NOTE ON THE "ASYMPTOTIC MINIMUM VARIANCE PROPERTY" OF LEAST SQUARES ESTIMATORS IN TIME-SERIES REGRESSION

MRS. V. MUKHERJEE, POONA[1]

(Received in March, 1967)

1. In his paper on "Estimation in time-series regression" Durbin[2] has given the following results for the least squares estimators of the parameters β in the regression equation

$$Y = X\beta + \epsilon$$

where Y is an $n \times 1$ vector

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

X is the $n \times k$ matrix

$$\begin{bmatrix} X_{11} \ldots\ldots\ldots X_{1k} \\ X_{n1} \ldots\ldots\ldots X_{nk} \end{bmatrix}$$

of n sets of observations of the k variates $X_1, \ldots, X_k$ some of which may be lagged Y's the rank of X not being less than k ;

and ϵ is the error-variable vector with

$$E(\epsilon) = 0 \qquad \ldots(1)$$

and

$$E(\epsilon\epsilon') = \sigma^2 I.$$

Prior to obtaining the results Durbin first considers the set of unbiased linear estimating equations

$$t_1 b + t_2 = 0 \qquad \ldots(2)$$

1. At present at Shivaji University Kolhapur.

2. Durbin, J. : Estimation in Time Series Regression. J. Royal Stat. Soc, Ser. B., Vol. 22, No. 1, pp. 139-153, 1960.

for obtaining the estimators b of the parameters β where t_1 and t_2 are a $k \times k$ matrix and $k \times 1$ vector respectively of functions of all the variables involved and

$$E(t_1)=I$$

and

$$E(t_1\beta+t_2)=0 \qquad \ldots(3)$$

He shows that

$$V(t_1\beta+t_2)-g^{-1}$$

is a positive definite or semi-definite matrix, g being the information matrix. g^{-1} is therefore the minimal variance-matrix for all $t_1\beta+t_2$ satisfying conditions (3).

If $\tau_1\beta+\tau_2$ is any other estimating equation satisfying condition (3) and if

$$V(t_1\beta+t_2)-V(\tau_1\beta+\tau_2)$$

is a negative definite or semidefinite matrix for all such $\tau_1\beta+\tau_2$, Durbin calls the equation (2) the best unbiased linear estimating equation of β.

As a particular case Durbin considers the case when the regression relationship (1) holds and

(*i*) ϵ's are distributed normally,

or (*ii*) ϵ's are non-normal and have finite moments.

In case (*i*) (1) the maximum-likelihood estimates are given by

$$M^{-1}(X'X\beta-X'Y)=0 \qquad \ldots(4)$$

where

$$E(X'X)=M,$$

and the information-matrix is given by M/σ^2.

Further the variance-matrix of $M^{-1}(X'X\beta-X'Y)$ is

$$\sigma^2 M^{-1}=g^{-1}.$$

Hence according to Durbin's definitions the maximum-likelihood equations are the best unbiased linear estimating equations for all sample sizes. (2) If further $M^{-1}X'X$ converges stochastically to I and a diagonal matrix δ_n depending on n only exists such that

$$\delta_n M^{-1}\delta_n \rightarrow W, \text{ a finite (p.d.) matrix}$$

as

$$n \rightarrow \infty$$

$$\delta_n(b-\beta) \quad \text{and} \quad -\delta_n M^{-1}(X'X\beta-X'Y)$$

have the same limiting distributions as $n \to \infty$ provided they exist. In particular $\delta_n(b-\beta)$ has 0 as mean and $\sigma^2 W$ as variance-matrix in the limit as $n \to \infty$.

$$S^2=(y-Xb)'\,(y-Xb)/n-k$$

is again an asymptotically unbiased estimator of σ^2.

In case (*ii*) when the ϵ's are non-normal it is still true that the least squares estimators are given by the unbiased linear estimating equations

$$M^{-1}(X'Xb-X'Y)=0 \qquad \ldots(5)$$

with $E[M^{-1}(XX'\beta-X'Y)]=0$

and $V[M^{-1}(XX'\beta-X'Y)]=\sigma^2 M^{-1}$.

Durbin further shows that results as of case (*i*) are valid here also.

2. In this note a further property of asymptotic minimum variance is proved for the least squares estimators of β in the case of non-normal errors with finite moments.

The estimating equations for β given by (5) are a particular case of linear estimating equations given by

$$Bb-Ay=0 \qquad \ldots(6)$$

with $E(B)=\text{I}$

where A and B are matrices of functions of elements in X. Let these equations be unbiased estimating equations.

Then

$$E(B\beta-Ay)=0$$

or $E[(B-AX)\beta-A\epsilon]=0 \qquad \ldots(7)$

Hence $E(B-AX)=0$ or $E(AX)=E(B)=1 \qquad \ldots(8)$

and $E(A\epsilon)=0 \qquad \ldots(9)$

Conditions (9) are satisfied if ϵ_r is uncorrelated with elements in the r^{th} column of A for all r.

Let us further assume that $B=AX \qquad \ldots(10)$

and AX converges to I stochastically $\qquad \ldots(11)$

and ϵ_r is distributed independently of elements in the first r columns of A. $\qquad \ldots(12)$

Because of assumptions (10)

$$(B-AX)\beta-A\epsilon=-A\epsilon.$$

Hence $$V(B\beta - Ay) = V(-A\epsilon) = \sigma^2 E(AA') \quad ...(13)$$

because of assumptions (12).

But $E(AA')$

$$= E[((A - M^{-1}X') + M^{-1}X')((A - M^{-1}X') + M^{-1}X')']$$
$$= E[(A - M^{-1}X')(A - M^{-1}X')' + (A - M^{-1}X')XM^{-1} + M^{-1}X'(A' - XM^{-1}) + M^{-1}X'XM^{-1}$$
$$= E[(A - M^{-1}X')(A - M^{-1}X')'] + M^{-1},$$

since $E(AX) = 1$

and $E(X'X) = M$.

$(A - M^{-1}X')(A - M^{-1}X')'$ has non-negative diagonal elements. Hence

$E[(A - M^{-1}X')(A - M^{-1}X')']$ has non-negative diagonal elements.

Hence $E(AA')$ has minimum diagonal elements when $A - M^{-1}X^1 = 0$ and by condition (10) $B = M^{-1}X'X$, so that the equations (6) reduce to the l.h.s. of equation (5). Hence the elements of $B\beta - Ay$ have minimum variance when $A = M^{-1}X'$ and $B = M^{-1}X'X$.

Hence since by condition (11)B converges stochastically to I, elements in $\delta_n(B\beta - Ay)$ and hence in $-\delta_n(b - \beta)$ have minimum variances asymptotically when $A = M^{-1}X'$ and $B = M^{-1}X'X$ that is when b are the least squares estimators. It can similarly be shown that $\delta_n C(b - \beta)$ when C is a $k' \times k$ matrix of arbitrary constants and $k' \leqslant k$ to have asymptotically minimum diagonal elements when b are the least squares estimators of β among estimators given by (6) satisfying conditions (8) to (12).

Since the mini mum-variance property considered here is only asymptotic it follows that if

$$A_1 = M^{-1}X' + A_2$$

and

$$B_1 = A_1 X$$

where $E(A_2 X) = 0$ and $A_2 X$

converges stochistically to 0 then the estimators given by $B_1 b - A_1 y = 0$ have the same asymptotic properties as the least square estimators.

It should be noted that A in equations (6) need not have functions of elements in X only. The argument and the results are still valid if A contains some instrumental variables satisfying conditions (8), (9), (10), (11) and (12).