No. 8407

Small Sample Efficiency Gains From A First Observation Correction for Hatanaka's Estimator Of The Lagged Dependent Variable-Serial Correlation Regression Model

by Thomas B. Fomby

Research Paper

Federal Reserve Bank of Dallas

No. 8407

Small Sample Efficiency Gains From A First Observation Correction for Hatanaka's Estimator Of The Lagged Dependent Variable-Serial Correlation Regression Model

by Thomas B. Fomby

September 1984

SMALL SAMPLE EFFICIENCY GAINS FROM A FIRST OBSERVATION CORRECTION FOR HATANAKA'S ESTIMATOR OF THE LAGGED DEPENDENT VARIABLE-SERIAL CORRELATION REGRESSION MODEL

Thomas B. Fomby
Department of Economics
Southern Methodist University
Dallas, TX 75275*

Abstract: Evidence presented by Fomby and Guilkey (1983) suggests that Hatanaka's estimator of the coefficients in the lagged dependent variable-serial correlation regression model performs poorly, not because of poor selection of the estimate of the autocorrelation coefficient, but because of the lack of a first observation correction. This study conducts a Monte Carlo investigation of the small sample efficiency gains obtainable from a first observation correction suggested by Harvey (1981). Results presented here indicate that substantial gains result from the first observation correction. However, in comparing Hatanaka's procedure with first observation correction to maximum likelihood search, it appears that ignoring the determinantal term of the full likelihood function causes some loss of small sample efficiency. Thus, when computer costs and programming constraints are not binding, maximum likelihood search is to be recommended. In contrast, users who have access to only rudimentary least squares programs would be well served when using Hatanaka's two-step procedure with first observation correction because of the ease of calculating consistent standard errors of the estimates.

* This research was supported by the Federal Reserve Bank of Dallas. Mr. Frank Berger of the Research Department wrote the computer program for this Monte Carlo study. Frank Berger, Andrew Harvey and Jim Pearce provided useful comments on the initial draft of this paper. Any opinions expressed herein are my own and are not necessarily those of the Federal Reserve Bank of Dallas of of the Federal Reserve System. Any errors that may exist are, of course, my responsibility.

SMALL SAMPLE EFFICIENCY GAINS FROM A
FIRST OBSERVATION CORRECTION FOR HATANAKA'S ESTIMATOR

1. Introduction

Hatanaka (1974) proposed a two-step estimator of the coefficients of a regression model with lagged dependent variables and serially correlated errors. Hatanaka's estimator, unlike a two-step feasible generalized least squares estimator studied by Wallis (1972), is asymptotically efficient. This follows since Hatanaka's estimator is the two-step Gauss-Newton estimator where the first observation is omitted (Harvey, 1981). Though asymptotically efficient, the small sample performance of Hatanaka's estimator leaves much to be desired relative to asymptotically inefficient two-step estimators (Hatanaka, 1974). Evidence presented by Fomby and Guilkey (1983) suggests that Hatanaka's estimator performs poorly, not because of poor selection of the estimate of the autocorrelation coefficient in the first step, but because of the lack of a first observation correction. This study conducts a Monte Carlo investigation of the small sample efficiency gains obtainable from a first observation correction for the Hatanaka procedure suggested by Harvey (1981). The competing estimators examined are maximum likelihood search, Hatanaka's estimator, first with and then without a first observation correction, and the asymptotically inefficient two-step estimator investigated by Wallis (1967, 72) and Maddala (1971).

The outline of this paper is as follows. The details of the competing estimators are presented in the next section while the design of the Monte Carlo experiment is described in Section 3. The results of the experiment are summarized in Section 4. Conclusions are developed in the final section.

2. Description of Competing Estimators

For expository convenience the model under consideration is chosen to be

$$y_t = \beta x_t + \alpha y_{t-1} + e_t \tag{1}$$

$$e_t = \rho e_{t-1} + u_t, \; t = 2, 3, \cdots, T, \tag{2}$$

where x_t is nonstochastic, $|\alpha|<1$, $|\rho|<1$, and the u_t are independent and identically distributed normal random errors having zero mean and finite variance σ_u^2, the initial value of u_t being realized in the infinite past. The estimators to be discussed are easily derivable for more general models involving additional lagged dependent and exogenous variables.

Method 1 (Maximum Likelihood Search)

The log likelihood function of model (1)-(2) is

$$L(\beta, \alpha, \rho, \sigma_u^2 \mid \underset{\sim}{y}, X) = -\frac{(T-1)}{2}\ln(2\pi) - \frac{(T-1)}{2}\ln \sigma_u^2 + \frac{1}{2}\ln(1-\rho^2) - \frac{1}{2\sigma_u^2}(\underset{\sim}{y}^* - X^*\underset{\sim}{\beta})'(\underset{\sim}{y}^* - X^*\underset{\sim}{\beta}) \quad , \tag{3}$$

where $\underset{\sim}{y}' \equiv (y_2, y_3, \cdots, y_T)$, $\underset{\sim}{\beta}' = (\beta, \alpha)$,

$$X' = \begin{bmatrix} x_2 & x_3 \cdots & x_T \\ y_1 & y_2 \cdots & y_{T-1} \end{bmatrix} \quad ,$$

$\underset{\sim}{y}^* \equiv P\underset{\sim}{y}$, $X^* = PX$, and P is the Prais-Winsten (1954) matrix

$$P = \begin{bmatrix} (1-\rho^2)^{\frac{1}{2}} & 0 & \cdot & \cdot & \cdot & 0 \\ -\rho & 1 & 0 & \cdot & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 \;\; -\rho & 1 \end{bmatrix} \tag{4}$$

of dimension (T-1) x (T-1). Maximizing (3) partially with respect to $\underset{\sim}{\beta}$ and σ_u^2 produces the conditional estimators

$$\hat{\underset{\sim}{\beta}}(\rho) = (X^{*\prime}X^*)^{-1}X^{*\prime}\underset{\sim}{y}^* \tag{5}$$

$$\hat{\sigma}_u^2(\rho) = \frac{1}{(T-1)}\ [\underset{\sim}{y}^* - X^*\hat{\underset{\sim}{\beta}}(\rho)]'[\underset{\sim}{y}^* - X^*\hat{\underset{\sim}{\beta}}(\rho)]\ , \tag{6}$$

where the notation $\hat{\underset{\sim}{\beta}}(\rho)$ and $\hat{\sigma}_u^2(\rho)$ emphasizes that they are functions of the parameter ρ.

Substituting (5) and (6) into (3) produces the concentrated log likelihood function

$$L^*(\rho|\underset{\sim}{y}, X) = -\frac{(T-1)}{2}\ [\ln(2\pi) + 1] - \frac{(T-1)}{2}\ \ln\left[\frac{\sigma_u^2(\rho)}{(1-\rho^2)^{1/(T-1)}}\right]\ . \tag{7}$$

Thus, the maximum likelihood estimates of $\underset{\sim}{\beta}$, σ_u^2, and ρ are obtained as follows. Grid the admissable range of ρ, (-1,1), by points ρ_i, i=1, 2, ···, n. For each ρ_i evaluate the function $\sigma_u^2(\rho)/(1-\rho^2)^{1/(T-1)}$ and choose the estimator of ρ, say $\hat{\rho}$, which provides the global minimum. The maximum likelihood estimates of $\underset{\sim}{\beta}$ and σ_u^2 are then obtained by evaluating (5) and (6) at $\rho = \hat{\rho}$. That is, maximum likelihood estimates of $\underset{\sim}{\beta}$ and σ_u^2 are $\hat{\underset{\sim}{\beta}}(\hat{\rho})$ and $\hat{\sigma}_u^2(\hat{\rho})$, respectively.

Since $\lim_{T\to\infty}(1-\rho^2)^{1/(T-1)} = 1$, the estimates of ρ, $\underset{\sim}{\beta}$, and σ_u^2 obtained by minimizing the transformed sum of squared errors

$$\begin{aligned} S(\beta,\ \alpha,\ \rho) = {} & ((1-\rho^2)^{\frac{1}{2}}y_2 - \hat{\beta}(\rho)(1-\rho^2)^{\frac{1}{2}}x_2 - \hat{\alpha}(\rho)(1-\rho^2)^{\frac{1}{2}}y_1)^2 \\ & + \sum_{t=3}^{T}[(y_t-\rho y_{t-1}) - \hat{\beta}(\rho)(x_t-\rho x_{t-1}) - \hat{\alpha}(\rho)(y_{t-1}-\rho y_{t-2})]^2, \end{aligned} \tag{8}$$

where $(\hat{\beta}(\rho),\ \hat{\alpha}(\rho))' = \hat{\underset{\sim}{\beta}}(\rho)$ in equation (5), have the same asymptotic distribution as the maximum likelihood estimates and, thus, are likewise asymptotically efficient. In addition, ignoring the initial observation $(y_2,\ x_2,\ y_1)$

makes no difference, asymptotically, as its contribution to the sum of squared errors $S(\beta, \alpha, \rho)$ is inconsequential in infinite samples. Choosing to minimize the sum of squared errors function

$$S_o(\beta, \alpha, \rho) = \sum_{t=3}^{T} [(y_t - \rho y_{t-1}) - \hat{\beta}(\rho)(x_t - \rho x_{t-1}) - \hat{\alpha}(\rho)(y_{t-1} - \rho y_{t-2})]^2 \quad (9)$$

thus provides estimates of ρ, $\underset{\sim}{\beta}$ and σ_u^2 which are also asymptotically efficient.

Method 2 (Hatanaka Estimator with First Observation Correction)

Before proceeding to the discussion of the Hatanaka estimator, a brief discussion of the Gauss-Newton estimation scheme will prove useful. Suppose that the minimization of the sum of squares function $S(\underset{\sim}{\psi}) = \sum_t u_t^2$, where $u_t = u_t(\underset{\sim}{\psi})$ is a "residual" which depends on the value taken by $\underset{\sim}{\psi}$, provides an estimator for the parameter vector $\underset{\sim}{\psi}$. The two-step Gauss-Newton estimator of $\underset{\sim}{\psi}$, say $\underset{\sim}{\psi}^*$, is obtained by computing

$$\underset{\sim}{\psi}^* = \hat{\underset{\sim}{\psi}} - \left[\sum_t \frac{\partial u_t}{\partial \underset{\sim}{\psi}} \cdot \frac{\partial u_t}{\partial \underset{\sim}{\psi}'}\right]^{-1} \sum_t \frac{\partial u_t}{\partial \underset{\sim}{\psi}} \cdot u_t \quad , \qquad (10)$$

where $\hat{\underset{\sim}{\psi}}$ is an initial consistent estimate of $\underset{\sim}{\psi}$ and the gradient vector $(\partial u_t/\partial \underset{\sim}{\psi})$ and u_t are evaluated at $\underset{\sim}{\psi} = \hat{\underset{\sim}{\psi}}$. Should the minimization of $S(\underset{\sim}{\psi})$ with respect to $\underset{\sim}{\psi}$ provide asymptotically efficient estimates of $\underset{\sim}{\psi}$, the Gauss-Newton estimator, $\underset{\sim}{\psi}^*$, will likewise be asymptotically efficient.

With respect to the lagged dependent variable-serial correlation model (1)-(2), let

$$u_2 = (1-\rho^2)^{\frac{1}{2}} y_2 - \beta(1-\rho^2)^{\frac{1}{2}} x_2 - \alpha(1-\rho^2)^{\frac{1}{2}} y_1 \qquad (11)$$

and

$$u_t = (y_t - \rho y_{t-1}) - \beta(x_t - \rho x_{t-1}) - \alpha(y_{t-1} - \rho y_{t-2}) \qquad (12)$$

$$t = 3, 4, \cdots, T.$$

Harvey (1981, pp. 269-270) has shown that the application of the two-step Gauss-Newton method to the sum of squares function

$$S_o(\underset{\sim}{\psi}) = \sum_{t=3}^{T} u_t^2 \quad , \tag{13}$$

where $\psi' = (\beta, \alpha, \rho)$, is numerically equivalent to Hatanaka's estimator (1974). Note the first observation is not treated here. In light of this correspondence, Harvey (1981, pp. 270-271) suggested a treatment of the first observation by applying the two-step Gauss-Newton method to the sum of squares function

$$S(\psi) = \sum_{t=2}^{T} u_t^2 \quad . \tag{14}$$

This treatment is described in the following steps.

The Hatanaka estimator of (β, α) using a first observation correction in the model (1)-(2) is obtained in the following manner:

1. Form the instrumental variable matrix

$$Z' = \begin{bmatrix} x_2 & x_3 & \cdots & x_T \\ & & & \\ x_1 & x_2 & \cdots & x_{T-1} \end{bmatrix}$$

   where x_{t-1} serves as an instrument for y_{t-1}.
   Compute the instrumental variable estimator of $(\beta, \alpha)'$

$$\underset{\sim}{b} = (Z'X)^{-1} Z'\underset{\sim}{y} = (b_1, b_2)' \quad . \tag{15}$$

   This estimator is consistent.

2. Compute the residuals $\hat{\underset{\sim}{e}}' = (\hat{e}_2, \hat{e}_3, \cdots \hat{e}_T) = (\underset{\sim}{y} - X\underset{\sim}{b})'$ and construct a consistent estimator of ρ,

$$\tilde{\rho} = (\sum_{t=3}^{T} \hat{e}_t\hat{e}_{t-1} / \sum_{t=3}^{T} \hat{e}_{t-1}^2). \tag{16}$$

3. Form the regressand vector

$$\hat{\underset{\sim}{y}}^* = [(1-\tilde{\rho}^2)^{\frac{1}{2}} y_2,\ y_3 - \tilde{\rho} y_2,\ \cdots,\ y_T - \tilde{\rho} y_{T-1}]' \tag{17}$$

and the matrix of regressors

$$X_e' = \begin{bmatrix} (1-\tilde{\rho}^2)^{\frac{1}{2}} x_2 & x_3 - \tilde{\rho} x_2 & \cdots & x_T - \tilde{\rho} x_{T-1} \\ (1-\tilde{\rho}^2)^{\frac{1}{2}} y_1 & y_2 - \tilde{\rho} y_1 & \cdots & y_{T-1} - \tilde{\rho} y_{T-2} \\ \tilde{\rho} \hat{e}_2 / (1-\tilde{\rho}^2)^{\frac{1}{2}} & \hat{e}_2 & \cdots & \hat{e}_{T-1} \end{bmatrix} . \tag{18}$$

4. The Hatanaka estimates of β and α, say $\tilde{\beta}$ and $\tilde{\alpha}$, <u>using a first observation correction</u>, are obtained by computing the least squares estimates

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\alpha} \\ \bar{\rho} \end{pmatrix} = (X_e' X_e)^{-1} X_e' \hat{\underset{\sim}{y}}^* \quad . \tag{19}$$

The Hatanaka estimate of ρ, say $\hat{\hat{\rho}}$, <u>using a first observation correction</u> is $\hat{\hat{\rho}} = \bar{\rho} + \tilde{\rho}$.

<u>Method 3 (Hatanaka Estimator Without First Observation Correction)</u>

The Hatanaka estimator derived without a first observation correction is obtained by duplicating the steps for method 2 except deleting the first observation in the matrices $\hat{\underset{\sim}{y}}^*$ and X_e. This is the original form of the estimator suggested by Hatanaka (1974).

<u>Method 4 (Two-Step Feasible Generalized Least Squares Estimator Investigated by Wallis (1972))</u>

Wallis (1967) first presented the following consistent estimator of (ρ,α). Maddala (1971) provided a general proof of its asymptotic inefficiency while Wallis (1972), using the specific assumption that the x_t's follow a stationary autoregressive process, determined quantitatively the extent of the estimator's inefficiency. The estimator's asymptotic inefficiency results because of the lack of independence in the estimation of α and ρ as reflected by the nondiagonality of the information matrix of the model (1)-(2). That is, only efficient estimates of ρ will provide, via feasible generalized least squares, asymptotically efficient estimates of (β,α). Not just any consistent estimate of ρ will suffice.

1. Using the first two steps of Method 2, obtain the instrumental variable estimate of ρ, $\tilde{\rho}$, of equation (16).
2. Form the regressand vector $\hat{y}^*$ of step 3 of Method 2.
3. Form the matrix of regressors

$$\hat{X}^{*\prime} = \begin{bmatrix} (1-\tilde{\rho}^2)^{\frac{1}{2}} x_2 & x_3 - \tilde{\rho} x_2 & \cdots & x_T - \tilde{\rho} x_{T-1} \\ (1-\tilde{\rho}^2)^{\frac{1}{2}} y_1 & y_2 - \tilde{\rho} y_1 & \cdots & y_{T-1} - \tilde{\rho} y_{T-2} \end{bmatrix} \tag{20}$$

4. Calculate the estimates of β and α using

$$\begin{pmatrix} \beta^+ \\ \alpha^+ \end{pmatrix} = (\hat{X}^{*\prime}\hat{X}^*)^{-1}\hat{X}^{*\prime}\hat{y}^* \quad . \tag{21}$$

The estimates β^+ and α^+, though consistent, have been shown by Wallis (1972) to be asymptotically inefficient. However, in previous Monte Carlo studies (Hatanaka (1974) and Fomby and Guilkey (1983)), these estimates have performed quite well relative to the Hatanaka estimates derived without a first observation correction for sample sizes as large as T = 200.

In the next section, the design of a Monte Carlo experiment used to examine the relative efficiencies of these estimators and the contribution of a first observation correction for the Hatanaka procedure is discussed.

3. Design of Monte Carlo Experiment

The experimental design of the present Monte Carlo experiment consists of the following model:

$$y_t = \beta x_t + \alpha y_{t-1} + e_t \tag{22}$$

$$e_t = \rho e_{t-1} + u_t \tag{23}$$

$$x_t = \lambda x_{t-1} + v_t \tag{24}$$

where $|\alpha|<1$, $|\rho|<1$, $|\lambda|<1$, the u_t's are independent and identically distributed normal random errors with zero mean and variance σ_u^2, the v_t's are similarly defined with variance σ_v^2 and u_t and v_s are assumed to be mutually independent for all s and t. The autoregressive process on the x_t's has been introduced to allow the analysis of the effects of collinearity on estimator efficiency.

The design points were chosen to be:

$$\beta = 1.0,\ \alpha = \{0.2,\ 0.4,\ 0.6,\ 0.8\}$$

$$\rho = \{0.2,\ 0.4,\ 0.6,\ 0.8\},\ \lambda = \{0.0,\ 0.4,\ 0.8\}$$

$$R^2 = \{0.50,\ 0.75,\ 0.90.\ 0.99\}\ ,$$

$$T = \{15,\ 30,\ 50,\ 100\}\ ,$$

$$\sigma_x^2 = \sigma_v^2/(1-\lambda^2) = 1.0\ .$$

For given values of α, λ, ρ, and R^2, the value of σ_u^2 was chosen so that

$$\sigma_u^2 = \frac{(1-R^2)}{R^2} \cdot \frac{(1+\lambda\alpha)(1-\rho\alpha)(1-\rho^2)}{(1-\lambda\alpha)(1+\rho\alpha)}\ .$$

To maintain the spirit of "an infinite past starting point," the following sampling procedure was used to generate observations on y_t, x_t, and y_{t-1}. Two <u>independent</u> random realizations of u_t's and v_t's were generated by a random number generator such that $u_t \sim \text{niid}(0, \sigma_u^2)$ and $v_t \sim \text{niid}(0, \sigma_v^2)$ with $t = -49, -48, \cdots, 0, 1, 2, \cdots, T$.

With initial starting values of $x_{-50} = e_{-50} = y_{-50} = 0$, realizations of x_{-49}, $\cdots$, x_T, $e_{-49}, \cdots e_T$, and y_{-49}, $\cdots$, y_T were generated using the u_t's and v_t's and assumed values of λ, ρ, β, and α. The first 50 observations on (y_t, x_t, y_{t-1}) were then discarded and the observations associated with $t = 2, 3, \cdots, T$ were retained. The sample size is thus equal to T with (T-1) observations on the dependent and independent variables being available for the purpose of estimation.

Once the random sample was generated, the above four estimators were calculated (the grid width of the maximum likelihood search was chosen to be 0.01) and the following squared errors were computed: $(\hat{\beta}^{(i)}-\beta)^2$, $(\hat{\alpha}^{(i)}-\alpha)^2$, and $(\hat{\beta}^{(i)}-\beta)(\hat{\alpha}^{(i)}-\alpha)$ where $\hat{\beta}^{(i)}$ and $\hat{\alpha}^{(i)}$ represent the i-th method's estimates of β and α, i=1, 2, 3, 4. The completion of this task represents one iteration given a specific setting of the parameter values, β, α, ρ, λ, and R^2.

For methods 2, 3, and 4, nine hundred and ninety-nine additional random samples were drawn using the same parameter settings (a total of 1000 iterations), resulting in the accumulated values

$$\sum_{1}^{1000}(\hat{\beta}^{(i)}-\beta)^2/1000 \equiv a^{(i)}, \qquad \sum_{1}^{100}(\hat{\alpha}^{(i)}-\alpha)^2/1000 \equiv b^{(i)}$$

$$\sum_{1}^{1000}\frac{(\hat{\beta}^{(i)}-\beta)(\hat{\alpha}^{(i)}-\alpha)}{1000} \equiv c^{(i)}, \text{ and } d^{(i)} = a^{(i)} \cdot b^{(i)} - (c^{(i)})^2,$$

$$i = 2, 3, 4.$$

The value $d^{(i)}$ represents the empirical generalized mean square error of the i-th method given a specific setting of the parameter values. Because of computational cost, the maximum likelihood estimates were calculated only over the first 100 iterations. The measures $a^{(1)}$, $b^{(1)}$, and $c^{(1)}$ were thus averaged over the first 100 iterations.

Finally, for a specific parameter setting, the relative efficiencies of the second, third, and fourth estimators relative to maximum likelihood search were calculated as

$$RE_2 = \frac{d^{(2)}}{d^{(1)}}, \quad RE_3 = \frac{d^{(3)}}{d^{(1)}}, \quad RE_4 = \frac{d^{(4)}}{d^{(1)}} \quad .$$

A RE value greater than one represents inefficiency relative to maximum likelihood search whereas a value less than one represents relative efficiency.

The direct simulation method was used rather than some of the recently proposed efficient Monte Carlo methods [e.g. Makhail (1972, 1975) and Hendry and Harrison (1974)] for the reasons cited in Fomby and Guilkey (1983, p. 298).

4. Results of Monte Carlo Experiment

The results of the Monte Carlo experiment are summarized in Tables 1 and 2 below. These tables are aggregative in nature. More detailed tables are available from the author upon request.

Table 1 contains the relative efficiencies of the estimators averaged over all possible settings for λ, R^2, and T for given values of ρ and α. In each cell of this table, the three entries are (from top to bottom), RE_2, RE_3, and RE_4, the efficiencies of methods 2, 3, and 4, respectively, relative to maximum likelihood estimation. Several points are noteworthy.

1. Overall, the maximum likelihood search procedure performed best. There were several instances, however, when maximum likelihood estimation was not necessarily the best method, particularly in the regions of simultaneously low or high values of ρ and α . The small sample inefficiency incurred by maximum likelihood estimation was, nevertheless, generally quite small and, when efficient, gains more than compensated for losses elsewhere.

2. The first observation correction for Hatanaka's procedure improved its performance substantially. On average over all design points, the first observation correction yielded a 10.7% improvement in generalized mean square error.

3. The Hatanaka estimator without first observation correction exhibited uniformly poor performance. Only in rare instances did it minimally dominate maximum likelihood estimation for parameter settings not shown here (e.g. $\lambda = 0.0$, $R^2 = 0.75$, $T = 15$, $\alpha = 0.2$, and $\rho = 0.6$). For almost every possible configuration, it was dominated by the Hatanaka estimator with first observation correction and method 4.

4. The estimator investigated by Wallis (1967, 72), method 4, performed surprisingly well overall despite the fact that it is asymptotically inefficient. There were particular parameter settings where the Hatanaka estimator with first observation correction was superior to it but these instances were rare.

Table 2 contains the relative efficiencies of the estimators averaged over all values of ρ and α for given values of T, R^2 and λ. The entries in each cell of this table have the same interpretations as in Table 1. Points of interest are:

1. Increased sample size does not provide a quick matching of the performance of the Gauss-Newton methods (2 and 3) with maximum likelihood estimation.

2. As expected, the contribution of a first observation correction for Hatanaka's estimation procedure diminished with sample size though the benefit of the correction remained substantial even when $T = 100$.

3. The asymptotically inefficient method 4 performed very well for $T = 15$. In addition, the inefficiency in method 4 developed only slowly with increased sample size and was not great even when $T = 100$.

4. Consider aggregation over all T and λ. Increased R^2 tended to improve the performance of maximum likelihood estimation relative to the other methods.

5. Aggregating over T and R^2, increased collinearity ($\lambda \to 1.0$) seemed to improve the performance of the Hatanaka estimator with first observation correction and method 4 relative to maximum likelihood search. Also, increased collinearity tended to accentuate the gain offered by a first observation correction for Hatanaka's procedure.

5. Conclusions of Study

When estimating the coefficients of the lagged dependent variable - serial correlation model, there is no loss of asymptotic efficiency in ignoring the determinantal term $1/(1-\rho^2)^{1/(T-1)}$ in the concentrated log likelihood function and proceeding to apply the Gauss-Newton method to the transformed sum of squared errors. Recognition of Hatanaka's procedure as a Gauss-Newton method allows the development of a first observation correction. As in the previous autocorrelation literature (e.g. Beach and MacKinnon (1978) and Maeshiro (1979)), the results here indicate the importance of utilizing the first observation even in moderate to large sample sizes (T = 50 and 100). However, the first observation correction for Hatanaka's procedure is not a panacea. In the Monte Carlo results presented here, it appears that some loss of small sample efficiency arises when ignoring the determinantal term of the full likelihood function. Interestingly, these results parallel the conclusions obtained in the nonstochastic regressor model with autocorrelated errors (see Beach and MacKinnon (1978)).

Thus, in the absence of considerations of computational expense, the lagged dependent variable-serial correlation model should probably be estimated by the maximum likelihood method. However, maximum likelihood estimation can, when using a global search with grid width of 0.01, be almost 200 times as expensive as two-step methods. In addition, care should be used in calculating the asymptotic standard errors of the maximum likelihood coefficient estimates since the information matrix is not diagonal and the usual generalized least squares formula for the variance-covariance matrix overstates the significance of the estimates. See Dhrymes (1971, pp. 199-201) for a derivation of the correct variance-covariance matrix expression.

When relegated to a computer program with only least squares capabilities the problem becomes one of selecting among the three two-step methods presented here. Given the poor performance of Hatanaka's procedure without first observation correction, the choice narrows to Hatanaka's procedure with first observation correction and method 4, the asymptotically inefficient two-step feasible generalized least squares method. Though their performances are quite similar, with method 4 possibly having a slight edge, secondary considerations probably dictate the choice of Hatanaka's procedure with first observation correction. As with maximum likelihood estimation, the usual generalized least squares formula for the variance-covariance matrix of the coefficient estimates of method 4 are inappropriate and overstate the actual significance of the estimates. Thus, the variance-covariance matrix available with naive computer printouts is incorrect and instead the correct matrix formula that should be used is displayed in Dhrymes (1971, pp. 205-206). In contrast, the Hatanaka procedure, implemented by using ordinary least squares on the appropriate residual adjusted, transformed data, provides, automatically, consistent standard errors through the conventional ordinary least squares formula.

TABLE 1

Relative Efficiencies of Estimation

Averaged Over All Values of λ, R^2 and T

For Given ρ and α

ρ \ α	0.2	0.4	0.6	0.8	Row Average
0.2	0.9508 1.0364 0.9262	0.9897 1.0903 0.9565	1.0811 1.2252 1.0327	1.0714 1.2464 1.0271	1.0233 1.1496 0.9856
0.4	0.9547 1.0491 0.9365	1.0279 1.1279 0.9983	1.0353 1.1714 0.9897	1.0905 1.2751 1.0243	1.0271 1.1559 0.9872
0.6	1.0613 1.1574 1.0561	1.0317 1.1309 1.0179	1.0680 1.2058 1.0297	1.0228 1.2048 0.9661	1.0460 1.1747 1.0175
0.8	1.0664 1.1329 1.0838	1.0890 1.1851 1.1044	1.0326 1.1425 1.0339	0.9974 1.1762 0.9673	1.0464 1.1590 1.0474
Column Average	1.0083 1.0940 1.0007	1.0346 1.1336 1.0193	1.0543 1.1862 1.0215	1.0455 1.2256 0.9962	1.0357 1.1598 1.0094

Note: Entries from top to bottom are the relative efficiencies of: Hatanaka with first observation correction, Hatanaka without first observation correction and the two-step estimator analyzed by Wallis (1972).

TABLE 2

Relative Efficiencies of Estimators

Averaged Over All Values of ρ and α for given λ, R^2 and T

T=15

λ \ R^2	0.50	0.75	0.90	0.99	Row Average
0.0	0.9087 1.0962 0.8449	1.1029 1.3178 1.0392	1.0806 1.2837 1.0311	1.0769 1.2811 1.0463	1.0423 1.2447 0.9904
0.4	0.9968 1.2547 0.8984	0.9453 1.1773 0.8887	0.9229 1.1670 0.8813	1.0340 1.3155 0.9905	0.9748 1.2286 0.9147
0.8	1.0946 1.5993 0.9417	0.9663 1.3995 0.8728	1.0073 1.4329 0.9140	1.0037 1.3766 0.9591	1.0180 1.4521 0.9219
Column Average	1.0000 1.3167 0.8950	1.0048 1.2982 0.9336	1.0036 1.2945 0.9421	1.0382 1.3244 0.9986	1.0117 1.3085 0.9423

T=30

λ \ R^2	0.50	0.75	0.90	0.99	Row Average
0.0	1.0393 1.1233 1.0100	1.0763 1.1527 1.0519	1.1512 1.2474 1.1353	1.0812 1.1753 1.0742	1.0870 1.1747 1.0679
0.4	1.0725 1.1783 1.0301	1.0565 1.1638 1.0271	0.9589 1.0544 0.9407	1.0838 1.1853 1.0744	1.0429 1.1455 1.0181
0.8	1.0160 1.1610 0.9656	0.8848 1.0325 0.8523	1.0570 1.2027 1.0267	1.1065 1.2385 1.0965	1.0161 1.1587 0.9853
Column Average	1.0426 1.1542 1.0019	1.0059 1.1163 0.9771	1.0557 1.1682 1.0342	1.0905 1.1997 1.0817	1.0487 1.1596 1.0238

T=50

λ \ R^2	0.50	0.75	0.90	0.99	Row Average
0.0	1.1158 1.1506 1.1000	1.0037 1.0441 0.9999	1.0895 1.1457 1.0833	1.0251 1.0693 1.0265	1.0585 1.1024 1.0524
0.4	1.0431 1.0891 1.0254	0.9990 1.0465 0.9869	1.0205 1.0741 1.0157	1.0555 1.1143 1.0575	1.0295 1.0810 1.0214
0.8	1.0710 1.1510 1.0406	1.0770 1.1633 1.0610	0.9975 1.0782 0.9892	1.0496 1.1287 1.0628	1.0488 1.1303 1.0384
Column Average	1.0766 1.1302 1.0553	1.0266 1.0846 1.0159	1.0358 1.0993 1.0294	1.0434 1.1041 1.0489	1.0456 1.1046 1.0374

T=100

λ \ R^2	0.50	0.75	0.90	0.99	Row Average
0.0	1.0493 1.0684 1.0553	1.0821 1.1054 1.0836	0.9945 1.0101 0.9956	1.0871 1.1081 1.0894	1.0533 1.0730 1.0560
0.4	1.0596 1.0830 1.0462	1.0778 1.1081 1.0763	1.0844 1.1096 1.0851	1.0251 1.0547 1.0246	1.0617 1.0889 1.0581
0.8	0.8915 0.9451 0.8774	0.9709 1.0067 0.9648	1.1393 1.1800 1.1314	0.9798 1.0142 0.9816	0.9954 1.0365 0.9888
Column Average	1.0001 1.0322 0.9930	1.0436 1.0734 1.0416	1.0727 1.0999 1.0707	1.0307 1.0590 1.0319	1.0368 1.0661 1.0343

T=ALL

λ \ R^2	0.50	0.75	0.90	0.99	Row Average
0.0	1.0283 1.1096 1.0026	1.0663 1.1550 1.0437	1.0790 1.1717 1.0613	1.0676 1.1585 1.0591	1.0603 1.1487 1.0417
0.4	1.0430 1.1513 1.0000	1.0197 1.1239 0.9948	0.9967 1.1013 0.9807	1.0496 1.1675 1.0368	1.0273 1.1360 1.0031
0.8	1.0183 1.2141 0.9563	0.9748 1.1505 0.9377	1.0503 1.2235 1.0153	1.0349 1.1895 1.0250	1.0196 1.1944 0.9836
Column Average	1.0299 1.1583 0.9863	1.0203 1.2431 0.9921	1.0420 1.1655 1.0191	1.0507 1.1718 1.0403	1.0357 1.1598 1.0094

Note: For interpretation of entries see note of Table 1.

References

Beach, C.M. and J.G. MacKinnon, 1978, "A maximum likelihood procedure for regression with autocorrelated errors," Econometrica, 46, 51-58.

Dhrymes, P.J. , 1971, Distributed Lags (Holden-Day, San Francisco).

Fomby, T.B. and D.K. Guilkey, 1983, "An examination of two-step estimators for models with lagged dependent variables and autocorrelated errors," Journal of Econometrics, 22, 291-300.

Harvey, A.C., 1981, The Econometric Analysis of Time Series (Halsted, New York).

Hatanaka, M., 1974, "An efficient two-step estimator for the dynamic adjustment model with auto-correlated errors," Journal of Econometrics, 2, 199-220.

Hendry, D.F. and R.W. Harrison, 1974, "Monte Carlo methodology and the small sample behavior of ordinary and two-stage least squares," Journal of Econometrics, 2, 151-174.

Maddala, G.S., 1971, "Generalized least squares with an estimated variance-covariance matrix," Econometrica, 39, 23-33.

Maeshiro, A., 1979, "On the retention of the first observation in serial correlation adjustment," International Economic Review, 20, 259-265.

Makhail, W.M., 1972, "Simulating the small smaple properties of econometric estimators," Journal of the American Statistical Association, 67, 620-624.

Makhail, W.M., 1975, "A comparative Monte-Carlo study of the properties of econometric estimators," Journal of the American Statistical Association, 70, 91-104.

Prais, S.J. and C.B. Winsten, 1954, "Trend estimators and serial correlation," Cowles Commision discussion paper no.383 (Chicago, IL).

Wallis, K.F., 1967, "Lagged Dependent Variables and Serially Correlated Errors: A Reappraisal of Three-Pass Least Squares," Review of Economics and Statistics, 69, pp. 555-567.

Wallis, K.F., 1972, "The efficiency of the two-step estimator," Econometrica, 40, 769-770.