NBER WORKING PAPER SERIES

AGGREGATION EFFECTS AND PANEL DATA ESTIMATION PROBLEMS: AN INVESTIGATION OF THE R&D INTENSITY DECISION

Ariel Pakes

Working Paper No. 344

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

May 1979

The research reported here is part of the NBER's research program in Productivity. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper 344
May 1979

Aggregation Effects and Panel Data Estimation Problems: An Investigation of the R&D Intensity Decision

SUMMARY

This paper considers why the determinants of the inter- and intra-industry variance in R&D intensity in U.S. manufacturing differ markedly even though response parameters are similar across industries. A similar aggregation effect is noted by Grunfeld and Griliches (1960), and this paper gives that effect operational content in terms of grouped data estimation procedures. Observationally equivalent aggregation results can be generated by errors in variables models (see Aigner and Goldfeld [1974]). A later section considers specifications which identify the empirical importance of both these problems. Finally, a summary of the empirical results on the determinants of R&D intensity is provided.

Ariel Pakes
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

(617) 868-3921

economic causes of these effects. Section 2 provides an economic explanation and a simple testing procedure for each type of aggregation effect, as well as presenting examples of the economic importance of the Grunfeld-Griliches aggregation effect. In Section 3 a specification is presented which permits empirical analysis of its nature and magnitude. Aggregated and disaggregated regressions (to be referred to as macro and micro) are compared in terms of the underlying parameters, and an estimation procedure is presented which provides consistent and efficient estimates of them. Section 4 investigates the R&D intensity equation at different levels of aggregation. The results indicate that only a single aggregation effect is present and that it is only operative at a particular level of aggregation. However, the magnitude of the effect is such that it dominates all other observable relationships in the data. Aigner and Goldfeld (1974) present an alternative, and, in their view, more plausible, explanation of aggregation results such as those reported by Grunfeld and Griliches in terms of a misspecification in the independent variable of the regression. Section 5 considers the relationship between these two causes of observed aggregation effects, presents methods for dealing with both problems simultaneously, and uses them to analyze the R&D intensity equation. Finally, Section 6 summarizes the empirical information on the characteristics of the aggregation effect in the R&D intensity decision.

Aggregation Effects and Panel Data Estimation Problems:
An Investigation of the R&D Intensity Decision[1]

This paper is concerned with differences that occur between micro and macro regression estimates of the same relationship. A preliminary study of the choice of R&D intensities by American manufacturing firms indicates that the determinants of inter and intra-industry variance in R&D intensity differ markedly. This result points to the presence of an aggregation effect in the R&D intensity decision. Grunfeld and Griliches (1960) distinguish two types of aggregation problems and consider their effects on the R^2 of micro and macro regressions.[2] However, they do not carry the analysis beyond this. In particular, they fail to give their aggregation problem operational content in terms of testing procedures and parameter specifications. As a result it has remained quite common to ignore the possibility that the Grunfeld-Griliches effect exists when interpreting aggregate relationships and little has been learned about the

---

1. This paper is a revised version of part of my thesis and has benefited from a series of comments by my supervisors, Zvi Griliches and Gary Chamberlain. Useful comments were also received from M.A. Schankerman, Manuel Trajtenberg, and Shlomo Yitzchaki. Financial assistance from NSF grant 73-05374 and the Falk Institute in Jerusalem is gratefully acknlowedged. All errors are mine.

2. Mention should also be made of an earlier paper by Kuh (1959) which alludes to many of the problems analyzed by Grunfeld and Griliches.

1. Preliminary Results

In a separate study (see Pakes [1978] Chapter 1) a simultaneous equation model was developed in order to analyze the variance in R&D intensity among American manufacturing firms. The leading equation in that model had the log of the R&D intensity of the firm explained by its average past growth rate and several latent variables. The model was estimated separately on the firms in each of four industries using a rather comprehensive data set supplied by the Census Bureau and the NSF to Zvi Griliches. The estimates of the parameters of interest to this paper are presented in the first four rows of Table 1. It is clear that past growth rates do not account for much of the intraindustry differences in R&D intensity (.02 to .04 per cent). The intraindustry growth-rate coefficient varied between 2.2 and 5.5 and accorded rather well with exogenous information on that parameter. Row five of Table 1 presents the results of regressing the mean of the R&D intensities of the four industries against the means of their average past growth rates. The interindustry or macro results are markedly different from the micro results. Growth rates account for 99 per cent of the interindustry variance in R&D intensity and the macro growth rate coefficient is far greater than any of the micro coefficients. In fact an χ_3^2 test of the null hypothesis that the growth rate coefficients in the different industries were the same (subject, of course, to sampling error) resulted in an observed test statistic of .74 which is well below the expected value of an χ_3^2 deviate, and a pooled intraindustry micro coefficient of about 4.0 with a standard deviation of about 1.0. It is obvious that no reasonable confidence interval for the pooled intraindustry coefficient would intersect the confidence interval for the interindustry

Table 1. Estimates of the Relationship Between Average Past Growth Rates and R & D Intensity in the Griliches-NSF data.[a/]

Regression	Units of Observation	Coefficient of Growth Rates (1)	Percentage of variance in R & D intensity attributable to growth rates (2)	Number of Observations (3)
Micro, Industry 1	firms	3.94 2.26	0.03	110
Micro, Industry 2	firms	2.16 1.30	0.02	187
Micro, Industry 3	firms	5.64 2.40	0.03	102
Micro, Industry 4	firms	5.58 3.40	0.04	34
Macro, Industry Means	industries	107.26 8.44	0.99	4

a/ The results reported above are maximum likelihood estimates from a three-equation model. Similar results were derived in Pakes (1978) for a six-equation model which allowed for a more complicated unobservable structure. Small numerals are standard errors.

Source: Pakes (1978) p. 35.

coefficient.

Table 1 points to two equally important problems. First why does the response of the aggregated value of the dependent variable seem to differ so markedly from the response of the firms which comprise the aggregate? Sections 2, 3, and 5 discuss this issue in the general setting of estimating economic relationships at different levels of aggregation. Second, both theory (see Arrow [1962]) and empirical work (see Griliches [1958]) have indicated that market size, the quantity of output in which the innovation is embodied, is a primary determinant of the social rate of return to knowledge-producing or research activities. The market size relevant for today's research activities is determined by today's output and future growth rates. Since the R&D intensity equation measures research effort relative to today's output, further differences in market size per unit of research are determined by future growth rates, a variable unknown to the firm at the time it formulates its research policy. However, past growth rates are one indicator of future growth rates, and it is, therefore, of considerable interest to determine if, and precisely why, private industry reacts to this incentive. Table 1 gives us mixed signals on this issue and the results of sections 4 and 5 will serve to clarify the source of the problem.

2. Two Types of Aggregation Problems

The macro and micro results are, of course, logically consistent and taken together they point to an important and often forgotten problem in interpreting aggregate relationships.

The general model underlying both the micro and the macro regression is

$$y_{ij} = \beta_{oj} + x_{ij}\beta_j + \varepsilon_{ij} \qquad (i = 1, \ldots, N, j = 1, \ldots, J) \; (1)$$

where i, j indexes individual i in group j; the x_{ij} are determinants of y_{ij} which are uncorrelated with ε_{ij}, and ε_{ij} are a set of independent and identically distributed random variables with zero expectations.[3]

The intergroup or aggregate relationship is derived by summing over i and dividing by N:

$$y_{\cdot j} = \beta_0 + x_{\cdot j}\beta_j + A_j + \varepsilon_{\cdot j}, \qquad (j = 1, \ldots, J) \; (2)$$

where $\beta_0 = J^{-1} \sum_{j=1}^{J} \beta_{0j}$ and $A_j = \beta_{0j} - \beta_0$.

Since the number of parameters in equation (2), $2J$, exceeds the number of observations (J) the equation itself cannot be estimated. Hence one looks for realistic, and one hopes testable, simplifications which can help us interpret aggregate relationships.

---

3. The problem will be described assuming that the groups are balanced, i.e., $N_j = N$ for all j, that the error variance between groups is constant, and that there is only one observed independent variable. None of these assumptions are necessary.

Partition the set of feasible structural constraints into two groups. First one may be able to model the distrubution of responses of individuals in different groups to changes in the independent variable. The extreme assumption would be

$$H^0: \quad \beta_j = \beta \qquad (j = 1, \ldots, J \text{ and } k = 1, \ldots, K). \quad (3)$$

Given micro data, H^0 is a testable hypothesis. In order to focus attention on the aggregation problems in the R&D intensity equation, I would like to anticipate some of my results. The preceding section reported an χ^2_3 test which clearly indicated acceptance of (3) for the R&D intensity equation on the Griliches-NSF data (the GD). A similar test was applied to observations for 530 firms in 17 industries provided by Business Week (these data are described in Section 4); this resulted in an χ^2_{16} test statistic of 14.71, which again indicates acceptance of H^0. That is, firms respond in the same manner to a change in their growth rate no matter which industry they belong to but that response is much lower than the response of the industry aggregate to changes in its growth rate.

This bring us to the second type of constraint. All individuals may respond in the same manner to a change in every independent variable but the constant terms may still differ between groups. They will differ if there is some independent variable which has the same value for all members of the same group but different values for different groups. As noted by Griliches and Grunfeld [1960], the existence of such a variable implies some misspecification in the original model. One should be careful, however. The specification error dealt with here has no implications for the properties of the micro or intraindustry regressions. Any omitted variable which has the same value for all members of a given group is submerged in the constant term in the within-group regression and has no other effect on these regression results.

Consider now the aggregate or between-group regression. Assuming that hypothesis H^0 is accepted, the aggregate relationship is written as

$$y_{\cdot j} = \beta_0 + x_{\cdot j}\beta + \varepsilon_{\cdot j} + A_j \qquad (j = 1, \ldots, J). \qquad (4)$$

The properties of equation (4) depend on those of the vector A. If the sample covariance of A with the rest of the independent variables is zero then an ordinary least-squares (OLS) repression will produce unbiased estimates of β (the other properties of these estimates cannot be discussed without imposing more structure on A). If A is correlated with at least one independent variable, then the regression will produce biased estimates of all coefficients.

A few examples will illustrate the potential importance of this type of aggregation effect in economics. In his analysis of investment demand, Grunfeld (1960) was one of the first to point out the existence of an aggregation effect which was not explicable by differences in response parameters. Consider the investment demand of a cross section of firms in different industries. Since capital is a long-lived asset, the demand for it will depend on the expected output of the firm. Generally, evaluations of the "health" of the industry will affect the output expectations of all firms, that is $A \neq 0$. Further, it is likely that these evaluations will depend on the past trend in the industry's output. In such a case, as long as the firm's output is an included independent variable, the micro, the macro, and the pooled regression with a single intercept will each estimate a different set of coefficients and have different measures of goodness of fit. Moreover, the nature of the relationship between A and industry ouput may be of considerable interest in the analysis of investment. A similar analysis, of course, can be applied to the difference between the equation determining industry investment demand and the equation determining investment demand of the economy.

Perhaps the most obvious example of this aggregation effect occurs when there are externalities in production or consumption. Consider the case of an industry in which the output of one firm in a given region creates externalities for other firms in the region. These externalities may be a result of the spread of training facilities or equipment suppliers, of learning by doing, or of infrastructure investment. Say one is investigating the production function of the industry; for simplicity assume both that all firms produce on the same Cobb-Douglass function and that the externalities are a function of the sum of the region's output. The production function is then written as

$$y_{ij} = \alpha_0 + \alpha_\ell \ell_{ij} + \alpha_c c_{ij} + \alpha_y N y_{\cdot j} + \varepsilon_{ij} \quad ,$$

where i, j indexes firm i in region j; y, ℓ , and c represent the logarithms of output, labor, and capital; N is the number of firms in each region, and

ε is an error which may have a region-specific component but which is uncorrelated with the rest of the independent variables.[5]

Since $\alpha_y N y_{.j}$ is the same for all members of the region, the intra-regional or micro regression will estimate labor and capital coefficients which will be unbiased efficient estimators of α_ℓ and α_c respectively. On the other hand, the intraregional or macro regression is written as

$$y_{.j} = \frac{\alpha_0}{1-\alpha_y N} + \frac{\alpha_\ell}{1-\alpha_y N}\,\ell_{.j} + \frac{\alpha_c}{1-\alpha_y N}\,c_{.j} + \frac{\varepsilon_{.j}}{1-\alpha_y N} \quad .$$

That is, the coefficients of labor and capital from the macro regression will provide consistent estimates of $\alpha_\ell/(1-\alpha_y N)$ and $\alpha_c/(1-\alpha_y N)$. A pooled regression of all the observations will, of course, estimate a weighted average of all these parameters. Here again one is not simply concerned with "taking care" of the influence of $y_{.j}$ in order to derive consistent estimates of α_ℓ and α_c. The coefficient α_y will be of considerable policy interest in itself. Note, however, the difference between this and the previous example. In this case the aggregation effect does not arise because of an omitted independent variable but because the group mean of the dependent variable is itself a determinant of individual outputs.

---

5. This example is taken from, and analyzed in greater detal in Grunfeld and Levhari [1962]. Note that the logarithm of the sum of industry outputs is written as N <u>times</u> the mean logarithmic output of the industry. This is justified so long as factor markets are competitive.

Finally, consider a simple case from elementary micro theory. The short-run factor-supply curves facing different industries may be fairly price elastic since the increase in the price offered by one industry will draw resources away from others. An increase in the offer price of the economy, however, is not likely to elicit a large output response, because of overall fixed endowments which enter the supply curves facing different industries in a similar manner.

These examples should suffice to show that the aggregation effect described above has relevance for a broad range of problems. The next section discusses a method for analyzing them.

3. Mixed Effects

To analyze the Grunfeld-Griliches aggregation effect, that is, differences in estimated relationships at different levels of aggregation when response parameters do not differ between individuals, hypothesis H^0 of the last section is imposed and equation (1) is rewritten as:

$$y_{ij} = x_{ij}\beta + A_j + \varepsilon_{ij} \qquad (i=1 \ldots N, j=1 \ldots J) \qquad (5)$$

where it is understood that all variables are written as deviations from their sample means and it is assumed that the augmented matrix (X, Z), (where Z is a matrix on qualitative or dummy variables, one for each group, and X is the vector of observations of x_{ij}) is of full column rank.

Models of this type have become increasingly familiar to econometricians in a slightly different context, the estimation of relationships involving panel data (data which follow a cross section of economic units over time). In these applications the joint i, j index refers to unit j in period i and A represents a vector of individual effects which remain constant over time. As will be shown in the next section, the idea may be extended to a model which puts each observation into two or more groups, each with its own effect, without changing the nature of the estimation problem.

The method of estimating (5) depends upon what one is willing to assume about A. Three specifications have received most of the attention of econometricians. The fixed-effects (FE) model simply assumes that A contains a set of numbers to be estimated. The random-effects (RE) model imposes more structure. This model assumes both that the observations on (A, X) are random drawings from a common population and that the covariance of A with X in this population is zero.[6] It then proceeds to estimate the parameters of the distribution of A. I shall impose the prior condition that the drawings on (A, X) are in fact random.[7] In that case the RE model is a special case of the FE model. The third alternative is that the distribution of the effects is degenerate, or that there are no effects (NE), and it is therefore a special case of the RE model. The efficient

6. Strictly speaking the RE model makes the stronger assumption that the conditional distribution of A given X does not depend on X. See also footnote 8.

7. When there is reason to believe that the drawings are not random one should consider adding a sample selection equation to the model described below.

estimators for these three specifications are reviewed in Mundlak [1978) and Pakes [1978].

Briefly the FE estimators are found by adding a dummy variable for each group to (5) and then performing OLS on the resulting equation. It produces the within group covariance estimator of β and an estimate of the vector A which is unique up to a normalization. Under RE consistent estimators of the distribution of $(A + \varepsilon)$ are obtained and then used to perform generalized least squares (GLS) on (5). Since the RE estimator of β uses part of the between group variance in X, it will generally have a lower variance then the FE estimator of that parameter. However, if A is correlated with X, the between group covariance of y with X will pick up part of Cov(A,X), and the RE estimator of β will be biased. The NE model amounts to a standard regression of y on X and therefore uses all of the between group variance in X to estimate β. If the specification is correct the OLS estimates have all the familiar desirable properties. If not, it yields inefficient (under RE), or inconsistent (under FE), estimates of the β coefficients.

Faced with the alternatives the econometrician must choose an estimator. Since the various models are ordered with respect to their generality, the obvious procedure is to specify an overall model which includes all the alternatives as special cases and to test the constraints implied by each. To do so, return to equation (5) and note that the vector A can always be partitioned into a linear function of the group mean of the independent variable and a residual which is, by construction uncorrelated with it, i.e.,

$$A_j = x_{\cdot j}\phi + \eta_j \qquad (j = 1, \ldots, J) \qquad (6)$$

where $\text{Cov}(X, \eta) = 0$, by construction. I will assume that the η_j are random drawings from a common population with variance σ^2_η.[8]

(4) can then be rewritten as

$$y_{ij} = x_{ij}\beta + x_{\cdot j}\phi + \mu_{ij} \qquad (7)$$

where $\mu_{ij} = \eta_j + \varepsilon_{ij}$. Since the vectors η and ε are uncorrelated with each other and with the vectors X and ΓX, where $\Gamma = Z(Z'Z)^{-1}Z'$ so that ΓX is just the vector of observations on $x_{\cdot j}$, we have; $\text{Cov}(\mu, X) = \text{Cov}(\mu, \Gamma X) = 0$, and $\text{Var}(\mu) = \sigma^2_\varepsilon I_{NJ} + \sigma^2_\eta \Gamma N$, where I_q is the identity matrix of order q.

Since the covariance of A with ΓX is equal to the covariance of A with X, (6) partitions the variance in A into a correlated effect ($\Gamma X\phi$) and a part that is uncorrelated with X (η). If $\phi = 0$ the effects are not correlated with X and the RE model is relevant, and if $\phi = 0$ and $\sigma^2_\eta = 0$ there are no effects.

The aggregate or between-group regression can now be given an explicit interpretation. Sum (7) over i and divide by N, i.e.,

$$y_{\cdot j} = x_{\cdot j}(\beta + \phi) + \eta_j + \varepsilon_{\cdot j} \qquad (j = 1, \ldots, J). \qquad (8)$$

8. In order to specify the regression function of y given X in (5) one needs to specify $E[A|X]$. In general, $E[A|X]$ will depend on all the x_{ij} as well as powers of these variables so that (6) does impose a non-trivial prior on the model. Statistically, the appeal in the structure imposed by (6) is that its estimate of β, and that parameter's standard error will be precisely the (cont.)

(8) is the traditional between-group or macro regression. Recall that the within-group or micro regression correctly estimates the resonse of the individual unit to changes in the independent variable, given the value of the effect. The aggregate regression estimates the sum of the values of this resonse plus the response of the aggregation effect to a unit increase in the group mean of the independent variable. Except for the special case where $\phi = 0$, the micro and the aggregate equations are estimating different parameters. A comparison of the within and between group parameter estimates gives us information on the relationship between the aggregation effect and the independent variable which, as noted above, may be of considerable interest in itself. The comparison does not, however, provide two sources of information on the same parameter vector.

This latter point is illustrated rather nicely by the form of the GLS estimators for (7). If we let $X^d = X - \Gamma X$ and $y^d = y - \Gamma y$, that is, if we let the superscript d denote deviations from group means, then a simple partitioning of the inverse matrix for the transformed independent variables proves that these estimators are independent of the error structure and equal to:

$$\hat{\beta} = (X'^d X^d)^{-1} X'^d y^d ,$$

and (7a)

$$\hat{\phi} = (X'\Gamma X)^{-1} X\Gamma y - (X'^d X^d)^{-1} X'^d y^d$$

---

same as those derived from a model using any more general structure for $E[A|X|$. Chamberlin (1978) discusses this point, while p. 17 below considers the interpretive benefits of using the structure in (6).

$\hat{\beta}$, is, of course, just the within-group covariance estimator (b_w). Referring to equation (8) and noting that Γ is idempotent one finds that $(X'\Gamma X)^{-1}X'\Gamma y$ is just the parameter vector estimated in the aggregate regression, or the between-group covariance estimator (b_b).[9] Hence $\hat{\phi}$ is calculated as the between-group estimator minus the within-group estimator. That is, there are no fixed aggregations effects iff the difference between the within-group and between-group estimators is attributable to sampling error.

The analogy between the within and between group covariance estimators may be carried further in order to determine the variance-covariance matrix of the estimated parameters. Recall that $\hat{\beta} = b_w$ and $\hat{\phi} = b_b - b_w$. Since $\text{cov}(b_b, b_w) = 0$, it follows that $\text{cov}(\hat{\beta}, \hat{\phi}) = -\text{var}(b_w)$ and $\text{var}(\hat{\phi}) = \text{var}(b_b) + \text{var}(b_w)$. This variance-covariance matrix may conveniently be written as

$$\text{var}(\hat{\beta},\hat{\phi}) = \sigma_\varepsilon^2 \begin{bmatrix} (X'^dX^d)^{-1} & \\ -(X'^dX^d)^{-1} & \dfrac{(\sigma_\eta^2 + N^{-1}\sigma_\varepsilon^2)}{\sigma_\varepsilon^2}(\frac{1}{N}X'\Gamma X)^{-1} + (X'^dX^d)^{-1} \end{bmatrix} \qquad (9)$$

The assymptotic properties of these estimators are provided in the appendix and are worth noting here. Under the traditional assumptions on population moment matrices $(\hat{\beta}, \hat{\phi})$ is, of course, consistent when both N and J approach infinity. Since successive increments in either the number of groups, holding the number of members in each group constant, or in the number of members in each group holding the number of groups constant, both provide successively more information on the within-group estimator, $\hat{\beta}$

---

9. If the groups are unbalanced the between-group regression must be weighted in order for the equivalences to be discussed here to hold. For details see the next section.

is consistent when either N or J approach infinity. Consider now the variance of ϕ which may be written as: $\text{var}(\hat{\phi}) = \sigma_\eta^2 N(X'\Gamma X)^{-1} + \sigma_\varepsilon^2 (X'\Gamma X)^{-1} + \sigma_\varepsilon^2 (X'^d X^d)^{-1}$. The first expression is constant as N increases. That is, η is a random error which is group specific and will not be averaged out unless the number of groups grows large. Hence, $\text{var}(\hat{\phi})$ does not approach zero as the number of members in each group grows when the number of groups is held constant. The implication of this point is that it may be difficult to derive precise estimators of ϕ when the number of groups (J) is not large. Since the number of groups in panel-data analysis is usually quite large this should not be an important problem in panel-data applications of the procedure. However, in the context of analyzing aggregation effects the number of groups is often limited by the nature of the problem. In our example there is a finite number of industries in the economy, and one may need to add another dimension to the data (e.g., time or economies) in order to derive precise estimates of ϕ.

Finally, a brief comparison of the new mixed effects (or ME) model with the models already available is in order. The ME and the FE models provide exactly the same estimates of β and its variance. The ME model, however, has two advantages. First, it summarizes the information in the FE estimates of the qualitative or dummy variables in terms of a small number of interpretable parameters [ϕ and σ_η^2] which provide the links necessary to compare and interpret aggregate and micro relationships and, hence, are likely to be of considerable interest in an aggregation context. Moreover, it is easy to think of cases where these parameters would be of interest in panel data estimation. Indeed, Grunfeld and Griliches [1960] originally pointed out the existence of their type of aggregation effect in two panel-data estimation problems.

Second, much of the literature on panel-data estimation has, perhaps wrongly, been concerned solely with the properties of the estimators of the within-group coefficients, β (see Balestra and Nerlove, 1966). In this context the ME model has the advantage of providing a direct test of the RE structure. If $\phi = 0$ within sampling error one may wish to impose an RE specification and re-estimate the model in order to derive more precise estimates of β.[10] Of course, in multivariate regressions some but not all of the elements of ϕ may be close to zero in which case the between-group variance in some of the independent variables may be used in order to estimate β.

I would like to raise one further problem here but postpone discussion of it until Section 4.5. As noted above most of the theoretical discussion of the relative advantages of the FE, RE, and NE estimators has concentrated on the relationship between the bias caused by an omitted variable with group structure in the estimates of β and their variance. However, users of these techniques are equally worried about the properties of the alternative estimators of β when there exists the possibility of an error in one or more of the independent variables of the regression. Once one admits the possibility of omitted variables without group structure, or errors in variables, as well as those with group structure, the performance of the alternative estimators of β differs markedly. Section 5 considers models which allow for both omitted variables with group structure and errors in variables.

---

10. The ME model, of course, can also be used to provide a direct test of the NE structure; that is, one could test if both ϕ and σ^2_η equal zero. However, a simple comparison of the FE to the NE regression results pro- the same information; see Section 4.

Given the conceptual and econometric framework for analyzing aggregation effects presented in the last two section, we can now return to the original problem of analyzing the R&D intensity equation at different levels of aggregation.

4. Aggregation Effects in the R&D Intensity Decision: Empirical Analysis

This section used two additional data sets and the framework outlined above to analyze the nature of aggregation effects in the R&D intensity decision.

The first data base to be examined is drawn from the periodical Business Week (June 27, 1977) which contains the 1976 ratio of company-financed R&D to sales and five year undeflated growth rates for firms who performed the vast majority of all privately financed R&D in the U.S. in 1976. The Securities and Exchange Commission's (1976) annual report was used to classify 530 of these firms into two and three-digit SIC industries, which in turn can be used to put the firms into the NSF industrial classification used throughout this paper. The growth rates were deflated by 2 1/2 digit industry-specific price deflators.[11] The major difference between the Business Week

---

11. These data are described in more detail and compared to the other data sets used in this paper in Pakes (1978). Business Week claims that the original data set covered firms which accounted for 98% of all privately financed R&D in the U.S., but NSF calculations indicate that the data cover only 82% of total privately financed R&D when NSF definitions of R&D are used (see NSF 78-303). The Business Week sample contains 598 firms. The firms in the "other manufacturing," "all manufacturing" and oil industries were dropped from the data used here; the first two on the grounds that these categories confuse inter and intraindustry differences in R&D intensity, and the latter because no adequate price deflator could be found for the oil industry.

Data (BWD) and the other data sets used in this paper are that the BWD have observations on company-financed R&D and total sales while the others use total R&D to sale ratios. The difference consists of publicly-financed R&D performed in the private sector which, according to Business Week (June 27, 1977), accounted for 39 percent of all privately performed R&D expenditures in 1976.

Table 2 presents the results. The first question one would like to answer is whether there is evidence of any aggregation effect in the R&D intensity decision. Under the null hypothesis of no aggregation effects the OLS regression provides the Gauss-Markov estimators while under the alternative that there are effects the ordinary least squares with industry-specific constant terms (OLSC) regression estimates the value of the aggregation effect for each industry. A test of the null hypothesis is simply a test of the joint significance of these effects. Columns (1) and (2) indicate that the observed value of the $F^{16, 512}$ test statistic is 24.20 while the 5 and 1 percent critical value of an $F^{16, 512}$ deviate are 1.67 and 2.03. There are highly significant aggregation effects and the ME model is now used to summarize their characteristics. The estimation procedure used for the ME model is as described in Section 3 except that since there are unbalanced groups one must weight the observations from the between regression by $[\sigma_\eta^2 + \sigma_\varepsilon^2/N_j]^{-1/2}$. σ_ε^2 and σ_η^2 are consistently estimated by the σ^2 of the OLSC regression, and by the σ^2 of an OLS regression of the firm's R&D intensity against the firm's and industry growth rate [labelled first-stage σ^2 of column (3)] minus the σ^2 of the OLSC regression. The variance of the random firm effect (σ_ε^2) was estimated to be .39 while that of the random industry effect (σ_η^2) was .12.

Table 2. Aggregation Effects in the BWD[a]

Coefficient of	Regression			
	No effects (OLS)	Industry-specific effects (OLSC)	Mixed effects	Weighted aggregate between industry[b]
	(1)	(2)	(3)	(4)
Micro growth rate, α		1.04 0.31	1.04 0.31	10.96 1.88 (Micro and Macro combined)
Macro growth rate, ϕ			9.93 1.91	
Pooled data growth rate	3.08 0.36			
σ^2	0.66	0.39[c]	n.r.[d]	n.r.
R^2	0.12	0.02[e]	n.r.	0.66
Degrees of freedom	528	512	527	15

a/ Small numerals are standard errors: n.r. means not relevant.

b/ The weight for industry j is $(\sigma_\eta^2 + \sigma_\varepsilon^2/N_j)^{-1/2}$.

c/ $0.39 = \sigma_\varepsilon^2$.

d/ First stage $\sigma^2 = 0.51$. It follows that $\sigma_\eta^2 = (0.51 - \sigma_\varepsilon^2) = 0.12$.

e/ Between firms, within industries.

It is clear from the results in column (3) that the industry's growth rate is a significant determinant of the aggregation effect.[12] Indeed, the effect of a unit increase in the industry growth rate on the firm's R&D (9.93) intensity is about 10 times the effect of a unit increase in the firm's own growth rate on its R&D intensity (1.04). As a result growth rates account for only .02 percent of the within-industry variance in R&D intensity but for 66 percent of the between-industry variance. The weighted between-industry coefficient [column (4)] estimates the sum of the micro and the aggregation effect growth rate coefficients, while the OLS regression estimates a weighted average of the between-group and within-group coefficients and has no meaning in itself. Note also that the standard error of the coefficient of the industry growth rate (1.91) is relatively large owing to the small number of industries in the sample.

The qualitative results from the BWD and the GD are strongly complementary.[13] They both indicate that there is a highly significant aggregation effect in going from firm to industry aggregation in the R&D intensity decision and that the value of the aggregation effect is in large part determined by the industry's past growth rate. The NSF (1958, Table C-1) provides another source of information which can be used to investigate one other aspect of the aggregation effect in the R&D intensity decision. That table lists the ratio of total R&D to sales of 54 "product groups" in

---

12. It should be noted that the phrase industry growth rate refers to the mean growth rate of the firms in the industry. The difference between the latter and the growth rate of industry output can be shown to be a second order term.

13. The differences between the parameter values estimated from the BWD and the GD are not highly significant and will be discussed briefly in Section 6.

15 industries in U.S. manufacturing in 1958, which, when combined with past growth rates of sales, can be used to ask if there is an aggregation effect in going from the product-group classification to the industry classification. The growth rates were obtained from the three data points on sales taken from the Censuses of 1958, 1954, and 1947 and the price deflators mentioned earlier.[14] Table 3 presents the results.

The test of the null hypothesis that there are no aggregation effects in going from product group to industry level aggregation results in an observed value of an $F^{14,38}$ test statistic of 6.72. This is to be compared with the 5 and 1 percent critical values of an $F^{14,38}$ deviate of 1.96 and 2.59. There are highly significant aggregation effects in going from porduct groups to industries. The estimates of the random product group (σ_ε^2) and industry (σ_ε^2) effects (derived in the manner described on p.20) of, respectively, 0.36 and 0.34 were used as the variance components in the estimates of the ME model reported in column (3). Again, the value of the product group's own past growth rate has very little to do with its R&D intensity decision while the value of the industry's past growth rate is a highly significant determinant of the product group's choice of R&D intensity. Therefore, the R^2 from the micro or between-product-group within-industry regression is zero while that from the macro or between-industry regression 0.60.

---

14. The product groups are a 3-digit NSF classification which is slightly more aggregative than the 3 digit SIC; see also Section 6. The original data contained observations on 56 product groups but a reclassification of census industries forced the dropping of two of them. The analysis was also done using four-year past average growth rates with similar qualitative results.

Table 3. Aggregation Effects in the ND[a/]

Coefficient of	Regression			
	No effects (OLS)	Industry-specific effects (OLSC)	Mixed effects	Weighted aggregate between-industry[b/]
	(1)	(2)	(3)	(4)
Product group growth rate, α		0.02 2.85	0.02 2.85	23.44 6.38
Macro growth rate, ϕ			23.43 6.07	
Pooled data growth rate	10.85 3.41			
σ^2	0.92	0.36[c/]	n.r.[d/]	n.r.
R^2	0.16	0.00[e/]	n.r.	0.60
Degrees of freedom	52	38	51	13

a/ Small numerals are standard errors: n.r. means not relevant.

b/ The weight for industry j is $(\sigma_\eta^2 + \sigma_\varepsilon^2/N_j)^{-\frac{1}{2}}$.

c/ $\sigma_\varepsilon^2 = 0.36$.

d/ First stage $\sigma^2 = 0.70$. It follows that $\sigma_\eta^2 = 0.70 - \sigma_\varepsilon^2 = 0.34$.

e/ Between product groups, within industries.

Three independent data sets for three years (1958, 1963, and 1976) have all indicated that there is a highly significant aggregation effect in the R&D intensity decision in going to industry aggregation, and that the value of the aggregation effect is strongly and positively correlated with that of the industry's past growth rate. The new information in the ND, however, is contained in its estimate of the product-group growth rate coefficient (0.02). Considering the standard errors, this is similar to, if anything, a little below the micro growth rate coefficients estimated from the BWD (1.03) and GD (4.00) which are both based on firm-level observations. This finding suggests that there are no significant correlated aggregation effects in going from firm to product-group aggregation.

To investigate this possibility further the BWD were analyzed again allowing for both product-group and industry aggregation effects, the variance of each set of effects being partitioned into portions accounted for and unaccounted for by the aggregate's growth rate. Statistically, this is just a two-way classification ME model and it can be shown that the GLS estimates of the firm's, the product group's, and the industry's growth rate coefficients will be identical to the within-product group covariance estimator of the growth rate coefficient, the between-product group within industry minus the within-product group covariance estimator, and the between-industry minus the between-product group within industry covariance estimator, respectively. Further, the standard errors of the three OLS regression coefficients are correct and they are distributed independently of each other.[15] Briefly the $F^{28,484}$ test statistic for product-group aggregation effects was 1.98 which is just significant, the product-group's growth rate coefficient was .03 with a standard error of 1.31, and the other parameters were almost identical to those reported in table (2). The product-group growth rate does not have a significant effect on the firm's choice of R&D intensities.

The ND leads to essential the same conclusion as did the GD and the BWD. There is a consistent significant industry-wide aggregation effect on the firm's R&D intensity decision, where an industry is defined by the NSF's classification of groups of firms utilizing similar technologies. The value of the aggregation effect for all firms in a given industry is a significant and increasing function of that industry's past growth rate. On the other hand, there is little evidence of an aggregation effect at a lower (3-digit or product-group) level of aggregation and what effect there is seems not to be correlated with the past growth rate of the product group.

5. Mixed Effects, Errors in Variables, and the Aggregation Problem in the R&D Intensity Decision

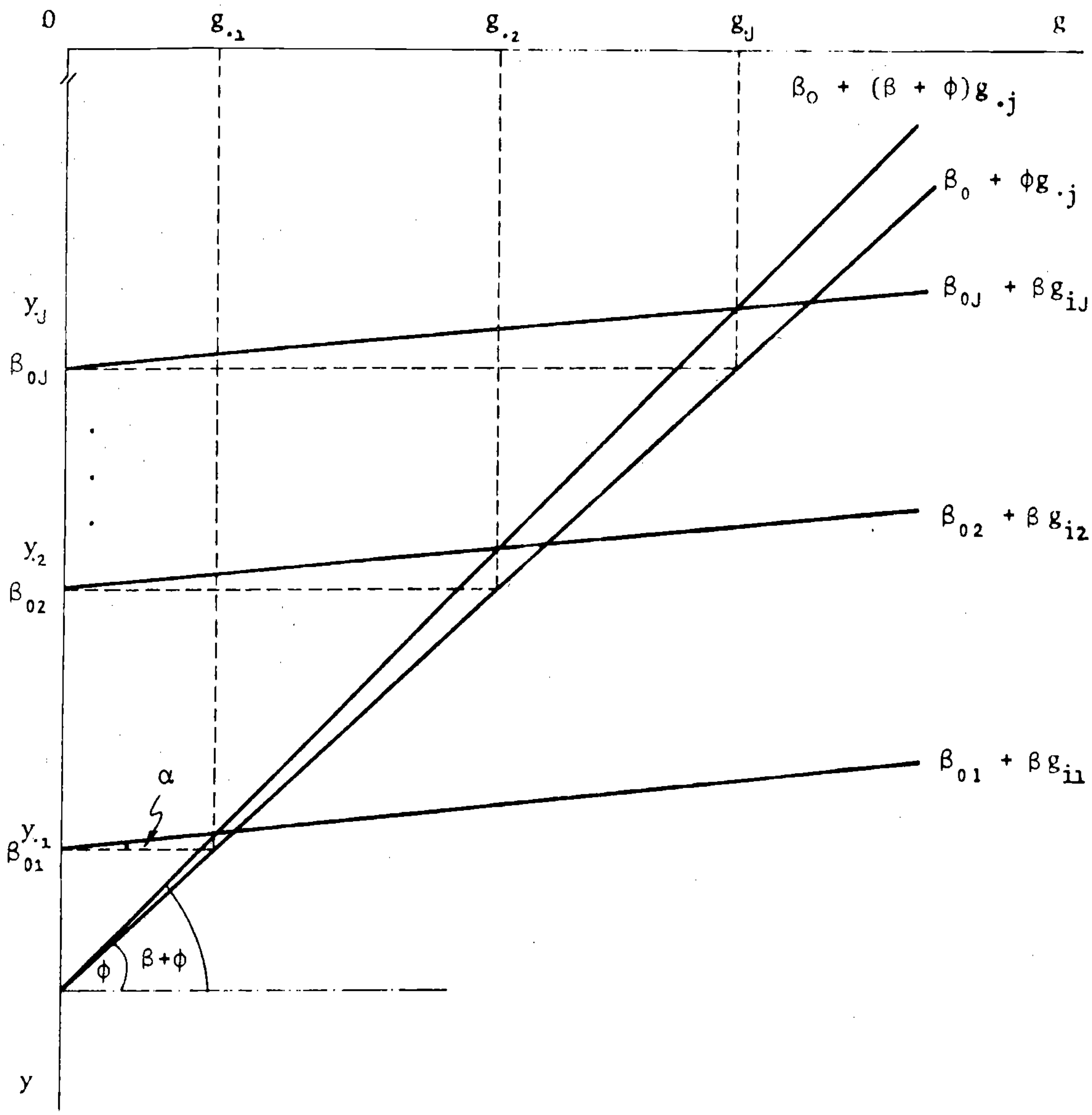
It is worthwhile to pause here and consider whether the reults of the last section could be explained by alternative model structures. Figure 2.1 illustrates the use of the Grunfeld-Griliches aggregation

15. For cases with unbalanced groups, such as ours, the within-product group between industry and the between-industry regressions must be weighted. Details of the estimation procedure are to be found in Pakes (1978).

effect in this context. The lines $\beta_{0j} + \beta g_{ij}$ for $j = 1, \ldots, J$ represent the within-industry regressions. Differences in R&D intensity between two firms within any one industry are caused by differences in their growth rates, and the response to a unit difference in growth rates, ϕ, is the same no matter which industry the two firms belong to. The intercepts of the within-industry regressions, $\beta_{01}, \ldots, \beta_{0j}$, differ because of differences in a stimulus which has the same value for all firms within the industry. This stimulus may be partitioned into a part accounted for by differences in the growth rates of the various industries, the line labelled $\beta_0 + \phi g_{.j}$, and a disturbance which is uncorrelated with $g_{.j}$ and, therefore, with g. The effect of industry growth on the average R&D intensity is given by the sum of the average effect of industry growth on the firm's choice of R&D intensitities given the level of aggregation effect ($\beta g_{.j}$) and the effect of industry growth on the level of aggregation effect ($\phi g_{.j}$). It follows that the between-industry regression is the line labelled $\beta_0 + (\phi + \beta)g_{.j}$ in the diagram.

Now consider the aggregation properties of the model proposed by Friedman (1957, see, in particular, the figure on p. 64). If the observed growth rate (g) measures the sum of the growth rate to which the firm responds (g^*) and an independent and identically distributed error (v), and if, in fact, $\phi = 0$ or there are no correlated aggregation effects, then both the within-and between-industry regression coefficients will estimate a weighted average of the response of the firm to changes in g^* (β) and to changes in v (zero). The weight given to zero will, in the limit, equal the fraction of the variance in the observed independent variable attributable to v, say λ_{wg} and λ_{bg} in the within- and between-industry regressions, respectively. Since v is assumed to have no group

Figure 2.1. The Grunfeld-Griliches Aggregation Effect



structure its variance will tend to be averaged out in the between-industry regression, and provided that g^* is correlated with the industry dummies, λ_{bg} is expected to be less than λ_{wg}.[16] In fact, $\operatorname{plim}_{N\to\infty} \lambda_{bg} = 0$, where N is the number of observations in each group, so that for large N and $\phi = 0$, b_{bg} provides a consistent estimate of β, whereas, a comparison of the b_{wg} with b_{bg} provides a consistent estimate of λ_{wg}, the noise total variance ratio in the within-group regression, and, therefore, of σ_v^2, the variance in measurement error.

This idea was used by Aigner and Goldfeld (1974) to show that an error-in-variable model provides an alternative, and in their view more reasonable, explanation of the Grunfeld-Griliches aggregation effect than that provided by the latter authors themselves. As noted by Aigner and Goldfeld, under their assumptions, that is, assuming there are no aggregation effects and allowing for errors in variables, a regression of the micro dependent variable on the micro and macro values of the independent variables, the ME regression, provides a consistent estimate of β. To see this for the simpler case when N grows large recall that the estimate of β from the ME model equals $b_{bg} - b_{wg}$ and that of ϕ equals b_{wg}. From the discussion above it follows that

$$\operatorname{plim}_{N\to\infty} (\hat{\phi} + \hat{\beta}) = \beta$$

and

$$\operatorname{plim}_{N\to\infty} \hat{\phi} = \lambda_{wg}\beta .$$

16. Actually all that is required for b_{bg} to have a larger probability limit than b_{wg} is for the intraclass correlation in v to be less than the intraclass correlation in g^*.

Therefore, provided that $\beta \neq 0$, a test result of $\phi \neq 0$ under Aigner and Goldfeld's assumption implies that $\sigma_v^2 \neq 0$, that is, that there are errors in variables. Recall from Sections 3 and 4 that in the Grunfeld-Griliches case, that is, assuming that $\sigma_v^2 = 0$ but allowing for aggregation effects, a test result that $\phi \neq 0$ implies that there are correlated effects. The errors-in-variables and the correlated aggregation effects models are alternative explanations of the same observed phenomena.[17] To distinguish between them one needs a model capable of identifying and testing for the existence of both mixed aggregation effects and errors in variables (MEEV).

Moreover, MEEV models are also needed to bridge a gap which has developed between the theoretical literature on grouped-data estimation (including most of the discussion in Section 3) and the factors considered by empirical researchers who must choose estimators for grouped-data estimation problems. As noted earlier both theory and empirical analysis have focused on the properties of the within-group coefficient estimators. The theoretical literature has concentrated on comparing the bias caused by an omitted variable with group structure with the variance of the within-group coefficients resulting from the alternative FE, RE, and NE models. In this context, the FE, or equivalently the ME, estimator of β , is, in general, the only estimator which is unbiased, a point which is forcibly made by Mundlak (1978). Empirical research, however, often has to contend with poorly measured or ill-defined independent variables and as a result is as worried about omitted variables without group structure (or errors in variables) as those with group structure (see, for example, the discussion in Griliches, forthcoming). Once one

17. If β and ϕ have different signs one can distinguish between the special cases of $\phi = 0$, $\sigma_v^2 \neq 0$, and $\sigma_v^2 = 0$, $\phi \neq 0$, but β, ϕ and σ_v^2 remain unidentified.

admits the possibility of errors in variables it is not at all clear which of the FE, RE, or NE estimators of β has a smaller asymptotic bias. The problem arises because the more one uses the between-group variance in the independent variable to estimate the within-group coefficients the larger will be the bias caused by omitted variables with group structure but the smaller will be the bias caused by errors in variables. As noted above the FE model uses none, the NE model uses all, and the RE model uses some of the between-group variance in the dependent variable to estimate the within-group coefficients. Clearly, one would like to consider models which estimates β , ϕ and σ_v^2. After identifying each of these parameters it still may be worthwhile to trade off the bias against the variance of, say, the within-group coefficient estimator, but only after the magnitude of the problem caused by both types of biases are determined.

The MEEV model may be written as:

$$y = \beta_0 + \beta g + A - \beta v + \varepsilon$$

where $g = g^* + v$, $A = \Gamma g^* + \eta$, η, g^*, v and ε are mutually independent, and $\mathrm{var}(v) = \sigma_v^2 I$. The other properties of ε and η are found on page 14.

The model contains two unobservables that are correlated with the observed independent variable, the error in measurement, v, and the group effect, A. The macro or between-group regression averages out the effect of v on g but includes the bias in β caused by the left out variable with group structure, A. Therefore, the between group regression coefficient provides an assymptotically unbiased, as N grows large, and consistent, as both N and J grow large, estimate of $\beta + \phi$.[18] The within-group regression differences out the effect of A but contains the errors in variable bias so that its regression coefficient provides a consistent, as either N or J grow large, estimate of $\beta\lambda_{wg}$. It is clear, then, that first and second order within and between sample moments do not suffice to separate out, or identify, β, ϕ, or σ_v^2. Since when the observations on both the within-group pairs (y^d, g^d) and the between-group pairs distribute joint normally, first and second order moments suffice to determine the entire within- and between-distributions, a model with normal deviates cannot be identified without adding more information then is contained in these variables.[19] As N grows large, the between group moments do not contain the effect of v, and do not distinguish between β and ϕ. Therefore, if, in the non-normal case, higher order moments are to be used to identify an additional parameter, they must be within moments.

To see how higher order moments can be used for identification note that as either N or J grows large plim $1/NJ \sum (y^d)^2 g^d = \beta^2 \sigma^3_{g*d}$, where σ^3_{g*d} is the third order moment of g^{*d}, and plim $1/NJ \sum (g^d)^2 y^d = \beta\sigma^3_{g*d}$, so that provided neither β nor σ^3_{g*d} are zero, the ratio of these two-third order moments

18. The variance in the between group regression coefficient will not go to zero with N unless $\sigma_\eta^2 = 0$.

19. This is a simple extension of a result due to Riersöl (1950). As N and J grow large the five moments which determine the entire joint between distribution will be estimated precisely and will suffice to identify β_o, $(\beta+\phi)$, the variances of η and $g_{\cdot j}$, and the mean of $g_{\cdot j}$. Since the between and within variables have zero covariance by construction the only information that is left in the data is contained in the joint distribution of y^d and g^d, and Riersöl (1950) has shown that in the normal case this distribution does not identify either β or σ_v^2.

provide a consistent estimate of β . In fact, it is clear that as N grows large the model determining the joint within distribution, i.e, $y^d = \alpha g^d - \alpha v^d + \varepsilon^d$, is a simple error in variable model, and it is well known that in this case the method of cumulants (see Maliavaud [1970]) provides an easy way of using higher order moments for identification. The variance of moments of order r contains moments of order 2r. As a result early experiments with the cumulant method, most of which were based on small sample sizes, concluded that the variance of the cumulant estimator was too large for it to be of practical use (see, for example, Madansky [1959]). This variance is, however, of order 1/NJ, where NJ is sample size, and the larger data sets now available in economics may, in fact, provide more precise cumulant estimators for the errors in variable model. Our sample is moderately large, and, hence, the first attempt at identifying the MEEV model will use the cumulant method on the within-group sample moments.

In addition, the average past growth of employment (g_n) was gathered for each of the sample firms. Recall that g is calculated as the growth rate of sales minus the growth rate of prices, and, therefore, the error in g may be a result of errors in either of the latter variables. For the most part price indexes are not adjusted for changes in the quality of the goods sold, and are, therefore, likely to be particularly bad for technologically progressive firms such as those that dominate our sample (see Griliches [1979] for a discussion). Since g_n is calculated independently of both price and sales data it is reasonable to assume that Cov $(g_n v) = 0$. Moreover, previous estimates indicate that $\text{Cov}(\varepsilon g_n)$ and $\text{Cov}(\eta g_n)$ are small enough to be ignored.[20]

20. See Pakes (1978) appendix 1D.

Since the output elasticity of labor is non-zero, g_n can be used as an instrument in both the within- and between-regressions to produce consistent estimates of β and $\beta+\phi$, respectively. Note that, unlike the cumulant estimates, the instrumental variable (IV) estimators do not depend on the mutual independence of g^*, v, η and ε for their consistency properties.

Though the within-group moments were clearly non-normal, the standard error of the third order cumulant estimator (132.0) was extremely large due to the underlying symmetry of the distribution function, and that of the fourth order cumulant estimator was too large (4.09) for us to put any faith in its point estimate of -.26.[21] It appears that even in highly non-normal cases cumulant estimators require sample sizes a great deal larger than 500.

We now move on to the IV estimators of the MEEV model. The parameters of this model are estimated in precisely the same way as the parameters of the ME model except that the independent variables in the between- and within-regressions become the predicted values from the projection of the growth rate of output onto that of employment, instead of the growth rate of output itself, and the variance of estimate is corrected for the error in this projection. Table 5 presents the results. The first question to ask is whether there is evidence of significant errors in variables at all in the data. Letting q be the difference between the parameter values estimated in the MEEV and ME models and V(q) be the difference in their variance-covariance

---

21. The probability of the observed value of the Kolmogorov-Smirnov test statistic for normality, under the null hypothesis that the within-moments were normal, was zero. The third and fourth order estimators were calculated as K(2,1)/K(1,2) and K(2,2)/K(3,1), respectively, where K(p,q) is the sample cumulant of order p in y^d and order q in g^d. These statistics are discussed in detail in Kendell and Stuart (1969).

Table 5: Instrumental Variable Estimates for the MEEV Model. a/

Coefficient of	Regression		
	Industry specific effects	Mixed effects	Weighted aggregate between industry b/
	(1)	(2)	(3)
Micro growth rate, α	1.55 0.40	1.55 0.40	10.77 2.01 (applies to α and ϕ jointly)
Macro growth rate, ϕ		9.22 2.05	
σ^2	0.39 c/	n.r. d/	n.r.
Degrees of freedom	512	527	15

a/ Small numerals are standard errors: n.r. means not relevant.

b/ The weight for industry j is $(\sigma_\eta^2 + \sigma_\varepsilon^2/N_j)^{-\frac{1}{2}}$.

c/ $0.39=\sigma_\varepsilon^2$.

d/ First stage $\sigma^2 = 0.51$. It follows that $\sigma_\eta^2 = (0.51 - \sigma_\varepsilon^2) = 0.12$.

e/ Between firms, within industries.

matrix, then, Hausman (1978) has shown that under the null hypothesis that there are no errors in variables $q'[V(q)]^{-1}q$ distributes, assymptotically, as an X_2^2 deviate. The observed value of the X_2^2 statistic was 6.16 which is significant at the .05 level. Similar tests when applied to the within- and between-regressions separately resulted in X_1^2 deviates of 4.5, which is significant at the .05 level, and .03, which is clearly insignificant. That is, though there is a significant error variance in g_{ij} the averaging procedure leaves $g_{.j}$ relatively error free. Since g_{ij} and $g_{.j}$ are positively correlated, a simple left out variable argument indicates that the MEEV estimate β should be higher then the estimate from the ME model, while its estimate of ϕ should be lower. This is, in fact, what occurs as the estimate of β moves from 1.05 to 1.55 while that of ϕ from 9.93 to 9.22. A comparison of the MEEV to the ME estimator of β indicates that the error to total variance ratio in g^d is about 33 percent which implies that the value of the ratio in g is 27 percent. In this problem, then, use of within-group moments to estimate β does not aggravate the errors in variable bias in that coefficient markedly. In sum, the results from the MEEV model are quite clear. There is an error in variable problem in the BWD but it is not the major factor in explaining the observed differences between the macro and micro regressions. The macro coefficient is still seven times as large as the micro coefficient, indicating that there is a large aggregation effect in the R&D intensity decision, which is highly, and positively, correlated with the industry's past growth rate.

A slightly different test for the influence of misspecification in the growth rate measure on the results reported in the last section can be performed on the ND. The errors in variable model assumes that the unobserved growth rate (g) measures the true independent variable plus an uncorrelated measurement error (v). In the context of the micro model underlying the R&D intensity decision, the firm changes its R&D intensity in response to a change in its expected future growth rate (g^e) and hence the errors model assumes that $g = g^e + v$. The ND can be used to check if a rational expectations model of the formulation of g^e can account for the observed aggregation effect in the R&D intensity decision.

Rewriting the equation determining the firm's R&D intensity in terms of g^e, allowing for mixed aggregation effects at an industry level, and aggregating to determine the R&D intensity of the product group, one has

$$y_{.pj} = \beta_0 + \beta g^e_{.pj} + \phi g_{..j} + \eta_j + \varepsilon_{.pj} , \qquad (11)$$

where p indexes product groups and j indexes industries.

Since g^e is not observable, (10) cannot be estimated without more information on either indicators or determinants of its value. Consider the realized future growth rate, g^r. It can always be partitioned into a linear function of the variables, X, known to the firm in period t, and an uncorrelated error, ξ ;

$$g^r = X\delta + \xi \ . \qquad (12)$$

The firm uses the information available to it at t to choose an estimator of g^r. The minimum vsriance unbiased predictor of g^r, given X, is, of course:

$$g^e = X\delta \ . \qquad (13)$$

Substituting (12) into (11):

$$g^r = g^e + \xi \ . \tag{14}$$

That is, in the rational expectations model g^r is an indicator which measures g^e subject to an uncorrelated disturbance. Hence, substitution of g^r into the R&D intensity decision for g^e produces the classical errors in variable estimation problem. Any variable which is a member of the set of variables described by X and which is uncorrelated with the other error components in the R&D intensity decision is a suitable instrument for a two-stage least-squares estimation procedure which allows for the covariance structure of the errors in (10).[22]

Past growth rates of the product group and the industry for 4 and 11 years were used as instruments while g^r was set equal to the product-group growth rate for the 5-year period following the R&D intensity decision. Estimation of the model is done in several steps. First g^r is projected onto the instruments and the value predicted by the instruments is derived. The product group's

---

22. It is a suitable instrument because X is uncorrelated with ξ by construction, and is correlated with g^r provided that $\delta \neq 0$. I thank Zvi Griliches for suggesting to me the basic idea underlying this model which is contained in an unpublished paper by J.F. Muth. Of course, one does not know the actual list of variables in X. However, any set of variables, W, which are both known to the firm in period t and are uncorrelated with the other error components in the R&D intensity decision, may be used as instruments, since a non-zero correlation between W and ξ would imply that the econometrician knows more about the future demand conditions of the firm than the firm itself.

R&D intensity is then regressed once on the predicted value of g^r and the industry growth rate, and once on the predicted value and the industry dummies. The variance of estimate from the latter regression provides a consistent estimate of the variance of the error without group or industry structure, which was 0.37 while the difference in the variance of estimate from the two regressions provides a consistent estimate of the variance of the error with group structure, which was 0.33. Finally, $y_{\cdot pj}$ is regressed against the predicted value of g^r and the industry growth rate, using the estimated variance components to do the required GLS transformation of the data. This latter regression, after correction for the standard error of estimate, provides consistent estimates of all parameters and their standard errors.

The estimated micro coefficient ($\hat{\beta}$) was 1.91 which, as expected, is slightly higher than the ME estimate of this parameter from the ND (0.02) and more in line with the estimates from the firm level data sets. The standard error of the estimate of β, however, was 7.83 which indicates that the ND and the rational-expectations models do not provide very precise information on the micro coefficient. On the other hand, the estimate of ϕ was 23.51 which is almost identical to the ME estimate of ϕ from the ND (23.47), and had a standard error of 5.35, which is lower than the ME standard error (6.07). Though the ND and the rational-expectations model do not define β very precisely, they do provide strong evidence of an aggregation effect in the R&D intensity which is highly correlated with the industry growth rate.

Both the MEEV and the rational expectations models indicate that the results reported in the last section cannot be explained via misspecification in the growth rate measure. Rather, there is an aggregation effect in the R&D intensity decision whose variance is mostly (about 62 percent of it) accounted for by the variance in the indutry's growth rate.

6. Some Characteristics of the Aggregation Effect in the R&D Intensity Decision

The preceding five sections used the R&D intensity equation to illustrate a farmework for analyzing differences in estimated relationships at different levels of aggregation when all observations have the same response parameters. Given the descriptive and policy implications of knowledge producing activities, one would like to know more about the causal factors underlying the aggregation effect in the R&D intensity decision. Though an explicit analysis of this issue is beyond the scope of the present paper, we are able to summarize some of the characteristics of this aggregation effect.

First, it is associated with the two digit industry to which the firm belongs. The two-way ME model indicated that the firm's three digit industry had little effect on its choice of R&D intensity, and what effect there was, was uncorrelated with the product groups past growth rate. When instrumental variables were used to estimate a two-way MEEV model the product group growth rate coefficient did increase, to 1.51 with a standard error of 1.49, but was still insignificant and small compared to the firm's response to a unit increase in its two digit industry's growth rate. Either the factors underlying the 3-digit classification scheme are not particularly relevant to the firm's choice of R&D intensity, or those factors are not adequately represented in the classification that exists. One should consider this fact when choosing data sets for analyzing issues related to R&D. One possibility which was not examined here is that the factors underlying the aggregation effect do not differentiate between all different two-digit industries but only between groups of them.

Second, the aggregation effect is highly correlated with the industry's past growth rate. Since the interindustry variance in R&D intensity is about 49 percent of the total variance in this variable, this makes the industry growth rate an important determinant of interfirm differences in R&D intensity even though the firm's past growth rate is not. Moreover, experiments were run using longer term growth rates on the BWD (8-year) and shorter term growth rates on the ND (4-year). In all cases the larger the term of the past growth rate used, the larger the R&D intensity response to an increase in the industry's growth rate, and the smaller the response to the firm's growth rate. Moreover, the values of the aggregation effects for the same industries in the different data sets (which, recall, were many years apart) were highly correlated. Apparently it is sustained, long-term, industry growth which accounts for the simultaneous movement of the R&D intensities of the industry's firms.

In a recent comparison of micro and macro estimates of investment functions for traditional capital goods, Eisner (1978) proposed that differences in estimated parameters could be explained in terms of the firm using past industry output in its predictions of its own future output. The rational expectations and errors in variable models of the last section indicate that this is not the major factor underlying the aggregation effect in the demand for research resources. Further, the values of the aggregation effect were not highly correlated with some rough measures of industry concentration presented in Gort (1962), so that explanations of the observed results based on rivalry, say the average value of the industry's R&D intensity being a determinant of the firm's choice of R&D intensities, do not seem promising.

One determinant of R&D demand which is widely discussed in the literature and whose impact seems to be consistent with the results reported above is technological opportunity. That is the scientific base of society may make R&D investments in one part of the economy more productive than in another. Technological opportunities are likely to be similar for firms within an industry, and should be involved in a causal nexus relating sustained increases in industry growth to R&D demand [see Pakes (1978)]. Further investigation of this possibility seems warranted.

Appendix: The Assymptotic Variances of $\hat{\beta}$ and $\hat{\phi}$ in Dimensions N and J.

The following assumptions are made on the population moment matrices:

$$\lim_{N\to\infty} N^{-1}X'\Gamma X = B_J \qquad \lim_{N\to\infty} N^{-1}X'{}^dX^d = W_J$$

$$\lim_{J\to\infty} J^{-1}X'\Gamma X = B_N \qquad \lim_{J\to\infty} J^{-1}X'{}^dX^d = W_N$$

where B_J, B_N, W_J and W_N are of full column rank and do not depend on the index excluded in their subscripts.

It, then, follows trivially from equation (9) in the text that $\lim_{N\to\infty} Var(\hat{\beta}) = \lim_{J\to\infty} Var(\hat{\beta}) = 0$. Also,

$$\lim_{J\to\infty} Var\ (\hat{\phi}) = \lim_{J\to\infty} \frac{N\sigma_2^2}{J}\ B_N^{-1} + \lim_{J\to\infty} \frac{\sigma_\varepsilon^2}{J}\ (B_N^{-1} + N^{-1}W_N^{-1}) = 0\ ,$$

but:

$$\lim_{N\to\infty} Var(\hat{\phi}) = \lim_{N\to\infty} \frac{\sigma_\varepsilon^2}{N}\ (B_J^{-1} + N^{-1}W_J^{-1}) + \lim_{N\to\infty} \sigma_\eta^2 B_J^{-1} = \sigma_\eta^2 B_J^{-1}\ .$$

References

Aigner, D.J. and S.M. Goldfeld, 1974, Estimation and prediction from aggregate data when aggregates are measured more accurately than their components, Econometrica XLII, 113-134.

Arrow, Kenneth J, 1962, Economic welfare and the allocation of resources for invention, in: R.R. Nelson, ed., 1962, The rate and direction of economic activity: economic and social factors (Princeton University Press, Princeton, NJ) 609-625.

Balestra, Pietro and M. Nerlove, 1960, Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas, Econometrica XXXIV, 585-612.

Chamberlin, Gary, 1978, Analysis of covariance with qualitative data, Harvard Institute of Economic Research Discussion Paper no. 678 (Harvard University, Cambridge, MA).

Eisner, Robert, 1978, Cross section and time series estimates of investment functions, in: The Econometrics of Panel Data, Annales de l'inséé, (Collogue International Du CNRS, Inséé, Paris) 99-130.

Friedman, Milton, 1957, A theory of the consumption function, National Bureau of Economic Research, General Series no. 63 (Princeton University Press, Princeton, NJ).

Gort, Michael, 1962, Diversification and integration in American industry, National Bureau of Economic Research, General Series no. 77 (Princeton University Press, Princeton, NJ).

Griliches, Zvi, 1958, Research costs and social returns: hybrid corn and related innovations, Journal of Political Economy LXVI, 419-431.

__________, forthcoming, Issues in assessing the contribution of research and development to productivity growth, in: Bell Journal of Economics, forthcoming.

Grunfeld, Yehuda, 1960, The determinants of corporate investment, Arnold C. Harberger, ed., The Demand for Durable Goods (Univeristy of Chicago Press, Chicago, IL).

Grunfeld, Yehuda and Zvi Griliches, 1962, Is aggregation necessarily bad?, The Review of Economics and Statistics XLII, 1-13.

Grunfeld, Yehuda and David Levahri, 1962, A note on external economics, unpub. mimeograph (Jerusalem, Israel).

Hausman, Jerry, 1978, Specification tests in econometrics, Econometrica 46, Nov., 1251-1273.

Kendall, M.G. and A. Stuart, 1969, The advanced theory of statistics, 3rd ed., (Charles Griffin & Company Limited, London).

Kuh, Edwin, 1959, The validity of cross-sectionally estimated behavior in time-series applications, Econometrica XXVII, 197-214.

Mundlak, Yair, 1978, On the pooling of time series and cross section data, Econometrica XLVI, 69-86.

National Science Foundation, 1958, Research and development in industry (U.S. Government Printing Office, Washington).

__________, 1978, Review of data on science resources, NSF78-303 (U.S. Government Printing Office, Washington).

Pakes, Ariel, 1978, Economic Incentives in the production and transmission of knowledge; an empirical analysis, unpublished Ph.D. dissertation (Harvard University, Cambridge, MA).

Reiersöl, O., 1950, Identifiability of a linear relation between variables which are subject to error, Econometrica XVIII, 375-399.