TR # 60

A TECHNICAL REPORT

FROM

The Laboratory for Social Research

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Constrained and Unconstrained Maximum Likelihood Estimation of a Variance Components Model of Cross-sections Pooled over Time

Nancy B. Tuma
Dept. of Sociology
Stanford University

Alice A. Young
Graduate School of Industrial Administration
Carnegie-Mellon University

October, 1976

Technical Report #60

Laboratory for Social Research
Stanford University

*The order in which the authors' names are listed is alphabetical. Our thanks to Michael T. Hannan for useful comments and to Alan J. Smith for numerous helpful and gracious contributions.

ABSTRACT

This paper reports the results of Monte Carlo experiments on the quality of different estimators of parameters in a variance components model with both a lagged endogenous and an exogenous variable. Methods of estimation include ML (with and without the imposition of natural constraints on parameters) and three least squares techniques, including MGLS. We conclude that the relative quality of ML and MGLS estimates varies according to the amount of serial error correlation over time and to the relative strength of the effects of the lagged endogenous and the exogenous variable. We find that there is a slight, but usually negligible, advantage to constrained ML estimation as compared with unconstrained ML estimation.

1. INTRODUCTION

The variance components model of cross-sections pooled over a few discrete points in time has attracted considerable interest among social scientists (e.g., Balestra and Nerlove [1]; Nerlove and Schultz [15]; Hannan and Nielsen [7]). This interest has been heightened by the increasing availability of panel data and the accompanying need for a model and estimation technique that take full advantage of the richness of this information.

Maximum likelihood (ML) estimation of a variance components model has intuitive appeal for at least two reasons. First, ML estimators have desirable asymptotic properties under quite unrestrictive conditions (see, e.g., Kendall and Stuart [9]). Second, the ML method incorporates all assumptions about the model in a single estimation step, unlike modified generalized least squares procedures that require two estimation steps.

Previous evidence on the quality of ML estimates of regression coefficients in a variance components model is equivocal. Using ML to estimate regression coefficients in a variance components model containing both a lagged endogenous variable and an exogenous variable and simulated data, Nerlove [14] concluded that ML estimates were poor in quality.[1] On the other hand, Maddala and Mount's [11] application of ML to a variance components model containing only an exogenous variable (and no lagged endogenous variable), which also used simulated data, found that the quality of ML estimates was similar to that of estimates obtained by various other techniques, even when data were generated with lognormally-distributed errors rather than the normally-distributed errors assumed for the likelihood function.

There are two main possibilities why these two studies reached different conclusions concerning the quality of ML estimates of a variance components model. First, Nerlove's study, unlike Maddala and Mount's, included a lagged endogenous variable, and the joint presence of lagged endogenous and exogenous regressors may adversely affect the quality of ML estimates. Second, Nerlove's study constrained ρ to lie in the [0,1] interval by substituting a trigonometric function for it in the likelihood equation. This method of imposing constraints, which has been recommended by Box [2], leads to multiple maxima of the likelihood function; it can also increase the nonlinearity of the likelihood function and cause the matrix of second derivatives of the likelihood function to be singular or ill-conditioned (Murray [12]). Consequently the poor performance of ML in Nerlove's study may result from the way that constraints on ρ were imposed.

The research reported here uses simulated data to study the small-sample properties of estimates of a variance components model in which the right-hand side includes both a lagged enodgenous and an exogenous variable. We focus on three main issues. First, we compare the quality of estimates obtained by least squares alternatives with the quality of ML estimates obtained by imposing constraints on ρ. We obtained the constrained ML estimates with a recently-developed method (Gill and Murray [4]) that avoids the problems accompanying the use of a trigonometric function to impose constraints. Second, we compare the quality of constrained and unconstrained ML estimates of the slope coefficients and of ρ. We are unaware of any previous studies concerning the relative quality of constrained and unconstrained ML estimates of these parameters. Third, we consider how the quality of ML estimates depends on the effects of the lagged endogenous and exogenous variables and on ρ, the intra-class correlation of an individual's observations over time.

Section 2 gives an outline of the model that we study. Section 3 describes the method of generating the Monte Carlo data. In Section 4 we discuss the methods of estimation utilized. The results are presented in Section 5 and further discussed in the conclusion, Section 6.

2. THE MODEL

The model that we have studied can be represented as follows:

$$y_{it} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{it} + u_{it}, \qquad (2.1)$$

$$(i = 1,\ldots,N; t = 1,\ldots,T)$$

where the u_{it}'s are unobserved variables such that:

(a) $u_{it} = \mu_i + \nu_{it}$

(b) $E(\mu_i) = E(\nu_{it}) = 0$ For all i and t

(c) $E(\mu_i, \nu_{jt}) = 0$ For all i,j and t

(d) $E(\mu_i, \mu_j) = \begin{cases} \sigma_\mu^2 & i = j \\ 0 & i \neq j \end{cases}$

(e) $E(\nu_{it}, \nu_{js}) = \begin{cases} \sigma_\nu^2 & i = j, t = s \\ 0 & \text{Otherwise} \end{cases}$

Let:

$$\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2, \qquad (2.2)$$

$$\rho = \sigma_\mu^2 / \sigma^2 \qquad (2.3)$$

and:

$$u' = (u_{11}, \ldots, u_{iT}, u_{21}, \ldots, u_{2T}, \ldots, u_{N1}, \ldots, u_{NT}).$$

(We arrange the u_{it}'s in vector form, first by individual and then by time period.) Then the error variance-covariance matrix can be represented as:

$$E(uu') = \sigma^2\Omega,$$

where $\Omega = A \otimes I_N$, (2.4)

$$A = \begin{bmatrix} 1 & \rho & \cdot & \cdot & \cdot & \rho \\ \rho & 1 & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & 1 & \rho \\ \rho & \cdot & \cdot & \cdot & \rho & 1 \end{bmatrix},$$

and I_N is an N by N identity matrix. For a more detailed discussion of these assumptions see Nerlove [14] or Hannan and Young [8].

To comply with the ML method's requirement that a particular probability distribution be assumed, we have also assumed that the u_{it}'s are normally distributed random variables.

3. DATA GENERATION

We have followed Nerlove's [14] procedure for generating data with four exceptions: First, we have chosen the number of individuals N as fifty and the number of time periods T as five, whereas Nerlove chose twenty-five and ten, respectively. We chose the former values of N and T because they are representative of many available data sets. Second, we have generated pseudo-random variates by Marsaglia's rectangle-wedge-tail algorithm, recommended as best by Knuth [10], rather than the method described by Nerbove [14]. Third, we have studied somewhat different combinations of parameter values. In each combination we set $\beta_0 = 0.0$ and $\sigma^2 = 1.0$. We selected five values for ρ: 0.0, 0.25, 0.50, 0.75, and 0.90. To examine the dependence of estimator quality on the relative strength of the effects of the lagged endogenous and the exogenous variables, we chose three combinations of β_1 and β_2: $(\beta_1,\beta_2) = (0.3,1.0)$, $(0.8,1.0)$, and $(0.8,0.5)$. Thus, we examined a total of fifteen combinations of parameter values. Fourth, for each combination of parameter values we generated 100 sets of data, where Nerlove generated 50. The additional data sets give increased confidence about the properties of an estimator.

4. METHODS OF ESTIMATION

We have employed five methods of estimation. Two are variants on maximum likelihood: one with constraints on ρ and σ^2, one without. The other three are variations of least squares; they are included to provide baseline evaluations of the ML estimators. A fuller discussion of the three least squares estimators may be found in Nerlove [14] and in Hannan and Young [8].

We list each of these five methods:

(1) Ordinary Least Squares (OLS): We used ordinary least squares to estimate β_0, β_1 and β_2 in equation (2.1). This method ignores the nonzero error covariances across time periods and consequently leads to inconsistent estimates of the regression coefficients.

(2) Least Squares with Dummy Variables (LSDV): We obtained least squares estimates of the parameters in:

$$y_{it} = \beta_1 y_{i,t-1} + \beta_2 x_{it} + \sum_{i=1}^{N} \mu_i \ell_i + \nu_{it},$$

where the μ_i's are the individual specific constants and ℓ_i is a dummy variable which equals one for the i^{th} individual and zero otherwise.

The LSDV procedure provides the following sample estimate of ρ:

$$r_{LSDV} = \frac{\sum_{i=1}^{N} \left\{\hat{\mu}_i - \sum_{i=1}^{N} \hat{\mu}_i/N\right\}^2}{N} / \hat{\sigma}^2$$

where:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \left\{\mu_i - \sum_{i=1}^{N} \mu_i / N\right\}^2}{N} + \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \hat{y}_{it})^2}{N}$$

(3) Modified ("Two-round") Generalized Least Squares (MGLS): We calculated generalized least squares estimates of the regression coefficients, using estimates of the error variance-covariance matrix Ω in equation (2.4) based on r_{LSDV}.

(4) Maximum Likelihood Constrained (MLC): Maximum likelihood estimates of the parameters β_0, β_1, β_2, ρ and σ^2 can be found by maximizing the log likelihood function:

$$L = -\frac{NT}{2} \log 2\pi - 1/2 \log |\Omega| - 1/2\, u' \Omega^{-1} u.$$

The constraints are that $0 \leq \rho \leq 1$ and $\sigma^2 > 0$.

(5) Maximum Likelihood Unconstrained (MLU): Maximum likelihood estimates were obtained as above, except that constraints on ρ or σ^2 were not imposed.

An initial set of parameter estimates must be provided to find the ML estimates in both methods (4) and (5). We compared the performance of two types of starting values for five different parameter combinations (a total of 500 data sets) using unconstrained ML: the LSDV estimates and the true values used to generate the data. The two types of initial estimates produced nearly identical final estimates for the four combinations in which $\rho > 0$. For $\rho = 0$

the two sets of parameter estimates differed in only a handful of cases, and by a negligible amount. Therefore, because of the cost involved in obtaining the LSDV estimates, we used the true parameter values as starting estimates in all remaining ML estimations. We report only the results obtained from using this latter type of initial estimates.

Whereas Nerlove [14] used the Fletcher-Powell algorithm [3] programmed by Wells [16] to maximize L, we used the Gill-Murray algorithm [4] programmed by Wright [17]. Both algorithms are iterative procedures and are based on modified steepest descent methods of function minimization. Gill, Murray and their coworkers [5,6] have shown that the Gill-Murray algorithm converges more rapidly and more reliably than the Fletcher-Powell algorithm. However, when both converge, they report that the two algorithms give extremely similar estimates of the function optimum for a variety of functions.

Our treatment of constraints on parameter values for σ^2 and ρ departed markedly from Nerlove's [14]. Nerlove constrained σ^2 to be positive by maximizing L with respect to σ rather than σ^2. He imposed a nonnegativity constraint on ρ by equating it with $\sin^2\theta$ and maximizing L with respect to θ rather than ρ. As Nerlove acknowledges, this method of applying constraints causes L to have multiple maxima with respect to θ since $\sin\theta$ is a periodic function. Murray [12] warns against employment of trigonometric constraints. Such a procedure can increase the nonlinearity of the function being maximized and cause the matrix of second derivatives (which must be negative definite at the maximum of the likelihood function) to become singular or ill-conditioned.

The Gill-Murray algorithm that we have used for ML estimation utilizes a projection method of optimization that permits any feasible equality or inequality constraints to be imposed on parameter values. For a detailed discussion of this constrained optimization procedure, see [4]. This method does not increase the nonlinearity of the function being optimized or the number of local maxima.

To our knowledge there is no previous evidence indicating the magnitude of the effects of constraining $\hat{\rho}$ and $\hat{\sigma}^2$ on ML parameter estimates for the model we have simulated. Thus, we do not know whether the mean squared errors (MSE's) of the constrained estimates of ρ and σ^2 will be appreciably smaller than the unconstrained versions. Further, we do not know the effects of constraining $\hat{\rho}$ and $\hat{\sigma}^2$ on the quality of the estimates of β_1 and β_2. Finally, it is important to learn whether the poor performance of the ML method in Nerlove [14] results from the small-sample properties of ML estimation of this model or from the implementation of parameter constraints.

5. RESULTS

In analyzing the results of our simulation study three points are critical. First, we must evaluate the success and practicality of our implementation of the two ML methods, since they affect the credibility of our results and their potential usefulness. Second, we need to assess the quality of the two sets of ML estimates relative to each other and to estimates obtained by the least squares alternatives. Properties of estimates used to assess quality are their mean squared error (MSE), bias and variance. Third, we want to know in what way the quality of the ML estimates of β_1, β_2, σ^2 and ρ depends on the true parameter values.

With regard to the first point, we found our implementation of the ML methods to be both successful and practical. Not only did ML estimation converge to a solution for every data set, but also the time required for this was short. On the average the ML solution was found in four to ten iterations, depending on the particular combination of parameter values. The MLC and MLU methods required nearly identical numbers of iterations to converge. For both methods several more iterations were usually needed for high values of ρ, especially when ($\beta_1 = 0.8$, $\beta_2 = 0.5$). These higher numbers of iterations occur together with poor quality of the ML estimates of β_1, σ^2 and ρ, as described more fully later in this section.

Before comparing the MLC and MLU estimates with each other and with estimates obtained by least squares procedures, it is helpful to know which parameter combinations led to activation of constraints. Obviously for the cases in which no constraints were activated, the MLC and MLU estimates are identical. The constraints that σ^2 be positive and that ρ be less than or equal to one were never brought into play (cf. Nerlove [14]). However, the constraint that ρ be nonnegative was activated in about sixty percent of the cases in which either ($\rho = 0.0$) or ($\beta_1 = 0.8$, $\beta_2 = 0.5$, $\rho < 0.9$). Thus, the quality of the MLC and MLU estimates is unlikely to differ except for these parameter combinations.

Since β_1 and β_2, the coefficients of the lagged dependent and exogenous variables, respectively, are the primary interest of most social scientists, we begin our assessment of the two ML methods by comparing the overall MSE's of $\hat{\beta}_1$ and $\hat{\beta}_2$ obtained from the MLU, MLC, MGLS, LSDV and OLS methods (see Table 1). In terms of MSE the three most preferable methods are clearly MGLS and the two ML methods. The two ML methods lead to estimates with almost identical MSE's; for the parameter combinations where the two ML estimates differ, the MSE's of estimates are slightly smaller for the MLC method than for the MLU method. Overall the LSDV and OLS methods are worse than the MGLS and ML methods.

Table 1 about here

Across all methods the MSE for $\hat{\beta}_1$, the coefficient of the lagged dependent variable, is generally larger than that for $\hat{\beta}_2$, the coefficient of the exogenous variable. Furthermore, across combinations of β_1 and β_2, each of the five methods produces fairly constant MSE's for $\hat{\beta}_2$ but widely varying MSE's for $\hat{\beta}_1$. For all five methods the largest MSE for $\hat{\beta}_1$ occurs when the effect of the exogenous variable is small in comparison to the effect of the lagged endogenous variable.

The relative quality of the ML and MGLS estimates varies according to the size of the ratio of β_1, the coefficient of the lagged endogenous variable, to β_2, the coefficient of the exogenous variable. We find that ML is superior when the effect of the lagged endogenous variable is small in comparison to the effect of the exogenous variable, while MGLS is best when the opposite is true. As we report below, the dependence of the relative quality of the ML and MGLS estimates of regression coefficients on the relative effects of β_1 and β_2 becomes even more apparent when the simulation results are not aggregated over values of ρ.

We now turn our attention to a more detailed examination of the performance of the MLU and MLC estimates, contrasted to each other and to the best of the least squares methods, MGLS. First we use the measure:

$$\% \text{ bias}(\hat{\theta}) = 100\% * \text{bias}(\hat{\theta})/\theta.$$

For both β_1 and β_2 the % biases of the MLC and MLU estimates are very similar across all parameter combinations (see Tables 2 and 3, respectively). For $\rho = 0.0$, the MLC estimates of β_1 and β_2 have slightly larger biases than the MLU estimates. The reverse is true for $\rho \geq 0.25$, but again the differences are small.

Tables 2 and 3 about here

Both the ML and MGLS methods display consistently low % biases in $\hat{\beta}_2$ across all parameter combinations. However, both methods of estimation produce widely varying % biases in $\hat{\beta}_1$. As might be expected from the large MSE of $\hat{\beta}_1$ when the effect of the exogenous variable is relatively small ($\beta_1 = 0.8$, $\beta_2 = 0.5$) (see Table 1), the ML estimates of $\hat{\beta}_1$ have upward biases approaching 25% for this parameter combination. For each combination of β_1 and β_2 the % bias in ML estimates of β_1 tends to become worse as ρ increases. However, for the first two combinations of β_1 and β_2 there is a downturn in the % bias for very high values of ρ. On the other hand, the MGLS estimates of β_1 are downwardly biased for low values of ρ but the % bias increases monotonically as ρ increases, approaching a negligible % bias for $\rho = 0.9$.

The MSE's of $\hat{\beta}_1$ and $\hat{\beta}_2$ depend on the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ (see Tables 4 and 5, respectively), as well as their biases. The variances of the MLU and MLC estimates of both β_1 and β_2 differ only slightly, the greatest differences occurring for estimates of β_1 when $\rho = 0.0$. In these cases the variance of $\hat{\beta}_1$ is smaller for MLC than MLU.

Tables 4 and 5 about here

The variances of the ML and MGLS estimates of β_1 and β_2 are also fairly comparable. For some parameter combinations these variances are smaller for ML than MGLS, for others the opposite is true. For both ML and MGLS estimates, the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$

are usually the smallest for the highest values of ρ.

By comparing Tables 2 and 4 we can learn whether there are compensating trade-offs between size of bias and size of variance for the ML and MGLS estimates of β_1.[2] For the ML estimates of β_1, large % bias and large variance are not systematically associated. In fact, the very sizeable % bias in $\hat{\beta}_1$ found with the ($\beta_1 = 0.8$, $\beta_2 = 0.5$) parameter combinations are accompanied by very small variances in $\hat{\beta}_1$. In contrast, for MGLS estimates of β_1, both % bias and variance tend to increase in magnitude as ρ decreases. Thus there does appear to be a certain degree of compensating trade-off between bias and variance in ML estimates of β_1, but not for MGLS.

The qualities of estimates of ρ are of special interest both because of social scientific concern with the serial correlation of omitted causal variables over time, and because of methodological concern with occasional negative estimates of ρ when MLU is the method of estimation. In both MLU and MLC estimates of ρ the biases are usually negative and very similar (see Table 6). The magnitude of the bias in $\hat{\rho}$ is somewhat smaller for MLC than for MLU when ($\beta_1 = 0.8$, $\beta_2 = 0.5$, $0.25 \leq \rho \leq 0.75$) but slightly larger for MLC than for MLU when $\rho = 0.0$. The size of the bias in ML estimates of ρ tends to increase as ρ increases; however, for two of the three combinations of the regression coefficients, there is a downturn in the bias in $\hat{\rho}$ as the value of ρ becomes very large.[3]

Table 6 about here

The ML and MGLS methods perform optimally at opposite ends of ρ continuum. Whereas ML estimates of ρ are almost always downwardly[4] biased, the MGLS estimates of ρ are almost always upwardly biased. And, as we found in our examination of the % biases of $\hat{\beta}_1$, the performance of the ML method tends to be best when MGLS is at its worst, and vice versa. Thus, we find that while the bias in ML estimates of ρ is greatest for high values of ρ and least for low values of ρ, just the opposite is true for MGLS. The MGLS estimates of ρ are most biased when ρ is near zero and least biased when ρ is near unity.

Nonetheless, the ML and MGLS methods have two obvious similarities: (1) there is an inverse relationship between % bias in $\hat{\beta}_1$ and bias in $\hat{\rho}$, and (2) absolute values of the biases in $\hat{\beta}_1$ and $\hat{\rho}$ are positively associated. These similarities are curious becuase the ML and MGLS methods have opposite signs to the biases of their estimates of ρ and of β_1. Though the two methods differ dramatically in their tendencies to attribute stability in the dependent variable to serial correlation of residuals for individual units rather than to inertia in the dependent variable, for both methods there are compensating effects. That is, for both methods error in one direction in estimating the strength of serial correlation of residuals is accompanied by error in the opposite direction in estimating the strength of the lagged endogenous variable.

6. CONCLUSION

The Monte Carlo experiments reported in this paper provide evidence on several questions of interest to social scientists planning to estimate a variance components model containing a lagged endogenous variable and an exogenous variable when an individual's errors over time have the intra-class correlation coefficient ρ .

The first question for which our research has implications is the following: Under what circumstances--if any--should maximum likelihood (ML) estimation be used? We assume that cost and reliability of the estimation technique and the quality of the estimates (in terms of their mean squared error, bias, and variance) are the most important criteria for selecting an estimation technique.

With our implementation of ML, both constrained and unconstrained ML estimates of parameters were obtained rapidly. Thus we find little difference in the practicality of using ML rather than a two-stage least squares procedure, such as modified generalized least squares (MGLS).

More importantly, we find that under certain circumstances ML estimates of the parameters of most interest to social scientists (i.e., the regression coefficients and ρ) were superior in quality to those obtained by MGLS. Overall both ML and MGLS estimates had smaller mean squared errors than estimates obtained by ordinary least squares or by least squares with individual constants (the other least squares methods that we studied). Our results indicate that ML estimation is a better choice than MGLS when ρ is small, but that MGLS is better than ML when ρ is large, especially when the effect of the lagged endogenous variable is large relative to the effect of the exogenous variable.

Since in practice social scientists do not know the true parameter values, they cannot know which estimation technique is perferable for their particular problem. In this situation there are good reasons for using both ML and MGLS to estimate parameters in a variance components model.

First of all, by using both methods, investigators should be able to determine the approximate range within which the true parameter values fall. This is because the biases of ML and MGLS estimates of parameters whose biases vary greatly with the true parameter values (namely, ρ and the coefficient of the lagged endogenous variable) are almost always opposite in sign. Thus, $\hat{\rho}$ is downwardly biased with ML but upwardly biased with MGLS, while $\hat{\beta}_1$ is upwardly biased with ML but downwardly biased with MGLS. Usually the true values of ρ and β_1 will lie between their ML and MGLS estimates. Of course, these guides, like all our results, undoubtedly depend upon proper specification of the model used in estimation.

Second, a comparison of the ML and MGLS estimates provides information that can be useful in judging whether the model assumed in estimation is correct. Our results indicate that for correctly specified models the ML and MGLS estimates of the coefficient of the exogenous variable should be nearly identical. This suggests that if the ML and MGLS estimates of β_2 differ greatly, the model may be misspecified.

Third, a comparison of the ML and MGLS estimates provides information that may be useful in judging which set of estimates is probably better and thus deserves greater weight. Our results suggest that: (1) The smaller the MGLS estimate of ρ, the more weight should be given to the ML estimates of the regression coefficients. (2) The smaller the difference between the ML and MGLS estimates of ρ, the more likely that the effect of the exogenous variable is large relative to the effect of the lagged endogenous variable, and that ML estimates have high quality. Or, to state it differently, the larger the difference between the ML and MGLS estimates of ρ, the more likely that the lagged endogenous variable has a strong effect relative to that of the exogenous variable, and that ML estimates are poor in quality. Further research is needed to determine whether these two relationships hold more generally.

Our results also have implications for a second question that investigators using ML estimation may ask: What conclusions should be drawn from obtaining a boundary solution for ρ in constrained ML estimation or a negative estimate for ρ in unconstrained ML estimation?

We find that a boundary solution for ρ in constrained ML (or a negative estimate of ρ in unconstrained ML) is not associated with poor quality in the ML estimates of regression coefficients. In fact, when ρ is truly zero, ML not only leads to a boundary estimate of ρ in about sixty percent of our simulated data sets, but also produces estimates of regression coefficients that have very small biases and variances, and are more accurate than the MGLS estimates. On the other hand, when there is a boundary solution for ρ and the true value of ρ is large, the ML estimate of the coefficient of the lagged endogenous variable tends to be poor in quality. Consequently, our results suggest that a boundary solution for the ML estimate of ρ is not a good clue to the quality of the ML estimates of the parameters in the model studied here. Better clues about estimator quality can probably be obtained by comparing ML and MGLS estimates of regression coefficients and of ρ, as outlined above.

Investigators who have chosen to use ML estimation of a variance components model will want to know the implication of our research for a final question: Should natural constraints on parameters be imposed? We find, as did Nerlove [14], that in practice only the nonnegativity constraint on ρ is at issue because other natural constraints are never violated.

Our results show that in terms of the mean squared error of $\hat{\beta}_1$ and $\hat{\beta}_2$, ML estimation with constraints on ρ has a slight advantage over that without constraints. Clearly constrained ML estimation gives more reasonable estimates of ρ because it prevents $\hat{\rho}$ from having a negative value, which is contrary to the assumptions of the model. In addition, the constrained ML estimates of the regression coefficients always have a smaller variance than the unconstrained ones, and this usually compensates for occasionally larger biases in the constrained estimates. Still, the differences between the constrained and unconstrained ML estimates are never large--and always negligible for those parameter combinations in which ML estimates are superior in quality to MGLS estimates. Consequently, our research provides no evidence that omitting constraints on ρ will seriously damage the quality of ML estimates of regression coefficients in the model.

REFERENCES

(1) Balestro, Pietro and Nerlove, Marc, "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, 34 (1966), 585-612.

(2) Box, M. J., "A Comparison of Several Current Optimization Methods, and The Use of Transformations in Constrained Problems," Computer Journal, 9, (January 1966), 67-77.

(3) Fletcher, R. and Powell, M. J. D., "A Rapidly Convergent Descent Method for Minimization," Computer Journal, 6 (July 1963), 163-68.

(4) Gill, P. E. and Murray, W., "Two Methods for the Solution of Linearly Constrained and Unconstrained Optimization Problems," National Physical Laboratory Report NAC 25 (November 1972).

(5) Gill, P. E.; Murray, W.; and Picken, S. M., "The Implementation of Two Modified Newton Algorithms for Unconstrained Optimization," National Physical Laboratory Report NAC 24 (August 1972).

(6) Gill, P. E.; Murray, W.; and Pitfield, R. A., "The Implementation of Two Revised Quasi-Newton Algorithms for Unconstrained Optimization," National Physical Laboratory Report NAC 11 (April 1972).

(7) Hannan, M. T. and Nielsen, F., "The Expansion of National Educational Systems: Tests of a Population Ecology Model," unpublished manuscript.

(8) Hannan, M. T. and Young, A. A., "Estimation in Panel Models: Results on Pooling Cross-Sections and Time Series," forthcoming in Sociological Methodology 1977.

(9) Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, Vol. 2, London: Charles Griffin, 1961, 36-46.

(10) Knuth, D. E., *The Art of Computer Programming*, Vol. 2, Reading, Mass.: Addison-Wesley Publishing Co., 1969.

(11) Maddala, G. S. and Mount, T. D., "A Comparative Study of Alternative Estimators for Variance Components Models Used in Econometric Applications," *Journal of the American Statistical Association*, 68 (June 1973), 324-28.

(12) Murray, W., ed., *Numerical Methods for Unconstrained Optimization*, New York: Academic Press, 1972, ch. 7.

(13) Nerlove, M., "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Economic Studies Quarterly*, 18 (December 1967), 42-74.

(14) Nerlove, M., "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, 39 (March 1971), 359-82.

(15) Nerlove, M., and Schultz, H., "Love and Life Between the Censuses," in A. S. Goldberger and O. D. Duncan, eds., *Structural Equation Models in The Social Sciences*, New York: Seminar Press, 1973.

(16) Wells, M., "Function Minimization," Algorithm 251, in *Collected Algorithms from CACM*, New York: Association for Computing Machinery, 1967.

(17) Wright, M., "LCMNA: A Set of FORTRAN Subroutines for Minimization of a Set of Non-Linear Functions Subject to Linear Inequality or Equality Constraints," Stanford University: Stanford Computation Center Documentation, 1975.

FOOTNOTES

[1] Balestra and Nerlove [1] used ML on data on natural gas demand to estimate a variance components model with both a lagged endogenous and an exogenous variable. They also concluded that ML estimates were poor in quality, but their conclusion may result from a poor fit between the model and their data rather than their use of ML.

[2] There is so little variation in the bias of $\hat{\beta}_2$ across parameter combinations that a similar comparison of Tables 3 and 5 is not very informative.

[3] Not surprisingly the variance of $\hat{\rho}$ is smaller for MLC than MLU for parameter combinations in which the nonnegativity constraint on ρ was frequently activated, i.e., for $\rho = 0.0$ and ($\beta_1 = 0.8$, $\beta_2 = 0.5$). The variance of the ML estimates of ρ and the magnitude of the bias in $\hat{\rho}$ appear to covary positively across values of ρ for each combination of β_1 and β_2. Our results indicate that the bias and variance of the ML estimates of σ^2 vary across parameter combinations in a way closely parelleling the results for the ML estimates of ρ.

[4] By MGLS estimates of ρ we mean the estimates of ρ used in the second stage of the MGLS estimation, calculated from the results of the LSDV regression, as described in Section 4 above.

TABLE 1

Mean Squared Error of Estimates

(Cases averaged over all values of ρ; each entry based on 500 sets of estimates)

	$\hat{\beta}_1$	$\hat{\beta}_2$
$\beta_1 = 0.3,\ \beta_2 = 1.0$		
MLU	.175*	.182
MLC	.169	.182
MGLS	.226	.194
LSDV	.822	.239
OLS	6.449	.348
$\beta_1 = 0.8,\ \beta_2 = 1.0$		
MLU	.722	.199
MLC	.720	.199
MGLS	.146	.198
LSDV	.748	.228
OLS	1.592	.341
$\beta_1 = 0.8, \beta_2 = 0.5$		
MLU	2.415	.218
MLC	2.352	.208
MGLS	.925	.194
LSDV	3.865	.220
OLS	2.420	.228

*All entries in this table have been multiplied by 10^2.

TABLE 2

Percent Bias in $\hat{\beta}_1$

(Each entry based on 100 sets of estimates)

ρ = :	0.0	0.25	0.5	0.75	0.9
$\beta_1 = 0.3, \beta_2 = 1.0$					
MLU	0.0%*	3.7	3.9	2.3	0.9
MLC	-1.2	3.7	3.9	2.3	0.9
MGLS	-22.4	-12.2	-5.9	-1.5	-0.1
$\beta_1 = 0.8, \beta_2 = 1.0$					
MLU	0.2%	9.3	13.6	14.2	6.4
MLC	-0.8	9.2	13.6	14.2	6.4
MGLS	-7.3	-2.4	.3	1.9	1.7
$\beta_1 = 0.8, \beta_2 = 0.5$					
MLU	-0.4	16.7	21.4	23.4	23.9
MLC	-1.5	15.8	21.0	23.3	23.9
MGLS	-19.0	-12.4	-7.7	-1.8	1.9

*All entries in this table have been rounded off to the nearest tenth of a percent.

TABLE 3

Percent Bias in $\hat{\beta}_2$

(Each entry based on 100 sets of estimates)

ρ= :	0.0	0.25	0.5	0.75	0.9
$\beta_1 = 0.3$, $\beta_2 = 1.0$					
MLU	-0.0%*	-0.0	0.1	0.1	-0.0
MLC	0.1	-0.0	0.1	0.0	-0.0
MGLS	-0.3	-0.4	-0.3	-0.2	-0.1
$\beta_1 = 0.8$, $\beta_2 = 1.0$					
MLU	0.0	-1.6	-1.6	-0.4	0.4
MLC	0.1	-1.6	-1.6	-0.4	0.4
MGLS	-0.1	-0.3	-0.2	0.0	0.0
$\beta_1 = 0.8$, $\beta_2 = 0.5$					
MLU	0.0	-2.5	-3.1	-2.8	-1.5
MLC	0.2	-2.0	-2.6	-2.6	-1.5
MGLS	-0.9	-1.4	-1.1	-0.5	0.0

*All entries in this table have been rounded off to the nearest tenth of a percent.

TABLE 4

Variance of $\hat{\beta}_1$

(Each entry based on 100 sets of estimates)

ρ = :	0.0	0.25	0.5	0.75	0.9
$\beta_1 = 0.3$, $\beta_2 = 1.0$					
MLU	.168*	.197	.165	.087	.035
MLC	.133	.195	.165	.087	.035
MGLS	.146	.141	.117	.072	.033
$\beta_1 = 0.8$, $\beta_2 = 1.0$					
MLU	.061	.055	.046	.076	.084
MLC	.055	.054	.046	.076	.084
MGLS	.078	.087	.074	.047	.021
$\beta_1 = 0.8$, $\beta_2 = 0.5$					
MLU	.133	.043	.019	.009	.006
MLC	.116	.041	.019	.009	.006
MGLS	.221	.247	.211	.139	.074

*All entries in this table have been multiplied by 10^2.

TABLE 5.

Variance of $\hat{\beta}_2$

(Each entry based on 100 sets of estimates)

ρ = :	0.0	0.25	0.5	0.75	0.9
$\beta_1 = 0.3$, $\beta_2 = 1.0$					
MLU	.283*	.272	.202	.107	.044
MLC	.285	.272	.202	.107	.044
MGLS	.330	.282	.203	.107	.044
$\beta_1 = 0.8$, $\beta_2 = 1.0$					
MLU	.284	.267	.211	.128	.050
MLC	.287	.266	.211	.128	.052
MGLS	.338	.290	.208	.109	.045
$\beta_1 = 0.8$, $\beta_2 = 0.5$					
MLU	.281	.233	.166	.098	.052
MLC	.282	.238	.169	.097	.052
MGLS	.331	.281	.199	.105	.044

*All entries in this table have been multiplied by 10^2.

TABLE 6.

Bias of $\hat{\rho}$

(Each entry based on 100 sets of estimates)

ρ = :	0.0	0.25	0.5	0.75	0.9
$\beta_1 = 0.3$, $\beta_2 = 1.0$					
MLU	-.006	-.040	-.048	-.025	-.008
MLC	.017	-.040	-.048	-.025	-.008
MGLS	.254	.215	.145	.064	.021
$\beta_1 = 0.8$, $\beta_2 = 1.0$					
MLU	-.007	-.169	-.345	-.398	-.101
MLC	.016	-.166	-.344	-.398	-.101
MGLS	.325	.320	.219	.092	.027
$\beta_1 = 0.8$, $\beta_2 = 0.5$					
MLU	-.005	-.273	-.519	-.728	-.766
MLC	.017	-.242	-.491	-.718	-.766
MGLS	.445	.477	.340	.156	.047