NBER WORKING PAPER SERIES

THE VARIANCES OF REGRESSION COEFFICIENT ESTIMATES USING AGGREGATE DATA

Roy E. Welsch*
Edwin Kuh*

Working Paper No. 60

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

October 1974

Preliminary: not for quotation

*NBER Computer Research Center and Massachusetts Institute of Technology. Research supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research, Inc.

Abstract

This paper considers the effect of aggregation on the variance of parameter estimates for a linear regression model with random coefficients and an additive error term. Aggregate and micro variances are compared and measures of relative efficiency are introduced. Necessary conditions for efficient aggregation procedures are obtained from the Theil aggregation weights and from measures of synchronization related to the work of Grunfeld and Griliches.

Contents

1. Introduction

In economics and other social sciences we are often confronted by aggregate (macro) data with little hope of recovering all of the micro data used to obtain the aggregate. In other cases we can obtain some of the micro data, for instance individual company data for firms listed on the stock exchanges, and then we must consider whether to use the limited micro data or the aggregate data.

It is also possible to obtain representative bodies of micro data in order to analyse their aggregation properties with techniques to be developed here. In this way, it should be possible to canvass a broad range of situations and (empirically) arrive at a general understanding of how aggregation influences estimates from aggregative data, since econometricians are frequently obliged to use aggregates (e.g., in macro model building) when micro data are unavailable.

In an earlier paper (Kuh (1974)) it was shown under certain assumptions that the variances of the estimated macro coefficients of a particular regression model decrease as the number of individuals in the aggregate increases. This suggests that in some cases it is plausible that the aggregate data rather than the limited micro data should be used for estimation.

In this paper we expand the class of regression models considered and sharpen the results obtained in Kuh (1974). We also propose a measure of the relative efficiency of aggregation and examine in more detail what conditions should hold in order to make aggregate estimates useful competitors to estimates obtained from limited micro data.

The results presented are related to those of Theil (1968 and 1971) and by Swamy (1971, pp. 15-16) in somewhat more general form, who indicated

that the population coefficient variances of a certain macro equation with random micro coefficients could tend to zero as the aggregate grows, whereas we treat a more complete model as well as its estimation properties. At the end of this paper, we derive some interesting inequalities relating to the aggregation weights of Theil (1954).

Grunfeld and Griliches (1960) have compared the power and degree of "explanation" obtained from micro data with that obtained from aggregate data. Our approach emphasizes estimated coefficient variances rather than explanatory power, but we are able to show that the grouping or "synchronization" effect first noticed by Grunfeld and Griliches continues to play an important role.

Feige and Watts (1972) have studied the problems of a data collecting organization which is trying to protect privacy by partial cross-sectional aggregation of data for individuals to the state level, and trying to minimize the information loss at the same time. While the problems we attack are related to the use of aggregate data, rather than creating the aggregates, some of the procedures Feige and Watts recommend have proven useful to us.

Aigner and Goldfeld (1974) consider the problems of estimation and prediction when the independent variables can be measured more accurately with aggregate data than with micro data. In the beginning of their paper Aigner and Goldfeld derive several results for the case of no error in the independent variables. The results in our paper are generalizations to the case of more than one independent variable, more than two micro units, and to a model with stochastic parameter variation.

It is important to emphasize that in a situation where substantially all of the micro data is available, aggregation is not particularly appealing when the performance criterion is the variance of the estimated coefficient. In such cases maximum likelihood and related procedures applied to the micro data use information that is ignored when estimates are based in the aggregate.

2. The Regression Model

The micro data is assumed to be centered in order to avoid purely technical details in later sections. We have, for i=1,...,N:

$\underline{y}_i$, a T x 1 vector with $\sum_{t=1}^{T} y_{it} = 0$, and

$\underline{\underline{x}}_i$, a T x K matrix with $\sum_{t=1}^{T} x_{it\ell} = 0$ for $\ell=1,\ldots,K$.

For the micro equations, we assume the existence of the regression structures

$$(2.1) \qquad y_{it} = \underline{x}_{it}\underline{\beta}_{it} + \varepsilon_{it} \qquad \begin{matrix} i=1,2,\ldots,N \\ t=1,2,\ldots,T \end{matrix}$$

where

y_{it} is the dependent variable,

$\underline{x}_{it}$ is a 1 x K vector of nonstochastic "explanatory" variables,

$\underline{\beta}_{it}$ is a K x 1 vector of regression parameters,

ε_{it} is the additive "error" component,

and

a. $E(\varepsilon_{it}) = 0$

b. $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij}\delta_{st}$ where δ is the Kronecker delta. For each i we allow $_{ij} \neq 0$ only for a set of subscripts of j having no more than L_1 elements where for N sufficiently large, $L_1 < N$, and where L_1 is functionally independent of N. When only j = i is included in each set of subscripts we have the special case of uncorrelated errors, i.e. $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ii}\delta_{ij}\delta_{st}$.

c. $\underline{\beta}_{it}$ is the realization of a multivariate wide-sense stationary stochastic process with $E(\underline{\beta}_{it}) = \underline{\beta}$ and

$$(2.2) \qquad E(\underline{\beta}_{it}-\underline{\beta})(\underline{\beta}_{jt+s}-\underline{\beta})' = \underline{\underline{r}}_{ij}(s).$$

For each i we allow the elements of $\underline{\underline{r}}_{ij}(s)$ to be non-zero only for a set of

subscripts j having no more than L_2 elements where for N sufficiently large, $L_1 < N$, and L_1 is functionally independent of N. Of course $j = i$ is always included in this set and when only $j = i$ is included we have the special case of $\underline{\beta}_{it}$ and $\underline{\beta}_{jt}$ uncorrelated for $i \neq j$.

d. $\underline{\beta}_{it}$ and ε_{js} are uncorrelated for all s and t and i and j. For each i this model is related to that proposed by Burnett and Guthrie (1970).

The above assumptions require comment. (2.2a) is a standard assumption in regression analysis that we retain here. (2.2b) permits some contemporaneous correlation in the additive error variance among individuals and allows these variances to differ across micro units.

(2.2c) allows for what, in principle, could be a complex autocorrelated random process in the micro coefficients. Relaxation of the assumption that the population micro parameters are fixed for all time represents a substantial increase in realism. Individual firms or persons may often react according to a stable underlying set of parameters, but that behavior often departs from its basic (i.e. average) modus operandi in more complicated ways than can be represented by additive errors. Some correlation among the micro random coefficient processes is permitted. This should be sufficient to allow e.g., for geographic interactions, or taste dependence among individuals. However, where strong oligopolistic dependence exists among firms the condition governing correlation among parameter vectors could easily be violated. In general, however, there does

not appear to be a greater departure from reality in this instance than in other assumptions made in the estimation of economic or social behavior relationships.

Finally (2.2d) asserts that the two sources of randomness are uncorrelated, a proposition that is convenient and does not appear to be a particular cause for concern.

In summary, the random coefficient model allows for much richer behavioral variations that should be considered in an aggregation context. Since there are two sources of random variation which are assumed to be independent, results from the following analysis hold for either alone, or both. Thus the reader can choose which aspects are most appealing for his immediate estimation concerns.

The above model overlaps with the one considered by Aigner and Goldfeld (1974) when $K = 1$, $N = 2$ and $\underline{\underline{r}}_{ij}(s) \equiv 0$ for all s, i and j. Our assumption that $E(\beta_{it}) = \underline{\beta}$ then implies that $\underline{\beta}_{it} = \underline{\beta}$ for all i and t. Hence in the non-stochastic parameter case we do not allow $\underline{\beta}$ to vary among the micro units. Aigner and Goldfeld relax this assumption and consider the implication of $\underline{\beta}_i \neq \underline{\beta}_j$ in their special non-stochastic parameter model.

The macro data will be represented by $\underline{Y} = \sum_{i=1}^{N} \underline{y}_i$ and $\underline{\underline{X}} = \sum_{i=1}^{N} \underline{\underline{x}}_i$. Throughout what follows we shall assume that $\underline{\underline{X}}$ is of full rank. We propose to estimate $\underline{\beta}$ by

$$(2.3) \qquad \underline{b} = (\underline{\underline{X}}'\underline{\underline{X}})^{-1}\underline{\underline{X}}'\underline{Y}.$$

Using an argument analogous to one in [69] we have

$$(2.4) \qquad E(Y_t) = \sum_{i=1}^{N} E(y_{it}) = \sum_{i=1}^{N} \underline{x}_{it} E(\underline{\beta}_{it}) + \sum_{i=1}^{N} E(\epsilon_{it})$$

$$= \sum_{i=1}^{N} \underline{x}_{it}\underline{\beta} = \underline{X}_t\underline{\beta}$$

which implies that $E(\underline{Y}) = \underline{\underline{X}}\underline{\beta}$.

Given the complicated regression structure, the estimator in (2.3) is certainly open to improvement. Under simplifying assumptions there are operational methods for an improved estimate of the micro parameters. Details are available in Rosenberg [1973a,b].

We chose to look at a simple estimator, ordinary least squares, for two reasons. First, we are primarily interested in studying aggregation and second, we feel that given the present state of the theory, a lot of data will still be analyzed using ordinary least squares.

3. Variance Properties of the Aggregate Estimator

For each N we can compute the covariance matrix of $\underline{b}$, denoted by $\underline{\underline{V}}_N(\underline{b})$. We are interested in finding conditions so that the elements of $\underline{\underline{V}}_N(\underline{b})$ will remain bounded (or go to zero) as N increases. Let $\underline{\underline{Z}}_i$ be the T x KT matrix

$$\underline{\underline{Z}}_i = \begin{bmatrix} \underline{X}_{i1} & \underline{0} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{X}_{i2} & \underline{0} & \cdots & \vdots \\ \underline{0} & \underline{0} & \underline{X}_{i3} & \underline{0} & \vdots \\ \cdots & \cdots & \underline{0} & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \underline{X}_{iT} \end{bmatrix}$$

and set

$$\underline{\beta}_i = \begin{bmatrix} \underline{\beta}_{i1} \\ \vdots \\ \underline{\beta}_{iT} \end{bmatrix} \qquad \underline{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \qquad \beta^* = \begin{bmatrix} \underline{\beta} \\ \vdots \\ \underline{\beta} \end{bmatrix}$$

$$\underline{\underline{G}} = (\underline{\underline{X}}'\underline{\underline{X}})^{-1}\underline{\underline{X}}'.$$

Then

$$\underline{Y} = \sum_{i=1}^{N} (\underline{\underline{Z}}_i\underline{\beta}_i + \underline{\varepsilon}_i) \tag{3.1}$$

and

$$(\underline{b}-\underline{\beta}) = \underline{\underline{G}} \left(\sum_{i=1}^{N} [\underline{\underline{Z}}_i(\underline{\beta}_i-\underline{\beta}^*) + \underline{\varepsilon}_i] \right). \tag{3.2}$$

It is now convenient to define the KT x KT matrix

$$\underline{\underline{V}}_{ij} = \begin{bmatrix} \underline{\underline{r}}_{ij}(0) & \underline{\underline{r}}_{ij}(1) & \underline{\underline{r}}_{ij}(T-1) \\ \vdots & \vdots & \vdots \\ \underline{\underline{r}}_{ij}(T-1) & \underline{\underline{r}}_{ij}(T-2) & \underline{\underline{r}}_{ij}(0) \end{bmatrix} \qquad i,j,=1,2,\ldots,N$$

which represents the covariance structure of the stochastic process described in (2.2c). If we use assumptions b, c, and d of (2.2) then

$$(3.3) \qquad \underline{\underline{V}}_N(\underline{b}) = E(\underline{b}-\underline{\beta})(\underline{b}-\underline{\beta})' = \sum_{1\le i,j\le N} [\underline{\underline{G}}\underline{z}_i \underline{\underline{V}}_{ij} \underline{z}_j' \underline{\underline{G}}' + \sigma_{ij}(\underline{\underline{X}}'\underline{\underline{X}})^{-1}]$$

Since $\underline{\underline{V}}_N(\underline{b})$ is a covariance matrix, the Cauchy-Schwarz inequality implies that in order to obtain bounds on the magnitude of the elements of $\underline{\underline{V}}_N(\underline{b})$ we need only examine the diagonal elements.

Theorem 1. If

$$(3.4) \qquad \begin{aligned} &(a) \quad \sup_{\substack{\text{all } i \\ 1\le \ell \le K \\ 1\le t\le T}} |x_{i\ell t}| \le M_1 \\ &(b) \quad \sup_{\substack{\text{all } i,j \\ 1\le p\le TK \\ 1\le q\le TK}} |V_{ijpq}| \le M_2 \\ &(c) \quad \sup_{\text{all } i\ne j} |\sigma_{ij}| \le M_3 \end{aligned}$$

then

$$(3.5) \qquad [\underline{\underline{V}}_N(\underline{b})]_{\ell\ell} \leq [BT + E + \sigma^2(N)]N(\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{\ell\ell}$$

where B and E are constants independent of N, the subscript ℓ designates a typical explanatory variable or its coefficient and $\sigma^2(N) = \sum_{i=1}^{N} \sigma_{ii}/N$.

Proof. Let $|\underline{g}^{(\ell)}|$ denote the vector whose components are the absolute value of the components of $\underline{g}^{(\ell)}$, the ℓ^{th} row of $(\underline{\underline{X}}'\underline{\underline{X}})^{-1}\underline{\underline{X}}'$, and $\underline{\underline{1}}$ denote the $T \times T$ matrix with each component equal to 1. Conditions (a) and (b) of (3.4) imply that

$$(3.6) \qquad \sup_{\substack{\text{all } i,j \\ 1\leq p\leq T \\ 1\leq q\leq T}} |(\underline{\underline{z}}_i\underline{\underline{V}}_{ij}\underline{\underline{z}}'_j)_{pq}| \leq B' < \infty$$

where B' is a function of M_1 and M_2. Now

$$(3.7) \qquad [\underline{\underline{V}}_N(\underline{b})]_{\ell\ell} = \sum_{i,j}^{N} \underline{g}^{(\ell)}\underline{\underline{z}}_i\underline{\underline{V}}_{ij}\underline{\underline{z}}'_j\underline{g}'^{(\ell)} + \Big(\sum_{i=1}^{N} \sigma_{ii} + \sum_{i\neq j}^{N} \sigma_{ij}\Big)(\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{\ell\ell} .$$

and

$$(3.8) \qquad \sum_{i,j}^{N} \underline{g}^{(\ell)}\underline{\underline{z}}_i\underline{\underline{V}}_{ij}\underline{\underline{z}}'_j\underline{g}'^{(\ell)} = \sum_{i,j}^{N} \operatorname{tr} \underline{\underline{z}}_i\underline{\underline{V}}_{ij}\underline{\underline{z}}'_j\underline{g}'^{(\ell)}\underline{g}^{(\ell)}$$

$$\leq NB \operatorname{tr} \underline{\underline{1}}|\underline{g}'^{(\ell)}||\underline{g}^{(\ell)}|$$

with $B = B'L_2$.

But by the Cauchy-Schwarz inequality

$$(3.9) \qquad \operatorname{tr} \underline{\underline{1}}|\underline{g}'^{(\ell)}||\underline{g}^{(\ell)}| = \Big[\sum_{t=1}^{T} |g_t^{(\ell)}|\Big]^2 \leq T \sum_{t=1}^{T} (g_t^{(\ell)})^2$$

and

$$(3.10) \qquad \sum_{t=1}^{T} (g_t^{(\ell)})^2 = |\underline{g}^{(\ell)}||\underline{g}'^{(\ell)}| = (\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{\ell\ell} .$$

Returning to the second term of (3.7) we have

$$\left(\sum_{i=1}^{N} \sigma_{ii} + 2 \sum_{i<j}^{N} \sigma_{ij}\right)(\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{\ell\ell} \leq (N\sigma^2(N) + NM_3L_1)(\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{\ell\ell}$$

$$= N(\sigma^2(N) + E)(\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{\ell\ell}$$

where $E = M_3L_1$. The inequality (3.5) follows immediately and the proof of Theorem 1 is complete.

Thus we have shown that under rather plausible conditions we can examine $\underline{V}_N(\underline{b})$ by looking at $N(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$. Condition (3.4a) merely states that all elements of the explanatory variables should be bounded, (3.4b) imposes the same mild restriction on the covariance structure of the stochastic processes generating the β_{it} and (3.4c) places an upper bound on the micro equation additive error variance.

4. The Structure of $(\underline{X}'\underline{X})^{-1}$

From the above discussion it is clear that $N(\underline{X}'\underline{X})^{-1}$ plays a crucial role in determining the reduction in the estimated parameter variances that might be obtained from aggregation. We can always compute $N(\underline{X}'\underline{X})^{-1}$ but it is useful to see what conditions imposed on $N(\underline{X}'\underline{X})^{-1}$ imply about the structure of the micro data. Let $\underline{X}_\ell$, $\ell = 1,\ldots,K$ denote the columns of $\underline{X}$. (Recall that all the data are centered.) We shall use S_ℓ^2 to denote $\sum_{t=1}^{T} X_{\ell t}^2/T$ (it is assumed that $S_\ell^2 > 0$ for all ℓ), $S_{\ell..}^2$ to denote the error variance for the regression of X_ℓ on the remaining K-1 explanatory macro-variables and $R_{\ell..}^2$ the corresponding multiple correlation. From page 166 of Theil (1971) we have that

$$(4.1) \qquad (\underline{X}'\underline{X})^{-1}_{\ell\ell} = \frac{1}{TS_{\ell..}^2} = \frac{1}{T\,S_\ell^2(1-R_{\ell..}^2)} \qquad \ell = 1,\ldots,K.$$

5. Limiting Properties of Macrovariances

The previous analysis has provided bounds for $\underline{\underline{V}}_N(\underline{b})$, the macro parameter variance in terms of $(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$ and conditions on the microvariables and microparameter variances, when the macroparameters are defined simply as least squares estimates based on the macro data. We now discuss under what conditions $\underline{\underline{V}}_N(\underline{b})$ tends to zero as the number of elements in the aggregate increases. We shall always make the plausible assumption that $\sup_N \sigma^2(N) < \infty$.

It is then clear from Theorem 1 that

$$\lim_{N\to\infty} \underline{\underline{V}}_N(\underline{b}) = 0 \tag{5.1}$$

if

$$\lim_{N\to\infty} N(\underline{\underline{X}}'\underline{\underline{X}})^{-1} = 0 \tag{5.2}$$

and conditions (3.4a) to (3.4c) are satisfied.

The formulas in (4.1) imply that (5.2) will hold if

$$\lim_{N\to\infty} \frac{N}{TS_\ell^2(1-R_{\ell\cdot\cdot}^2)} = 0, \quad \ell = 1,2,\ldots,K. \tag{5.3}$$

In order for $(\underline{\underline{X}}'\underline{\underline{X}})$ to be invertible we must have $R_{\ell\cdot\cdot}^2 < 1$ for all ℓ. In most applications related to economic data it is reasonable to assume that

a $\delta > 0$ exists so that

$$\sup_N R^2_{\ell\cdot\cdot} < 1 - \delta. \tag{5.4}$$

If (5.4) holds then having

$$\lim_{N\to\infty} \frac{N}{TS^2_\ell} = 0 \tag{5.5}$$

is enough to imply (5.3).

In the spirit of Grunfeld and Griliches (1960), we define the average variance among microvariables as

$$s^2_\ell = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} (x_{i\ell t})^2 \tag{5.6}$$

and the average simple correlation among microvariables as

$$r_\ell = \frac{2}{N(N-1)} \sum_{i<j} \frac{1}{T} \sum_{t=1}^{T} \frac{(x_{i\ell t})\,(x_{j\ell t})}{s^2_\ell}. \tag{5.7}$$

Then it follows that

$$\frac{N}{TS^2_\ell} = \frac{1}{Ts^2_\ell[1+(N-1)r_\ell]}. \tag{5.8}$$

In any cases of conceivable interest, we would expect s^2_ℓ to be bounded away from zero. In many but not all economic applications, r_ℓ can be expected to be positive and bounded away from zero. If, however, the aggregate is constrained, we can have $r_\ell < 0$, and other such instances could arise. (Since $S^2_\ell > 0$ we must at least have $r_\ell > -1/(N-1)$.) Thus, for given T, positive s^2_ℓ and r_ℓ, condition (5.5) will hold so that the

macrovariances can be expected to shrink as the aggregate grows. A meaningful industrial aggregate is normally composed of firms with common production methods and similar customers. While in the short run, one firm's gain may be another's loss, fluctuations in market demand will ordinarily be shared in rough proportion to each firm's productive capacity. While some firms grow in periods of declining demand and others fade when demand is growing, this "maverick" behavior is unlikely to dominate. Clearly, however, effectiveness from the point of view of reduction of parameter variance depends on the strength of the average correlation among entities comprising the aggregate as well as collinearity among the explanatory variables reflected in $R^2_{\ell..}$. This effect has also been discussed by Aigner and Goldfeld (1974) for their model.

We have examined conditions which imply that $\underline{\underline{V}}_N(\underline{b}) \to 0$ with increasing N. These conditions can be weakened if we only require that the elements of $\underline{\underline{V}}_N(\underline{b})$ remain bounded as N becomes large. One might conjecture that the larger most aggregates become, even in a well designed aggregation procedure, the more dissimilar the components will be, thereby placing definite limits on the amount of variance reduction that can in fact be achieved. For example, because of the dissimilarities introduced as the number of components increases it might be that $(N-1)r_\ell$ in equation (5.8) is bounded (i.e. r_ℓ is not bounded away from zero). In this case N/TS^2_ℓ would be bounded (if s^2_ℓ is bounded away from zero) and of course $\underline{\underline{V}}_N(\underline{b})$ would not necessarily approach zero.

The condition (5.2) also implies that the aggregation weights introduced by Theil (1954) must vanish. Theil used the aggregation weights to discuss the aggregation bias that occurs when $E(\underline{\beta}_{it}) = \underline{\beta}_i$ rather than $E(\underline{\beta}_{it}) = \underline{\beta}$ as we have assumed (2.2c) ; there is no bias under our assumptions. However, the Theil weights play an interesting role when estimated parameter variances are considered, even in the case $E(\underline{\beta}_{it}) = \underline{\beta}$.

The Theil weights are defined by

$$\underline{w}_{i\ell} = (\underline{\underline{X}}'\underline{\underline{X}})^{-1}\,\underline{\underline{X}}'\underline{x}_{i\ell} \tag{5.9}$$

where $\underline{x}_{i\ell}$ $\ell=1,\ldots,K$ denote the columns of the matrix $\underline{\underline{x}}_i$. Using a proof similar to the one used for Theorem 1 (but not reproduced here) we have the following result.

Theorem 2. If (3.4a) holds then

$$\sum_{i=1}^{N} (w_{ilp})^2 \leq M_1^2\, T\, N\, [(\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{pp}]. \tag{5.10}$$

Finally from (5.10) it is possible to show that if (5.2) holds

$$\lim_{N\to\infty} w_{ip\ell} = 0 \quad \ell=1,\ldots,K;\ p=1,\ldots,K.$$

In the special case where $\sigma_{ij} = 0$ $i\neq j$, $\sigma_{ii} = \sigma^2$ for all i, and $\underline{\beta}_{it} = \underline{\beta}$ for all i and t we have from (3.3) that

$$\underline{\underline{V}}_N(b) = N\sigma^2(\underline{\underline{X}}^1\underline{\underline{X}})^{-1}$$

and therefore

$$\sum_{L=1}^{N} (w_{ip\ell})^2 \leq M_1^2\, T\sigma^{-2}\, [V_N(b)]_{pp} \quad . \tag{5.11}$$

If we view the diagonal weights $w_{i\ell\ell}$ as resembling proportions [cf. Kuh 1974] which represent the relative contribution of the micro components $\underset{\sim}{x}_{i\ell}$ to $\underset{\sim}{X}_{\ell}$, then (5.11) provides a necessary condition for a reduction in aggregate parameter variance toward zero: no micro explanatory variable can be a large proportion of the aggregate of that variable. This is of interest because size distritutions of most extensive measures of firm or household activity are reasonably stable and quite heavily skewed. Thus conditions most favorable to the swift attenuation of proportional shares as members of an aggregate increase - approximate size equality - are notably absent. As a result, distinct limits to the shrinkage of the macroparameter variances are likely to be imposed by the behavior of the underlying size distributions.

6. Comparisons of Micro and Macro Data.

In section 5 we examined the macro parameter variances and attempted to discover some conditions which would cause them to decrease as N increased. If only the macro data is available we could check these conditions to see how well we might be doing but we would probably use the macro data anyway because nothing else would be available.

If all of the micro data is available, we still have several problems. The data could be completely pooled (that is, the i and t subscripts could be treated as replications) or for each i we could compute a micro regression and then average the resulting estimates to obtain an estimate of $\underline{\beta}$. Other approaches are clearly possible, including generalized least squares [Swamy (1970), Swamy (1971) and Swamy and Arora (1972), Amemiya (1971)].

These same techniques are available if we have only a portion of the micro data. We can, of course, try these and compare the results with those obtained from the aggregate data. Resource constraints, however, may render this infeasible.

We prefer to adopt the view that only a relatively small percentage of the micro data is actually obtainable. To keep the notation simple we will assume that there is just one piece of micro data available. Thus our theoretical measure of relative efficiency is defined as

$$(6.1) \qquad E_{i\ell} = \frac{\text{variance of } i^{th} \text{ individual}}{\text{macro variance}} = \frac{[\underline{\underline{V}}_i(\underline{b})]_{\ell\ell}}{[\underline{\underline{V}}_N(\underline{b})]_{\ell\ell}}$$

where $\underline{\underline{V}}_N(\underline{b})$ is defined in (3.3),

$$(6.2) \qquad \underline{V}_i(\underline{b}) = \underline{G}_i \underline{z}_i \underline{V}_{ii} \underline{z}_i' \underline{G}_i' + \sigma_{ii}(\underline{x}_i'\underline{x}_i)^{-1}$$

and

$$\underline{G}_i = (\underline{x}_i'\underline{x}_i')^{-1}\underline{x}_i' \ .$$

Using (3.5) we have

$$(6.3) \qquad E_{i\ell} \geq \frac{\sigma_{ii}[1+(N-1)r_\ell]}{((\sum_k \sigma_{kk}/N)+BT+E)} \cdot \frac{\sum_k^N S_{k\ell}^2/N}{S_{i\ell}^2} \cdot \frac{(1-R_{\ell..}^2)}{(1-R_{i\ell..}^2)} \ .$$

Therefore the right-hand side of (6.3) is a conservative estimate of $E_{i\ell}$. It has many interesting qualitative properties. If r_ℓ is small or negative (i.e. there is not much correlation among the exogenous micro-variables or negative correlation) then the relative efficiency is reduced. A positive r_ℓ coupled with a reasonable N indicates that aggregate estimates might be better when compared to estimates based on this particular portion of the micro data.

In the special case where the β_{it} are fixed independent of i and t, as in the standard regression case, B = 0. Thus a positive B decreases the relative efficiency, giving us a warning that the relative efficiency of aggregation could be severely reduced by a time-varying or cross-sectionally varying parameter structure. This case is also discussed by Aigner and Goldfeld [1974] from a different point of view. A similar result holds when E=0.

The remaining parts of (6.3) relate the particular micro data we have to the aggregate. If these ratios are greater than 1, then we may want to consider aggregation as a reasonable alternative to the use of the micro data.

Note that we are comparing the efficiency of inference from one sample of one micro unit with inference based on the aggregate data. This is an approximation to the case of limited micro data and is designed to point out the major factors that could make the aggregate estimate a useful competitor to the limited data micro estimate. A detailed simulation study is planned to more fully understand the measurement of relative efficiency in this case.

References

[1] Aigner, D. J. and Goldfeld (1974). Estimation and Prediction from Aggregate Data when Aggregates are Measured More Accurately than their Components. To appear in Econometrica.

[2] Amemiya, T. (1971). The Estimation of the Variances in a Variance-Components Model. International Economic Review 12, pp. 1-13.

[3] Burnett, T. and Guthrie, D. (1970). Estimation of Stationary Stochastic Regression Parameters. JASA 65, pp. 1517-1573.

[4] Feige, E. L. and Watts, H. W. (1972). An Investigation of the Consequences of Partial Aggregation of Micro-Economic Data. Econometrica 40, pp. 343-360.

[5] Grunfeld, Y. and Griliches, Z. (1960). Is Aggregation Necessarily Bad? The Review of Economics and Statistics 42, pp. 1-13.

[6] Klock, T. (1961). Note on a Convenient Notation in Multivariate Statistical Analysis and in the Theory of Linear Aggregation. International Economic Review 2, pp. 351-360.

[7] Kuh, E. (1974). An Essay on Aggregation Theory and Practice in Essays in Honor of Jan Tingergen (W. Sellekaerts, editor). New York: Macmillan, (forthcoming).

[8] Rosenberg, B. (1973a). A Survey of Stochastic Parameter Regression. Annals of Economic and Social Measurement 2, pp. 381-398.

[9] Rosenberg, B. (1973b). The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression. Annals of Economic and Social Measurement 2, pp. 398-428.

[10] Swamy, P. A. V. B. (1970). Efficient Inference in a Random Coefficient Regression Model. Econometrica 38, pp. 3-23.

[11] Swamy, P.A.V.B. (1971). Statistical Inference in Random Coefficient Regression Models, New York: Springer-Verlag.

[12] Swamy, P.A.V.B. and Arora, S.S. (1972). The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Model. Econometrica 40, pp. 261-75.

[13] Theil, H. (1954). Linear Aggregation of Economic Relations. Amsterdam: North Holland Publishing Company.

[14] Theil, H. (1968). Consistent Aggregation of Micromodels with Random Coefficients. Report 6816, Chicago, Center for Mathematical Studies in Business and Economy (April 1968). (A more accessible and somewhat more extensive treatment can be found in H. Theil [15].)

[15] Theil, H. (1971). Principles of Econometrics. New York: Wiley.

[16] Zellner, A. (1969). On the Aggregation Problem: A New Approach to a Troublesome Problem. Economic Models, Estimation and Risk Programming (Fox, K. A. et al., eds.), New York: Springer-Verlag.