On the Pooling of Time Series and Cross Section Data

Yair Mundlak

Econometrica, Vol. 46, No. 1 (Jan., 1978), 69-85.

Econometrica, Vol. 46, No. 1 (January, 1978)

ON THE POOLING OF TIME SERIES AND CROSS SECTION DATA[1]

BY YAIR MUNDLAK

In empirical analysis of data consisting of repeated observations on economic units (time series on a cross section) it is often assumed that the coefficients of the quantitiative variables (slopes) are the same, whereas the coefficients of the qualitative variables (intercepts or effects) vary over units or periods. This is the constant-slope variable-intercept framework. In such an analysis an explicit account should be taken of the statistical dependence that exists between the quantitative variables and the effects. It is shown that when this is done, the random effect approach and the fixed effect approach yield the same estimate for the slopes, the "within" estimate. Any matrix combination of the "within" and "between" estimates is generally biased. When the "within" estimate is subject to a relatively large error a minimum mean square error can be applied, as is generally done in regression analysis. Such an estimator is developed here from a somewhat different point of departure.

1. INTRODUCTION

The use of a sample consisting of time series observations on a cross section constitutes an important problem of empirical research in economics. A simple version of this problem is concerned with the estimation of a vector of parameters β in the relation.

$$(1.1)\qquad \underset{\sim}{Y} = X\beta + \underset{\sim}{\varepsilon}$$

where $\underset{\sim}{Y}$ and $\underset{\sim}{\varepsilon}$ are n-vectors, X is a $n \times k$ matrix of full rank and β is a k vector of parameters to be estimated. The error term is decomposed into:

$$(1.2)\qquad \varepsilon_{it} = m_i + s_t + u_{it}$$

where m_i and s_t are the systematic components, or effects, associated with the ith economic unit and the tth period (year) respectively; $i = 1, \ldots, N$; $t = 1, \ldots, T$ and $n = NT$. Thus, it is recognized that $X\beta$ does not account for all the systematic variations in $\underset{\sim}{Y}$.

The question is what effect should the decomposition (1.2) have on the method of estimation. Basically, two alternative approaches have been suggested, the "fixed effects" (FE) and the "random effects" (RE) of the analysis of variance. Each of the two models has been associated with a different estimator, the FE has resulted in the "within" estimator of covariance analysis [**14**] whereas the RE has led to a GLSE [**2**, **23**]. Knowing the variances in question, it is generally true that the GLSE is BLUE and therefore the current thinking among some writers has been to prefer this estimator. Furthermore, it has been explained that the gain in efficiency results from the utilization of the "between" estimator in addition to the within estimator. Since the GLSE is associated with the RE, its use had to be

[1] This is a revised and shorter version of [**17**]. At points, reference is made to [**17**] for more details. This work has been supported by an NSF Grant #SOC73-05374AO1. I have greatly benefited from the insight of Gary Chamberlain in discussions of the model. This, as well as the helpful comments of Zvi Griliches are reflected in the paper.

justified by arguing that economic effects are indeed random and not fixed. This position is well presented by Maddala [**11**].

The present state of thinking is unsatisfactory for two major reasons: first, the suggested rules for deciding whether an effect is fixed or random are at best inadequate. Second, the proposed GLSE approach has completely neglected the consequences of the correlation which may exist between the effects and the explanatory variables. Such a correlation leads to a biased estimator and it is the elimination of this bias that has originally led to the use of the covariance analysis estimator [**14**].

This paper proposes to remedy the situation by first indicating that the whole approach which calls for a decision on the nature of the effect, whether it is random or fixed, is both, arbitrary and unnecessary. Without a loss in generality, it can be assumed from the outset that the effects are random and view the FE inference as a conditional inference, that is, conditional on the effects that are in the sample. It is up to the user of the statistics to decide whether he wants inference with respect to the population of all effects or only with respect to the effects that are in the sample.[2] This view unifies the two approaches in a well defined form and eliminates any arbitrariness in deciding about "nature," in a way which is influenced by the subsequent choice of a "desirable" estimator.

If the foregoing approach is accepted the question is why would a uniform approach lead to two competing estimators for β, the coefficients which do not vary over individuals. That brings us to the second point which can be stated very simply: when the model is properly specified, the GLSE is identical to the "within" estimator. Thus there is only one estimator. The whole literature which has been based on an imaginary difference between the two estimators, starting with Balestra and Nerlove [**2**] is based on an incorrect specification which ignores the correlation between the effects and the explanatory variables.

It is thus argued that there is a uniform approach and a unique estimator. Furthermore, to obtain the correct GLSE of β, it is not necessary to know the components of variance. If this is the case, the old question of what to do if the within estimator has a large variance still remains but it is not different in nature from the question of having too many variables in a regression. One way to deal with this question is to use a mean square error estimator (MSEE). This is not a new idea but it is integrated into the discussion here.

The foregoing comments summarize the main points of the paper. In the remainder of this section we outline the plan of the paper and give some more detailed results. The model is outlined in Section 2. The formulation takes an explicit account of the relationships between the effects and the explanatory variables. Section 3 evaluates the performance of the alternative estimators under the RE set up. It is shown that the GLSE of β is the within estimator. Furthermore, when the effects are not correlated with the explanatory variables, the within and the between estimators are the same and therefore any weighted matrix combination thereof will be the same. What has been known in the literature as the GLSE for the error component model is actually a restricted

[2] The move to the unconditional inference requires that the sample be randomly drawn.

estimator, and when the restriction does not hold it is a biased estimator. A similar analysis is conducted under the FE model in Section 4.

The MSE estimator is introduced and discussed in Section 5. Basically, this estimator minimizes the MSE of the estimate of any linear combination of $\underset{\sim}{\beta}$. It requires a two stage procedure. The whole motivation for introducing the GLSE has been to gain efficiency. The question of efficiency is particularly important in small samples. The variance of the within estimator declines with the size of the sample as determined by increasing either the number of observations per unit or the number of units. Thus, any alternative estimator which increases the precision in small samples at the expense of unbiasedness should have the property of converging to the within estimator or simply be consistent. As shown in Section 6, the restricted GLS estimator is inconsistent and asymptotically biased when the sample increases by increasing the number of units rather than the number of observations per unit. This is of course the relevant process for increasing the sample size in economics. In contrast, the MSE estimator converges to the within estimator when the sample increases either by increasing the number of units or the number of observations per unit.

Section 7 outlines the analysis for a two way layout where "time effects" are added to individual effects. Section 8 outlines the estimation of the variance components which are necessary for statistical inference and the computation of the restricted GLS as well as the MSE estimators.

2. THE MODEL

Let us rewrite the basic equation to be estimated:

$$(2.1)\qquad \underset{\sim}{Y} = X\underset{\sim}{\beta} + Z\underset{\sim}{\alpha} + \underset{\sim}{u}$$

and assume

$$(2.2)\qquad \underset{\sim}{u} = (\underset{\sim}{0}, \sigma^2 I_n), \qquad E(\underset{\sim}{u}'X) = E(\underset{\sim}{u}'Z\underset{\sim}{\alpha}) = 0,$$

where $\underset{\sim}{Z}$ is a matrix of qualitative variables, or dummies, and $\underset{\sim}{\alpha}$ is a vector of effects. We now proceed under the assumption that there is no time effect and therefore we can write $Z = I_N \otimes \underset{\sim}{e}_T$ where $\underset{\sim}{e}_T$ is a T-vector on ones. However some of the discussion is not restricted to qualitative Z and it applies as well to quantitative Z. To simplify the discussion at this point, it is assumed that the X's are deviations from their sample means and the matrix (X, Z) is of full rank. A more general formulation is taken up in Section 7 below.

The properties of the various estimators to be considered depend on the existence and extent of the relations between the X's and the effects. In order to take an explicit account of such relationships we introduce the auxiliary regression:[3]

$$(2.3)\qquad \alpha_i = \underset{\sim}{X}_{it}\underset{\sim}{\pi} + w_{it};$$

averaging over t for a given i:

$$(2.4)\qquad \alpha_i = \underset{\sim}{X}_{i\cdot}\underset{\sim}{\pi} + w_{i\cdot}$$

[3] $(E\alpha_i|X)$ need not be linear. However, only the linear expression is pertinent for the present analysis.

It is assumed that

(2.5) $w_{i\cdot} \sim (0, \omega_1^2)$.

Clearly, $\underset{\sim}{\pi} = \underset{\sim}{0}$ if and only if the explanatory variables are uncorrelated with the effects. Let the projection matrix on the column space of Z be denoted as $K(Z) = Z(Z'Z)^{-1}Z'$ and its orthogonal complement by $M(Z) = I - K(Z)$. For the present definition of Z, $K(Z) \equiv K = I_N \otimes \bar{J}_T$ where $\bar{J}_T = (1/T)\underset{\sim}{e}_T\underset{\sim}{e}'_T$. Equation (2.4) can now be written as an NT vector:

(2.6) $Z\underset{\sim}{\alpha} = K(X\underset{\sim}{\pi} + \underset{\sim}{W})$

where $\underset{\sim}{W}$ is the NT-vector of w_{it}.

Combining (2.6) and (2.1) yields:

(2.7) $\underset{\sim}{Y} = X\underset{\sim}{\beta} + K(X\underset{\sim}{\pi} + \underset{\sim}{W}) + \underset{\sim}{U}$,

(2.8) $\underset{\sim}{\varepsilon} \equiv \underset{\sim}{U} + K\underset{\sim}{W} \sim (\underset{\sim}{0}, \sigma^2 I_{NT} + T\omega_1^2 K)$.[4]

We are now in a position to differentiate between the two models. Under the random effects we are concerned with the expectation of Y conditional on X and the grouping, to be denoted by Z. This is given by the systematic part of (2.7):

(2.9) $E(Y|\cdot) = X(\beta + K\pi)$

where $E(Y|\cdot) \equiv E(Y|X, Z)$. On the other hand, the FE model calls for the expectation of Y conditional on X and the effects to be denoted by $Z\underset{\sim}{\alpha}$. This is given by the systematic part of (2.1):

(2.10) $E(Y|\cdot\cdot) = X\beta + Z\underset{\sim}{\alpha}$

where $E(Y|\cdot\cdot) = E(Y|X, Z\underset{\sim}{\alpha})$.

This is the framework for the subsequent analysis. In the following two sections we show that the within estimator is the GLS estimator for both models. At the same time we evaluate the moments of alternative estimators. The various estimators can all be generated by the expression:[5]

(2.11) $\underset{\sim}{b}_F = A(F)\underset{\sim}{Y},\ A(F) = (X'FX)^{-1}X'F$.

In what follows we consider the following estimators:

Notation	Name	F
$\underset{\sim}{b}_o$	OLS	I_{NT}
$\underset{\sim}{b}_b$	Between	K
$\underset{\sim}{b}_w$	Within	M
$\underset{\sim}{b}_g$	GLS	Σ^{-1}

[4] A more general version of assumption (2.5) would call for a decomposition of w_{it} into systematic and random components with respect to the ith unit. Consequently, var $\underset{\sim}{w} = \omega_0^2 I_{NT_2} + \omega_1^2 K$. For a given T, the two components are non-distinguishable and ω_0^2 can be ignored without affecting the analysis. However, a comparison of samples with different T may reveal such a decomposed variance structure.

[5] Unless otherwise indicated, it is assumed that rank $F \geq$ rank $X = k$. When this assumption is violated, a GI should replace the inverse in (2.11).

Σ is the variance of the error term and will be explicitly specified for each of the cases under consideration. In addition, the estimators will be differentiated by the restrictions which are imposed on the coefficients.

3. ESTIMATION UNDER THE RE MODEL (CONDITIONAL ON X AND Z)

Under the RE model $\underset{\sim}{\beta}$ has to be estimated from equation (2.7). Starting with the GLSE, the relevant variance is given in (2.8). Consequently

$$\begin{aligned} \Sigma^{-1} &= \gamma_2 I_{NT} + \gamma_1 TK, \\ \gamma_1 &= -\omega_1^2/\sigma^2(\sigma^2 + T\omega_1^2), \qquad \gamma_2 = 1/\sigma^2. \end{aligned} \tag{3.1}$$

Then the GLSE of $\underset{\sim}{\beta}$ and $\underset{\sim}{\pi}$ in (2.7) is given by:

$$\begin{bmatrix} \underset{\sim}{b}_g \\ \hat{\underset{\sim}{\pi}}_g \end{bmatrix} = \left[\begin{pmatrix} X' \\ X'K \end{pmatrix} \Sigma^{-1}(X, KX)\right]^{-1} \Sigma^{-1}(X, KX)\underset{\sim}{Y}. \tag{3.2}$$

Utilizing (3.1) and the expression for the inverse of a partitioned matrix, we can obtain after some simplications:

$$\begin{aligned} \underset{\sim}{b}_g &= \underset{\sim}{b}_w, \\ \hat{\underset{\sim}{\pi}}_g &= \underset{\sim}{b}_b - \underset{\sim}{b}_w, \end{aligned} \tag{3.3}$$

where $\underset{\sim}{b}_b$ and $\underset{\sim}{b}_w$ are defined in Section 2. Thus, the GLSE is the within estimator and as such it does not depend on the knowledge of the variance components; in the present framework it is invariably BLUE.

The present analysis differs from previous discussions on the subject in that $KX\underset{\sim}{\pi}$ appears in (2.7). The question is how is this estimator affected by when $\underset{\sim}{\pi} = 0$ and conversely what happens to estimators which restrict $\pi = 0$ when such restriction is violated. Starting with the first question, if $\underset{\sim}{\pi} = 0$, $\underset{\sim}{b}_b$ and $\underset{\sim}{b}_w$ have the same expectation and, therefore, their difference will only reflect sampling errors. Consequently, any matrix combination of them will also have the same expectation, a point of subsequent pertinence. The term $KX\underset{\sim}{\pi}$ can also vanish when $KX = 0$ which implies that there are no between individuals variations in the X's (recall that the X's are measured from their sample means). In this case, no between regression can be calculated.

We turn now to the second question, the consequence of imposing $\underset{\sim}{\pi} = 0$. We refer to such estimates as restricted estimates. We start with the restricted GLSE (RGLSE):

$$\underset{\sim}{b}_{rg} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underset{\sim}{Y}. \tag{3.4}$$

This is the Balestra-Nerlove estimator.

Utilizing (3.1) we obtain, as shown in [**3**],

$$\underset{\sim}{b}_{rg} = \underset{\sim}{\lambda}_{rg}\underset{\sim}{b}_b + (I - \underset{\sim}{\lambda}_{rg})\underset{\sim}{b}_w \tag{3.5}$$

where $\lambda_{rg} = (X'KX + (1/\sigma^2(\gamma_1 + \gamma_2))X'MX)^{-1}X'KX$. Thus when $\underset{\sim}{b}_b = \underset{\sim}{b}_w$, $\underset{\sim}{b}_{rg} = \underset{\sim}{b}_w$, but in general:

(3.6) $E(\underset{\sim}{b}_b|\cdot)=\underset{\sim}{\beta}+\underset{\sim}{\pi}$

and therefore, by (3.5) and (3.6),

(3.7) $E(\underset{\sim}{b}_{rg}|\cdot)=\underset{\sim}{\beta}+\lambda_{rg}\underset{\sim}{\pi}$.

Thus the estimator is biased.

Another restricted estimator to be considered is the restricted LSE (RLSE), to be denoted as $\underset{\sim}{b}_{ro}$:

(3.8) $\underset{\sim}{b}_{r0}=(X'X)^{-1}X\underset{\sim}{Y}$.

It can be written as a matrix weighted combination of the between and within estimators and it is therefore generally biased. Let $\lambda_0=(X'X)^{-1}X'KX$; then $E(\underset{\sim}{b}_{r0}|\cdot)=\underset{\sim}{\beta}+\lambda_{r0}\underset{\sim}{\pi}$.

The reason for considering the restricted estimators is that restrictions are likely to decrease the variance of the estimators. As we have seen, the price for such possible reduction is the bias. There is therefore a trade off between bias and variance and the choice of an estimator depends on the weights to be assigned to the two components. In Table I we summarize the variances and MSE of the alternative estimators.

TABLE I

RE—VARIANCES AND MEAN SQUARE ERRORS OF ALTERNATIVE ESTIMATORS, CONDITIONAL ON (X, Z)

Estimator	Variance	MSE
	Unrestricted Estimators	
$\underset{\sim}{b}_w$, $\underset{\sim}{b}_g$, $\underset{\sim}{b}_0$	$V_w\equiv\sigma^2(X'MX)^{-1}$	V_w
$\underset{\sim}{b}_b$	$V_b=(\sigma^2+T\omega_1^2)(X'KX)^{-1}$	$M_b=V_b+\underset{\sim}{\pi}\underset{\sim}{\pi}'$
	Restricted Estimators	
$\underset{\sim}{b}_{r0}$	$V_{r0}=\lambda_{r0}V_b\lambda'_{r0}+(I-\lambda_{r0})V_w(I-\lambda_{r0})'$	$M_{r0}=\lambda_{r0}M_b\lambda'_{r0}+(I-\lambda_{r0})V_w(I-\lambda_{r0})'$
$\underset{\sim}{b}_{rg}$	$V_{rg}\equiv\lambda_{rg}V_b\lambda'_{rg}+(I-\lambda_{rg})V_w(I-\lambda_{rg})'$	$M_{rg}=\lambda_{rg}M_b\lambda'_{rg}+(I-\lambda_{rg})V_w(I-\lambda_{rg})'$

Clearly, none of the terms in the last column of Table I dominates the others for all possible values V_w, V_b, and $\underset{\sim}{\pi}$. By dominance it is meant that any quadratic form in the difference between two M's will be uniquely signed for all admissible values of the matrices in question. We return to this question below.

4. ESTIMATOR UNDER THE FE MODEL (CONDITIONAL ON X AND $Z\alpha$)

The FE model can be viewed as an end by itself so that the conditional inference, given the particular effects which appear in the sample, is all that matters. In that case equation (2.3) simply represents the design of the experiment. On the other hand, if the sample is a random sample, the conditional inference can be also considered as a step in deriving the unconditional inference.

Under the FE, $\underset{\sim}{\beta}$ is to be estimated from (2.1). The conditional variance is given in (2.2) and clearly the GLSE of (2.1) is identical to the OLSE. The OLSE of $\underset{\sim}{\beta}$ in (2.1) is simply the within estimator. This remark concludes the statement about the BLUE. We now turn to examine properties of some of the restricted estimators.

The RLSE is obtained by omitting $Z\underset{\sim}{\alpha}$ from (2.1), and it is the same as (3.8). The moments are:

(4.1) $$E(\underset{\sim}{b}_{r0}|\cdot\cdot)=\underset{\sim}{\beta}+\hat{\underset{\sim}{\pi}}_{r0}$$

where $E(\underset{\sim}{b}_{r0}|\cdot\cdot)$ represents conditional expectation given X and $Z\underset{\sim}{\alpha}$,

(4.2) $$\hat{\underset{\sim}{\pi}}_{r0}=A(I)Z\underset{\sim}{\alpha},$$

(4.3) $$V(\underset{\sim}{b}_{r0})=\sigma^2(X'X)^{-1}.$$

Again, the restriction is likely to decrease the variance of the estimator. This however is done at the expense of obtaining a biased estimator, which may have a larger MSE than that of the unrestricted estimator. To compare the MSE of the restricted and unrestricted estimator we write:[6]

(4.4) $$M(\underset{\sim}{b}_w|\cdot\cdot)-M(\underset{\sim}{b}_{r0}|\cdot\cdot)=A(I)Z(V_{\hat{\alpha}}-\underset{\sim}{\alpha}\underset{\sim}{\alpha}')Z'A(I)'$$

where $V_{\hat{\alpha}}$ is the variance of the unrestricted L.S. estimator of $\underset{\sim}{\alpha}$. Equation (4.4) constitutes a special case of a result obtained by Toro–Vizcarondo and Wallace [**22**] who also show that (4.4) is positive semidefinite if and only if $\underset{\sim}{\alpha}'V_{\hat{\alpha}}^{-1}\alpha \leqslant 1$.[7] Thus, when effects exist and the variance of their estimate is not excessively large the unrestricted L.S. dominates the restricted L.S. in MSE. This result is repeated here in order to emphasize that the omission of the variables is not priceless. We return to this point below.

As in Section 3, RLSE can be written as a matrix weighted sum of the between and within estimators. In a more general form this expression can be written as

(4.5) $$\underset{\sim}{b}_p=\lambda_p\underset{\sim}{b}_b+(I-\lambda_p)\underset{\sim}{b}_w,$$

(4.6) $$\lambda_p=[(X'KX)+\theta(X'MX)]^{-1}X'KX.$$

When $\theta=1$ the estimator in (4.5) becomes the RLSE. Thus, (4.5) has a general appearance and it looks like a GLSE. But as noted above, in view of (2.2) the GLSE is the within estimator. We therefore refer to (4.5) as a pseudo GLSE (PGLSE). In Section 5 we deal with the optimal selection of weights such as λ_p so as to minimize the variance of the resulting estimator. It turns out that $\underset{\sim}{b}_p$ has a minimum variance when $\theta=1$ and for that value $\underset{\sim}{b}_p=\underset{\sim}{b}_{r0}$. Thus, the scope for the PGLS is rather limited.

We are left with the two estimators whose MSE are compared in (4.4). It is possible to dominate these estimators in a MSE sense by deliberately selecting an estimator to do it. We outline the derivation of such an estimator from two points of view. However, it should be indicated that such an estimator requires

[6] Referring to $(\underset{\sim}{b}_w)$, we can write a symmetric expression: $\hat{\alpha}=(Z'M(X)Z)^{-1}Z'M(X)\underset{\sim}{Y}$ and $V_{\alpha}=\sigma^2(Z'M(X)Z)^{-1}$. Using $(X'X)^{-1}+A(I)Z(Z'M(X)Z)^{-1}Z'A(I)=(X'M(Z)X)^{-1}$ leads to (4.4).

[7] Feldstein [**5**] obtained a similar result for the case of a simple regression with one left out variable.

knowledge of σ^2 and $\underset{\sim}{\alpha}$ and therefore, in practical applications it can only be computed in two stages in the same spirit as the RGLSE. In fact, we start the presentation by deriving a GLS-like estimator the weights of which are obtained not from the variance of the error term but rather from its MSE. Let $\underset{\sim}{m} \equiv Z\underset{\sim}{\alpha}$,

(4.7) $\tilde{\Sigma} \equiv E\{(\underset{\sim}{Y}-X\underset{\sim}{\beta})(\underset{\sim}{Y}-X\underset{\sim}{\beta})'|X, \underset{\sim}{m}\} = \underset{\sim}{m}\underset{\sim}{m}' + \sigma^2 I_{NT},$

and obtain

(4.8) $\underset{\sim}{b}_m = (X'\tilde{\Sigma}^{-1}X)^{-1}X'\tilde{\Sigma}^{-1}\underset{\sim}{Y}.$

It can be shown that[8]

(4.9) $E(\underset{\sim}{b}_m|\cdot\cdot) = \underset{\sim}{\beta} + \hat{\underset{\sim}{\pi}}_{r0}(1-\tilde{\rho})$

where $1-\tilde{\rho} = \sigma^2(\sigma^2 + \hat{\underset{\sim}{w}}'\hat{\underset{\sim}{w}})^{-1}$, $\underset{\sim}{w} = M(X)\underset{\sim}{m}$,

(4.10) $M(b_m|\cdot\cdot) = \sigma^2(X'X)^{-1} + \hat{\underset{\sim}{\pi}}_{r0}\hat{\underset{\sim}{\pi}}'_{r0}(1-\tilde{\rho}) \equiv M_m.$

This estimator dominates $\underset{\sim}{b}_{r0}$ since

(4.11) $M_{r0} - M_m = \tilde{\rho}\hat{\underset{\sim}{\pi}}_0\hat{\underset{\sim}{\pi}}'_0$

and (4.11) is positive semidefinite. Thus, we conclude that if some variables are omitted and their coefficients are known, it would be better to add the omitted part to the error term and use a GLS-like estimator rather than restricted LS. This is a general result and it is not limited to qualitative variables.

A comparison of $\underset{\sim}{b}_m$ with $\underset{\sim}{b}_w$ indicates that it also dominates the within estimator. This can be seen by comparing (4.10) and V_w using the result of footnote 6:

(4.12) $M(\underset{\sim}{b}_w|\cdot\cdot) - M(\underset{\sim}{b}_m|\cdot\cdot) = A(I)Z[V_{\hat{\alpha}} - \underset{\sim}{\alpha}\underset{\sim}{\alpha}'(1-\tilde{\rho})]Z'A(I)'$

where (4.12) is a positive semidefinite matrix.[9] The implication is that if $\underset{\sim}{m}$ is known, then it need not be estimated and it could be used in deriving a MSE estimator.

The main purpose of introducing this MSE estimator is suggestive. It cannot be used as such since $\underset{\sim}{m}$ is not known. If it were known, it could be used for deriving a modified LS estimator which dominates all the others:

(4.13) $\underset{\sim}{b}_f = A(I)(\underset{\sim}{Y} - \underset{\sim}{m}) \sim (\underset{\sim}{\beta}, \sigma^2(X'X)^{-1}).$

An alternative approach to the construction of the MSE estimator is discussed in the next section.

5. AN ALTERNATIVE VIEW OF THE VARIOUS ESTIMATORS

In introducing the MSEE it is helpful first to consider the GLSE from a different point of view. Assume that we want to estimate a linear function in $\underset{\sim}{\beta}$, $\psi = \underset{\sim}{c}'\underset{\sim}{\beta}$, with

[8] Note that for large samples $1-\hat{\rho}$ is close to zero.

[9] (4.12) is positive semidefinite if and only if $\underset{\sim}{\alpha}' V_{\hat{\alpha}}^{-1}\underset{\sim}{\alpha} < 1/(1-\tilde{\rho})$. Using the definitions, $\underset{\sim}{\alpha}' V_{\hat{\alpha}}^{-1}\underset{\sim}{\alpha} = \sigma^{-2}\underset{\sim}{w}'\underset{\sim}{w} = \tilde{\rho}/(1-\tilde{\rho}) \leq 1/(1-\tilde{\rho})$.

$\underset{\sim}{c}$ given. The problem is to select λ so as to minimize $V(\hat{\psi}_g) \equiv \text{var}\,(\underset{\sim}{c}'\underset{\sim}{b}_g)$. Since $\text{cov}\,(\underset{\sim}{b}_b, \underset{\sim}{b}_w) = 0$, we can write $V_g = \lambda V_b \lambda' + (I-\lambda)V_w(I-\lambda)'$ where $V_g = \text{var}\,\underset{\sim}{b}_g$, etc. The result is given by:

$$(5.1) \qquad \lambda = V_w(V_b + V_w)^{-1} = (V_b^{-1} + V_w^{-1})^{-1}V_b^{-1}.$$

The resultant value of λ is the same as that obtained by GLS. We can therefore consider the GLS as the estimator which combines the various orthogonal estimators of $\underset{\sim}{\beta}$ so as to minimize the variance of $\hat{\psi}$. The proof is obtained by considering an alternative estimator, $b_H = H\underset{\sim}{b}_b + (I-H)\underset{\sim}{b}_w$ to be used in $\hat{\psi}_H = \underset{\sim}{c}'\underset{\sim}{b}_H$. Comparing the variances of the two estimators, we obtain after some simplifications:

$$(5.2) \qquad V(\hat{\psi}_H) - V(\hat{\psi}_g) = \underset{\sim}{c}'\{(H-\lambda)(V_b + V_w)(H-\lambda)' - 2(H-\lambda)[V_w - (V_b + V_w)\lambda']\}\underset{\sim}{c}.$$

Selecting λ according to (5.1) annihilates the second term on the right-hand side and makes (5.2) nonnegative for any H and $\underset{\sim}{c}$.

In a similar way it is possible to construct a minimum MSE estimator. The problem can be formulated as follows:

Select λ_m such that the MSE of $\hat{\psi}_m = \underset{\sim}{c}'\underset{\sim}{b}_m$ as an estimator of $\underset{\sim}{c}'\underset{\sim}{\beta}$ is minimized for any $\underset{\sim}{c}$ and where

$$(5.3) \qquad \underset{\sim}{b}_m = \lambda_m \underset{\sim}{b}_b + (I-\lambda_m)\underset{\sim}{b}_w.$$ [10]

The result is:

$$(5.4) \qquad \lambda_m = (M_b^{-1} + M_w^{-1})^{-1}M_b^{-1} = M_w(M_b + M_w)^{-1}.$$

To apply (5.4) in the model under consideration, assuming the RE, Table I and other results of Section 3 are used. It can then be shown that:

$$(5.5) \qquad E(b_m) = \underset{\sim}{\beta} + \lambda_m \underset{\sim}{\pi},$$

$$(5.6) \qquad V_m = (M_b^{-1} + V_w^{-1})^{-1}(M_b^{-1}V_b M_b^{-1} + V_w^{-1})(M_b^{-1} + V_w^{-1})^{-1}.$$

By writing M_b explicitly and simplifying it can be shown that:

$$(5.7) \qquad \underset{\sim}{c}'(M_{rg} - M_m)\underset{\sim}{c} \geqslant 0$$

and the superiority of $\underset{\sim}{b}_m$ in the MSE sense is demonstrated.

The estimator $\underset{\sim}{b}_m$ cannot be utilized directly since the variances and biases in question are not known . However, it is possible to follow the two stage procedure as used also in obtaining the RGLS estimator. In addition to the estimators of the variance we need also an estimate of $\underset{\sim}{\pi}$. Such an estimate can be obtained from:

$$(5.8) \qquad \hat{\underset{\sim}{\pi}} = \underset{\sim}{b}_b - \underset{\sim}{b}_w$$

and the weight for the MSE estimator is derived from:

$$(5.9) \qquad \hat{\lambda}_m = \hat{V}_w(\hat{M}_b + \hat{V}_w)^{-1}$$

[10] Cf. Feldstein [**5**].

where $\hat{}$ indicates an estimator and:[11]

(5.10) $\quad \hat{M}_b = \hat{V}_b + \hat{\pi}\hat{\pi}'$.

We do not deal here with the distribution of this two stage estimator. However its limiting value is considered in the next section.

The RGLS and the MSE estimators can be considered as parts of a more general framework. Suppose, we have r estimators b_j of β, $j = 1, \ldots, r$. Let b be the rk vector of such estimators. Then, $E(b)$ can be written as

(5.11) $\quad E(b) = A_1\beta + A_2\pi$.

The variance of b is denoted by Σ_b, a square nonsingular matrix of order kr. The problem is how to combine the components of b in order to obtain a final estimator. Assuming first a knowledge of Σ_b, we can derive the maximum likelihood estimator (ML):

(5.12) $$\begin{matrix}\hat{\beta} \\ \hat{\pi}\end{matrix} = (A'\Sigma_b^{-1}A)^{-1}A'\Sigma_b^{-1}b.$$

This estimator is unbiased as can be verified immediately in view of (5.11). Consequently, we already know from the foregoing discussion that it is not necessarily the most efficient MSE estimator.

In the problem under consideration we have $A_1 = e_2 \otimes I_k$ and $A_2 = \binom{I}{0}k)$ and

(5.13) $\quad \hat{\beta} = b_w, \qquad \hat{\pi} = b_b - b_w,$

and we are back with the within estimator.

The present formulation of the problem allows us to utilize the discussion in Section 4 above. Since we are mainly interested in estimating β, we can omit $A_2\pi$ to gain precision and obtain

(5.14) $\quad \hat{\beta} = (A_1'\Sigma_b^{-1}A_1)^{-1}A_1'\Sigma_b^{-1}b.$

With $A_1 = e_r \otimes I_k$, (5.14) is the RGLS estimator. From the discussion in Section 4 we also know that the MSE can be reduced by adding the omitted term $A_2\pi$ to the error and replacing Σ_b in (5.14) by M_b:

(5.15) $\quad \hat{\beta} = (A_1'M_b^{-1}A_1)^{-1}A_1'M_b^{-1}b$

where $M_b = \Sigma_b + A_2\pi\pi'A_2'$. The estimator in (5.15) is simply our MSE estimator.

We have thus produced a framework which yields the three estimators, the within, as a ML, the RGLS and the MSE as special cases. This approach can be further generalized as it is shown in [**17**, Section 6].

6. INCREASING THE SIZE OF THE SAMPLE

We now examine the properties of the estimators as the size of the sample increases. In so doing, we differentiate between an increase in the number of observations, T, taken on each individual, and the increase in the number of

[11] $\hat{\pi}$ can be used to correct b_{rg} so as to make it unbiased; the result is b_w.

individuals, N. It is assumed that the design matrices are bounded and their limits exist:

$$\lim_{T\to\infty}\left(\frac{1}{T}X'KX\right)=\bar{B}_N,\qquad \lim_{T\to\infty}\left(\frac{1}{T}X'MX\right)=\bar{W}_N,$$

(6.1)

$$\lim_{N\to\infty}\left(\frac{1}{N}X'KX\right)=\bar{B}_T,\qquad \lim_{N\to\infty}\left(\frac{1}{N}X'MX\right)=\bar{W}_T,$$

where all the limits are positive definite matrices. We can now obtain:

(6.2) $$\lim_{T\to\infty} V(\underset{\sim}{b}_b|\cdot)=\omega_1^2\bar{B}_N^{-1},$$

(6.3) $$\lim_{N\to\infty} V(\underset{\sim}{b}_b|\cdot)=0.$$

Consequently, the between variance can be reduced only by increasing the number of individuals and not by increasing the number of observations per individual. Referring to (3.6) it is seen that $\underset{\sim}{b}_b$ is asymptotically biased and from (6.3), $P_{N\to\infty}\lim \underset{\sim}{b}_b=\underset{\sim}{\beta}+\underset{\sim}{\pi}$. Consequently

$$\lim_{T\to\infty} M(\underset{\sim}{b}_b|\cdot)=\omega_1^2\bar{B}_N^{-1}+\underset{\sim}{\pi}\underset{\sim}{\pi}',$$

$$\lim_{N\to\infty} M(\underset{\sim}{b}_b|\cdot)=\underset{\sim}{\pi}\underset{\sim}{\pi}'.$$

The variance of the within estimator decreases with either T or N:

(6.4) $$\lim_{T\to\infty} V(\underset{\sim}{b}_w|\cdot)=0=\lim_{N\to\infty} V(\underset{\sim}{b}_w|\cdot).$$

Therefore $\underset{\sim}{b}_w$ is asymptotically unbiased and consistent.

These results make it possible to evaluate the plim of the other estimators which are expressed as weighted combination of the within and between estimators. The following comments can be made, omitting the technical details.

1. The weight λ_0 of the RLSE has a limit. Therefore $\text{plim}_{N\to\infty}\underset{\sim}{b}_{r0}\neq\underset{\sim}{\beta}$ if $\text{plim}_{N\to\infty}\underset{\sim}{b}_b\neq\underset{\sim}{\beta}$. On the other hand, $\underset{\sim}{b}_{r0}$ does not converge with T.

2. The RGLSE estimator converges to $\underset{\sim}{b}_w$ with T since $\lim_{T\to\infty}\lambda_{rg}=0$. On the other hand, $\lim_{N\to\infty}\lambda_{rg}\neq 0$ and therefore $\text{plim}_{N\to\infty}\underset{\sim}{b}_{rg}\neq\underset{\sim}{\beta}$ if $\text{plim}\,\underset{\sim}{b}_b\neq\underset{\sim}{\beta}$.

3. The MSE converges to $\underset{\sim}{b}_w[\text{plim}\,(\underset{\sim}{b}_m-\underset{\sim}{b}_w)=0]$ with both N and T since in both cases $\lim\lambda_m=0$.

Of the two limits considered, the one generated by increasing N is by far more important for two reasons. First, in general, the number of observations per individual (T) is limited and relatively small, and second, if it were not small then it would be inappropriate to assume that the effects α_i remain constant. Since the observations are periods, usually a year, it would not be reasonable to assume that individuals do not change. In fact, a more realistic approach would be to assume that individuals constantly change but when observed for short time intervals such changes could be neglected [**16**]. However, it is in this process that the RGLSE fails and the MLSE survives. They are both biased for finite samples, but by increasing N the bias of the MLSE approaches zero whereas that of the RGLSE does not.

A similar evaluation can now follow conditional on the FE. However in this case it does not make sense to trace the effect of $N \to \infty$. If N becomes very large, one would be interested not in the specific effect of each individual but rather in the characteristics of the population and will therefore carry the analysis within the RE framework.

Increasing T results in the decline of both variances:

$$(6.5) \qquad \lim_{T\to\infty} V(\underset{\sim}{b}_b|\cdot\cdot)=0, \qquad \lim_{T\to\infty} V(\underset{\sim}{b}_w|\cdot\cdot)=0.$$

It is therefore concluded that the between estimator is asymptotically biased and inconsistent. Similar properties are attributed to $\underset{\sim}{b}_{r0}$ and the pseudo-GLS estimator $\underset{\sim}{b}_p$ defined in (4.5) and (4.6) since the weights are invariant to the sample size. Different results are obtained for the MSE estimator $\underset{\sim}{b}_m$. Since $\lim_{T\to\infty}(1-\tilde{\rho})=0$, we have

$$(6.6) \qquad \lim_{T\to\infty} V(b_m|\cdot\cdot)=0 \qquad \text{and} \qquad \lim_{T\to\infty} E(b_m|\cdot\cdot)=\underset{\sim}{\beta}$$

and the estimator is consistent and asymptotically unbiased.

7. INTRODUCING TIME EFFECT

The introduction of time effects does not introduce conceptual problems and this is primarily due to the fact that time will only represent here another "lay out." It is introduced here briefly in order to give a complete technical framework which is utilized in the next section for estimating the components of variance.

The basic equation is still given by (2.1) except that we now decompose $Z\underset{\sim}{\alpha} = Z_1\underset{\sim}{\alpha}_1 + Z_2\underset{\sim}{\alpha}_2$ where $\underset{\sim}{m} \equiv Z_1\underset{\sim}{\alpha}_1$ and $\underset{\sim}{s} \equiv Z_2\underset{\sim}{\alpha}_2$ are the vectors representing unit and time effects respectively. The observations are arranged by units, beginning with the T readings on the first unit, etc.

Then $Z_1 = I_N \otimes \underset{\sim}{e}_T$, $Z_2 = \underset{\sim}{e}_N \otimes I_T$. We now have to be more specific about the intercept. Let $Z_0 = \underset{\sim}{e}_{NT}$ and rank $X = k-1$.[12] The projecting matrices on the vector spaces generated by the columns of Z_1, Z_2, and Z_0 respectively are:

$$(7.1) \qquad K_1 + K_0 = I_N \otimes \bar{J}_T, \qquad K_2 + K_0 = \bar{J}_N \otimes I_T, \qquad K_0 = \bar{J}_{NT}.$$

Note that $K_1 + K_0$ is the same as K in the previous sections. Also

$$(7.2) \qquad K_1K_2 = K_1K_0 = K_2K_0 = 0$$

and therefore

$$(7.3) \qquad K_1Z_2 = K_1Z_0 = K_2Z_1 = K_2Z_0 = 0.$$

[12] There are $k-1$ columns in X and the requirement of zero means is eliminated.

Rewrite the basic equations:

(7.4) $\underset{\sim}{Y}=X\underset{\sim}{\beta}+Z_0\alpha_0+Z_1\underset{\sim}{\alpha}_1+Z_2\underset{\sim}{\alpha}_2+\underset{\sim}{u}, \qquad \underset{\sim}{u}\sim(\underset{\sim}{0},\sigma^2 I_{NT}),$

(7.5) $\underset{\sim}{m}\equiv Z_1\underset{\sim}{\alpha}_1=(K_1+K_0)X\underset{\sim}{\pi}_1+(K_1+K_0)\underset{\sim}{w}_1, \qquad \underset{\sim}{w}_1\sim[\underset{\sim}{0}, T\omega_1^2(K_1+K_0)],$

(7.6) $\underset{\sim}{s}\equiv Z_2\underset{\sim}{\alpha}_2=(K_2+K_0)X\underset{\sim}{\pi}_2+(K_2+K_0)\underset{\sim}{w}_2, \qquad \underset{\sim}{w}_2\sim[\underset{\sim}{0}, N\omega_2^2(K_2+K_0)],$

where we have used $(K_j+K_0)Z_j=Z_j$, $j=1,2$. We also assume that the error components $\underset{\sim}{u}$, $\underset{\sim}{w}_1$, and $\underset{\sim}{w}_2$ are independent for all i and t.

Combining (7.4) and (7.6) and following the procedure of Section 3 we can derive Table II for the RE model.

TABLE II

BIAS AND VARIANCE CONDITIONAL ON X AND Z OF VARIOUS WITHIN AND BETWEEN ESTIMATORS[a]

Estimator	F	Bias	Variance
(1) $\underset{\sim}{b}_w^{it}$	M_{12}	0	$\sigma^2(X'M_{12}X)^{-1}$
(2) $\underset{\sim}{b}_b^i$	K_1	$\underset{\sim}{\pi}_1$	$(\sigma^2+T\omega_1^2)(X'K_1X)^{-1}$
(3) $\underset{\sim}{b}_b^t$	K_2	$\underset{\sim}{\pi}_2$	$(\sigma^2+N\omega_2^2)(X'K_2X)^{-1}$
(4) $\underset{\sim}{b}_w^i$	M_1	$A(M)_1K_2X\underset{\sim}{\pi}_2$	$(X'M_1X)^{-1}[\sigma^2(X'M_1X)+N\omega_2^2(X'K_2X)](X'M_1X)^{-1}$
(5) $\underset{\sim}{b}_w^t$	M_2	$A(M)_2K_1X\underset{\sim}{\pi}_1$	$(X'M_2X)^{-1}[\sigma^2(X'M_2X)+T\omega_1^2(X'K_1X)](X'M_2X)^{-1}$
(6) $\underset{\sim}{b}_0$	M_0	$A(I)(K_1X\underset{\sim}{\pi}_1+K_2X\underset{\sim}{\pi}_2)$	$(XM_0X)^{-1}X'M_0[\sigma^2I_{NT}+T\omega_1^2K_1+N\omega_2^2K_2]M_0X(X'M_0X)^{-1}$

[a] $M_0=I_{NT}-K_0$, $M_1=I_{NT}-K_1-K_0$, $M_2=I_{NT}-K_2-K_0$, $M_{12}=I_{NT}-K_1-K_2-K_0$; when rank $F<k-1$, the particular estimator is ignored.

A similar evaluation now follows for the FE. This is done by applying F to (7.4), recalling (7.2). The results are summarized in Table III.

TABLE III

BIAS AND VARIANCE CONDITIONAL ON X, $\underset{\sim}{m}$ AND $\underset{\sim}{s}$ OF VARIOUS WITHIN AND BETWEEN ESTIMATORS

Estimator	F	Bias	Variance
b_w^{it}	M_{12}	0	$\sigma^2(X'M_{12}X)^{-1}$
b_b^i	K_1	$A(K_1)\underset{\sim}{m}$	$\sigma^2(X'K_1X)^{-1}$
b_b^t	K_2	$A(K_2)\underset{\sim}{s}$	$\sigma^2(X'K_2X)^{-1}$
b_w^i	M_1	$A(M_1)\underset{\sim}{s}$	$\sigma^2(X'M_1X)^{-1}$
b_w^t	M_2	$A(M_2)\underset{\sim}{m}$	$\sigma^2(X'M_2X)^{-1}$
b_0	M_0	$A(M_0)(\underset{\sim}{m}+\underset{\sim}{s})$	$\sigma^2(X'M_0X)^{-1}$

See footnote to Table II.

The MSE examination can now be written as a matrix weighted combination of the alternative estimators. Assuming $T, N>k-1$, then:

(7.7) $\underset{\sim}{b}_m=\lambda_1\underset{\sim}{b}_b^i+\lambda_2\underset{\sim}{b}_b^t+(I-\lambda)\underset{\sim}{b}_w^{it}$

where according to [**17**, Section 6] we have

$$(7.8)\qquad (\lambda_1\lambda_2)=(V_wV_w)\begin{bmatrix} M_{11}+V_w M_{12}+V_w \\ M_{12}+V_w M_{22}+V_w \end{bmatrix}^{-1}$$

where $M_{11}=M(\underline{b}_b^i)$, $M_{22}=M(\underline{b}_b^t)$, $V_w=V(\underline{b}_w^{it})$, and $M_{12}=\pi_1\pi_2'$. Note that we utilize the fact that the three estimators are orthogonal and that the within estimator is also unbiased.

If $\underline{b}_b^t$ cannot be computed we can use $\underline{b}_w^i$ instead. However, $\underline{b}_w^i$ is orthogonal to $\underline{b}_b^i$ but not to $\underline{b}_w^{it}$; consequently the simple form (7.8) cannot be used and the weight matrixes will have to be computed from a somewhat more detailed form [**17**].

The examination of the behavior of the estimators as the sample size increases follows directly the analysis of Section 6. Assuming the limits exist, the following remarks can be made. The RGLS estimator based on the first three estimators in Table II does not converge in distribution to the within estimator unless both T and $N\to\infty$. Since it is unlikely to have a large T, $\underline{b}_{rg}$ will be inconsistent. On the other hand, the MSEE tends to the within estimator in large samples, regardless of whether the increase is in N or T.[13]

As indicated in Section 6, it is not particularly relevant to consider the limits under the FE since as the size of the sample increases in a particular dimension, interest would shift toward characterization of the propulation in terms of a fewer parameters.

The discussion has been conducted for the FE and RE models. It is also possible to consider mixed models where some effects are random whereas others are fixed. Such a specification simply dictates the conditional variables and as such the foregoing results are immediately applicable.

Finally when t stands for time and T is sufficiently large, it would be unrealistic to assume that the individuals do not change in a differential way as the model assumes. As indicated in the previous section, it is more realistic to assume that individuals do change differentially but at a pace that can be ignored for short time intervals. Under this assumption, it would be desirable to allow for interaction between i and t. Such interaction introduces too many parameters and a simplifying form has to be used. A possible formulation for the effects, i and t, is:

$$\mu_{it}=\gamma_{0i}+\gamma_{1i}t.$$

Such a formulation was used empirically in Mundlak [**16**] and Evenson and Kislev [**4**, Ch. 5].

[13] Nerlove [**18**, p. 395] raises the question why treating the effects as fixed rather than random should become asymptotically unimportant. "After all as N, $T\to\infty$ there are infinite number of such parameters, their number increases just as fast as the number of pieces of new information available as the sample size increases. The solution to the puzzle is in fact that we are not estimating them but only β . . . " It is not quite clear what is meant here by pieces of information. However it should be noted that under the FE there are $T-1$ degrees of freedom in estimating α_i and $N-1$ degrees of freedom in estimating α_T. Obviously those increase with N and T. The degree of freedom in estimating β increase with the product $(N-1)(T-1)$. The reason why $\text{plim}_{N,T\to\infty}(\underline{b}_{rg}-\underline{b}_w^{it})=0$ under the assumption of $\pi_1=\pi_2=0$, which corresponds to the model examined by Nerlove, is that b_w^{it}, b_b^i, b_b^t are all unbiased and converge in quadratic mean to $\underline{\beta}$.

8. ESTIMATING THE VARIANCE COMPONENTS

In order to make statistical reference in the RE model it is necessary to estimate the components of variance. Such estimators are also required for obtaining the RGLSE and the MSEE.[14] In what follows we present unbiased estimators based on the residuals of the various regressions.

To derive those, let $F\hat{Y} = FX\underline{b}_F$, and the residual is

$$(8.1) \qquad \underline{V}_F = F(\underline{Y} - \hat{\underline{Y}}) = M(FX)F\underline{Y}$$

where $M(FX) = I_{NT} - FX(X'FX)^{-1}X'F$. Under the RE,

$$(8.2) \qquad \underline{V}_F = M(FX)F\underline{\varepsilon}$$

where $\underline{\varepsilon}$ is the combined error term. Then

$$(8.3) \qquad E(V_F'V_F) = E(\underline{\varepsilon}'FM(FX)F\underline{\varepsilon}) = \operatorname{tr} M(FX)F\Sigma_\varepsilon.$$

The degrees of freedom in each case are given by $\operatorname{tr} M(FX)F = \operatorname{tr} F - (k-1)$, assuming of course that $\operatorname{tr} F \geqslant k$. The results are presented in Table IV. The last column of the Table gives the expected value of the error mean square, denoted by s_F^2.

TABLE IV
ESTIMATORS FOR COMPONENTS OF VARIANCE[a]

F	$\operatorname{tr} F$	$E\left[s_F^2 = \dfrac{V_F'V_F}{\operatorname{tr} F - (k-1)}\right]$
(1) M_{12}	$(N-1)(T-1)$	σ^2
(2) K_1	$N-1$	$\sigma^2 + T\omega_1^2$
(3) K_2	$T-1$	$\sigma^2 + N\omega_2^2$
(4) M_1	$N(T-1)$	$\sigma^2 + \dfrac{N}{N(T-1)-(k-1)}[(T-1) - \operatorname{tr}(X'M_1X)^{-1}(X'K_2X)]\omega_2^2$
(5) M_2	$T(N-1)$	$\sigma^2 + \dfrac{T}{T(N-1)-k-1}[(N-1) - \operatorname{tr}(X'M_2X)^{-1}(X'K_1X)]\omega_1^2$
(6) M_0	$NT-1$	$\sigma^2 + \dfrac{1}{NT-k}\{[(N-1) - \operatorname{tr}(X'M_0X)^{-1}(X'K_1X)]T\omega_1^2 + [(T-1) - \operatorname{tr}(X'M_0X)^{-1}(X'K_2X)]N\omega_2^2\}$

[a] In cases where $\operatorname{tr} F < k$, ignore the corresponding line in the table.

For $N > k$, ω_1^2 is estimable from lines (1) and (2) of the table. Such an estimate is independent of ω_2^2 and holds true also when $\omega_2^2 = 0$. Consequently, it is also the appropriate estimator for the one way layout with no time effect. Similarly, if $T \geqslant k$, ω_2^2 is estimable from lines (3) and (1). If however T is small, ω_2^2 can be estimated from lines (4) and (1).

[14] Alternative estimators exist for estimating the variance components. Maddala and Mount [12] examine the effect of using alternative estimators of the components on the MSE of the resulting GLS estimator, using the Monte Carlo technique. They find the results in general to be insensitive to most of the alternatives.

The results of Table IV are basically the analysis of variance results modified for the model under consideration. This modification has an important implication. Note that we estimate ω_j^2 rather than the unconditional variance $\sigma_{\alpha_j}^2$. The difference reflects the true correlation coefficients of equations (7.5) and (7.6). Let for instance $1-\alpha_1^2=\omega_1^2$. When the systematic component of (7.5) constitutes an important role, $1-\rho_1^2$, and therefore ω_1^2 will be relatively small. Thus, the estimate based on lines (1) and (2):

$$\hat{\omega}_1=\frac{1}{T}(s_{K_1}^2-s_{M_{1,2}}^2) \tag{8.4}$$

is an unbiased estimator of a small number and the probability of such a number to be negative increases with ρ_1^2. This finding bears on the negative values that are sometimes obtained for estimators like (8.4). To avoid this problem Nerlove **[19]** used as an estimator of the between variance component the $\Sigma\hat{\alpha}_1^2$ where $\hat{\alpha}_i$ is the LS estimate of α_i, or simply the estimate of the fixed effects α_i. Such an estimate constitutes an upper limit for est ω_1^2 and, aside from some correction factor, it is appropriate only for the case of $\rho_1^2=0$, that is when there is no auxiliary regression. The relative importance of the auxiliary regression in the total variance of σ_α^2 was computed for a specific problem in Mundlak [**14**, p. 53; **15**, p. 76]. The results vary depending on the estimator, between 0.4 to 0.5.

It is of some interest to obtain $E(S_F^2|X, \underset{\sim}{m}, \underset{\sim}{s})$, that is under the FE structure. The result has the following structure:

$$E(S_F^2X, \underset{\sim}{m}, s)=\sigma^2+\frac{(\underset{\sim}{\alpha}'Z')F'M(FX)F(Z\underset{\sim}{\alpha})}{\operatorname{tr} F-(k-1)}. \tag{8.5}$$

For $F=K_1$ and K_2 the second term on the right-hand side of (8.5) has a simple interpretation; it is equal to the sum of the computed residuals from the particular auxiliary regression. For instance,

$$\underset{\sim}{m}'K_1M(K_1X)K_1m = T\sum_{i=1}^{N}\hat{w}_i^2 \tag{8.6}$$

and we can then write

$$T\hat{\tilde{\omega}}_1^2=\frac{T\Sigma_{=1}^N\hat{w}_i^2}{N-k} \tag{8.7}$$

and a similar expression can be obtained for $N\omega_2^2$ by letting $F=K_2$. Of course, when $F=M_{12}$, the second term on the right-hand side of (8.5) vanishes. (8.7) can be considered as an estimate of $T\omega_1^2$ only if α_i were random.

Hebrew University of Jerusalem

Manuscript received February, 1976; final revision received March, 1977.

REFERENCES

[1] AMEMIYA, TAKESHI: "The Estimation of the Variance in a Variance-Component Model," *International Economic Review,* 12 (1971), 1–13.

[2] BALESTRA, P., AND M. NERLOVE: "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica,* 34 (1966), 585–612.

[3] CHAMBERLAIN, GARY: *Unobservables in Econometric Models,* unpublished Ph.D. dissertation, Harvard University, Cambridge, Ma., 1975.

[4] EVENSON, ROBERT E., AND YOAV KISLEV: *Agricultural Research and Productivity.* New Haven: Yale University Press, 1975.

[5] FELDSTEIN, MARTIN S: "Multicollinearity and the Mean Square Error of Alternative Estimators," *Econometrica,* 41 (1973), 337–346.

[6] FRIEDMAN, MILTON: *Theory of Consumption Function.* Princeton, N.J.: Princeton University Press, 1957.

[7] GRILICHES, ZVI: "Specification Bias in Estimates of Production Functions," *Journal of Farm Economics,* 39 (1957), 8–20.

[8] ———: "Estimating the Returns to Schooling," Harvard Institute of Economic Research, Discussion Paper #433, 1975.

[9] HENDERSON, CHARLES R., JR.: "Comment on the Use of Error Components Models in Combining Cross Section With Time Series Data," *Econometrica,* 39 (1971), 397–401.

[10] KUH, EDWIN: "The Validity of Cross Sectionally Estimated Behavior Equations in Time Series Applications," *Econometrica,* 27 (1959), 197–214.

[11] MADDALA, G. S.: "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica,* 39 (1971), 341–358.

[12] MADDALA, G. S., AND T. D. MOUNT: "A Comparative Study of Alternative Estimators for Variance Components Models Use in Econometric Applications," *Journal of the American Statistical Association,* 68 (1973), 324–328.

[13] MCELROY, F. W.: "A Necessary and Sufficient Condition that Ordinary Least Squares Estimators be Best Unbiased," *Journal of the American Statistical Association,* 62 (1967), 1302–1304.

[14] MUNDLAK, YAIR: "Empirical Production Function Free of Management Bias," *Journal of Farm Economics,* 43 (1961), 44–56.

[15] ———: *An Economic Analysis of Established Family Farms in Israel.* Jerusalem: Falk Project for Economic Research in Israel, 1964.

[16] ———: "Empirical Production Functions with a Variable Firm Effect," Working Paper #7005, The Center for Agricultural Economic Research, Rehovot, Israel, 1970.

[17] ———: "On the Pooling of Time Series and Cross Section Data," Harvard Institute of Economic Research, Discussion Paper #457, 1976.

[18] NERLOVE, M.: "A Note on Error Components Models," *Econometrica,* 39 (1971), 383–396.

[19] ———: "Further Evidence on the Estimation of Dynamic Economic Relations From a Time Series of Cross Section Data," *Econometrica,* 39 (1971), 359–382.

[20] RAO, R. C., AND S. K. MITRA: *Generalized Inverse of Matrices and its Applications.* New York: John Wiley & Sons, Inc., 1971.

[21] THEIL, H.: "Specification Errors and the Estimation of Economic Relationships," *Review of International Statistical Institute,* 25 (1957), 41–51.

[22] TORO-VIZCARONDO, CARLOS, AND T. D. WALLACE: "A Test of The Mean Square Error Criterion for Restrictions in Linear Regression," *Journal of the American Statistical Association,* 63 (1968), 558–572.

[23] WALLACE, T. D., AND ASHIQ HUSSAIN: "The Use of Error Components Models in Combining Cross Section With Time Series Data," *Econometrica,* 37 (1969), 55–68.

[24] ZYSKIND, GEORGE: "Parametric Augmentation and Error Structures Under Which Certain Simple Least Squares and Analysis of Variance Procedures Are Also Best," *Journal of the American Statistical Association,* 64 (1969), 1353–1368.