Journal of the Korean Statistical Society (1998), **27**: *3, pp 319–329*

The Efficiency of the Cochrane-Orcutt Estimation Procedure in Autocorrelated Regression Models

Seuck Heun Song[1], Myoungshic Jhun[2] and Byoung Cheol Jung[3]

Abstract

In the linear regression model with an autocorrelated disturbances, the Cochrane-Orcutt estimator (*COE*) is a well known alternative to the Generalized Least Squares estimator (*GLSE*). The efficiency of *COE* has been studied empirically in a Monte Carlo study when the unknown parameters are estimated by maximum likelihood method. In this paper, it is theoretically proved that the *COE* is shown to be inferior to the *GLSE*. The comparisons are based on the difference of corresponding information matrices or the ratio of their determinants.

Keywords: Linear regression model; GLS estimator; Cochrane-Orcutt estimator; Maximum likelihood estimation; Information matrix; Autocorrelation

1. INTRODUCTION

For the linear regression model with autocorrelated disturbances, a variety of estimators for regression coefficients have been proposed in the literature (Judge et al. (1985) and Greene (1997)). One of most commonly used estimator for this situation is the Cochrane-Orcutt estimator (*COE*) as a well known alternative to Generalized Least Squares estimator (*GLSE*) due to its intuitive and computational simplicity. However, in the first-order autoregressive (AR(1)) case, several authors including Kadiyala (1968), Maeshiro (1979), Krämer (1982), Oxley and Robert (1983), Puterman (1988) and Stemann and Trenkler (1993) have studied the effect of omitting the first transformed observation and they have pointed out that the efficiency of *COE* can be very adversely affected if the sample is relatively small. On the other hand, Hoque (1989) has found that the *COE* can be more efficient than the Ordinary Least Squares estimator.

[1]Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.
[2]Department of Statistics, Korea University, Seoul, 136-701, Korea.
[3]Department of Statistics, Korea University, Seoul, 136-701, Korea.

Beach and MacKinnon (1978a) developed an iterative Maximum Likelihood Estimation (MLE) procedure for linear regression model with AR(1) disturbances, and they achieved Monte Carlo results which suggest that their MLE-procedure yields substantially more efficient estimator than CO-procedure. Beach and MacKinnon (1978b) extended the above approach to models which have a second-order autoregressive, AR(2), disturbances. Such models are often encountered in practice when dynamics of model behaviour are of some concern, as for example in Bruce (1975). The efficiency of COE has been previously demonstrated empirically in a Monte Carlo study. It does not appear to be known that the COE is inferior to the $GLSE$ theoretically. In this paper we will show that the COE is shown to be inferior to the $GLSE$ in terms of the information matrix criterion when the unknown parameters are estimated by maximum likelihood method.

This paper is organized as follows; In section 2, we consider a linear regression model with second order autoregressive scheme and define the transformation matrices for weighted least squares estimators and covariance matrices of estimators. In section 3, we present the likelihood function and investigate the efficiency of COE relative to $GLSE$ in terms of the difference of corresponding information matrices or the ratio of their determinants.

2. MODEL AND ESTIMATORS

We consider the linear regression model

$$y = X\beta + u, \tag{2.1}$$

where y is $T \times 1$ vector of observations on the dependent variable, X is a $T \times k$ matrix of observations on the independent variables (non-stochastic, of rank $k < T$) and the $k \times 1$ vector β is contained the unknown regression coefficients to be estimated. u is a $T \times 1$ vector of unobservable disturbances with $E(u) = 0$.

The disturbances are assumed to follow a second-order autoregressive process

$$u_t = \theta_1 u_{t-1} + \theta_2 u_{t-2} + \varepsilon_t, \qquad t = 1, 2, \cdots, T, \tag{2.2}$$

where $E(\varepsilon_t) = 0$, $E(\varepsilon_t\ \varepsilon_s) = 0$ for $t \neq s$, and $E(\varepsilon_t^2) = \sigma_\varepsilon^2$. This process will be stationary if $\theta_1 + \theta_2 < 1, \theta_2 - \theta_1 < 1$ and $-1 < \theta_2 < 1$. Under these assumptions the elements of the covariance matrix $E(uu') = \sigma_\varepsilon^2 V$ can be found from the variance

$$\sigma_u^2 = \frac{(1-\theta_2)\,\sigma_\varepsilon^2}{(1+\theta_2)\{(1-\theta_2)^2-\theta_1^2\}} \tag{2.3}$$

and the autocorrelation coefficients

$$\rho_1 = \frac{\theta_1}{1-\theta_2}, \tag{2.4}$$

$$\rho_2 = \theta_2 + \frac{\theta_1^2}{1-\theta_2} \tag{2.5}$$

$$\rho_s = \theta_1\rho_{s-1} + \theta_2\rho_{s-2}, \qquad s > 2. \tag{2.6}$$

The inverse of V is given by

$$V^{-1} = \begin{bmatrix} 1 & -\theta_1 & -\theta_2 & \cdots & 0 & 0 \\ -\theta_1 & 1+\theta_1^2 & -\theta_1+\theta_1\theta_2 & \cdots & 0 & 0 \\ -\theta_2 & -\theta_1+\theta_1\theta_2 & 1+\theta_1^2+\theta_2^2 & \cdots & 0 & 0 \\ 0 & -\theta_2 & -\theta_1+\theta_1\theta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\theta_1^2 & -\theta_1 \\ 0 & 0 & 0 & \cdots & -\theta_1 & 1 \end{bmatrix}, \tag{2.7}$$

a $T \times T$ matrix such that $V^{-1} = R_1'R_1$ is given by (see Lempers and Kloek (1973))

$$R_1 = \begin{bmatrix} \sigma_\varepsilon/\sigma_u & 0 & 0 & \cdots & 0 & 0 \\ -\rho_1\sqrt{1-\theta_2^2} & \sqrt{1-\theta_2^2} & 0 & \cdots & 0 & 0 \\ -\theta_2 & -\theta_1 & 1 & \cdots & 0 & 0 \\ 0 & -\theta_2 & -\theta_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\theta_1 & 1 \end{bmatrix}, \tag{2.8}$$

where $\dfrac{\sigma_\varepsilon}{\sigma_u} = \left\{\dfrac{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]}{1-\theta_2}\right\}^{1/2}$.

When θ_1 and θ_2 are known the $GLSE$ $\widetilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ can be obtained as the least squares estimator in the transformed model

$$y^* = X^*\beta + u^*, \tag{2.9}$$

where $y^* = R_1y$, $X^* = R_1X$ and $u^* = R_1u$.

The covariance matrix of $GLSE$ is given by

$$Cov(\widetilde{\beta}) = \sigma_\varepsilon^2 (X'R_1'R_1X)^{-1}. \tag{2.10}$$

As in the AR(1) case, let R_2 be the $(T-2)\times T$ matrix obtained from R_1 by deleting its first two top rows. Cochrane and Orcutt (1949) suggested the model

$$y^+ = X^+\beta + u^+, \tag{2.11}$$

where $y^+ = R_2y, X^+ = R_2X$ and $u^+ = R_2u$ which amounts to dropping the first and second observations from the transformed model in (2.9). Their approximate $GLSE$ then is given by

$$\widetilde{\beta}_{co} = (X'R_2'R_2X)^{-1}X'R_2'R_2y \tag{2.12}$$

with following dispersion matrix

$$Cov(\widetilde{\beta}_{co}) = \sigma_\varepsilon^2 (X'R_2'R_2X)^{-1}. \tag{2.13}$$

Since $\widetilde{\beta} = (X'R_1'R_1X)^{-1}X'R_1'R_1y$ is the best linear unbiased estimator (BLU E) of β, the $GLSE$ is a better estimator than the COE when the parameters θ_1, θ_2 are known.

However, both estimators heavily depend on the unknown parameters θ_1 and θ_2. Under the assumption of normality we may consider maximum likelihood estimation of the unknown parameters which will be done in the following section.

3. MAXIMUM LIKELIHOOD ESTIMATION

Assume that in the transformed model in (2.9) the disturbances obey $u^* \sim N(0, \sigma_\varepsilon^2 I_T)$. Then the loglikelihood function is equal to (cf. Judge et al., 1985, p 297)

$$\begin{aligned} L_1(\beta, \theta_1, \theta_2, \sigma_\varepsilon^2) = & - \frac{T}{2}log2\pi - \frac{T}{2}log\sigma_\varepsilon^2 + \frac{1}{2}log\Big\{(1+\theta_2)^2[(1-\theta_2)^2 - \theta_1^2]\Big\} \\ & - \frac{1}{2\sigma_\varepsilon^2}\Big\{\frac{\sigma_\varepsilon^2}{\sigma_u^2}u_1^2 + (1-\theta_2^2)u_2^{*2} + \sum_{t=3}^{T} u_t^{*2}\Big\}, \end{aligned} \tag{3.1}$$

where $u_2^{*2} = (u_2 - \rho_1 u_1)^2$ and $\sum_{t=3}^{T} u_t^{*2} = \sum_{t=3}^{T}(u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2})^2$.

Ignoring the first and second transformed observations we may alternatively look at the model $y^+ = X^+\beta + u^+$, where $u^+ = R_2 u \sim N(0, \sigma_\varepsilon^2 I_{T-2})$. In this case the loglikelihood function is

$$L_2(\beta, \theta_1, \theta_2, \sigma_\varepsilon^2) = -\frac{T-2}{2} log 2\pi - \frac{T-2}{2} log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=3}^{T} u_t^{*2}. \qquad (3.2)$$

The information matrix about the paremeter vector $\psi = (\beta, \theta_1, \theta_2, \sigma_\varepsilon^2)'$ contained in the preceding estimation procedures is characterized by the following information matrices

$$I_j = -E\left[\frac{\partial^2 L_j(\beta, \theta_1, \theta_2, \sigma_\varepsilon^2)}{\partial\psi\partial\psi'}\right], \quad j = 1, 2. \qquad (3.3)$$

The inverse of the information matrices provides a lower bound for the sampling precision of the maximum likelihood estimators in model $y^* = X^*\beta + u^*$ and $y^+ = X^+\beta + u^+$, respectively.

In the following we show that the CO-procedure is inferior to the method that incorporates the first and second transformed observations. Only in Beach and MacKinnon (1978b) comparable simulation study results were presented. In their papers they developed an iterative MLE procedure for the model (2.1) and suggested the MLE-procedure yields substantially more efficient estimators than CO-procedure through a Monte Carlo study. Our theoretical investigations confirm their findings.

Theorem 3.1. *The difference $I_1 - I_2$ of information matrics is n.n.d. for all value of ψ.*

Proof: Since $L_1 := L_1(\beta, \theta_1, \theta_2, \sigma_\varepsilon^2)$ is obtained from the joint density $f(y_1, \cdots, y_T)$, whereas $L_2 := L_2(\beta, \theta_1, \theta_2, \sigma_\varepsilon^2)$ is based on the conditional density $f(y_3, \cdots, y_T | y_1, y_2)$ it follows that

$$\begin{aligned} L_1 - L_2 = &- log 2\pi - log \sigma_\varepsilon^2 + log(1+\theta_2) + \frac{1}{2} log\{(1-\theta_2)^2 - \theta_1^2\} \\ &- \frac{1}{2\sigma_\varepsilon^2}\left\{\frac{\sigma_\varepsilon^2}{\sigma_u^2} u_1^2 + (1-\theta_2^2) u_2^{*2}\right\} \\ = &- log 2\pi - log \sigma_\varepsilon^2 + log(1+\theta_2) + \frac{1}{2} log\{(1-\theta_2)^2 - \theta_1^2\} \\ &- \frac{1}{2\sigma_\varepsilon^2}\left\{(1-\theta_2^2)(u_1^2 + u_2^2) - 2\theta_1(1+\theta_2) u_1 u_2\right\}. \qquad (3.4) \end{aligned}$$

As a matter of straightforward calculations we obtain

$$
\begin{aligned}
\frac{\partial\ (L_1 - L_2)}{\partial \beta} &= \frac{1}{\sigma_\varepsilon^2}\Big\{\frac{\sigma_\varepsilon^2}{\sigma_u^2}u_1x_1' + (1-\theta_2^2)(u_2 - \rho_1 u_1)(x_2 - \rho_1 x_1)'\Big\} \\
\frac{\partial\ (L_1 - L_2)}{\partial \theta_1} &= -\frac{\theta_1}{(1-\theta_2)^2 - \theta_1^2} + \frac{(1+\theta_2)u_1u_2}{\sigma_\varepsilon^2} \\
\frac{\partial\ (L_1 - L_2)}{\partial \theta_2} &= \frac{1}{1+\theta_2} - \frac{1-\theta_2}{(1-\theta_2)^2 - \theta_1^2} + \frac{1}{\sigma_\varepsilon^2}\Big\{\theta_2(u_1^2 + u_2^2) + \theta_1 u_1 u_2\Big\} \\
\frac{\partial\ (L_1 - L_2)}{\partial \sigma_\varepsilon^2} &= -\frac{1}{\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4}\Big\{(1-\theta_2^2)(u_1^2+u_2^2) - 2\theta_1(1+\theta_2)u_1u_2\Big\}.
\end{aligned}
$$

Then the derivatives of second order are

$$
\begin{aligned}
\frac{\partial^2\ (L_1 - L_2)}{\partial \beta \partial \beta'} &= -\frac{1}{\sigma_\varepsilon^2}\Big\{\frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1x_1' + (1-\theta_2^2)(x_2 - \rho_1 x_1)(x_2 - \rho_1 x_1)'\Big\} \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \theta_1^2} &= -\frac{(1-\theta_2)^2 + \theta_1^2}{\{(1-\theta_2)^2 - \theta_1^2\}^2} \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \theta_2^2} &= -\frac{1}{(1+\theta_2)^2} - \frac{(1-\theta_2)^2 + \theta_1^2}{\{(1-\theta_2)^2 - \theta_1^2\}^2} + \frac{1}{\sigma_\varepsilon^2}(u_1^2 + u_2^2) \\
\frac{\partial^2\ (L_1 - L_2)}{\partial (\sigma_\varepsilon^2)^2} &= \frac{1}{\sigma_\varepsilon^4} - \frac{1}{\sigma_\varepsilon^6}\Big\{(1-\theta_2^2)(u_1^2+u_2^2) - 2\theta_1(1+\theta_2)u_1u_2\Big\} \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \beta \partial \sigma_\varepsilon^2} &= -\frac{1}{\sigma_\varepsilon^4}\Big\{(1-\theta_2^2)(u_1x_1' + u_2x_2') - \theta_1(1+\theta_2)(u_1x_2' + u_2x_1')\Big\} \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \beta \partial \theta_1} &= -\frac{1}{\sigma_\varepsilon^2}\Big\{(1+\theta_2)(u_1x_2' + u_2x_1')\Big\} \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \beta \partial \theta_2} &= -\frac{1}{\sigma_\varepsilon^2}\Big\{2\theta_2(u_1x_1' + u_2x_2') + \theta_1(u_1x_2' + u_2x_1')\Big\} \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \sigma_\varepsilon^2 \partial \theta_1} &= -\frac{1+\theta_2}{\sigma_\varepsilon^4}u_1u_2 \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \sigma_\varepsilon^2 \partial \theta_2} &= -\frac{1}{\sigma_\varepsilon^4}\Big\{\theta_2(u_1^2+u_2^2) + \theta_1 u_1 u_2\Big\} \\
\frac{\partial^2\ (L_1 - L_2)}{\partial \theta_1 \partial \theta_2} &= -\frac{2\theta_1(1-\theta_2)}{\{(1-\theta_2)^2 - \theta_1^2\}^2} + \frac{1}{\sigma_\varepsilon^2}u_1u_2.
\end{aligned}
$$

After taking expectations we get the following expression for the difference of the information matrices.

$$
I_1 - I_2 = -E\left[\frac{\partial^2(L_1 - L_2)}{\partial\psi\ \partial\psi'}\right] \tag{3.5}
$$

$$= \begin{bmatrix} I(\beta) & 0 & 0 & 0 \\ 0 & A & B & C \\ 0 & B & D & E \\ 0 & C & E & F \end{bmatrix}, \tag{3.6}$$

where $I(\beta) = -E\Big[\dfrac{\partial^2(L_1-L_2)}{\partial\beta\,\partial\beta'}\Big] = \dfrac{1}{\sigma_\varepsilon^2}\Big\{\dfrac{\sigma_\varepsilon^2}{\sigma_u^2}x_1x_1' + (1-\theta_2^2)(x_2-\rho_1x_1)(x_2-\rho_1x_1)'\Big\}$
and

$$\begin{aligned}
A &= -E\Big[\frac{\partial^2(L_1-L_2)}{\partial(\sigma_\varepsilon^2)^2}\Big] = \frac{1}{\sigma_\varepsilon^4} \\
B &= -E\Big[\frac{\partial^2(L_1-L_2)}{\partial(\sigma_\varepsilon^2)\partial\theta_1}\Big] = \frac{\theta_1}{\sigma_\varepsilon^2\{(1-\theta_2)^2-\theta_1^2\}} \\
C &= -E\Big[\frac{\partial^2(L_1-L_2)}{\partial(\sigma_\varepsilon^2)\partial\theta_2}\Big] = \frac{2\theta_2(1-\theta_2)+\theta_1^2}{\sigma_\varepsilon^2(1+\theta_2)\{(1-\theta_2)^2-\theta_1^2\}} \\
D &= -E\Big[\frac{\partial^2(L_1-L_2)}{\partial\theta_1^2}\Big] = \frac{(1-\theta_2)^2+\theta_1^2}{\{(1-\theta_2)^2-\theta_1^2\}^2} \\
E &= -E\Big[\frac{\partial^2(L_1-L_2)}{\partial\theta_1\partial\theta_2}\Big] = \frac{2\theta_1(1-\theta_2^2)-\theta_1\{(1-\theta_2)^2-\theta_1^2\}}{(1+\theta_2)\{(1-\theta_2)^2-\theta_1^2\}^2} \\
F &= -E\Big[\frac{\partial^2(L_1-L_2)}{\partial\theta_2^2}\Big] = \Big\{\frac{[(1-\theta_2)^2-\theta_1^2]^2+(1+\theta_2)^2[(1-\theta_2)^2+\theta_1^2]}{(1+\theta_2)^2[(1-\theta_2)^2-\theta_1^2]^2}\Big\} \\
&\quad + \Big\{\frac{-2(1-\theta_2^2)[(1-\theta_2)^2-\theta_1^2]}{(1+\theta_2)^2[(1-\theta_2)^2-\theta_1^2]^2}\Big\}.
\end{aligned}$$

$I(\beta)$ is n.n.d. and the 3×3 matrix in the southeast corner of $I_1 - I_2$, $I(\sigma_\varepsilon^2,\theta_1,\theta_2)$, is easily seen to be n.n.d., because $det\{I(\sigma_\varepsilon^2,\theta_1,\theta_2)\} = 0$ and diagonal elements of $I(\sigma_\varepsilon^2,\theta_1,\theta_2)$ are also positive, where $det(.)$ denotes the determinant of a matrix. Therefore, the difference of the information matrices is n.n.d. for all ψ. □

Remark 3.1: In the AR(1) case, we get the following expression for the difference of the information matrices:

$$I_1 - I_2 = \begin{bmatrix} \dfrac{1-\theta_1^2}{\sigma_\varepsilon^2}x_1x_1' & 0 & 0 \\ 0 & \dfrac{2\theta_1^2}{(1-\theta_1^2)^2} & \dfrac{\theta_1}{\sigma_\varepsilon^2(1-\theta_1^2)} \\ 0 & \dfrac{\theta_1}{\sigma_\varepsilon^2(1-\theta_1^2)} & \dfrac{1}{2\sigma_\varepsilon^4} \end{bmatrix},$$

which is also easily seen to be n.n.d. It is a theoretical proof of empirical foundings of Beach and MacKinnon (1978a).

Another measure relating the content of information in two estimation methods is given by

$$Eff(\psi) = \frac{det(I_2)}{det(I_1)}. \tag{3.7}$$

In the following theorem we will show that $Eff(\psi)$ is always smaller than one, but for $T \to \infty$ this measure converges to its upper bound, provided some sufficient conditions are satisfied.

Theorem 3.2.

i) $Eff(\psi) < 1$ for all ψ. (3.8)

ii) If $\lim_{T\to\infty} \frac{1}{T} X'R_2'R_2X = Q$ and Q is positive definite, then $\lim_{T\to\infty} Eff(\psi) = 1$.

Proof: i) By a similar reasoning which led to (3.6) it follows that

$$det(I_1) = (\sigma_\varepsilon^2)^{-(k+3)} \; det(X'R_1'R_1X) \; det(I_{GLS}) \tag{3.9}$$

and

$$det(I_2) = (\sigma_\varepsilon^2)^{-(k+3)} \; det(X'R_2'R_2X) \; det(I_{CO}) \tag{3.10}$$

which gives

$$[Eff(\psi)]^{-1} = \frac{det(X'R_1'R_1X) \cdot det(I_{GLS})}{det(X'R_2'R_2X) \cdot det(I_{CO})},$$

where I_{CO} and I_{GLS} are 3×3 matrices in the south east corner of I_2 and I_1, respectively.

Therefore, $\frac{det(I_{GLS})}{det(I_{CO})} =$

$$\left\{1 + \frac{1}{T-2}\left[\frac{1-\theta_2}{1+\theta_2} + \frac{2(1-\theta_2^2)}{(1-\theta_2)^2-\theta_1^2}\right] + \frac{2(1+\theta_2^2)}{(T-2)^2[(1-\theta_2)^2-\theta_1^2]}\right\}. \tag{3.11}$$

Observe that $Eff(\psi)$ does not depend on β and σ_ε^2.

Using the following relation

$$X'R_1'R_1X = X'R_2'R_2X + \frac{\sigma_\varepsilon^2}{\sigma_u^2} x_1x_1' + (1-\theta_2^2)(x_2-\rho_1x_1)(x_2-\rho_1x_1)',$$

we obtain

$$\begin{aligned}
det(X'R_1'R_1X) \\
&= det\left[X'R_2'R_2X + \frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1x_1'\right] \\
&\quad \cdot \Big\{1 + (1-\theta_2^2)(x_2-\rho_1x_1)'\big[X'R_2'R_2X + \frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1x_1'\big]^{-1}(x_2-\rho_1x_1)\Big\} \\
&= det(X'R_2'R_2X)\Big\{1 + \frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1'(X'R_2'R_2X)^{-1}x_1\Big\} \\
&\quad \cdot \Big\{1 + (1-\theta_2^2)(x_2-\rho_1x_1)'\big[X'R_2'R_2X + \frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1x_1'\big]^{-1}(x_2-\rho_1x_1)\Big\}.
\end{aligned}$$

Therefore

$$\begin{aligned}
[Eff(\psi)]^{-1} \\
&= \Big\{1 + \frac{1}{T-2}\Big[\frac{1-\theta_2}{1+\theta_2} + \frac{2(1-\theta_2^2)}{(1-\theta_2)^2-\theta_1^2}\Big] + \frac{2(1+\theta_2^2)}{(T-2)^2[(1-\theta_2)^2-\theta_1^2]}\Big\} \\
&\quad \cdot\Big\{1 + \frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1'(X'R_2'R_2X)^{-1}x_1\Big\} \\
&\quad \cdot\Big\{1 + (1-\theta_2^2)(x_2-\rho_1x_1)'\big[X'R_2'R_2X + \frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1x_1'\big]^{-1}(x_2-\rho_1x_1)\Big\}. \qquad (3.12)
\end{aligned}$$

$[Eff(\psi)]^{-1}$ is the product of three factors each of which larger than 1, and the inequlity in (3.8) is established.

ii) If the $\lim_{T\to\infty}\frac{1}{T}X'R_2'R_2X = Q$, and Q is positive definite, the first factor of (3.12) converges to 1 if $T \to \infty$. Also

$$\lim_{T\to\infty}\Big\{1 + \frac{\sigma_\varepsilon^2}{\sigma_u^2}\frac{1}{T}x_1'\big(\frac{1}{T}X'R_2'R_2X\big)^{-1}x_1\Big\} = 1.$$

Thus

$$\lim_{T\to\infty}\Big\{1 + (1-\theta_2^2)(x_2-\rho_1x_1)'\frac{1}{T}\big[\frac{1}{T}(X'R_2'R_2X + \frac{\sigma_\varepsilon^2}{\sigma_u^2}x_1x_1')\big]^{-1}(x_2-\rho_1x_1)\Big\} = 1.$$

□

4. CONCLUSIONS

In the preceding chapter we analyze the relative efficiency of COE and $GLSE$ in terms of the information matrix criterion and the ratio of their determinants. In this paper we have shown that the COE is inferior to the $GLSE$ when the unknown parameters are estimated by maximum likelihood method. It has been also found that the COE has the same efficiency as the $GLSE$, as the sample size increases. However, it must be pointed out that a significant loss in efficiency may be incurred if the COE is used in finite sample.

REFERENCES

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

Beach, C.M. and MacKinnon, J.G. (1978b). "Full maximum likelihood estimation of second order autoregressive error models," *Journal of Econometrics,* **7**, 187-198.

Bruce, C.J. (1975). "The wage-tax spiral; Canada 1953-1970," *The Economic Journal,* **85**, 372-376.

Cochrane, D. and Orcutt, G.H. (1949). "Application of least squares regressions to relationships containing autocorrelated error terms," *Journal of American Statistical Association,* **44**, 32-61.

Greene, W.H. (1997). *Econometric Analysis,* Prentice-Hall, London.

Hoque, A. (1989). "The efficiency of the Cochrane-Orcutt procedure," *Australian Journal of Statistics,* **31**, 385-392.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepoel, H and Lee, T.C.(1985). *The Theory and Practice of Econometrics,* John Wiley, New York.

Kadiyala, K.R. (1968). "A transformation used to circumvent the problem of autocorrelation," *Econometrica,* **36**, 93-96.

Krämer, W. (1982). "Note on estimating linear trend when residuals are autocorrelated," *Econometrica,* **50**, 1065-1067.

Lempers, F.B. and Kloek, T. (1973). "On a simple transformation for second order autocorrelated disturbances in regression analysis," *Statistica Neerlandica,* **27**, 69-75.

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

Oxley, L.T. and Robert, C.T. (1982). "Pitfalls in the application of the Cochrane Orcutt technique," *Oxford Bulletin of Econometrics and Statistics,* **44** 227-240.

Puterman, M.L. (1988). "Leverage and influence in autocorrelated regression-models," *Applied Statistics,* **37**, 76-86.

Stemann, D. and Trenkler, G. (1993). "Leverage and Cochrane-Orcutt estimation in the linear regression," *Communications in Statistics: Theory and Method,* **22**, 1315-1333.