resources. It is important to realize that there will be no private benefits from the employment of research resources without some degree of monopoly power. The unique characteristics of knowledge as a commodity imply that the private rate of return to research resources must be determined jointly by the parameters of the production function for knowledge and the ability of the firm to internalize the benefits from the knowledge it produces.

We would like to clarify the relationship between our model and Schmookler's (1966) celebrated work on demand inducement. Schmookler argued that the *level* of inventive activity is directly related to the absolute size of the market for the output of such activity. By focusing on the determination of *research intensity*, our model normalizes by the current level of output, and further differences in expected market size are associated with the expected rate of growth in demand. Of course, one would not expect that equations relating research intensity to ex-

4. This should be distinguished from the role of market size in models of the demand for traditional capital. The level of investment in traditional capital is related to the expected growth of output (accelerator models), whereas in our model the level of investment in the stock of R & D depends on the expected level of output.

5. This does not mean, however, that an increase in θ raises the optimal R & D intensity, since θ affects both β_0 and β_1.

pected growth would fit as well as those relating the level of research to the current level of output. Nevertheless, we will show that the explanatory power of growth rates remains substantial, at least at higher levels of aggregation.

9.3 Stochastic Specification and Identification of the Model

The equations for the optimal intensities of research capital and research labor (10), together with that for traditional labor, form the basis of the model to be estimated. In this section we add appropriate disturbance terms and consider the identification of the model's parameters. Letting asterisks denote the optimal levels of each variable, we have

$$(11)\qquad \begin{aligned} \log(w_r L^*/Q^*) &= \log\beta_0 + \alpha g^* + k_i, \\ \log(p_c C^*/Q^*) &= \log\beta_1 + \alpha g^* + k_i, \\ \log(wN^*/Q^*) &= \log\gamma_1 . \end{aligned}$$

The variable denoted by Q^* is expected output, that is, the value of output on which input decisions are made. We follow Mundlak and Hoch (1965) in assuming "partial transmission" of the error in output to the input decision-making process. Letting the superscript o denote the observed value of a variable, η_q represent disturbances and firm-specific characteristics known before input decisions are made, and ν_q reflect transitory disturbances realized after inputs are chosen, we have[6]

$$(12)\qquad Q^o = \gamma_0 K N^{\gamma_1} H^{\gamma_2} e^{\eta_q + \nu_q}, \text{ and } Q^* = E(Q^o | \eta_q),$$

so that $Q^o = Q^* e^{\nu_q}$, and we define $\sigma_q^2 = E(\nu_q^2)$.

The observed level of each factor of production differs from its optimal level by an error which has two components: a decision component resulting from an inoptimal choice of factor levels and a pure measurement component. Letting ϵ_j be the sum of the two errors for factor j, we have

$$(13)\qquad C^o = C^* e^{\epsilon_c}, \; L^o = L^* e^{\epsilon_\ell}, \text{ and } N^o = N^* e^{\epsilon_n},$$

where $E(\epsilon_j) = 0$ and $V(\epsilon_j) = \sigma_j^2$ for $j = c, \ell, n$.

To complete the model two further assumptions need to be made, one on the structure of the covariance matrix of the error components (ν_q and ϵ_j for $j = \ell, c, n$) and one providing an empirical measure of the expected growth rate (g^*). For expositional clarity we first describe the identification scheme under the assumption that all the error components are

6. Familiar special cases of partial transmission are full transmission ($\nu_q = 0$; Marschak and Andrews 1944) and zero transmission ($\eta_q = 0$; Zellner, Kmenta, and Drèze 1966).

mutually uncorrelated. The model actually estimated allows for free covariances between the $\epsilon_j (j = \ell, c, n)$ and a test of whether it is reasonable to assume that ν_q is uncorrelated with them. The extensions required to estimate the more general model are briefly summarized at the end of this section. Finally, the empirical results in sections 9.4 and 9.5 are based on the assumption that a firm's expected growth rate equals its average past growth rate plus a component reflecting common expectational changes in the trend of industry demand, that is,

$$g_i^* = \Delta g + g_i \text{ for } i = 1, \ldots, n, \tag{14}$$

where g_i is the average past growth rate of firm i, and Δg is the commonly held, expected difference between the average past and the expected future growth rates. In section 9.5, where we use more flexible data seta than the ones used in section 9.4, we try alternative empirical specifications of g^*, but these alternatives do not change our basic conclusions.

For the remainder of this section it will prove convenient to redefine all variables as deviations from their sample means. With this understanding, substitution of (12), (13), and (14) into (11) yields the following system of factor share equations:

$$\log(w_r L^o/Q^o) = \alpha g + k_i + \epsilon_c - \nu_q, \tag{15a}$$

$$\log(p_c C^o/Q^o) = \alpha g + k_i + \epsilon_\ell - \nu_q, \tag{15b}$$

$$\log(wN^o/Q^o) = \epsilon_n - \nu_q. \tag{15c}$$

Assuming that the structural parameter, k_i, is uncorrelated with the various error components and with g (see below), maximum likelihood estimation provides consistent and asymptotically efficient estimates of α and of the variance-covariance matrix of disturbance Ω, where

$$\Omega = \begin{bmatrix} \sigma_k^2 + \sigma_c^2 + \sigma_q^2 & & \\ \sigma_k^2 + \sigma_q^2 & \sigma_k^2 + \sigma_\ell^2 + \sigma_q^2 & \\ \sigma_q^2 & \sigma_q^2 & \sigma_q^2 + \sigma_n^2 \end{bmatrix}. \tag{16}$$

The identification of the various components from (16) is straightforward. Any factor which affects the returns to the production of knowledge will affect the optimal intensities of both research resources. Consequently, the covariance between the disturbances in the two research intensity equations will capture σ_k^2. However, this covariance also picks up any measurement or expectational error in output, σ_q^2. Since the traditional labor intensity equation will also contain the error in output, σ_q^2 can be identified by the covariance between the research intensity and the traditional labor demand equations. Finally, the variances of the errors in the research resource variables are calculated as the residual portion of the research intensity equations.

The parameters from (15) and (16) permit a decomposition of the variance in research intensity into three components: (1) variance caused by differences in the expected growth rate of the internal appropriability base of the firm ($\alpha^2\sigma_g^2$), or demand inducement; (2) variance caused by differences in the structural parameter (σ_k^2), which determines the private benefits accruing to a cost-reducing innovation, given the internal appropriability base of the firm; and (3) variance caused by measurement and decision errors in research resources and in expected output.

We would like to explore briefly the economic interpretation of this decomposition and indicate caveats concerning the distinction between technological opportunity, appropriability, and demand inducement as determinants of R & D intensity. Put simply, the optimal R & D intensity depends on the supply of new knowledge and the effective demand for that knowledge. Given factor prices, differences across firms or industries in the supply curve for new knowledge reflect differences in the parameters of the underlying knowledge production function. This is the precise meaning we have given to technological opportunity in the earlier discussion. The effective demand for new knowledge depends on the current level (by which we normalize) and the expected rate of growth in the demand for products that embody the new knowledge (demand inducement) and on the ability of the firm to capture the benefits from the market (appropriability). If a measure of the expected shift in the product demand curve were available, it would serve to identify empirically the role of demand inducement separately from the joint contribution of technological opportunity and appropriability. However, the available measures are based on realized growth rates of output reflecting shifts in both the product supply and demand curves for the firm. Therefore, these growth rates will be positively correlated with the other structural determinants of R & D intensity in our model, namely, technological opportunity and appropriability. We have developed a model which endogenizes the firm's expected growth in output, and it generates structural equations similar to those in this paper. The main difference is the positive correlations referred to above, in particular, k may be correlated with g. As a result, our empirical estimates in section 9.4 represent the reduced form association between growth rates of output and R & D intensity and overstate the importance of pure demand inducement.[7] In section 9.5 we explore some aspects of the association among the structural determinants and demonstrate their empirical importance.

Our main focus is on total research intensity and we now derive an

7. If $E(kg) \neq 0$, the estimates of α derived from the models relying on zero correlation should differ from estimates based on other techniques. An assortment of exogenous information on the components of α (described in section 9.4) yields estimates of α similar to those obtained from our models, at least in the intraindustry regressions, and this may be interpreted as an indirect test of the assumption $E(kg) = 0$. At the interindustry level this problem has more empirical content, and we explore it in greater depth in section 9.5.

equation for that variable. The observed research expenditures of a firm are calculated as the sum of the firm's expenditures on research labor and research capital. That is,

$$R^o = w_r L^o + p_c C^o. \tag{17}$$

Since we define research capital to include all R & D expenditures other than payments to scientists and engineers, (17) is an identity. Analogously, we define the optimal level of research expenditures, R^*, as

$$R^* = w_r L^* + p_c C^*. \tag{18}$$

It follows from (15a), (15b), (17), and (18) that

$$\log(R^o/Q^o) = \alpha g + k_i + \epsilon_r - \nu_q, \tag{19}$$

where $\epsilon_r = \psi\epsilon_\ell + (1 - \psi)\epsilon_c$, and $\psi = a/(a + b)$. Since equations (15a), (15b), and (19) are definitionally related, only two of these equations contain independent information. The form of the data made it simpler to estimate (15a) and (19) together with (15c).

We have proceeded on the assumption that the various error components are mutually uncorrelated. As indicated earlier, we actually estimate a more complicated six-equation model which allows for free correlation among the ϵ_j ($j = c, \ell, n$) and a test of whether ν_q is correlated with them. Details of this model and its identification scheme are contained in Pakes (1978) and Schankerman (1979). Briefly, the six-equation model is constructed by adding the factor demand equations for research expenditures, research labor, and traditional labor in year $(t - 1)$ to those same equations for year t. Each error component is assumed to be generated by an arbitrary, stationary stochastic process. The model allows for a χ^2_{20} test of the stationarity assumptions (T_1), a χ^2_8 test of the assumption of no correlation between ν_q and the factor errors (T_2), and a χ^2_4 test of the intertemporal stability of the coefficient of g (T_3). T_1, T_2, and T_3 test for consistency between the data and the assumptions used to identify the model. There are also nonnegativity restrictions on all the estimated error variances. Since these restrictions are equivalent to a ranking of the elements of the covariance matrix and as such are not guaranteed by our estimating procedure, the nonnegativity conditions constitute an informal test of the model.

The six-equation model also allows us to investigate two aspects of the intertemporal stability of the unobservable structural parameter, k. First, we provide a test of whether the interfirm variance in k is constant over time. Second, we can estimate the correlation coefficient between the values of the structural parameter for a given firm between two adjacent years, which we denote by λ.

9.4 Empirical Results at the Intraindustry Level

The data were gathered jointly by the National Science Foundation and the Bureau of the Census (for a more complete description see Griliches 1980). They contain company information on R & D expenditures, the number of scientists and engineers, total employment, value added, and a variety of other company economic indicators. The data include observations on one level year value and a corresponding growth rate for most variables. The sample used here consists of 433 large firms which account for 48 percent of all R & D performed in American industry in 1963, and 78 percent of all R & D excluding aircraft and missiles.[8] The firms are broken down into four broad industry groups—chemicals and petroleum, electrical and communications equipment, fabricated metals products and machinery, and motor vehicles and other transport equipment—and the analysis is performed on each of these industries separately.

The six-equation version of equations (15a), (15c), and (19) is the model which we estimate.[9] Before presenting the empirical results, however, exogenous information is used to derive a plausible range for α. Recall that $\alpha = [1/(r+\delta)] + \theta$, where r,δ, and θ are the discount rate, the decay rate in appropriable revenues accruing to the innovation, and the mean lag between the outlay of research resources and the beginning of the associated revenue stream, respectively. A comparison of exogenous information on the value of α with the direct estimates here will provide an informal test of the assumptions of the model. The estimates of δ and θ (taken from Pakes and Schankerman, chap. 4 in this volume) range between 0.18–0.36 and 1.2–2.5 (years). Based on a discount rate of 0.15, these estimates provide an approximate range of $3 < \alpha < 5$.

The model was estimated using a full information maximum likelihood technique developed by Jöreskog (1973). A summary of the empirical results is presented in table 9.1. The computed values (pooled across industries) for the test statistics T_1 (χ^2_{20}), T_2 (χ^2_8), and T_3 (χ^2_4) are 29.80,

8. The original sample consists of 883 firms. We discarded the data for the "aircraft and missiles" and the "all others" industries. The first was dropped because of inconsistencies in the data and because it is dominated by government-financed R & D (74 percent versus 20 percent in the other industries). Our market-inducement model has limited applicability for government-financed R & D unless it were known that privately financed and government-financed R & D are close substitutes and that the supply of the latter is very elastic. The "all others" category was discarded on the grounds that it contains both intraindustry and interindustry variance in R & D intensity, a critical distinction as we show in section 9.5.

9. Two points should be noted. First, the data include both the average past growth rate in sales and in value added for each firm. To allow each variable to contain measurement error, we use both variables and identify σ^2_g from the covariance between the two (see Pakes and Schankerman 1977 for details). Second, the parameter $\psi = a/(a+b)$ is measured as the share of scientists and engineers in total R & D expenditures, constructed for each industry from information in National Science Foundation (1963).

Table 9.1 **Summary of Results of the Six-Equation Model**

	Industry			
Parameter	Chemicals and Petroleum	Metal Products and Machinery	Electrical and Communications Equipment	Motor Vehicles and Transport Equipment
1. α	4.10	2.62	5.13	3.49
2. Standard error of α	2.26	1.30	2.40	3.40
3. σ_r^2	0.33	0.20	0.37	0.58
4. $\sigma_r^2/\sigma^2_{\log R^o}$	0.12	0.07	0.08	0.09
5. $(\sigma_r^2+\sigma_q^2)/\sigma^2_{\log R^o/Q^o}$	0.23	0.28	0.21	0.40
6. $\alpha^2\sigma_g^2/\sigma^2_{\log R^*/Q^*}$[a]	0.04	0.04	0.06	0.03
7. $\sigma_k^2/\sigma^2_{\log R^*/Q^*}$[b]	0.96	0.96	0.94	0.97
8. λ	1.00	0.99	0.99	0.99
9. n	110	187	102	34

[a] $\sigma_g^2 = \text{cov}(g_1 g_2)$, where g_1 and g_2 are the measured average past growth rates of sales and value added, respectively.

[b] $\sigma^2_{\log R^*/Q^*} = \sigma^2_{\log R^o/Q^o} - \sigma_r^2 - \sigma_q^2$.

0.32, and 1.32, respectively. None of these values is surprising under the null hypothesis that the constraints are indeed satisfied. It is noteworthy that the value of T_2 indicates strong acceptance of the assumption of a zero covariance between the transitory error in output and the factor errors in this sample. The difference between the sum of squared residuals in the model using all three test constraints and in the totally unconstrained model can be used to produce a χ^2_{32} test of the validity of the model as a whole. The computed value of the χ^2_{32} statistics is 32.76, which is about equal to the expected value of a χ^2_{32} deviate under the null hypothesis. As noted above, there is an additional test of whether the interfirm variance in k is stable over time. The observed value of the χ^2_4 deviate (combined over the four industries) for this test is 6.64. While this indicates acceptance of the hypothesis at the 5 percent level, a sample with more than two time periods would be required to determine more conclusively whether the variance in the structural parameter is in fact constant over time. We also note that of the twenty-four error variances estimated in the model (σ_j^2 for $j = \ell, c, r, n, q, k$ in each industry), only two violated the nonnegativity restriction (see Pakes and Schankerman 1977 for details).

All of the estimated α coefficients are of the right sign, and three are statistically significant. Moreover, all of the four point estimates of α are within or very near the interval predicted by the prior information summarized earlier. To derive a summary measure of α, we tested the null hypothesis that the differences between the various estimates of α are simply a result of random differences in the estimators. The hypothesis is accepted. The value of α for the combined sample is 3.37 with a standard error of 0.95. On the whole, the data and the exogenous information provide mutually consistent information on the magnitude of the parameters determining α.

We now turn to the basic decomposition of the intraindustry variance in research intensity. Line 5 in table 9.1 indicates that an average of 28 percent of the variance in observed research intensity is attributable to errors (of measurement and decision), and hence, 72 percent of the variance is accounted for by the structural determinants in the model. Of special interest is the effect of the firm's past growth rate. Though this variable is neither statistically nor economically insignificant in determining the firm's R & D intensity,[10] line 6 indicates that differences in growth rates account for only a minor portion (3–6 percent) of the structural intraindustry variance in R & D intensity. Morevoer, this finding is robust to different specifications of the expected growth rate variable (see discussion in section 9.5). It is evident that a pure demand inducement mechanism does not do well in explaining the *intraindustry* variance in R & D intensity. As noted earlier, this finding does not contradict Schmookler's argument for demand inducement, which is cast in terms of the level of R & D and the absolute size of the market.

As indicated in line 7, over 95 percent of the structural variance in R & D intensity is picked up by differences in the firm-specific structural parameter, k, which we have interpreted as reflecting appropriability conditions facing the firm (but which may also include intraindustry differences in technological opportunity). It is also noteworthy that the value of the structural parameter associated with a given firm seems to be stable, at least over short periods of time, since the correlation coefficient (λ) between the values of k_t and k_{t-1} is essentially unity in all industries.

Finally, line 4 provides the fraction of the variance in measured R & D expenditures attributable to errors in research resources. The average value is quite large, 9 percent. Unfortunately, it is not possible at this stage to determine what fraction of this error variance is caused by pure measurement error, as opposed to other factors that cause inoptimal choices of R & D intensity in the context of our model.

9.5 Aggregation Effects and the Interindustry Variance in R & D Intensity

In the previous sections we presented a model of R & D intensity at the micro level and explored the empirical determinants of the intrain-

10. The elasticity of R & D intensity with respect to past growth rates, evaluated at the sample mean of the growth rate, is about 0.25.

dustry variance in research intensity. We now explicitly aggregate the micro equation and empirically examine the determinants of the interindustry variance in research intensity (i.e., the variance in the average R & D intensities of different industries). Based on the NSF industrial classification (which is roughly at the two-digit SIC level), the total variance in R & D intensity is about equally divided between intraindustry and interindustry variance.

The Griliches-Census data (GD) used in section 9.4 do not contain sufficient industrial detail to investigate the interindustry variance in research intensity. The main data set used here is constructed from information contained in *Business Week* (see Pakes 1979 for details). These data (BWD) contain the ratio of company-financed R & D to sales, five-year average past growth rates of sales, and a fairly detailed industrial classification for 536 firms, which account for the vast majority of company-financed R & D in the United States in 1976. The BWD do not contain the full set of variables required to estimate the entire model in section 9.3. However, the main points we wish to emphasize in this section can be demonstrated by focusing on the R & D intensity equation.[11]

To analyze these data we let the index (i, j) refer to firm i in industry j ($i = 1, \ldots, N_j$ and $j = 1, \ldots, J$), and for simplicity we define $y_{ij} = \log(R^o_{ij}/Q^o_{ij})$. Then the R & D intensity equation of the model (see eq. [19]) can be written

$$y_{ij} = \alpha_{0j} + \alpha_j g_{ij} + \mu_{ij}, \tag{20}$$

where $\mu_{ij} = k_{ij} + \epsilon_{r,ij} - v_{q,ij}$, and by assumption, $E(\mu_{ij}) = E(\mu_{ij} g_{ij}) = 0$, and $E(\mu_{ij}\mu_{i',j'}) = \sigma^2_\mu$, if $i = i'$ and $j = j'$, and zero otherwise.[12]

Recall that the micro (intraindustry) coefficient on the expected growth rate is $\alpha_j = 1/(r_j + \delta_j) + \theta_j$, and that the GD used in section 9.4 indicated acceptance of $H^0_1 : \alpha_j = \alpha$ for available j. Seventeen NSF industries are available in the BWD, and the computer χ^2_{16} test statistic for H^0_1 on the BWD is 14.71, also indicating acceptance of H^0_1. Hence, in the remainder of the discussion we maintain H^0_1. The parameter α_{0j} depends on the determinants of α_j and, in addition, on the indices of the technological opportunities and the average degree of appropriation (a_j, b_j, and $\bar{k}_j$) in the industry. Testing $H^0_2 : \alpha_{0j} = \alpha_0$ for $j = 1, \ldots, J$ on the BWD

11. In previous work we estimated the entire three-equation model on aggregate data (industry means) on a data set we constructed by combining information from the annual reports of the NSF and the Census of Manufactures (see Pakes and Schankerman 1977). There are no essential differences between those estimates and the ones reported in this section.

12. We are assuming that σ^2_μ does not vary across industries even though α_{0j} and α_j may. This simplifies the presentation without affecting our major results. In general we keep the discussion of technical details in this section very brief since they can be found in Pakes (1978, 1983) and the literature cited there.

yields an $F(16,512)$ test statistic of 24.20. The 5 percent critical value is 1.67, so H_2^0 is clearly rejected. We conclude that, though there are no perceptible interindustry differences in the micro growth rate coefficient, there are clear interindustry differences in the constant terms.

Since we accept H_1^0 and reject H_2^0, equation (20) is formally identical to econometric models that allow for group effects. Under the assumption that the average past growth rate provides a reasonable approximation to the expected growth rate relevant to R & D decisions (which we discuss below), the presence of a group effect indicates that there is a determinant of R & D intensity which is common to all firms in an industry but differs across industries. Since we show later that this group effect is an important determinant of the interindustry variance in R & D intensity, we now explore its characteristics in more detail. To do so we note that one can always define α_0 and ϕ such that

$$\alpha_{0j} = \alpha_0 + \phi g_{\cdot j} + \zeta_j , \tag{21}$$

where $g_{\cdot j} = N_j^{-1} \sum_{i=1}^{N_j} g_{ij}$, and the mean and the sample covariances of ζ_j with g_{ij} and $g_{\cdot j}$ are all zero by construction. We assume that ζ_j are random draws from a common population that satisfy the mean and the covariance restrictions stated above and define $\sigma_\zeta^2 = E[\zeta_j^2]$. Equation (21) partitions the group effect into a part correlated with the past industry growth rate $(\phi^2\sigma_{g_{\cdot j}}^2)$ and into a part not correlated (σ_ζ^2). The parameter ϕ may be interpreted as the reduced form response of R & D intensity of a firm to a unit increase in its industry growth rate, holding constant its own growth rate.

Using H_1^0 and substituting (21) into (20), the micro R & D intensity equation becomes

$$y_{ij} = \alpha_0 + \alpha g_{ij} + \phi g_{\cdot j} + v_{ij} , \tag{22}$$

where $v_{ij} = \zeta_j + \mu_{ij}$, $E(v_{ij}) = E(v_{ij}g_{ij}) = E(v_{ij}g_{\cdot j}) = E(\zeta_j\mu_{ij}) = 0$, and the covariance matrix of v_{ij} has a standard error components structure. Summing (22) over i and dividing by N_j, we obtain the corresponding interindustry R & D intensity equation:

$$y_{\cdot j} = \alpha_0 + (\alpha + \phi)g_{\cdot j} + \zeta_j + \mu_{\cdot j} , \quad i = 1, \ldots , J. \tag{23}$$

Clearly, the determinants of the interindustry variance in R & D intensity are a mixture of the determinants of the intraindustry variance and the determinants of the variance in α_{0j}. The growth rate coefficient from the intraindustry regression (20) (with $\alpha_j = \alpha$ for all j) provides an unbiased estimate of the firm's response in R & D intensity to an increase in its own growth rate, holding constant the group effect. The interindustry growth rate coefficient from (23) provides an unbiased estimate of the

Table 9.2 Aggregation Effects in the BWD[a]

Parameter	Intraindustry (industry specific constant terms) (1)	Weighted Aggregate between Industry[b] (2)	Mixed[c] Effects (3)
Micro growth rate, α	1.04 (0.31)	—	1.04 (0.31)
		10.96 (1.88)	
Industry growth rate, ϕ	—	—	9.93 (1.91)
σ^2	0.39[d]	n.r.	n.r.[e]
R^2	0.02	0.66	n.r.
Degrees of freedom	512	15	527

[a]Numbers in parentheses are standard errors. The letters n.r. mean "not relevant."

[b]The weight for industry j is $(\hat{\sigma}_\zeta^2 + \hat{\sigma}_\mu^2/N_j)^{-1/2}$.

[c]Estimated by generalized least squares using $\hat{\sigma}_\mu^2$ and $\hat{\sigma}_\zeta^2$ from notes d and e.

[d]$\hat{\sigma}_\mu^2 = 0.39$.

[e]First stage $\hat{\sigma}^2 = 0.51$. It follows that $\hat{\sigma}_\zeta^2 = 0.51 - \hat{\sigma}_\mu^2 = 0.12$.

sum of this response plus the response of the group effect to a unit increase in the industry's past growth rate. The intraindustry and interindustry coefficients would be similar only in the special case where the determinants of the α_{0j} are uncorrelated with the past industry growth rate.

Table 9.2 summarizes the empirical results for the intraindustry regression under H_1^0 (column [1]), the interindustry regression (column [2]) and the "mixed effects" model in equation (23) (column [3]). Column (1) indicates that the intraindustry growth rate coefficient is similar to (but somewhat smaller than) the estimates obtained with the GD in section 9.4.[13] It indicates that growth rates account for very little (about 2 percent) of the intraindustry variance in research intensity, which confirms our earlier results with the GD.

The interindustry regression yields very different results. The estimate of the aggregate mean response in R & D intensity to a unit increase in the industry growth rate is about ten times as large as the firm's response to its own growth rate in the intraindustry regression. As a consequence, growth rates account for over 65 percent of the interindustry variance in

13. There are several differences between the two data sets which could account for this difference: the BWD use the ratio of company-financed R & D to sales while the GD use the ratio of total R & D to sales; the GD are at a slightly higher level of aggregation than the BWD; the estimating technique on the GD allows for an error in the measurement of g (see note 9 and the discussion below); the g used for the GD is based on a slightly longer average of past years than the BWD; and the two data sets are for different years.

research intensity. As the results in the mixed effects model indicate, however, this is not a result of firms' responses to their own growth rates. Rather, it reflects the fact that the industry growth rate is highly correlated with factors at the industry level which stimulate research activity of all firms in the same industry (compare α and ϕ in column [3]).

The BWD also contain the three-digit industrial classification of the firm. We use this information to examine whether the three-digit assignment of the firm exerts any independent influence (beyond the two-digit classification) on its choice of R & D intensity and whether the factors underlying this influence are correlated with the past industry growth rate at the three-digit level. The procedure we use is a generalization of the one outlined in equations (20)–(23). We allow the constant term in the R & D intensity equation to vary across both two-digit and three-digit industries and partition the constant term into a part correlated with the past two- and three-digit industry growth rates (with coefficients ϕ and ϕ_*, respectively) and a part uncorrelated with those growth rates. The test of whether there is any variation in the constant terms across three-digit (within two-digit) industries yields a computed $F(25{,}484)$ test statistic of 1.98, marginally significant at the 1 percent level of significance. There is weak evidence of variation in the constant terms across three-digit industries. Estimation of the mixed effects model with both two-and three-digit industry growth rates yields estimates of ϕ and α which are almost identical to those reported in table 9.2, while the estimate of ϕ_* is 0.02 (standard error 1.31). Hence, there is not much evidence of a group effect varying among three-digit (within two-digit) industries, and whatever effect there is does not seem to be related to the past growth rate of the three-digit industry.

We next ask whether the results in table 9.2 could reflect nothing more than a misspecification of the expected growth rate relevant to a firm's R & D decisions, g^*. We consider two alternative specifications that may predict the type of discrepancy between the micro and aggregate growth rate coefficients which we observe in table 9.2.

In the first we allow the average past growth rate to measure expected growth with an independently distributed and uncorrelated error, $g_{ij} = g^*_{ij} + v_{ij}$, where v_{ij} has a zero mean and is uncorrelated with g^*_{ij}. This is the classical "errors in variable" model (EVM), suggested in the literature as an alternative explanation of differences in estimated coefficients at different levels of aggregation (Aigner and Goldfeld 1974; Eisner 1978). The motivation for the EVM is that under the stated assumptions $\operatorname*{plim}_{N_j \to \infty} \gamma_{\cdot j} = 0$, so that $g_{\cdot j}$ converges in distribution to g^*_j. In the interindustry regression the error in $g^*_{\cdot j}$ averages out so the estimated coefficient does not contain errors in variable bias, but the intraindustry coefficient is biased downward.

We use two approaches to this problem. The first is to use an instrumental variables estimator with past rates of growth of employment (g_n) as instruments (see Pakes 1979 for details). This approach relies on the assumption that $E(g_n\nu) = 0$. The second technique uses a three-digit industrial classification as a grouping device to average out the measurement error, and then obtains an estimate of ϕ from a comparison of the interindustry regression at the NSF (two-digit) level with the regression between three-digit (but within two-digit) industries (see Pakes 1983). This provides an asymptotically unbiased estimate of ϕ regardless of errors in variables and a test of the presence of such measurement error. The results indicate acceptance of the hypothesis that there are no errors in variables. Both estimation procedures yield estimates of ϕ which are nearly identical to the estimate in table 9.2. The instrumental variables estimate of α is 1.55 (standard error 0.40), which is slightly larger than the estimate in table 9.2 but does not change our basic conclusions.

The alternative specification we consider assumes that the firm forms a rational forecast by taking the expectation of its future growth rate conditional on the information available to it in period t, $g^* = E_t g_{t+1}$. This implies $g_{t+1} = g^* + \omega$, where $E_t\omega = 0$, so that the actual future growth rate measures expected growth subject to an error uncorrelated with all variables known to the firm in period t (including the firm's and the industry's past growth rates). To obtain asymptotically unbiased estimators, we substitute the future growth rate for g_{ij} in (23), and use g_{ij} and $g_{\cdot j}$ as instruments on the future growth rate. Since the BWD did not contain future growth rates, additional sources of information were used and the analysis was conducted at a somewhat different level of aggregation (see Pakes 1979 for details). Nonetheless, this rational expectations formulation of g^* yields the same basic results as those reported in table 9.2.

We conclude that the evidence does not support the hypothesis that the differences between the intraindustry and interindustry results reported in table 9.2 are the result of a misspecification in the measure of the expected growth rate. These experiments may not dispose of the issue entirely, but they do indicate that misspecifications which average out over firms (within an industry) and the use of the past industry growth rate by firms to predict their own growth do not explain the results in table 9.2.

We now summarize the empirical characteristics of the group effect, which appears to be the dominant determinant of the interindustry variance in research intensity. First, it is associated with the NSF (roughly two-digit) industry to which the firm belongs. The three-digit industrial classification of the firm has little independent influence on the firm's R & D intensity. Of course, there may be other more detailed classification schemes that group firms with common industry factors affecting their R & D intensities. Second, the factors in the industry environment affecting the firms' R & D intensities are highly and positively correlated with the industry's past growth rates. Various experiments on other data sets (not reported here) indicate that the industry growth rate coefficient is larger, the longer the term of the past growth rate used (we experimented with values from four to eight years), and that the values of the group effect for different industries are fairly stable over time. The factors affecting the R & D intensities of firms in an industry appear to be associated with sustained, long-term past growth. Third, the evidence does not support the hypothesis that the observed group effect is simply a result of the industry growth rate acting as an indicator of the firm's expected growth rate. Finally, the group effect provides an explanation for the basic empirical anomaly that growth rates account for only a minor portion of the intraindustry variance in R & D intensity but for about 65 percent of the interindustry variance.

We have established both the importance of and certain empirical characteristics of the group effect. At this stage, we cannot determine the underlying mechanisms generating it. Within the context of our model, the industry constant terms contain indices of technological opportunity and the average degree of appropriability. This suggests a reduced form, empirical association between technological opportunity, appropriability, and a broader concept of demand inducement. As Schmookler (1966, pp. 176–77) put it:

> . . . science and engineering appear as given, to be used to explain but not themselves to be explained. In the larger context, however, these too would require explanation. I believe that their explanation, at least for modern times, would probably not differ greatly from that advanced here for invention. The rate and direction of scientific and engineering process are probably greatly affected by demand, subject to the constraints imposed by man's innate abilities and by nature. . . . If this view is approximately correct, then even if we choose to regard the demand for new knowledge for its own sake as a non-economic phenomenon, the growth of modern science and engineering is still primarily a part of the economic process.

9.6 Concluding Remarks

In the literature on the determinants of research demand by private firms one can identify three leading hypotheses: expected market size for the output of R & D activities, the degree to which firms can appropriate the benefits from the industrial knowledge they produce, and the technological opportunities facing firms reflecting the set of production possibilities for transforming research resources into innovations. The model we presented provides a first step toward integrating these hypotheses into a formal framework capable of investigating the empirical determinants of R & D demand. Our framework is only a first step because of its partial equilibrium character and because it does not model explicitly the underlying mechanisms determining the degree of appropriability, technological opportunities, and the expected market size of firms. The model results in equations determining the intensity of use of research inputs as a function of these three factors and disturbance terms. The market size relevant to current R & D decisions depends on current output and the expected growth rate. Since R & D intensity measures research effort relative to current output, further differences in expected market size are associated in our model with differences in expected growth rates. The limitations of the model and of the available data require us to specify technological opportunity and appropriability as unobservable factors in a way that permits us to assess their (joint) empirical contribution to the observed variance in research intensity within and across industries.

The empirical results indicate that at the intraindustry level of aggregation about three-quarters of the large observed variance in research intensity is structural, in the sense that it is consistent across diffrent factor demand equations. However, while the growth rate coefficient is broadly consistent with the predictions of the micro model, the variance in growth rates accounts for very little (less than 5 percent) of the intraindustry structural variance in research intensity. Most of the variance is attributed to differences in appropriability and technological opportunities within industries. The results at the interindustry level of aggregation are strikingly different. Growth rates account for the majority of the interindustry variance in research intensity. The evidence suggests that this finding is not due to differences in firms' responses to their own growth rates. Rather, it appears to be due to factors in the industry environment that affect the research intensities of all firms within the industry and that are highly and positively correlated with past industry growth.

The theoretical framework and these stylized empirical facts suggest certain fruitful lines for future research. The first is to model the determinants and disentangle the empirical contributions of appropriability and technological opportunity at the intraindustry level of aggregation. Second, work is needed to understand the causal nexus underlying the empirical association between the industry effect on the choice of R & D intensity and past industry growth.

References

Aigner, Dennis, and Stephen Goldfeld. 1974. Estimation and prediction from aggregate data when aggregates are measured more accurately than their components. *Econometrica* 42:113–34.

Arrow, Kenneth. 1962. Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, ed. R. R. Nelson. Universities-NBER Conference Series no. 13. Princeton: Princeton University Press for the National Bureau of Economic Research.

Eisner, Robert. 1978. *Factors in business investment*. National Bureau of Economic Research General Series no. 102. Cambridge, Mass.: Ballinger.

Griliches, Zvi. 1973. Research expenditures and growth accounting. In *Science and technology in economic growth*, ed. B. R. Williams, 59–83. London: Macmillan.

———. 1980. Returns to research and development expenditures in the private sector. In *New developments in productivity measurement and analysis*, ed. J. W. Kendrick and B. N. Vaccara, 419–54. Conference on Research in Income and Wealth: Studies in Income and Wealth, vol. 44. Chicago: University of Chicago Press for the National Bureau of Economic Research.

Griliches, Zvi, and Jacob Schmookler. 1963. Inventing and maximizing. *American Economic Review* 53:725–29.

Jöreskog, Karl. 1973. A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasurable variables (LISREL). Princeton University.

Lucas, Robert. 1967. Tests of a capital-theoretic model of technological change. *Review of Economic Studies* 34:175–89.

Machlup, Fritz. 1962. *The production and distribution of knowledge in the United States*. Princeton: Princeton University Press.

Marschak, Jacob, and W. H. Andrews. Random simultaneous equations and the theory of production. *Econometrica* 12:143–205.

Mundlak, Yair, and I. Hoch. 1965. Consequences of alternative specifications in estimation of Cobb-Douglas production functions. *Econometrica* 33:814–28.

National Science Foundation. 1963. *Basic research, applied research, and development in industry*. Washington, D.C.: GPO.

Nordhaus, William D. 1969a. *Invention, growth, and welfare: A theoretical treatment of technological change*. Cambridge: MIT Press.

———. 1969b. An economic theory of technological change. *American Economic Review* 59:18–28.

Pakes, Ariel. 1978. Economic incentives in the production and transmission of knowledge: An empirical analysis. Ph.D. diss., Harvard University.

———. 1979. Aggregation effects and panel data estimation problems: An investigation of the R & D intensity decision. Discussion Paper 699. Harvard Institute of Economic Research.

———. 1983. On group effects and errors in variables in aggregation. *Review of Economics and Statistics* 64:168–73.

Pakes, Ariel, and Mark Schankerman. 1977. A decomposition of the intraindustry and interindustry variance in research intensity in American manufacturing: Demand inducement, technological opportunity, and appropriability. Harvard University. Manuscript.

Rosenberg, Nathan. 1963. Technological change in the machine tool industry, 1840–1910. *Journal of Economic History* 23:414–43.

———. 1969. The direction of technological change: Inducement mechanisms and focusing devices. *Economic Development and Cultural Change* 18, no. 1: 1–24.

———. 1974. Science, invention, and economic gowth. *Economic Journal* 84:90–108.

Schankerman, Mark. 1979. Essays on the economics of technical change: The determinants, rate of return, and productivity impact of research and development. Ph.D. diss., Harvard University.

Scherer, F. M. 1965. Firm size, market structure, opportunity, and the output of patented inventions. *American Economic Review* 55: 1097–125.

Schmookler, Jacob. 1966. *Invention and economic growth*. Cambridge, Mass.: Harvard University Press.

Schumpeter, Joseph. 1950. *Capitalism, socialism, and democracy*. 3d ed. New York: Harper and Row.

Zellner, A., J. Kmenta, and J. Drèze. 1966. Specification and estimation of Cobb-Douglas production function models. *Econometrica* 34:784–95.