ERROR COMPONENTS REGRESSION MODELS AND THEIR APPLICATIONS*

by Swarnjit S. Arora

This paper develops an operational method for estimating error components regression models when the variance-covariance matrix of the disturbance terms is unknown. Through Monte Carlo studies, the relative efficiency of the resulting pooled estimator is compared with (1) an OLS estimator based on data aggregated over time; (2) the covariance estimator; (3) the OLS estimator, and (4) a GLS estimator based on a known variance-covariance matrix. Overall, the pooled estimator performs best.

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

\[ y_{it} = \beta_0 + \sum_{k=1}^{K-1} \delta_k Z_{ik} + u_{it} \quad (i = 1, 2, \ldots, n; t = 1, 2, \ldots, T) \]

where \( y_{it} \) is an observation on the dependent variable for individual \( i \) in period \( t \), \( Z_{ik} \) is an observation on the \( k \)th independent variable, \( \beta_0 \) is an intercept term, \( \delta_k (k = 1, 2, \ldots, K - 1) \) 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 \( \mu_i \)), some are specific to the time period but are invariant over all individuals (say \( \lambda_t \)), and some are specific to both individual and time (say \( \nu_{it} \)). In this case we can write \( u_{it} \) as

\[ u_{it} = \mu_i + \lambda_t + \nu_{it} \]

Mundlak (10) and Hoch (5) analyzed this model, treating \( \mu_i \) and \( \lambda_t \) as unknown parameters and assuming \( \Sigma_{t=1}^{T} \mu_i = 0 \) and \( \Sigma_{i=1}^{n} \lambda_t = 0 \). 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).

* A major portion of this research was done during the author's stay at the National Bureau of Economic Research as a Research Fellow. Special thanks are due to Professors David Belsley, Murray Brown, Paul Holland, and Edwin Kuh for their valuable comments and suggestions, and to Professor P. A. V. B. Swamy for introducing me to this subject. 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.

** 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.

* Revised June 25, 1973
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. 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 intractability, 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 to 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.

2. ESTIMATION OF ERROR COMPONENTS REGRESSION MODELS

Let us assume that \( u_t = \mu_i + \epsilon_t \) and the components \( \mu_i \)s and \( \epsilon_t \)s are random such that

\[
E\mu_i = 0 \\
E\mu_i\mu_j = \begin{cases} 
\sigma_\mu^2 & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases} \\
E\epsilon_i = 0 \\
E\epsilon_i\epsilon_{i'} = \begin{cases} 
\sigma_\epsilon^2 & \text{if } i = j \text{ and } t = t' \\
0 & \text{otherwise} \end{cases}
\]

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

In (13) Swamy and I have analytically shown that the estimator based on the assumption of both \( \mu_i \) and \( \epsilon_t \) being random is more efficient than other 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 component model which assumes both \( \mu_i \) and \( \epsilon_t \) as random. Here we consider a model with \( \epsilon_t = \xi_t \)
Let us further assume that \( u_i \) and \( v_i \) are independent of each other. Furthermore, \( T > K \) and \( n > K \) and the variances \( \sigma_u^2 \) and \( \sigma_v^2 \) are unknown. For all the \( nT \) observations combined we can write (1.1)

\[
y = X\beta + u,
\]

where \( y = (y_1, \ldots, y_T, \ldots, y_{1T}, \ldots, y_{nT}) \) is an \( nT \times 1 \) vector of observations on the dependent variable, \( X = [u_i, Z] \) is an \( nT \times K \) matrix of explanatory variables. \( u_i \) is a vector of 1's of order \( nT \times 1 \), and \( Z \) is an \( nT \times K' \) matrix of independent variables given by

\[
Z = \begin{bmatrix}
Z_{11} & Z_{12} & \cdots & Z_{1K'} \\
Z_{21} & Z_{22} & \cdots & Z_{2K'} \\
\vdots & \vdots & \ddots & \vdots \\
Z_{nT,1} & Z_{nT,2} & \cdots & Z_{nT,K'}
\end{bmatrix}
\]

\( \delta \) is a \((K' \times 1)\) vector of slope coefficients. \( \beta \equiv (\beta_0, \delta)^T \). \( K' = K - 1 \), and \( u = (u_1, \ldots, u_T, \ldots, u_{1T}, \ldots, u_{nT}) \) is an \( nT \times 1 \) vector of disturbance terms. Under the above assumptions, it is readily verified that

\[
E(u'u) = \sigma_u^2[I_n \otimes I_1] + \sigma_v^2[I_{nT}].
\]

Since the variance-covariance matrix of \( u \) is not scalar, application of the ordinary least squares procedure will in general lead to an inefficient estimator of \( \beta \).

Let us consider an orthogonal matrix \( O_T \) of order \( T \) such that its first row is equal to \( \frac{1}{\sqrt{T}} \) \( \cdot \). Let \( O_T = \frac{1}{\sqrt{T}} \cdot [T_C] \) where \( C_1 \) is a \((T - 1) \times T\) matrix such that \( C_11_T = 0 \), \( C_1C_1^T = I_{T-1} \), and \( C_1^TC_1 = 1 - \frac{1}{T} \).

Define the transformations \( Q_1 = (I_n \otimes \frac{1}{\sqrt{T}}) \) and \( Q_2 = (I_n \otimes C_1) \). By applying the transformation \( Q_1 \) to all \( nT \) observations, we get

\[
y_1 = X_1\beta + u_1,
\]

where \( y_1 = Q_1y \) is an \( n \times 1 \) vector of transformed dependent variables, \( X_1 = Q_1X \) is an \( n \times K \) matrix of independent variables, and \( u_1 \) is an \( n \times 1 \) vector of transformed disturbances. The variance-covariance matrix of \( u_1 \) is

\[
E(u_1'u_1) = \sigma_u^2[I_n],
\]

Substituting for \( E(u'u) \) from (2.2) and simplifying we get

\[
E(u_1'u_1) = \sigma_u^2[I_n].
\]

\( ^4 \) A model in this form was also used by Kuh (7), except that he did not assume \( u_i \) and \( v_i \) are uncorrelated. Hussain (6) treats a model with \( \rho_i \)'s as parameters, and \( i_1 \) and \( v_i \) as random.
where \( \sigma_1^2 = \theta \sigma_2^2 + \sigma_1^2 \). Thus the variance-covariance matrix reduces to scalar form, a best linear unbiased estimator of \( \beta \) is an OLS estimator
\[
\hat{\beta}(1) = (X_1'X_1)^{-1}X_1'Y_1.
\]
The subvector of \( \hat{\beta}(1) \) corresponding to the slope coefficients only is given by
\[
\hat{\delta}(1) = (Z_1'N_iZ_1)^{-1}Z_1'N_iY_1,
\]
where \( Z_1 = Q_1Z \) is an \( n \times K' \) matrix and \( N = I_n - t_iC_i/n \): a subscript \( t \) is attached to \( \beta \) and \( \hat{\delta} \) to differentiate these estimators from the other estimators of \( \beta \) and \( \hat{\delta} \) to be described later.

The variance-covariance matrix of \( \hat{\delta}(1) \) is
\[
V[\hat{\delta}(1)] = \sigma_1^2(Z_1'N_iZ_1)^{-1}.
\]
Applying the transformation \( Q = (I \otimes C) \) to all \( uT \) observations we have
\[
y_2 = Z_2\delta + u_2,
\]
where \( y_2 = Q_2y \) 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, \( u_2 = Q_2u \) is an \( nT' \times 1 \) vector of transformed disturbances, and use is also made of the result \( Q_2u = 0 \). The variance-covariance matrix of \( u_2 \) is
\[
E(u_2u_2') = EQ_2uu'Q_2 = Q_2EuuQ_2;
\]
which can easily be reduced to \( \sigma_1^2I_{nT} \). Thus the variance-covariance matrix of \( u_2 \) is of scalar form. A best linear unbiased estimator of \( \delta \) is the OLS estimator given by
\[
\hat{\delta}(2) = (Z_2'Z_2)^{-1}Z_2'y_2.
\]
The variance-covariance matrix of \( \hat{\delta}(2) \) is \( \sigma_1^2(Z_2'Z_2)^{-1} \). Notice that \( Q_1Q_2 = 0 \). The rank of \( Q_1 \) is equal to the rank of \( t_i'\sqrt{T} \) multiplied by the rank of \( I_n \) because if \( A \) and \( B \) are any arbitrary matrices, the rank of \( (A \otimes B) \) 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 parameter vector and we can pool them in the following manner.
\[
\hat{\delta}(0) = \left[ \begin{array}{c}
Z_1'N_iZ_1
\end{array} + \frac{Z_2'Z_2}{\sigma_1^2} \right]^{-1} \left[ \begin{array}{c}
Z_1'N_iY_1
\end{array} + \frac{Z_2'y_2}{\sigma_1^2} \right]
\]
where \( \theta = [\sigma_1^2, \sigma_2^2] \). The estimator \( \hat{\delta}(0) \) is a generalized least squares estimator of \( \delta \). For given values of \( \sigma_1^2 \) and \( \sigma_2^2 \), it is a best linear unbiased estimator. Any other

\(^5\) 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 \( nT - 1 \) independent observations.
estimator of \( \delta \) 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.

An unbiased estimator of \( \sigma_2^2 \) is given by

\[
\hat{\sigma}_2^2 = y_i'M_jy_i/(n-K),
\]

where \( M_j = T - X_j'(X_j'X_j)^{-1}X_j \). Also, an unbiased estimator of \( \sigma_1^2 \) is

\[
\hat{\sigma}_1^2 = y_i'M_jy_i/(nT-K),
\]

where \( M_j = T - Z_j'(Z_jZ_j)^{-1}Z_j \). \( T = T - 1 \) and \( K = K - 1 \). An Aitken estimator of the slope coefficients based on the estimated values of \( \sigma_1^2 \) and \( \sigma_2^2 \) is given by

\[
\hat{\delta} = \left[ \begin{array}{c} Z_1'ny_1 + Z_2'ny_2 \\ Z_1'n\sigma_2^2 + Z_2'n\sigma_1^2 \end{array} \right] \left[ \begin{array}{c} \sigma_1^2 \\ \sigma_2^2 \end{array} \right]
\]

where \( \hat{\delta} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2) \).

We can readily show that the estimator \( \hat{\delta}(2) \) as obtained in (2.11) is equivalent to a covariance estimator (say \( \hat{\delta} \)) obtained by assuming \( \mu_i \)'s as fixed parameters. We can also show equivalence of \( \hat{\delta}(0) \) with an ordinary least squares estimator (\( \hat{\delta} \)) when \( \sigma_2^2 = 0 \) as follows:

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

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

where

\[
Q_4 = I_{nT} - \frac{I_nI_nT}{nT}.
\]

The Aitken estimator \( \hat{\delta}(0) \) in (2.15) when \( \sigma_2^2 = 0 \) is

\[
\hat{\delta}(0) = [Z_1'ny_1 + Z_2'ny_2]^{-1}[Z_1'n\sigma_1^2 + Z_2'n\sigma_2^2]
\]

= \( [Z(Q_1'Q_1 + Q_2'Q_2)Z]^{-1}[Z(Q_1'Q_1 + Q_2'Q_2)y] \).

Since

\[
Q_1'Q_1 = I_n \otimes \frac{I_T^2}{T} - \frac{I_TI_T^T}{nT},
\]

and

\[
Q_2'Q_2 = I_n \otimes \left( I_T - \frac{I_TI_T^T}{nT} \right),
\]

we can easily show that \( Q_1'Q_1 + Q_2'Q_2 = I_{nT} - \frac{I_TI_T^T}{nT} = Q_4 \), thus proving equality of \( \hat{\delta} \) and \( \hat{\delta}(0) \) when \( \sigma_2^2 = 0 \).

\* A similar pooled estimator of \( \hat{\delta}(0) \) can be obtained if \( \mu_i \) and \( \sigma^2_i \) are assumed to be random. The variance-covariance matrix of \( \eta \) is given by \( \sigma^2_2\sigma_1^2 \otimes I_T + \sigma^2_1\sigma_2^2 \). To reduce this to scalar form, we consider an orthogonal matrix \( Q_0 = [Q_0^1 \otimes \ldots \otimes Q_0^n] \) of order \( n \), and apply transformations \( Q_0 = [Q_0^1 \otimes \ldots \otimes Q_0^n] \) and \( Q_2 = [Q_0^1 \otimes I_T] \) to all the \( nT \) observations.

455
3. Properties of the Estimators

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

\[
(3.1) \quad \hat{\theta}(\hat{\theta}) = \left[ \begin{bmatrix} Z_1'N & Z_2' \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} N & \dagger \\ \dagger & \dagger \end{bmatrix} \begin{bmatrix} Z_1'N & Z_2' \end{bmatrix} \right]^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} N \dagger \\ \dagger \end{bmatrix} \]

If \( W = [Z_1'N, Z_2']', \hat{\gamma} = [\hat{\gamma}_1'N, \hat{\gamma}_2']', \) the equation (3.1) can be written as

\[
(3.2) \quad \hat{\theta}(\hat{\theta}) = (W'\hat{\gamma}^{-1}W)^{-1}W'\hat{\gamma}^{-1}\hat{\gamma}.
\]

Substituting \( \gamma = \theta + \epsilon \) where \( \epsilon = [u_1N, u_2]' \) in (3.2), we get

\[
(3.3) \quad \hat{\theta}(\hat{\theta}) = \delta + (W'\hat{\gamma}^{-1}W)^{-1}W'\hat{\gamma}^{-1}\epsilon.
\]

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

\[
(3.4) \quad \hat{\theta}(\hat{\theta}) = \hat{\delta} + (W'\hat{\gamma}^{-1}W)^{-1}W'\hat{\gamma}^{-1}\epsilon.
\]

If we assume that \( u \)'s are normally distributed, and since

\[
M_1X_1 = [1_n - X_1'(X_1'X_1)^{-1}X_1]X_1 = 0,
\]

we can show that the linear form \( Z_1'N\hat{u}_1 \) is distributed independently of the quadratic form \( \hat{\delta}_1^2 = \hat{y}_1'M_1\hat{y}_1 \). Similarly, we can show that the linear form \( Z_2'\hat{u}_2 \) is independently distributed of the quadratic form \( \hat{\delta}_2^2 = \hat{y}_2'M_2\hat{y}_2 \). With these results, we can show that

\[
(3.5) \quad E[\hat{\theta}(\hat{\theta}) - \delta] = 0,
\]

i.e., \( \hat{\theta}(\hat{\theta}) \) is an unbiased estimator of \( \delta \).

To establish the asymptotic properties of \( \hat{\theta}(\hat{\theta}) \), we assume that \( X \)'s are weakly non-stochastic and, for a fixed \( n \), \( \lim_{T \to \infty} (nT)^{-1}Z_1'NZ_1, \lim_{T \to \infty} (nT)^{-1}Z_2'Z_2 \)

are all finite positive definite matrices. For a fixed \( n \), under the above assumptions, we can show that \( \lim_{T \to \infty} n^{-2}T^{-2}(Z_1'N\hat{u}_1\hat{u}_1'NZ_1) = \lim_{T \to \infty} n^{-2}T^{-2}Z_2'\hat{u}_2\hat{u}_2'Z_2 = 0 \), thus insuring that \( \lim_{T \to \infty} (nT)^{-1}Z_1'N\hat{u}_1 = 0 = \lim_{T \to \infty} (nT)^{-1}Z_2'\hat{u}_2 \). Also, \( \lim_{T \to \infty} \hat{\delta}_1^2 = \delta_1^2; \lim_{T \to \infty} \hat{\delta}_2^2 = \delta_2^2 \). Under these conditions, we can easily show that

\[
(3.6) \quad \lim_{T \to \infty} \sqrt{nT}[\hat{\theta}(\hat{\theta}) - \delta] = 0,
\]

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

Under the above assumptions, we can readily show that \( V \) is a consistent estimator of \( V \) and that \( \hat{\theta}(\hat{\theta}), \) as obtained in (2.19), is asymptotically equivalent to \( \hat{\theta}(\hat{\theta}) \), in the sense that \( \sqrt{nT}[\hat{\theta}(\hat{\theta}) - \hat{\theta}(\hat{\theta})] \) converges in probability to zero as \( T \to \infty \).

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 idea of an unknown process is based on the realization of the \( Y \)'s in accordance with some fixed nature of the process. For the model specified by \( \Delta \), we are simplifying the model and assuming that it is asymptotically equivalent to the model specified by \( \Delta \), for large values of \( n \).

See also Wallace and Hussain (14), p. 55-72.
both coefficient estimators being asymptotically normally distributed with mean vector \( \bar{\delta} \) and covariance matrix \( \Sigma = \sigma^2 \). Since \( \text{plim}_{T \to \infty} \delta = \sigma^2 \), we can further show that \( \delta_0 \) is also asymptotically equivalent to the covariance estimator \( \hat{\delta} \), i.e.,

\[
\text{plim}_{T \to \infty} \sqrt{nT} |\hat{\delta}_0 - \hat{\delta}| = 0.
\]

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.

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. 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

\[
y_{it} = \beta_0 + \beta_1 X_{10} + \mu_i + \epsilon_{it}.
\]

The explanatory variable, \( X_{10} \), held fixed throughout the experiment, is generated as follows:

\[
X_{10} = 0.1(t - 1) + 1.05X_{10_{t-1}} + w_t,
\]

where \( w_t \) is uniformly distributed in the range from 0 to 2. Initial values of \( X_{10} \) are chosen at random from the uniformly distributed numbers in the range 0 to 100. To generate \( nT \) values of \( u_i \), independent normal variables with zero mean and unit variance, \( n \mu_i \)'s are first selected with \( \mu_i \sim N(0, \sigma^2) \), \( nT \) \( v_i \)'s are then selected with \( v_i \sim N(0, \sigma^2) \), and these are summed to give the \( u_i \)'s. Defining \( \rho \), the intra-class correlation coefficient, as \( \rho = \sigma^2 / \sigma^2 + \sigma^2 \), we can write

\[
\mu_i \sim N(0, \rho \sigma^2) \quad \text{and} \quad v_i \sim N(0, (1 - \rho) \sigma^2).
\]

Twelve sets of \( u_i \)'s were generated for various combinations of the parameter values \( \beta_0 = 0 \) and 5; \( \beta_1 = 0.5 \) and 0.8; \( \rho = 0, 0.2, 0.4, 0.6 \) 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 estimation 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.
(b) covariance estimator, \( \hat{\delta} \).
(c) ordinary least squares, \( \hat{\beta} \).
(d) pooled estimator based on the estimated variance-covariance matrix, \( \hat{\delta} \).
(e) generalized least squares based on known variance-covariance matrix, \( \hat{\delta} \).

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 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.

From Table 1, we find that the mean values of \( \beta_0 \) and \( \beta_1 \) 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 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 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 \( \rho \) except \( \rho = 0 \). In this case, all estimators have mean square error equal to that of the generalized least squares estimator. As \( \rho \) 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 \( \rho \) the mean square error of the pooled estimator nearly equals the mean square error of the GLS estimator. Further, for large values of \( \rho \), the OLS method gives a serious underestimate of \( \sigma^2 \), 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean ( \hat{\beta}_0 )</th>
<th>MSE ( \hat{\beta}_1 )</th>
<th>Mean ( \hat{\beta}_1 )</th>
<th>MSE ( \hat{\beta}_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OLS</td>
<td>4.99128</td>
<td>0.703589</td>
<td>0.499717</td>
<td>7.223E-05</td>
</tr>
<tr>
<td>Covariance</td>
<td>0.499688</td>
<td>1.791 E-05</td>
<td>0.499714</td>
<td>5.221E-05</td>
</tr>
<tr>
<td>GLS</td>
<td>4.99128</td>
<td>0.400882</td>
<td>0.499682</td>
<td>7.788E-05</td>
</tr>
</tbody>
</table>

Mean square error of an estimator \( \hat{\beta} \) of \( \beta \) is given by

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{\beta}_i - \beta_i)^2.
\]

See tables 1 and 2 in the appendix to this paper. Closeness of the estimates to the true values may also indicate that the error terms are relatively small as compared to the \( X_i \).
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 $p$'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 $p$. On the basis of the criterion of unbiasedness, this compares equally well with all other estimators.

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 small and large $T$, 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 $p$ 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!

*National Bureau of Economic Research, and The University of Wisconsin Milwaukkee*

**APPENDIX A-1**

**TABLE I**

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean $(\hat{b}_0)$</th>
<th>m.s.e. $(\hat{b}_0)$</th>
<th>Mean $(\hat{b}_1)$</th>
<th>m.s.e. $(\hat{b}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 0.0$: True Value</td>
<td>5</td>
<td>0</td>
<td>0.80</td>
<td>0</td>
</tr>
<tr>
<td>OLS - Agg</td>
<td>5.26665</td>
<td>0.266728</td>
<td>0.798293</td>
<td>0.564E-05</td>
</tr>
<tr>
<td>Covariance</td>
<td>5.14132</td>
<td>0.316076</td>
<td>0.792798</td>
<td>5.323E-05</td>
</tr>
<tr>
<td>GLS</td>
<td>5.12310</td>
<td>0.200412</td>
<td>0.795582</td>
<td>5.227E-05</td>
</tr>
<tr>
<td>OLS</td>
<td>5.14149</td>
<td>0.216064</td>
<td>0.799297</td>
<td>5.342E-05</td>
</tr>
<tr>
<td>$p = 0.4$: True Value</td>
<td>5</td>
<td>0</td>
<td>0.80</td>
<td>0</td>
</tr>
<tr>
<td>OLS - Agg</td>
<td>5.34402</td>
<td>0.39164</td>
<td>0.794023</td>
<td>1.970E-04</td>
</tr>
<tr>
<td>Covariance</td>
<td>5.28775</td>
<td>0.673243</td>
<td>0.794863</td>
<td>1.900E-04</td>
</tr>
<tr>
<td>GLS</td>
<td>5.18582</td>
<td>0.544231</td>
<td>0.796471</td>
<td>9.108E-05</td>
</tr>
<tr>
<td>OLS</td>
<td>5.16231</td>
<td>0.558673</td>
<td>0.796834</td>
<td>8.246E-05</td>
</tr>
<tr>
<td>$p = 0.8$: True Value</td>
<td>5</td>
<td>0</td>
<td>0.80</td>
<td>0</td>
</tr>
<tr>
<td>OLS - Agg</td>
<td>4.99134</td>
<td>0.70300</td>
<td>0.49717</td>
<td>7.725E-05</td>
</tr>
<tr>
<td>Covariance</td>
<td>4.99708</td>
<td>0.994746</td>
<td>0.497143</td>
<td>5.540E-05</td>
</tr>
<tr>
<td>GLS</td>
<td>4.99329</td>
<td>0.386660</td>
<td>0.499695</td>
<td>2.788E-05</td>
</tr>
</tbody>
</table>

459
APPENDIX A-2

RANDOM NUMBER GENERATING PROCEDURES

A desired sequence of random numbers X, is obtained by setting

\[ X_{n+1} = (aX_n + c) \mod m \quad m \geq 0 \]

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 \( m > X_0, \) where \( X_0 \) 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) cannot 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

¹ See Knuth (8), pp. 25-31.
Step 3: [Exchange] Output $V(j)$ and then set $V(j) \leftarrow X$.

This method gives an incredibly long period if the periods of $(X)$ and $(Y)$ 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

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

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

**REFERENCES**


\(^{1}\) To generate variable uniformly distributed between zero and one, we first generate some random number $X$, between zero and one as described above, and then the fraction $u = X / m$ will lie between zero and one.

\(^{2}\) For a comprehensive discussion of this method see Knuth (81) pp. 103-105; also see Nerlove (11) p. 388, footnote 11.