1982

Efficiency of Estimators in Regression Model with AR(1) Errors

Lonnie Magee

Aman Ullah

V. K. Srivastava

20056

ISSN: 0318-725X
ISBN: 0-7714-0397-6

RESEARCH REPORT 8224

EFFICIENCY OF ESTIMATORS IN REGRESSION MODEL WITH AR(1) ERRORS

by

L. Magee
A. Ullah*
V. K. Srivastava

December, 1982

*The research support from the SSHRC is gratefully acknowledged. The authors are thankful to L. Bauwens, B. Raj and members of the Econometrics Workshop at the University of Western Ontario for useful comments and suggestions.

1. Introduction

For a linear regression model with first-order autocorrelated disturbances, a variety of estimators for the regression coefficients have been proposed in the literature. One of the most commonly used estimators for this situation has been the Cochrane-Orcutt (1949) estimator (CO) due to its intuitive and computational simplicity. Since its introduction, several alternative estimators have also been proposed and their efficiency properties have been investigated. For example, Kadiyala (1968) showed that OLS is a better estimator than Cochran-Orcutt, for known autocorrelation coefficient ρ in $0 < \rho \leq 1$, for the model containing only an intercept.[1,2] Maeshiro (1976), then showed that the OLS is better than CO with known ρ for all $\rho > 0$ in a similar model where the matrix of explanatory variables contains an intercept and a strongly trended variable, even for sample sizes of $T = 100$. In addition, OLS is vindicated by Harvey and McAvinchey (1978), who suggest that it performs acceptably when the variable is trended; by Spitzer (1979), who recommends its use when the absolute value of the autocorrelation coefficient is ≤ 0.2; and by Kramer (1980), who proves that the efficiency (when measured by the trace of the variance-covariance matrix) of OLS with respect to generalized least squares (GLS) approaches one as ρ approaches one when the model includes a constant term. However, recently Taylor (1982) has pointed out that Maeshiro's result only applies to a special case: very strong tends in the explanatory variable, and where this variable is fixed, as opposed to the Monte Carlo studies of Griliches and Rao (1968) and Spitzer (1979) in which the explanatory variable is drawn from a prespecified stochastic process. In this latter case, the first observation no longer remains as important for large T, hence the improved performance of CO. In fact, Hoque (1980) has also shown that CO is better than OLS when the matrix of explanatory variables contains a column vector following a linear or geometric trend and ρ is known, thus contradicting Maeshiro.

In addition to the above mentioned work there are studies which compare CO with the Prais-Winsten (1954) estimator (PW). For example, Maeshiro (1979) and Park and Mitchell (1980) found that CO was always worse than PW. Maeshiro considered the known ρ case analytically, and both Maeshiro and Park and Mitchell used Monte Carlo methods for the ρ unknown case. The only mention of CO ever outperforming PW is made by Spitzer (1979) who finds CO slightly better than PW when the absolute value of the autocorrelation coefficient is close to one.

It should be noted that all the previous efficiency studies have either been analytical with the assumption that ρ is known a priori, or have used Monte Carlo methods. No expressions have been obtained specifically in the context of two step CO and PW estimators which use an estimated autocorrelation coefficient. This paper is an attempt in this direction. We consider classes of two step CO and PW estimators which arise due to various choices of the estimated autocorrelation coefficient. We have shown that these two step estimators are unbiased if their mean vectors exist and disturbances are symmetrically distributed. Further, taking the disturbances to be normal, we have presented in Section 2 the expressions for the large sample asymptotic approximations of the variance covariance matrices. Using these expressions, the efficiencies of the estimators are then analyzed with the help of a numerical experiment in Section 3. Some remarks are also placed in this Section. Finally, in Section 4 we have provided the proofs of results in Section 2.

2. The Estimators and Their Properties

Consider the linear regression model with first order autocorrelated disturbances:

$$(2.1) \qquad \begin{aligned} y_t &= x_t\beta + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \qquad (t=1,2,\ldots,T) \end{aligned}$$

where y_t is the t^{th} observation on the variable to be explained, x_t is a $1 \times k$ vector of observations on k explanatory variables, β is the coefficient vector associated with them, u_t is the disturbance term following a first order autoregressive scheme with unknown autocorrelation coefficient $\rho(|\rho| < 1)$ and

$$(2.2) \qquad \begin{aligned} E(\epsilon_t) &= 0 \qquad \text{for all } t \\ E(\epsilon_t \epsilon_{t+s}) &= \Psi \qquad \text{if } s=0 \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

Defining

$$(2.3) \qquad \begin{aligned} y' &= (y_1, y_2, \ldots, y_T) \\ X' &= (x_1', x_2', \ldots, x_T') \\ u' &= (u_1, u_2, \ldots, u_T) \end{aligned}$$

we can write

$$(2.4) \qquad y = X\beta + u$$

with

$$(2.5) \qquad \begin{aligned} E(u) &= 0 \\ E(uu') &= \sigma^2\Sigma \end{aligned}$$

where $\sigma^2 = \Psi(1-\rho^2)$ and Σ is a $k \times k$ symmetric matrix with the $(i,j)^{th}$ element equal to $\rho^{|i-j|}$.

The ordinary least squares (OLS) estimator of β is

(2.6) $$b = (X'X)^{-1}X'y$$

which is unbiased with variance-covariance matrix:

(2.7) $$E(b - \beta)(b - \beta)' = \sigma^2 (X'X)^{-1}X'\Sigma X(X'X)^{-1} .$$

The estimator b ignores the autocorrelated nature of disturbances and is therefore not efficient. This is accounted for in the generalized least squares (GLS) estimator given by

(2.8) $$\beta^* = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

which is unbiased with variance-covariance matrix:

(2.9) $$E(\beta^* - \beta)(\beta^* - \beta)' = \sigma^2 (X'\Sigma^{-1}X)^{-1} .$$

The estimator β^* can be construed as the ordinary least squares estimator in the transformed model

(2.10) $$Py = PX\beta + Pu$$

where P is a T x T triangular matrix given by

(2.11) $$P = \begin{bmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

Since ρ is generally not known, we can replace ρ by its consistent estimator in order to get the following operational estimator:

(2.12) $$\hat{\beta} = (X'\hat{P}'\hat{P}X)^{-1}X'\hat{P}'\hat{P}y$$

where $\hat{P}$ is the same as P except that ρ is replaced by its consistent estimator $\hat{\rho}$.

If $\hat{u}_t$ is the t^{th} element of the ordinary least squares residual vector $\hat{u} = (y - Xb)$, a simple choice of $\hat{\rho}$ is

$$(2.13) \qquad \hat{\rho}_1 = \frac{\sum_{t=1}^{T-1} \hat{u}_t \hat{u}_{t+1}}{\sum_{t=1}^{T} \hat{u}_t^2} .$$

Some alternative estimators for ρ have been suggested in the literature (see Judge et al [1980, p. 183]). For example, Theil (1971) modifies $\hat{\rho}_1$ as

$$(2.14) \qquad \hat{\rho}_2 = \frac{(T-k)}{T-1} \hat{\rho}_1 .$$

Next, an estimate derived from the Durbin-Watson statistic is given by

$$(2.15) \qquad \hat{\rho}_3 = 1 - \frac{1}{2} d$$

where

$$(2.16) \qquad d = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

is the Durbin-Watson statistic. Theil and Nagar (1961) suggested the following estimator

$$(2.17) \qquad \hat{\rho}_4 = \frac{T^2(1 - d/2) + k^2}{T^2 - k^2} .$$

Finally, the Durbin (1960) estimator, $\hat{\rho}_5$, is obtained by estimating ρ (the coefficient of y_{t-1}) in the following equation using OLS:

$$(2.18) \qquad y_t = \rho y_{t-1} + \beta_o(1 - \rho) + x_t^o \beta^o - \rho x_{t-1}^o \beta^o + \epsilon_t, \qquad t=2,\ldots,T$$

where x_t^o is the t^{th} row of X with the constant term deleted, i.e., $x_t = [1\, x_t^o]$ in (2.1). If $x_t = x_t^o$ then $\beta_o(1-\rho)$ gets dropped out in (2.18).

Now denoting by $\hat{P}_s$ the matrix P in (2.11) with ρ replaced by $\hat{\rho}_s$, s=1,...,5, we can write a class of Prais-Winsten type two-step estimators from (2.12) as

(2.19) $$\hat{\beta}_s = (X'\hat{P}_s'\hat{P}_s X)^{-1} X'\hat{P}_s'\hat{P}_s y\ , \qquad s=1,\ldots,5\ .$$

For s = 1, the estimator $\hat{\beta}_1$ is often termed the Prais-Winston (1954) estimator.

The Cochrane-Orcutt (1949) type two-step estimators are analytically similar to the Prais-Winsten type two step estimators, and can be constructed as the ordinary least squares estimator in the transformed model (corresponding to (2.10)):

(2.20) $$CPy = CPX\beta + CPu$$

where

(2.21) $$C = [0:\ I_{T-1}]$$

is a T-1 x T constant matrix such that $CPu = \varepsilon$ is a T-1 x 1 vector of disturbances and CP is the same as P in (2.11) except that the first row is deleted; 0 in (2.21) is a T-1 x 1 vector and I_{T-1} is a T-1 x T-1 identity matrix. With ρ replaced by $\hat{\rho}_s$, s=1,...,5, these estimators can be written as

(2.22) $$\tilde{\beta}_s = (X'\hat{P}_s' C_o \hat{P}_s X)^{-1} X'\hat{P}_s' C_o \hat{P}_s y\ ;\quad C_o = C'C\ .$$

For s = 1, the estimator $\tilde{\beta}_1$ is a well known Cochrane-Orcutt two step estimator. We note that the Cochrane-Orcutt type estimators differ from the Prais-Winsten type estimators in (2.19) with respect to C_o matrix.

We shall now present the variance-covariance matrices of $\hat{\beta}_s$ and $\tilde{\beta}_s$. However, before doing this we introduce the following matrices for the sake of simplicity in exposition.

Let D be a T x T diagonal matrix with first and last diagonal elements equal to 1 and the remaining diagonal elements equal to $1-\rho^2$, let D_o be a diagonal matrix with a first element of 1 and the other diagonal elements equal to zero, and let B be a T x T symmetric matrix with $(i,j)^{th}$ element equal to $-\rho$ if $i=j$, $\frac{1}{2}$ if $i=j+1$, and 0 otherwise. Further, define

$$
\begin{aligned}
&C_1 = [I_{T-1} \vdots\ 0]\ , && C_\rho = C - \rho C_1 \\
(2.23)\qquad &M = I - X(X'X)^{-1}X'\ , && M_Z = I - Z(Z'Z)^{-1}Z',\quad R = [(1-\rho^2)\Sigma^{-1} - D]/\rho \\
&N = \Sigma C_1' M_Z C_\rho\ , && \Omega = (X'\Sigma^{-1}X)^{-1}\ ,\quad Q = \Sigma - X\Omega X'
\end{aligned}
$$

where C_1 and C_ρ are T-1 x T matrices, M is a T x T matrix, and M_Z is a T-1 x T-1 matrix in which $Z' = [z_2',\ldots,z_T']$; $z_t = [1\ x_{t-1}^o\ x_t^o]$ when $x_t = [1\ x_t^o]$ in (2.1) and $z_t = [x_{t-1}^o\ x_t^o]$ when there is no constant in x_t.

It is assumed that the matrix $\frac{1}{T}X'X$ tends to a finite matrix as T tends to infinity. This assumption implies the absence of trend as an explanatory variable. We can now state the main results. These are proved in the following section.

Theorem 1: If the disturbances are symmetrically distributed the Prais-Winsten type two-step estimators $\hat{\beta}_s$, s=1,...,5, are unbiased provided their mean vectors exist. Further, if the distribution of disturbances is normal, the variance-covariance matrices to order $O(T^{-2})$ of the asymptotic distribution of $\hat{\beta}_s$ are given by

$$(2.24)\qquad V(\hat{\beta}_s) = \sigma^2\Omega + \frac{2\sigma^2(\mathrm{tr}\ MBM\ \Sigma\ MBM\ \Sigma)}{T^2\rho^2(1-\rho^2)^2}\Omega X' DQDX\Omega\ , \qquad s=1,\ldots,4$$

$$= \sigma^2\Omega + (\sigma^2/\rho^2(1-\rho^2)^2T^2)\Omega X' D[\{(\mathrm{tr}\ N)^2 + (\mathrm{tr}\ N^2) + \mathrm{tr}\ \Sigma N' \Sigma^{-1}N\}Q$$

$$+ (\mathrm{tr}\ N)(N\Sigma + \Sigma N') + 2(N\Sigma + \Sigma N')(\Sigma^{-1}N\Sigma + N')]DX\Omega\ , \qquad s=5$$

Theorem 2: If the disturbances are symmetrically distributed the Cochrane-Orcutt two-step estimator $\tilde{\beta}_1$ is unbiased provided its mean vector exists. Further, if the distribution of disturbances is normal, the variance-covariance matrix to order $O(T^{-2})$ of the asymptotic distribution of $\tilde{\beta}_1$ is given by

$$(2.25)\qquad V(\tilde{\beta}_1) = V(\hat{\beta}_1) + \sigma^2\Omega X' [D_o + \frac{\mathrm{tr}\ \Sigma MBM}{T(1-\rho^2)}\{4D_o + G + G'\}]X\Omega$$

where

$$(2.26)\qquad G = R[X\Omega X' - Q - (2/\mathrm{tr}\ \Sigma MBM)\Sigma MBM\Sigma]D_o\ .$$

From (2.9) and (2.24), we see that the variance-covariance matrix, to order $O(T^{-2})$, of $\hat{\beta}_1$, s=1,...,4, exceeds the corresponding matrix of β^* by

$$(2.27)\qquad \frac{2\sigma^2(\mathrm{tr}\ MBM\ \Sigma\ MBM\ \Sigma)}{T^2\rho^2(1-\rho^2)^2}\Omega X' DQDX\Omega$$

which can be attributed to replacement of ρ by $\hat{\rho}$ in β^*. Similarly, the variance-covariance matrix of $\tilde{\beta}_1$ exceeds that of β^* by (2.27) plus the second term on the right-hand side of (2.25).

The result in Theorem 2 has been presented for the Cochrane-Orcutt two-step estimator $\tilde{\beta}_1$. For the case of $\tilde{\beta}_s$, s = 2,3,4 the results can be developed in the similar manner. These were not found to be the same as for $\tilde{\beta}_1$. However, it should be noted from Theorem 1 that $V(\hat{\beta}_s)$ is the same for s = 1,2,3,4 in the case of PW type two-step estimators.

3. Numerical Experiment

In this section we evaluate the results of Theorems 1 and 2 of Section 2 using a numerical experiment. These results are stated in Tables 1 to 3 for both the trace and determinant of the variance-covariance matrix measures of relative efficiency. In particular we have obtained numerical values for the following measures:

(3.1) $e_1 = \det(b)/\det(\beta^*)$; $e_1^* = tr(b)/tr(\beta^*)$

(3.2) $e_2 = \det(\hat{\beta}_s)/\det(\beta^*)$; $e_2^* = tr(\hat{\beta}_s)/tr(\beta^*)$, $s=1,\ldots,4$

(3.3) $e_3 = \det(\hat{\beta}_s)/\det(\beta^*)$; $e_3^* = tr(\hat{\beta}_s)/tr(\beta^*)$, $s=5$

(3.4) $e_4 = \det(\tilde{\beta}_s)/\det(\beta^*)$; $e_4^* = tr(\tilde{\beta}_s)/tr(\beta^*)$, $s=1$

where det(b) and tr(b), for example, respectively represent the determinant and trace of the variance-covariance matrix of b.

The following three models are used:

$$y = X\beta + u,$$

where

(3.5) M_1: $X = [1\, x_1\, x_2]$

(3.6) M_2: $X = [x_2\, x_3]$

(3.7) M_3: $X = [1\, x_4]$,

β is the coefficient vector with appropriate dimension, and

$x_2' = [1.723, .022, 1.157, .504, 2.832, .902, .853, 1.816, 2.898, 1.019]$,

$x_3' = [.432, 1.376, 1.01, .005, 1.393, 1.787, .105, 1.339, 1.041, .279]$,

$x_4' = [1.809, 2.309, 2.691, 3.191, 4.0, 5.191, 6.691, 8.309, 9.809, 11.0]$

in the case $T=10$. The same models were used for $T=40$ by writing each element in x_2', x_3' and x_4' four times. For example, for x_2 in the $T=40$ case:

$$x_2' = [1.723,\ 1.723,\ 1.723,\ 1.723,\ .022,\ .022,\ldots,1.019,\ 1.019] \ .$$

This method of increasing sample size preserves the features of the trended variables while being consistent with the assumption of a finite $\lim(X'X)/T$ as T approaches infinity.

The results are invariant with respect to β's. The scale transformations of exogenous column vectors, however, will affect the trace ratios while they have no effect on the determinant ratios.

3.1 Main Results and Remarks

(i) The Prais-Winsten estimator $\hat{\beta}_1$, which uses the standard ρ estimate, $\hat{\rho}_1$, is almost always better than $\hat{\beta}_5$, the Prais-Winsten estimator which uses Durbin's ρ estimate, $\hat{\rho}_5$.[3,4] This can be seen by comparing e_2 with e_3, and e_2^* with e_3^* in Tables 1 to 3. The $\hat{\beta}_1$ has lower mean square errors (MSE's) than $\hat{\beta}_5$ except for some cases where $\rho \geq 0.8$ and X has no constant term.[5] This corresponds with the Monte Carlo evidence of Spitzer (1979), Kramer (1980, 1982) and Harvey and McAvinchey (1978) while contradicting the recommendation of Griliches and Rao (1968). The differences in MSE's between the two estimators is usually quite small.

(ii) b(OLS) is better than $\hat{\beta}_1$ (Prais-Winsten using $\hat{\rho}_1$) when $0 \leq \rho \leq 0.2$ for $T = 10$ and $0 \leq \rho \leq 0.1$ for $T = 40$, roughly speaking, otherwise $\hat{\beta}_1$ is better for positive ρ (seen by comparing e_1, e_1^* with e_2, e_2^* in Tables 1 to 3. This agrees with Monte Carlo findings of Griliches and Rao (1968) and Spitzer (1979).

(iii) b(OLS) is better than $\tilde{\beta}_1$ (Cochrane-Orcutt using $\hat{\rho}_1$) when a constant is included usually for $0 \leq \rho \leq 0.5$ when $T = 10$ and $0 \leq \rho \leq 0.2$ when $T = 40$, otherwise $\tilde{\beta}_1$ is better (by comparing e_1, e_1^* with e_4, e_4^*). This result holds for both trended and non-trended X. Without a constant, b is usually better than $\tilde{\beta}_1$ when $0 \leq \rho \leq 0.3$ and $T = 10$ with this zone shrinking towards zero as T increases.

(iv) $\hat{\beta}_1$ (Prais-Winsten using $\hat{\rho}_1$) is better than $\tilde{\beta}_1$ (Cochrane-Orcutt using $\hat{\rho}_1$) except when a constant is not included and ρ is large. This can be seen by comparing e_2, e_2^* with e_4, e_4^*. The superiority of $\hat{\beta}_1$ over $\tilde{\beta}_1$ corresponds with all of the Monte Carlo evidence. The exception found here is similar to Spitzer's (1979) finding that $\tilde{\beta}_5$ outperforms $\hat{\beta}_5$ for high ρ. It should be noted that the expansion approximation for $\tilde{\beta}_1$ becomes suspect as ρ approaches one since we also observe e_4 and e_4^* less than one in some cases for large ρ. This is impossible due to Aitken's theorem.

(v) When a constant is not included, b is very poor compared to $\hat{\beta}_1$ for high ρ. This is clear by comparing e_1, e_1^* with e_2, e_2^* in Tables 1 to 3. This is consistent with Kramer's (1980) proof that b is very poor compared to β^* (GLS) as ρ approaches one when there is no constant term and the data are not centered.

(vi) A useful extension of this study would be the derivation of large-sample approximations to the MSE matrices of iterative estimators such as iterated Prais-Winsten and maximum likelihood, which perform well in the Monte Carlo studies of Beach and MacKinnon (1978), Harvey and McAvinchey (1978), Spitzer (1979) and Park and Mitchell (1980). Such an extension would give a more complete analytical picture to complement the Monte Carlo results. The results of this paper can also be extended for the model with a higher order autoregressive process, say AR(2) or to a model with a moving average process.

TABLE 1

Alternative Efficiency Measures of Estimators for M1

Sample Size: 10

ρ	e_1	e_2	e_3	e_4	e_1^*	e_2^*	e_3^*	e_4^*
.05	1.005	1.097	1.123	1.299	1.002	1.030	1.038	1.076
.10	1.021	1.086	1.110	1.304	1.007	1.027	1.033	1.075
.20	1.084	1.064	1.084	1.317	1.026	1.019	1.024	1.074
.30	1.193	1.045	1.060	1.343	1.055	1.012	1.016	1.082
.40	1.360	1.030	1.041	1.392	1.093	1.007	1.010	1.109
.50	1.607	1.020	1.027	1.475	1.134	1.005	1.006	1.170
.60	1.971	1.014	1.017	1.604	1.170	1.004	1.005	1.289
.70	2.505	1.012	1.012	1.787	1.192	1.005	1.006	1.496
.80	3.269	1.012	1.008	2.021	1.182	1.008	1.006	1.806
.90	4.292	1.008	1.003	2.238	1.124	1.007	1.003	2.144
.95	4.874	1.002	1.001	2.250	1.072	1.002	1.000	2.214

Sample Size: 40

ρ	e_1	e_2	e_3	e_4	e_1^*	e_2^*	e_3^*	e_4^*
.05	1.004	1.043	1.054	1.107	1.002	1.018	1.023	1.033
.10	1.019	1.049	1.062	1.119	1.008	1.021	1.027	1.037
.20	1.082	1.061	1.080	1.143	1.034	1.026	1.033	1.044
.30	1.209	1.071	1.096	1.168	1.083	1.029	1.039	1.049
.40	1.440	1.076	1.105	1.190	1.162	1.029	1.041	1.053
.50	1.857	1.072	1.103	1.209	1.275	1.026	1.037	1.057
.60	2.631	1.057	1.086	1.231	1.423	1.018	1.027	1.067
.70	4.162	1.036	1.060	1.278	1.581	1.009	1.015	1.104
.80	7.603	1.018	1.036	1.404	1.662	1.005	1.010	1.231
.90	17.807	1.008	1.026	1.775	1.511	1.005	1.018	1.648
.95	31.244	1.008	1.022	2.175	1.313	1.007	1.019	2.111

TABLE 2

Alternative Efficiency Measures of Estimators for M2

Sample Size: 10

ρ	e_1	e_2	e_3	e_4	e_1^*	e_2^*	e_3^*	e_4^*
.05	1.007	1.145	1.187	1.281	1.003	1.068	1.093	1.107
.10	1.026	1.148	1.186	1.289	1.013	1.063	1.085	1.101
.20	1.113	1.154	1.183	1.296	1.054	1.054	1.071	1.089
.30	1.284	1.159	1.178	1.283	1.125	1.048	1.060	1.078
.40	1.594	1.160	1.168	1.241	1.241	1.045	1.053	1.066
.50	2.170	1.155	1.151	1.161	1.428	1.044	1.046	1.050
.60	3.322	1.152	1.127	1.042	1.745	1.045	1.039	1.025
.70	5.923	1.167	1.097	.906	2.329	1.051	1.031	.993
.80	13.064	1.221	1.064	.800	3.582	1.069	1.020	.964
.90	42.226	1.362	1.037	.832	7.475	1.108	1.009	.967
.95	> 100.0	1.596	1.018	1.058	15.309	1.170	1.005	1.023

Sample Size: 40

ρ	e_1	e_2	e_3	e_4	e_1^*	e_2^*	e_3^*	e_4^*
.05	1.003	1.029	1.032	1.074	1.002	1.020	1.022	1.037
.10	1.012	1.034	1.038	1.083	1.008	1.023	1.026	1.041
.20	1.053	1.046	1.052	1.101	1.036	1.029	1.033	1.048
.30	1.137	1.059	1.069	1.120	1.089	1.035	1.041	1.053
.40	1.292	1.074	1.089	1.138	1.180	1.039	1.047	1.055
.50	1.584	1.090	1.111	1.153	1.329	1.041	1.050	1.053
.60	2.180	1.104	1.135	1.158	1.576	1.039	1.051	1.049
.70	3.626	1.108	1.153	1.139	2.017	1.036	1.052	1.042
.80	8.519	1.084	1.140	1.069	2.978	1.027	1.047	1.024
.90	42.269	1.036	1.074	.952	6.381	1.012	1.025	.990
.95	> 100.0	1.045	1.034	.921	14.026	1.014	1.011	.981

TABLE 3

Alternative Efficiency Measures of Estimators for M3

Sample Size: 10

ρ	e_1	e_2	e_3	e_4	e_1^*	e_2^*	e_3^*	e_4^*
.05	1.001	1.028	1.056	1.509	1.001	1.015	1.032	1.463
.10	1.005	1.028	1.058	1.558	1.002	1.014	1.033	1.507
.20	1.018	1.028	1.063	1.664	1.009	1.014	1.036	1.604
.30	1.040	1.026	1.066	1.782	1.019	1.012	1.037	1.709
.40	1.069	1.024	1.068	1.906	1.032	1.010	1.037	1.820
.50	1.106	1.021	1.066	2.026	1.046	1.008	1.034	1.926
.60	1.146	1.017	1.058	2.122	1.057	1.005	1.027	2.013
.70	1.184	1.014	1.044	2.161	1.063	1.003	1.017	2.056
.80	1.209	1.010	1.026	2.109	1.058	1.001	1.007	2.032
.90	1.201	1.004	1.010	1.977	1.039	1.000	1.001	1.951
.95	1.176	1.001	1.003	1.937	1.023	1.000	1.000	1.931

Sample Size: 40

ρ	e_1	e_2	e_3	e_4	e_1^*	e_2^*	e_3^*	e_4^*
.05	1.000	1.005	1.006	1.090	1.000	1.003	1.003	1.084
.10	1.002	1.006	1.006	1.100	1.001	1.003	1.004	1.094
.20	1.009	1.007	1.008	1.126	1.005	1.004	1.005	1.117
.30	1.021	1.010	1.011	1.159	1.012	1.006	1.007	1.148
.40	1.041	1.013	1.016	1.204	1.025	1.008	1.010	1.189
.50	1.075	1.019	1.024	1.267	1.045	1.012	1.016	1.256
.60	1.133	1.027	1.038	1.356	1.080	1.018	1.026	1.323
.70	1.242	1.038	1.065	1.480	1.142	1.025	1.043	1.427
.80	1.482	1.045	1.114	1.628	1.267	1.028	1.073	1.539
.90	2.165	1.028	1.152	1.689	1.512	1.015	1.086	1.607
.95	3.016	1.013	1.095	1.767	1.584	1.006	1.044	1.772

4. Proof of Theorems in Section 2

Suppose B is a $T \times T$ symmetric matris with $(i,j)^{th}$ element b_{ij} defined in Section 2 as

$$
\begin{aligned}
b_{ij} &= -\rho \quad \text{if } i=j \\
&= \frac{1}{2} \quad \text{if } i=j\pm 1 \\
&= 0 \quad \text{otherwise .}
\end{aligned}
$$

From (2.13), we can express

$$
\begin{aligned}
(4.1) \qquad \hat{\rho}_1 &= \frac{\hat{u}'(\rho I_T + B)\hat{u}}{\hat{u}'\hat{u}} \\
&= \rho + \frac{\hat{u}' B\hat{u}}{\hat{u}'\hat{u}} \\
&= \rho + \frac{\frac{1}{T}u' MBMu}{\sigma^2 + (\frac{1}{T}u' Mu - \sigma^2)} \\
&= \rho + \frac{1}{T\sigma^2} u' MBMu[1 + \frac{1}{\sigma^2}(\frac{1}{T}u' Mu - \sigma^2)]^{-1}
\end{aligned}
$$

Expanding the expression in square brackets and retaining terms to order $O(T^{-1})$ in probability, we find

$$
(4.2) \qquad \hat{\rho}_1 - \rho = \theta_{-1/2} + \theta_{-1}
$$

where

$$
(4.3) \qquad
\begin{aligned}
\theta_{-1/2} &= \frac{1}{T\sigma^2} u' MBMu \\
\theta_{-1} &= -\frac{1}{T\sigma^4} u' MBMu(\frac{1}{T} u' Mu - \sigma^2)
\end{aligned}
$$

Here the suffixes of θ indicate the order in probability. Using (4.2) we have

$$(4.4)\qquad \hat{P}'\hat{P} = (1-\rho^2)\Sigma^{-1} - \frac{\theta_{-1/2}}{\rho}[D-(1-\rho^2)\Sigma^{-1}] + [\frac{\theta^2_{-1/2}}{\rho^2}(I_T - D) - \frac{2\theta_{-1}}{\rho}(D-(1-\rho^2)\Sigma^{-1}]$$

to order $O(T^{-1})$ in probability, where D is a T x T diagonal matrix (as defined in Section 2) with first and last diagonal elements equal to 1 and the remaining diagonal elements equal to $1-\rho^2$.

4.1 Proof of Theorem 1

From (2.1) and (2.19), we have

$$(4.5)\qquad (\hat{\beta}_1 - \beta) = (X'\hat{P}_1'\hat{P}_1X)^{-1}X'\hat{P}_1'\hat{P}_1u\ .$$

Since $(\hat{\beta}_1 - \beta)$ is an odd function of u, it follows that $E(\hat{\beta}_1 - \beta) = 0$ when the distribution of u is symmetrical and $E(\hat{\beta}_1)$ exists. This proves the first part of the theorem. For the second part, the distribution of u is assumed to be multivariate normal.

Substituting (4.4) into (4.5), we have, to order $O(T^{-3/2})$ in probability, after a little algebraic simplification[6]

$$(\hat{\beta}_1 - \beta) = \xi_{-1/2} + \xi_{-1} + \xi_{-3/2}$$

so that, to order $O(T^{-2})$,

$$(4.6)\qquad E(\hat{\beta}_1 - \beta)(\hat{\beta}_1 - \beta)' = E(\xi_{-1/2}\xi'_{-1/2}) + E(\xi_{-1}\xi'_{-1/2} + \xi_{-1/2}\xi'_{-1})$$

$$+ E(\xi_{-3/2}\xi'_{-1/2} + \xi_{-1/2}\xi'_{-3/2} + \xi_{-1}\xi'_{-1})$$

where

$$(4.7)\quad \xi_{-1/2} = \Omega X'\Sigma^{-1}u \qquad \xi_{-1} = -\frac{\theta_{-1/2}}{\rho(1-\rho^2)}\Omega X' DQ\Sigma^{-1}u$$

$$\xi_{-3/2} = \frac{1}{\rho^2(1-\rho^2)}\Omega X'[\theta^2_{-1/2}\{I_T - \frac{1}{(1-\rho^2)}DX\Omega X'D\} - \rho\theta_{-1}D]Q\Sigma^{-1}u\ .$$

Utilizing normality of disturbances, it is easy to see that

$$(4.8)\quad E(\xi_{-1/2}\xi'_{-1/2}) = \sigma^2\Omega$$

while the second, third, fourth and fifth terms on the right-hand side of (4.6) are equal to a null matrix. For the last term, we employ the following result which can be obtained from Srivastava and Tiwari (1976).

$$(4.9)\quad E[(u'Cu)^2uu'] = \sigma^6[\{(\mathrm{tr}\ C\Sigma)^2 + 2\ \mathrm{tr}\ C\Sigma C\Sigma\}\Sigma + 4(\mathrm{tr}\ C\Sigma)\Sigma C\Sigma + 8\ \Sigma C\Sigma C\Sigma]$$

where C is any symmetric matrix with nonstochastic elements.

Employing the above along with the results

$$(4.10)\quad QM = \Sigma M$$

$$Q\Sigma^{-1}Q = Q$$

and observing that $(\mathrm{tr}\ MBM\Sigma MBM\Sigma)$ is of order $O(T)$, we have

$$(4.11)\quad E(\xi_{-1}\xi'_{-1}) = \frac{1}{\rho^2(1-\rho^2)^2}\Omega X'DQ\Sigma^{-1}\cdot E[\theta^2_{-1/2}uu']\Sigma^{-1}QDX\Omega$$

$$= \frac{1}{\sigma^4T^2\rho^2(1-\rho^2)^2}\Omega X'DQ\Sigma^{-1}[E(u'MBMu)^2uu']\cdot\Sigma^{-1}QDX$$

$$= \frac{2\rho^2(\mathrm{tr}\ MBM\Sigma MBM\Sigma)}{T^2\rho^2(1-\rho^2)^2}\Omega X'DQDX\Omega$$

to order $O(T^{-2})$.

Combining (4.8) and (4.11), we find the result (2.24) stated in theorem 1 for $s=1$. The result (2.24) corresponding to the choice of $\hat{\rho}_s$, $s=2,3,4$, given in (2.14), (2.15) and (2.17), can be similarly verified to be the same as that for $s=1$. For $s=5$ we proceed as below.

The $\hat{\rho}_5$ (Durbin's $\hat{\rho}$) is derived from OLS on the equation

$$(4.12) \qquad y_t = \rho y_{t-1} + \beta_o(1-\rho) + x_t^o\beta^o - \rho x_{t-1}^o\beta^o + \epsilon_t; \quad t=2,\ldots,T.$$

It is given by

$$(4.13) \qquad \hat{\rho}_5 = \frac{y'_{-1}M_Z y_o}{y'_{-1}M_Z y_{-1}}$$

where $y_o = (y_2,\ldots,y_T)'$, $y_{-1} = (y_1,\ldots,y_{T-1})'$, $M_Z = I - Z(Z'Z)^{-1}Z'$ and $Z = [1\, X^o\, X^o_{-1}]$.

Now from (2.4) $y_{-1} = [1\ X^o_{-1}]\beta + u_{-1}$ and $y_o[1\ X^o]\beta + u_o$ and noting that $M_Z Z = 0$ it can be verified that

$$(4.14) \qquad \hat{\rho}_5 = \frac{u'_{-1}M_Z u_o}{u'_{-1}M_Z u_{-1}} = \frac{u'\,C_1'M_Z C u}{u'\,C_1'M_Z C_1 u}$$

where C and C_1 are as defined in (2.21) and (2.23), respectively. Following the steps similar to those for $\hat{\rho}_1$ in (4.1) we can obtain

$$(4.15) \qquad \hat{\rho}_5 - \rho = \theta^*_{-1/2} + \theta^*_{-1}$$

where $\theta^*_{-1/2} = u'\,C_1'M_Z C_\rho u/T\sigma^2$, $\theta^*_{-1} = -u'\,C_1'M_Z C_\rho u(u'\,C_1'M_Z C_1 u/_T - \sigma^2)/T\sigma^4$ and $C_\rho = C - \rho C_1$. Further

$$(4.16) \qquad \hat{\beta}_5 - \beta = \xi^*_{-1/2} + \xi^*_{-1} + \xi^*_{-3/2}$$

where $\xi^*_{-1/2}$, ξ^*_{-1} and $\xi^*_{-3/2}$ are the same as $\xi_{-1/2}$, ξ_{-1} and $\xi_{-3/2}$, respectively, given in (4.7) with $\theta_{-1/2}$ and θ_{-1} replaced by $\theta^*_{-1/2}$ and θ^*_{-1}. In fact $\xi^*_{-1/2} = \xi_{-1/2}$.

It is easy to verify that

$$E(\xi^*_{-1/2}\xi^{*\prime}_{-1/2}) = \sigma^2\Omega$$

(4.17) $$E(\xi^*_{-1}\xi^{*\prime}_{-1/2}) = E(\xi^*_{-1/2}\xi^{*\prime}_{-1}) = E(\xi^*_{-3/2}\xi^{*\prime}_{-1/2}) = E(\xi^*_{-1/2}\xi^{*\prime}_{-3/2}) = 0$$

$$E(\xi^*_{-1}\xi^{*\prime}_{-1}) = \frac{1}{\rho^2(1-\rho^2)^2}\Omega X'DQ\Sigma^{-1}[E\theta^{*2}_{-1/2}uu']\Sigma^{-1}QDX\Omega$$

where

(4.18) $$E(\theta^{*2}_{-1/2}uu') = \frac{1}{T^2\sigma^4}E[(u'C_1'M_ZC_\rho u)^2uu']$$

$$= \frac{1}{T^2\sigma^4}[\{(tr\Sigma C_1'M_ZC_\rho)^2 + tr\Sigma C_1'M_ZC_\rho\Sigma C_1'M_ZC_\rho$$

$$+ tr\,\Sigma C_\rho'M_ZC_1\Sigma C_1'M_ZC_\rho\}\Sigma + 2(tr\,\Sigma C_1'M_ZC_\rho)(\Sigma C_1'M_ZC_\rho\Sigma$$

$$+ \Sigma C_\rho'M_ZC_1\Sigma) + 2\Sigma(C_1'M_ZC_\rho + C_\rho'M_ZC_1)\Sigma(C_1'M_ZC_\rho + C_\rho'M_ZC_1)\Sigma]$$

has been obtained by modifying the result (4.9) for the nonsymmetric matrix $C_1'M_ZC_\rho$. Using (4.18) in (4.17) we get the result in Theorem 1 for $s = 5$ (Durbin's case).

4.2 Proof of Theorem 2

From (2.1) and (2.22) we can write the sampling error of the Cochrane-Orcutt estimator as

$$\tilde{\beta}_1 - \beta = (X'\hat{P}_1'C_o\hat{P}_1X)^{-1}X'\hat{P}_1'C_o\hat{P}_1y$$

where $C_o = C'C$ and

$$\hat{P}_1'C_o\hat{P}_1 = \hat{P}_1'\hat{P}_1 - D_o(1-\hat{\rho}_1^2) ;$$

D_o is a $T \times T$ diagonal matrix of zeros with a one in the upper left corner (defined in Section 2) and $\hat{P}_1$ is as used in Prais-Winsten estimator. Using (4.4) we can write

$$\hat{P}_1' C_o \hat{P}_1 = (1-\rho^2)(\Sigma^{-1} - D_o) + \theta_{-1/2}(R + 2\rho D_o) + \theta_{-1}(R + 2\rho D_o) + \theta_{-1/2}^2 (D_1 + D_o)$$

where $D_1 = (I - D)/\rho^2$ and $R = \frac{1-\rho^2}{\rho}\Sigma^{-1} - \frac{D}{\rho}$ is a $T \times T$ matrix.

Proceeding as for the Prais-Winsten estimator in (4.1) we get

$$\tilde{\beta}_1 - \beta = \eta_{-1/2} + \eta_{-1} + \eta_{-3/2}$$

where, using (4.7)

$$\eta_{-1/2} = \xi_{-1/2}, \quad \eta_{-1} = \xi_{-1} - \Omega X' D_o u$$

$$\eta_{-3/2} = \xi_{-3/2} + \frac{\Omega X'}{(1-\rho^2)} [\theta_{-1/2}\{2\rho + RX\Omega X'\} D_o u + (1-\rho^2) D_o X \Omega X' \Sigma^{-1} u] .$$

Now writing $(\tilde{\beta}_1 - \beta)(\tilde{\beta}_1 - \beta)'$ and taking expectation we get the result stated in Theorem 2 .

FOOTNOTES

[1]"Better" means smaller mean square error or, equivalently in this paper, smaller variance, since the estimators of β considered here are unbiased.

[2]When using CO, the use of the correct ρ does not necessarily result in the smallest mean square error (MSE). In fact, in the case of Kadiyala's model, the smallest MSE of CO estimators is obtained by using $\hat{\rho} = -1$ regardless of the true value of ρ; see Magee (1982).

[3]Due to the equivalence of the expansions for MSE's of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ to $0(1/T^2)$, the results stated here for $\hat{\beta}_1$ also hold for $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$.

[4]Expansion of $\hat{\rho}_1$ and $\hat{\rho}_5$ to order $1/T$ was used to estimate $E(\hat{\rho}_1)$ and $E(\hat{\rho}_5)$. It was found that the negative bias is almost always smaller for $\hat{\rho}_5$, as was noted by Rao and Griliches.

[5]Results are presented here for three models only, although there were many other models considered, as well as results for $T = 20$ and $T = 30$.

[6]Expansions for the class of estimators $(X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}y$ where $\hat{S}^{-1}$ is such that $X'(\hat{S}^{-1} - S^{-1})X$ has $0(T^{1/2})$ for some fixed S can be shown to be suitable for asymptotic expansion procedures using conditions described by Sargan (1974, p. 172).

REFERENCES

Beach, C. M. and J. G. MacKinnon (1978). "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," Econometrica 46, pp. 51-58.

Cochrane, D. and G. H. Orcutt (1949). "Application of Least Squares Regressions to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association 44, pp. 32-61.

Durbin, J. (1960). "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society Series B 22, pp. 139-153.

Griliches, Z. and P. Rao (1969). "Small-Sample Properties of Several Two Stage Regression Methods in the Context of Auto-Correlated Disturbances," Journal of the American Statistical Association 64, pp. 253-272.

Harvey, A. C. and D. McAvinchey (1978). "The Small Sample Efficiency of Two-Step Estimators in Regression Models with Autoregressive Disturbances," University of British Columbia, Economics Discussion Paper #78-10.

Hoque, A. (1980). "The Finite Sample Analysis of Three Dynamic Models," unpublished thesis, University of Western Ontario.

Judge, G. G., W. E. Griffiths, R. C. Hill, and T. C. Lee (1980). The Theory and Practice of Econometrics, John Wiley and Sons.

Kadiyala, K. R. (1968). "A Transformation used to Circumvent the Problem of Autocorrelation," Econometrica 36, pp. 93-96.

Krämer, W. (1980). "Finite Sample Efficiency of Ordinary Least Wquares in the Linear Regression Model with Autocorrelated Errors," Journal of the American Statistical Association 75, pp. 1005-1009.

Krämer, W. (1982). "Note on Estimating Linear Trend When Residuals are Autocorrelated," Econometrica 50, pp. 1065-1068.

Maeshiro, A. (1976). "Autoregressive Transformation, Trended Independent Variables and Autocorrelated Disturbance Terms," Review of Economics and Statistics 58, pp. 497-500.

Maeshiro, A. (1979). "On the Retention of the First Observations in Serial Correlation Adjustment of Regression Models," International Economic Review 20, pp. 259-265.

Magee, L. (1982). "Efficiency of Estimators in Regression Model with AR(1) Errors: An Analytical Study," mimeo, University of Western Ontario, Economics Department.

Park, R. E. and B. M. Mitchell (1980). "Estimating the Autocorrelated Error Model with Trended Data," Journal of Econometrics 13, pp. 185-201.

Prais, S. J. and C. B. Winsten (1954). "Trend Estimators and Serial Correlation," unpublished Cowles Commission Discussion Paper, Chicago.

Sargen, J. D. (1974). "The Validity of Nagar's Expansion for the Moments of Econometric Estimators," Econometrica 42, pp. 169-176.

Spitzer, J. J. (1979). "Small-Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors," Journal of the American Statistical Association 74, pp. 41-47.

Srivastava, V. K. and R. Tiwari (1976). "Evaluation of Expectation of Products of Stochastic Matrices," Scandinavian Journal of Statistics, Vol. 3, pp. 135-138.

Taylor, W. E. (1981). "On the Efficiency of the Cochrane-Orcutt Estimator," Journal of Econometrics 17, pp. 67-82.

Theil, H. (1971). Principles of Econometrics, John Wiley and Sons.

Theil, H. and A. L. Nagar (1961). "Testing the Independence of Regression Disturbances," Journal of the American Statistical Association 56, pp. 793-806.