NBER WORKING PAPER SERIES

ERROR COMPONENTS REGRESSION MODELS AND THEIR APPLICATIONS

Swarnjit S. Arora*

Working Paper No. 3

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

June 1973

*NBER Computer Research Center and University of Wisconsin, Milwaukee. Research supported in part by National Science Foundation Grant GJ-1154X2 to the National Bureau of Economic Research, Inc.

Abstract

In this paper, we have developed an operational method for estimating error components regression models when the variance-covariance matrix of the disturbance terms is unknown. Monte Carlo studies were conducted to compare the relative efficiency of the pooled estimator obtained by this procedure to (a) an ordinary least squares estimator based on data aggregated over time, (b) the covariance estimator, (c) the ordinary least squares estimator, and (d) a generalized least squares estimator based on a known variance-covariance matrix. For T small and large ρ, this estimator definitely performs better than the other estimators which are also based on an estimated value of the variance-covariance matrix of the disturbances. For ρ small and large T it compares equally well with the other estimators.

Acknowledgments

A major portion of this research was done during the author's stay, as a Research Fellow, at the NBER Computer Research Center for Economics and Management Science. Special thanks are due to Professors David Belsley, Murray Brown, Paul Holland, and Edwin Kuh for their valuable comments and suggestions. Thanks are also due to Mr. David Jones for programming assistance. The simulation for this paper was done at the NBER Computer Research Center using the TROLL system.

Contents

Section 1. Introduction 1

Section 2. Estimation of Error Components Regression Models 4

Section 3. Properties of the Estimators 11

Section 4. Design of the Experiment and the Comparative Properties of the Various Estimators 14

Section 5. Conclusion 18

Appendix A-1. Tables of Mean and Mean Square Error of the Coefficients for Various Estimating Procedures 19

Appendix A-2. Random Number Generating Procedures 21

References 23

ERROR COMPONENTS REGRESSION MODELS AND THEIR APPLICATIONS

Section 1

Introduction

In several recent studies, attempts have been made to analyze the problems involved in pooling cross section and time series data by error components (or variance components) regression models. These models can be formulated as

$$(1.1) \qquad y_{it} = \beta_o + \sum_{k=1}^{K-1} \delta_k Z_{kit} + u_{it} \qquad (i=1,2,\ldots,n;\ t=1,2,\ldots,T),$$

where y_{it} is an observation on the dependent variable for individual i in period t. Z_{kit} is an observation on the k^{th} independent variable, β_o is an intercept term, δ_k $(k=1,2,\ldots,K-1)$ are the fixed but unknown slope coefficients, and u_{it} is an error term. This disturbance term is supposed to represent the net effect of numerous individually unimportant, but collectively significant, variables which have been omitted from the analysis. Some of these are specific to the individual and remain invariant over time (say μ_i); some are specific to the time period but are

invariant over all individuals (say λ_t); and some are specific to both individual and time (say ν_{it}). In this case we can write u_{it} as

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \cdot \tag{1.2}$$

Mundlak (10) and Hoch (5) analyzed this model, treating μ_i and λ_t as unknown parameters and assuming $\sum_{i=1}^{n} \mu_i = 0$ and $\sum_{t=1}^{T} \lambda_t = 0$. Maddala (9) points out a principal weakness in this approach: it eliminates a major portion of the variation among both the explained and the explanatory variables when the between individuals and between time periods variation is large. This approach can also cause a substantial loss in degrees of freedom. An alternative approach is to treat all components as random. This case was analyzed by Wallace and Hussain (14), Maddala (9), Nerlove (12), and Swamy and the present author (13).[1]

Under the assumptions of weakly non-stochastic X's and normally distributed disturbance terms, both approaches yield asymptotically equivalent estimates with asymptotically equivalent variance-covariance matrices. In fact, it can be shown that there are an infinite number of estimators which have the same asymptotic variance-covariance matrices.[2]

---

1. Whether or not the individual effects may be treated as parameters or random components for the purpose of statistical analysis depends upon the underlying data generating mechanism assumed. For an illuminating discussion of such data generating mechanisms, see Nerlove (11) p. 364.

2. See Swamy and Arora (13) p. 267.

Asymptotic properties, however, are cold comfort to the econometrician for whom the choice of a practical estimator (and its related small sample properties) is a problem of crucial importance. Unfortunately, because of mathematical intractibility, small sample properties are often hard to obtain theoretically. We therefore employ Monte Carlo experiments to evaluate relative efficiency of the various estimators.

The plan of this paper is as follows: In section 2, a means of estimating error components regression models is developed for a case when the variance-covariance matrix of the disturbance term is unknown. We also show the equivalence of this estimator with an ordinary least squares estimator when inter-individual and inter-temporal variations are zero. In section 3, the asymptotic properties of this estimator are derived. Section 4 describes the design of the Monte Carlo experiments and compares the relative efficiency of this estimator with the ordinary least squares estimator, a covariance estimator, an ordinary least squares estimator based on data aggregated over time and a generalized least squares estimator based on a known variance-covariance matrix of the disturbance terms. Concluding remarks are presented in section 5. An efficient way of generating random numbers and independently distributed normal variates is described in an appendix to this paper.

Section 2

Estimation of Error Components Regression Models

Let us assume that $u_{it} = \mu_i + \nu_{it}$ and the components μ_i's and ν_{it}'s are random such that

$$E\mu_i = 0$$

$$E\mu_i\mu_j = \begin{cases} \sigma_\mu^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$E\nu_{it} = 0$$

$$E\nu_{it}\nu_{jt'} = \begin{cases} \sigma_\nu^2 & \text{if } i = j \text{ and } t = t' \\ 0 & \text{otherwise.}^3 \end{cases}$$

Let us further assume that μ_i and ν_{it} are independent of each other.[4] Furthermore $T > K$ and $n > K$ and the variances σ_μ^2 and

---

3. In (13) Swamy and I have analytically shown that the estimator based on the assumption of both μ_i and λ_t being random is more efficient than either the covariance estimator or the ordinary least squares estimator only if (a) n and T are sufficiently larger than 10, and (b) if the sum of squares due to variation over time exceeds the sum of squares due to remaining variation. If these conditions are not satisfied, random error components model with both components random may give results inferior to other estimators. For a case where either n or T is less than 10, we conjecture that the error components models with a random component (the other component being a parameter or zero) perform better than the model which assumes both μ_i and λ_t as random. Here we consider a model with $\lambda_t=\lambda$ for all t, but we can easily treat all λ_t's as different.

4. A model in this form was also used by Kuh (7), except that he did not assume μ_i and ν_{it} are uncorrelated. Hussain (6)

σ_ν^2 are unknown. For all the nT observations combined we can write (1.1)

$$(2.1) \qquad y = X\beta + u,$$

where $y \equiv (y_{11}, \ldots, y_{1T}, \ldots, y_{n1}, \ldots, y_{nT})'$ is an nT×1 vector of observations on the dependent variable, $X = [\iota_{nT}, Z]$ is an nT×K matrix of explanatory variables, ι_{nT} is a vector of 1's of order nT×1, and Z is an nT×K′ matrix of independent variables given by

$$Z = \begin{bmatrix} Z_{111} & Z_{211} & \cdots & Z_{K-1,11} \\ \vdots & \vdots & & \vdots \\ Z_{11T} & Z_{21T} & \cdots & Z_{K-1,1T} \\ \vdots & \vdots & & \vdots \\ Z_{1n1} & Z_{2n1} & \cdots & Z_{K-1,n1} \\ \vdots & \vdots & & \vdots \\ Z_{1nT} & Z_{2nT} & \cdots & Z_{K-1,nT} \end{bmatrix},$$

δ is a (K′×1) vector of slope coefficients, $\beta \equiv (\beta_o, \delta')'$, $K' = K - 1$, and $u = (u_{11}, \ldots, u_{1T}, \ldots, u_{n1}, \ldots, u_{nT})'$ is an nT×1 vector of disturbance terms. Under the above assumptions, it is readily verified that

$$(2.2) \qquad Euu' = \sigma_\mu^2(I_n \otimes \iota_T\iota_T') + \sigma_\nu^2 I_{nT}.$$

Since the variance-covariance matrix of u is not scalar, ap-

---

(4) treats a model with μ_i's a parameters, and λ_t and ν_{it} as random. His estimator is identical with the covariance estimator and does not utilize full data information.

plication of the ordinary least squares procedure might lead to an inefficient estimator of β.

Let us consider an orthogonal matrix, O_T, of order T such that its first row is equal to $\iota_T'/\sqrt{T}$. Let $O_T = [\iota_T/\sqrt{T},\ C_1']'$ where C_1 is a $(T-1)\times T$ matrix such that $C_1\iota_T = 0$, $C_1C_1' = I_{T-1}$, and $C_1'C_1 = I_T - \iota_T\iota_T'/T$.

Define the transformations $Q_1 = (I_n \otimes \iota_T'/\sqrt{T})$ and $Q_2 = (I_n \otimes C_1)$. By applying the transformation Q_1 to all nT observations, we get

$$(2.3) \qquad y_1 = X_1\beta + u_1 ,$$

where $y_1 = Q_1y$ is an $n\times 1$ vector of transformed dependent variables, and u_1 is an $n\times 1$ vector of transformed observations. The variance-covariance matrix of u_1 is

$$(2.4) \qquad Eu_1u_1' = Q_1Euu'Q_1 .$$

Substituting for Euu' from (2.2) and simplifying we get

$$(2.5) \qquad Eu_1u_1' = \sigma_1^2 I_n ,$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$.

Thus the variance-covariance matrix reduces to scalar form, a best linear unbiased estimator of β is an OLS estimator

$$(2.6) \qquad \hat{\beta}(1) = (X_1'X_1)^{-1}X_1'y_1 .$$

The subvector of $\hat{\beta}(1)$ corresponding to the slope coefficient only is given by

$$(2.7) \qquad \hat{\delta}(1) = (Z_1'NZ_1)^{-1}Z_1'Ny_1 ,$$

where $Z_1 = Q_1 Z$ is an $n \times K'$ matrix and $N = I_n - \iota_n \iota_n'/n$; a subscript 1 is attached to $\hat{\beta}$ and $\hat{\delta}$ to differentiate these estimators from the other estimators of β and δ to be described later.

The variance-covariance matrix of $\hat{\delta}(1)$ is

$$(2.8) \qquad V[\hat{\delta}(1)] = \sigma_1^2 (Z_1' N Z_1)^{-1}.$$

Applying the transformation $Q_2 = (I_n \otimes C_1)$ to all nT observations we have

$$(2.9) \qquad y_2 = Z_2 \delta + u_2 ,$$

where $y_2 = Q_2 y$ is an $nT' \times 1$ vector of transformed observations on the dependent variable, Z_2 is an $nT' \times K'$ matrix of transformed observations on K' independent variables, $u_2 = Q_2 u$ is an $nT' \times 1$ vector of transformed disturbances, and use is also made of the result $Q_2 \iota_{nT} = 0$. The variance-covariance matrix of u_2 is

$$(2.10) \qquad Eu_2 u_2' = EQ_2 uu' Q_2 = Q_2 Euu' Q_2 =$$
$$= (I_n \otimes C_1)[\sigma_\mu^2 (I_n \otimes \iota_T \iota_T') + \sigma_\nu^2 I_{nT}](I_n \otimes C_1') ,$$

which can easily be reduced to $\sigma_\nu^2 I_{nT'}$. Thus the variance-covariance matrix of u_2 is of scalar form. A best linear unbiased estimator of δ is the OLS estimator given by

$$(2.11) \qquad \hat{\delta}(2) = (Z_2'Z_2)^{-1}Z_2'y_2 \ .^{5}$$

The variance-covariance matrix of $\hat{\delta}(2)$ is $\sigma_\nu^2(Z_2'Z_2)^{-1}$. Notice that $Q_1Q_2' = 0$. The rank of Q_1 is equal to the rank of $\iota_T'/\sqrt{T}$ multiplied by the rank of I_n because if A and B are any arbitrary matrices, the rank of (A⊗B) is equal to the rank of A multiplied by the rank of B. Therefore the rank of $Q_1 + Q_2$ = $= n + nT - n = nT$, which is the total number of observations. This indicates that in estimating (2.3) and (2.9) we have used up all the orthogonal linear combinations of the available observations. $\hat{\delta}(1)$ and $\hat{\delta}(2)$ are two uncorrelated estimators of the same parametric vector and we can pool them in the following manner.

$$(2.12) \qquad \hat{\delta}(\theta) = \left[\frac{Z_1'NZ_1}{\sigma_1^2} + \frac{Z_2'Z_2}{\sigma_\nu^2}\right]^{-1}\left[\frac{Z_1'Ny_1}{\sigma_1^2} + \frac{Z_2'y_2}{\sigma_\nu^2}\right]$$

where $\theta = [\sigma_1^2, \sigma_\nu^2]'$. The estimator $\hat{\delta}(\theta)$ is a generalized least squares estimator of δ. For given values of σ_1^2 and σ_ν^2, it is a best linear unbiased estimator. Any other estimator of δ which is also linear in the vector y and is unbiased, has a variance-covariance matrix which exceeds that of $\hat{\delta}(\theta)$ by a positive semidefinite matrix.

---

5. It can be easily recognized that the estimator $\hat{\beta}(1)$ in (2.6) is an OLS estimator obtained by applying OLS to data aggregated over time and multiplied by $1/\sqrt{T}$; the estimator $\hat{\delta}(2)$ in (2.11) is obtained by applying OLS to nT observations, each observation expressed as a deviation from its time series mean and the overall mean. Please note that there are only n(T-1) independent observations.

An unbiased estimator of σ_1^2 is given by

$$(2.13) \qquad \hat{\sigma}_1^2 = y_1' M_1 y_1/(n-K)\,,$$

where $M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$. Also an unbiased estimator of σ_ν^2 is

$$(2.14) \qquad \hat{\sigma}_\nu^2 = y_2' M_2 y_2/(nT'-K')\,,$$

where $M_2 = I_{nT'} - Z_2(Z_2'Z_2)^{-1}Z_2'$, $T' = T-1$, and $K' = K-1$. An Aitken estimator of the slope coefficients based on the estimated values of σ_1^2 and σ_ν^2 is given by

$$(2.15) \qquad \hat{\sigma}(\hat{\theta}) = \left[\frac{Z_1'NZ_1}{\hat{\sigma}_1^2} + \frac{Z_2'Z_2}{\hat{\sigma}_\nu^2}\right]^{-1}\left[\frac{Z_1'Ny_1}{\hat{\sigma}_1^2} + \frac{Z_2'y_2}{\hat{\sigma}_\nu^2}\right]$$

where $\hat{\theta} = [\hat{\sigma}_1^2, \hat{\sigma}_\nu^2]'$.[6]

We can readily show that the estimator $\hat{\delta}(2)$ as obtained in (2.11) is equivalent to a covariance estimator ($\tilde{\delta}$) obtained by assuming μ_i's as fixed parameters. We can also show equivalence of $\hat{\delta}(\hat{\theta})$ with an ordinary least squares estimator ($\hat{\delta}$) when $\sigma_\mu^2 = 0$ as follows:

An ordinary least squares estimator of the slope coefficients in (1.1) is

---

6. A similar pooled estimator of $\hat{\delta}(\hat{\theta})$ can be obtained if λ_t and ν_{it} are assumed to be random. The variance-covariance matrix of u is given by $\sigma_\lambda^2(\iota_n\iota_n'\otimes I_T) + \sigma_\nu^2 I_{nT}$. To reduce this to scalar form we consider an orthogonal matrix $O_n = [\iota_n'/\sqrt{n},\ C_1]'$ of order n, and apply transformations $Q_1 = (\iota_n/\sqrt{n}\otimes I_T)$ and $Q_2 = (C_1\otimes I_T)$ to all the nT observations.

(2.16) $\hat{\delta} = (Z'Q_4Z)^{-1}Z'Q_4y$,

where $Q_4 = I_{nT} - \frac{l_{nT}l'_{nT}}{nT}$.

The Aitken estimator $\hat{\delta}(\hat{\theta})$ in (2.15) when $\sigma_\mu^2 = 0$ is

(2.17) $$\hat{\delta}(\hat{\theta}) = Z_1'NZ_1 + Z_2'Z_2]^{-1}[Z_1'Ny_1 + Z_2'y_2]$$
$$= [Z'(Q_1'NQ_1 + Q_2'Q_2)Z]^{-1}[Z'(Q_1'NQ_1 + Q_2'Q_2)y] .$$

Since $Q_1'NQ_1 = (I_n \otimes \frac{l_Tl'_T}{T}) - \frac{l_{nT}l'_{nT}}{nT}$, and $Q_2'Q_2 = I_n \otimes (I_T - \frac{l_Tl'_T}{T})$, we can easily show that $Q_1'NQ_1 + Q_2'Q_2 = I_{nT} - \frac{l_{nT}l'_{nT}}{nT} = Q_4$, thus proving equality of $\hat{\delta}$ and $\hat{\delta}(\hat{\theta})$ when $\sigma_\mu^2 = 0$.

Section 3

Properties of the Estimators

The estimator $\hat{\delta}(\hat{\theta})$ of δ can be written in the generalized least squares form as follows:

$$(3.1) \qquad \hat{\delta}(\hat{\theta}) = \left[\begin{bmatrix} Z_1'N & Z_2' \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\sigma}_1^2} & 0 \\ 0 & \frac{1}{\hat{\sigma}_\nu^2} \end{bmatrix} \begin{bmatrix} NZ_1 \\ Z_2 \end{bmatrix}\right]^{-1} \left[\begin{bmatrix} Z_1'N & Z_2' \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\sigma}_1^2} & 0 \\ 0 & \frac{1}{\hat{\sigma}_\nu^2} \end{bmatrix} \begin{bmatrix} Ny_1 \\ y_2 \end{bmatrix}\right] .$$

If $W = [Z_1'N, Z_2']'$, $\hat{V} = \text{diag}[\hat{\sigma}_1^2 I_n, \hat{\sigma}_\nu^2 I_{nT'}]$ and $\gamma = [y_1'N, y_2']'$, the equation (3.1) can be written as

$$(3.2) \qquad \hat{\delta}(\hat{\theta}) = (W'\hat{V}^{-1}W)^{-1}W'\hat{V}^{-1}\gamma .$$

Substituting $\gamma = W\delta + e$, where $e = [u_1'N, u_2']'$ in (3.2), we get

$$(3.3) \qquad \hat{\delta}(\hat{\theta}) = \delta + (W'\hat{V}^{-1}W)^{-1}W'\hat{V}^{-1}e .$$

A generalized least squares estimator for a given V, as obtained in (2.12), can also be expressed as follows:

$$(3.4) \qquad \hat{\delta}(\theta) = \delta + (W'V^{-1}W)^{-1}W'V^{-1}e .$$

If we assume that u's are normally distributed, and since $M_1Z_1N = M_1X_1 = [I_n - X_1(X_1'X_1)^{-1}X_1']X_1 = 0$, we can show that the linear form $Z_1'Nu_1$ is distributed independently of the quadratic form $\hat{\sigma}_1^2 = y_1'M_1y_1/(n-K)$. Similarly, we can show that the linear form $Z_2'u_2$ is independently distributed of the quadratic

form $\hat{\sigma}_\nu^2 = y_2' M_2 y_2/(nT'-K')$. With these results, we can show that

$$(3.5) \qquad E[\hat{\delta}(\hat{\theta})|\theta] = \delta .$$

Since the expectation of δ over the distribution of $\hat{\theta}$ is δ, i.e., $E_{\hat{\theta}}\delta = \delta$, this proves that $\hat{\delta}(\hat{\theta})$ is an unbiased estimator of δ.

To establish the asymptotic properties of $\hat{\delta}(\hat{\theta})$, we assume that X's are weakly non-stochastic and, for a fixed n, $\lim_{T\to\infty} (nT)^{-1} Z_1' N Z_1$, $\lim_{T\to\infty} (nT)^{-1} Z_2' Z_2$ are all finite positive definite matrices.[7]

For a fixed n, under the above assumptions, we can show that $\lim_{T\to\infty} n^{-2}T^{-2}(Z_1' N u_1 u_1' N Z_1) = \lim_{T\to\infty} n^{-2}T^{-2} Z_2' u_2 u_2' Z = 0$, thus insuring that $\operatorname{plim}_{T\to\infty} (nT)^{-1} Z_1' N u_1 = 0 = \operatorname{plim}_{T\to\infty} (nT)^{-1} Z_2' u_2$. Also, $\operatorname{plim}_{T\to\infty} \hat{\sigma}_1^2 = \sigma_1^2$, $\operatorname{plim}_{T\to\infty} \hat{\sigma}_\nu^2 = \sigma_\nu^2$. Under these conditions, we can easily show that

$$(3.6) \qquad \operatorname{plim}_{T\to\infty} \sqrt{nT}[\hat{\delta}(\hat{\theta}) - \delta] = 0 ,$$

i.e., $\hat{\delta}(\hat{\theta})$ is a consistent estimator of δ.

---

7. The assumption of non-stochastic implies that the time pattern of the variable is bounded by some finite limits, even though it is not necessary for the pattern of the variable to repeat itself. The meaning of non-stochastic X's is simply that the realization of the X's is in accordance with some fixed (albeit unknown) process. Since economic data are stochastic, whichever assumption we adopt about the nature of the fixity of X's, we are simplifying and possibly mis-specifying the model. See also Wallace and Hussain (14) pp. 55-72.

Under the above assumptions, we can readily show that $\hat{V}$ is a consistent estimator for V, and that $\hat{\delta}(\hat{\theta})$, as obtained in (2.15), is asymptotically equivalent to $\hat{\delta}(\theta)$, in the sense that $\sqrt{nT}[\hat{\delta}(\hat{\theta}) - \hat{\delta}(\theta)]$ converges in probability to zero as $T\to\infty$, both coefficient estimators being asymptotically normally distributed with mean vector δ and covariance matrix $(W'V^{-1}W)^{-1}$. Since $\operatorname{plim}_{T\to\infty} \hat{\sigma}_1^2/T = \sigma_\mu^2$, we can further show that $\hat{\delta}(\hat{\theta})$ is also asymptotically equivalent to the covariance estimator $\tilde{\delta}$, i.e.,

$$\operatorname{plim}_{T\to\infty} \sqrt{nT}|\hat{\delta}(\hat{\theta}) - \tilde{\delta}| = 0 . \tag{3.7}$$

In fact we can show that there is an infinitely large number of estimators which yield asymptotically equivalent estimates with asymptotically equivalent variance-covariance matrices. Thus asymptotic theory casts relatively little light on the comparative small sample properties of the estimators. In the next section, we evaluate relative efficiency of the various estimation procedures by using a Monte Carlo study.

Section 4

Design of the Experiment and the Comparative Properties of the Various Estimators

The design of the Monte Carlo experiments given here is similar to that of Nerlove, except that our model contains an intercept and we generate random numbers by a slightly different, but more efficient, method.[8] Since Nerlove has already done extensive Monte Carlo studies, we examine intensively only those cases with large inter-individual heterogeneity and varying T. The model is given by

$$(4.1) \qquad y_{it} = \beta_o + \beta_1 X_{1it} + \mu_i + \nu_{it} .$$

The explanatory variable, X_{1it}, held fixed throughout the experiment, is generated as follows:

$$(4.2) \qquad X_{1it} = 0.1(t-1) + 1.05X_{1it-1} + w_t ,$$

where w_t is uniformly distributed in the range from 0 to 2. Initial values of X_{1i0} are chosen at random from the uniformly distributed numbers in the range 0 to 100. To generate nT values of u_{it} independent normal variables with zero mean and unit variance, n μ_i's are first selected with $\mu_i \sim N(0, \sigma_\mu^2)$, nT ν_{it}'s are then selected with $\nu_{it} \sim N(0, \sigma_\nu^2)$, and these are summed to give the u_{it}'s. Defining ρ, the intra-class correla-

---

8. This method and a method to convert uniformly distributed random variables to normal variate is described in Appendix A-2. See also Nerlove (11), pp. 366-371.

tion coefficient, as $\rho = \sigma_\mu^2/\sigma^2$, where $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2$, we can write $\mu_i \sim N(0, \rho\sigma^2)$ and $\nu_{it} \sim N(0, (1-\rho)\sigma^2)$.

Twelve sets of y_{it}'s were generated for various combinations of the parameter values $\beta_o = 0$ and 5; $\beta_1 = 0.5$ and 0.8; $\rho = 0$, 0.4 and 0.8; and $\sigma^2 = 10$. Initially n is set at 25 and T at 6.[9] For each set of parameters, five estimating procedures were examined:

(a) OLS estimator based on data aggregated over time, $\hat{\delta}(1)$,

(b) covariance estimator, $\hat{\delta}(2)$,

(c) ordinary least squares, $(\hat{\delta})$,

(d) pooled estimator based on the estimated variance-covariance matrix, $\hat{\delta}(\hat{\theta})$,

(e) generalized least squares based on known variance-covariance matrix, $\hat{\delta}(\theta)$.

In each experiment, 20 repetitions are performed, from which the mean and the mean square error of the estimated coefficients are calculated. The entire set of experiments is repeated with T set at 15, giving 480 runs and 24 tables of

---

9. These parameters were selected from the initial set of parameters $\beta_o = 0.0$, 0.5, 1.0, 5.0 and 10.0; $\beta_1 = 0.1$, 0.5, 0.8; $\rho = 0.0$, 0.2, 0.4, 0.6 and 0.8; $\sigma^2 = 10$ and 20. For these 150 sets of parameters 2 repetitions were performed. On the basis of mean square error of the estimators in the various estimating procedures only 12 parameter sets were selected for intensive study. The choice of these parameter values may itself cause bias in our results, but the very consistency of the trend strengthens our belief that this is a representative set.

mean and mean square error of the coefficients for different estimating procedures. Table 1 presents the mean and the mean square error for one such experiment. Results of various other runs are presented in an appendix to this paper.[10]

Table 1

Mean and Mean Square Error of the Coefficients for Various Estimating Methods for $\sigma^2 = 10$, $N = 25$, $T = 6$, and $\rho = .8$

Method	Mean (β_0)	m.s.e. (β_0)	Mean (β_1)	m.s.e. (β_1)
True Value	5	0	0.5	0
OLS--Agg.	4.99184	0.703589	0.499717	7.725E-05
Covariance	--	--	0.499688	5.221E-05
OLS	4.99208	0.594746	0.499714	5.540E-05
Pooled	5.00708	0.400888	0.499682	2.791E-05
GLS	4.99329	0.386660	0.499695	2.788E-05

From Table 1, we find that the mean values of β_0 and β_1 for all estimating methods are finitely close to the true values, thus demonstrating that all estimators under consideration are unbiased, but that the mean square error for the different estimators varies considerably. The mean square error of OLS--Agg. is about three times as large as that of the generalized least squares estimator, while those of the covariance estimator and of the ordinary least squares estimator are

10. Mean square error of an estimator $\hat{\theta}$ of θ is given by
$$\text{m.s.e.} = \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i - \theta)^2.$$

only about twice as large. The mean square error of the pooled estimator is nearly the same as that of the generalized least squares.

This is true for all values of ρ except $\rho=0$. In this case, all estimators have mean square error equal to that of the generalized least squares estimator.[11] As ρ increases, so does the ratio of the mean square error of the OLS estimator to that of the GLS estimator, but for all values of ρ the mean square error of the pooled estimator nearly equals the mean square error of the GLS. Further, for large values of ρ, the OLS method gives a serious underestimate of σ^2, giving low standard errors of the estimates. In contrast, the standard errors for the pooled estimator and the GLS are nearly equal. As T increases, the mean square error of the covariance estimator declines, becoming almost equal to that of the pooled estimator and the GLS estimator.

Hence we see, on the basis of the criterion of minimum mean square error, that the pooled estimator compares favorably, for all T's and all ρ's, with all other estimators which do not require a prior knowledge of the variance-covariance matrix. Furthermore, this estimator shows definite superiority to other estimators for small T's and large ρ. On the basis of the criterion of unbiasedness, this compares equally well with all other estimators.

---

11. See tables 1 and 2 in the appendix to this paper.

Section 5

Conclusion

In this paper, we have developed an operational method for estimating error components regression models when the variance-covariance matrix of the disturbance terms is unknown. Monte Carlo studies were conducted to compare the relative efficiency of the pooled estimator obtained by this procedure to (a) an ordinary least squares estimator based on data aggregated over time, (b) the covariance estimator, (c) the ordinary least squares estimator, and (d) a generalized least squares estimator based on a known variance-covariance matrix. For T small and large ρ, this estimator definitely performs better than the other estimators which are also based on an estimated value of the variance-covariance matrix of the disturbances. For ρ small and large T it compares equally well with the other estimators. In this instance, therefore, we are able to give a definite unconditional answer to the question posed to Nerlove's Dodo, "But who has won?"--the pooled estimator, of course!

Appendix A-1

Table 1

Mean and Mean Square Error of the Coefficients for Various Estimating Procedures for $T = 6$, $N = 25$, and $\sigma^2 = 10$

	Method	Mean (β_0)	m.s.e. (β_0)	Mean (β_1)	m.s.e. (β_1)
ρ=0.0:	True Value	5	0	0.80	0
	OLS--Agg.	5.20645	0.286728	0.798293	6.564E-05
	Covariance	--	--	0.806431	3.370E-04
	OLS	5.14152	0.216076	0.799298	5.342E-05
	Pooled	5.12310	0.200142	0.799582	5.227E-05
	GLS	5.14149	0.216061	0.799297	5.342E-05
ρ=0.4:	True Value	5	0	0.80	0
	OLS--Agg.	5.34402	0.839164	0.794023	1.970E-04
	Covariance	--	--	0.800825	2.190E-04
	OLS	5.28975	0.673243	0.794863	1.900E-04
	Pooled	5.18582	0.544211	0.796471	9.108E-05
	GLS	5.16231	0.538673	0.796834	8.246E-05
ρ=0.8:	True Value	5	0	0.50	0
	OLS--Agg.	4.99184	0.70300	0.499717	7.725E-05
	Covariance	--	--	0.499688	5.221E-05
	OLS	4.99208	0.594746	0.499714	5.540E-05
	Pooled	5.00708	0.400888	0.499682	2.731E-05
	GLS	4.99329	0.386660	0.499695	2.788E-05

Table 2

Mean and Mean Square Error of the Coefficients for Various Estimating Procedures for $T = 15$, $N = 25$, and $\sigma^2 = 10$

	Method	Mean (β_0)	m.s.e. (β_0)	Mean (β_1)	m.s.e. (β_1)
ρ=0.0:	True Value	5.0	0	0.80	0
	OLS--Agg.	5.04207	0.214164	0.800005	1.527E-05
	Covariance	--	--	0.800616	1.241E-05
	OLS	5.00136	0.074410	0.800355	4.829E-06
	Pooled	5.00143	0.074891	0.800353	4.721E-06
	GLS	5.00150	0.074409	0.800352	4.827E-06
ρ=0.4:	True Value	5.0	0	0.80	0
	OLS--Agg.	4.88203	1.03832	0.800393	6.816E-05
	Covariance	--	--	0.799514	6.300E-06
	OLS	4.94057	0.346083	0.799890	1.593E-05
	Pooled	4.97141	0.217877	0.799625	6.210E-06
	GLS	4.97792	0.222983	0.799570	6.135E-06
ρ=0.8:	True Value	5.0	0	0.80	0
	OLS--Agg.	5.57304	2.61798	0.793599	2.010E-04
	Covariance	--	--	0.799879	1.413E-06
	OLS	5.15513	0.639253	0.797190	3.645E-05
	Pooled	4.85477	0.341444	0.799771	1.381E-06
	GLS	4.85116	0.344148	0.799802	1.368E-06

Appendix A-2

Random Number Generating Procedures

A desired sequence of random numbers X_n is obtained by setting

$$X_{n+1} = (aX_n + c) \bmod m \qquad m \geq 0$$

where a is the multiplier, c is the increment and m is the modulus, $a \geq 0$, $c \geq 0$, $m > c$, $m > a$, and $m > X_o$, where X_o is the starting value. This method is called linear congruential sequence. When c=0, the random generation process is slightly faster, but the maximum period length (length after which sequence starts repeating itself) can not be achieved. Nerlove (11) in order to avoid this problem, suggests mixing two random sequences into a third, so that the third one is extremely random. We use a method suggested by Maclaren and Marsaglia as described below.[1]

A quite random sequence Given methods for generating two sequences X_n and Y_n, this method produces a "considerably more random" sequence. We use an auxiliary table V(0), V(1), ..., V(k-1), where k is some number chosen for convenience, usually in the neighborhood of 100. Initially, the V-table is filled with the first k values of the X-sequence.

---

1. See also Knuth (8) pp. 25-31.

Step 1: [Generate X, Y] Set X, Y equal to the next number of the sequence (X_n, Y_n) respectively.

Step 2: [Extract j] Set $j \leftarrow [kY/m]$, where m is the modulus used in the sequence Y_n; i.e., j is a random value, $0 \leq j \leq k$ determined by Y.

Step 3: [Exchange] Output V(j) and then set $V(j) \leftarrow X$.

This method gives an incredibly long period if the periods of (X_n) and (Y_n) are relatively prime; and even if the period is of no consequence, there is very little relation between the nearby terms of the sequence. To generate independently normal variates we follow the Polar Method, which consists of generating two independent random variables (u_1 and u_2) uniformly distributed between zero and one.[2] A set of independent normal variates with mean zero and variance one is obtained by the transformation

$$w_1 = (-2 \log u_1)^{1/2} \cos (2\pi u_2)$$

$$w_2 = (-2 \log u_2)^{1/2} \cos (2\pi u_1) .$$[3]

---

2. To generate variable uniformly distributed between zero and one, we first generate some random number X_n between zero and m as described above, and then the fraction $u_1 = X_n/m$ will lie between zero and one.

3. For a comprehensive discussion of this method see Knuth (8) pp. 103-105; also see Nerlove (11), p. 368 footnote 11.

References

1. Anemiya, Takeshi (1971): "The Estimation of the Variances in a Variance-Components Model," International Economic Review, Vol. 12, No. 1, pp. 1-13.

2. Arora, Swarnjit S. (1971): "Error Components and Multipolar Human Flow Models," unpublished Ph.D. dissertation submitted to the State University of New York at Buffalo, September 1971.

3. Balestra, P. and M. Nerlove (1966): "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, Vol. 34, No. 4, pp. 585-612.

4. Henderson, Charles R., Jr. (1971): "Comment on the Use of Error Components Models in Combining Cross Section with Time Series Data," Econometrica, Vol. 39, No. 2, pp. 397-401.

5. Hoch, I. (1962): "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data," Econometrica, Vol. 30, No. 1, pp. 34-53.

6. Hussain, A. (1969): "A Mixed Model for Regression," Biometrika, Vol. 56, pp. 327-36.

7. Kuh, E. (1959): "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Application," Econometrica, Vol. 27, No. 2, pp. 197-214.

8. Knuth, Donald E. (1968): Seminumerical Algorithms--The Art of Computer Programming, Vol. 2, Addison-Wesley Publishing Company, Reading, Massachusetts, pp. 25-31 and pp. 103-105.

9. Maddala, G.S. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," Econometrica, Vol. 39, No. 2, pp. 341-58.

10. Mundlak, Y. (1963): "Estimation of Production and Behavioral Functions from a Combination of Cross-Section and Time Series Data," pp. 138-66 in Measurement in Economics, C.F. Christ, Ed., Stanford, California, Stanford University Press.

11. Nerlove, M. (1971a): "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Section," Econometrica, Vol. 39, No. 2, pp. 359-81.

12. _____ (1971b): "A Note on Error Components Models," Econometrica, Vol. 39, No. 2, pp. 382-96.

13. Swamy, P.A.V.B. and S.S. Arora (1972): "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," Econometrica, Vol. 40, No. 2, pp. 261-75.

14. Wallace, T.D. and A. Hussain (1969): "The Use of Error Components Models in Combining Cross Section with Time Series Data," Econometrica, Vol. 37, No. 1, pp. 55-72.