COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 368

Autocorrelation of the Disturbances in Linear Regression

John Gurland

November 13, 1951

1. Introduction.

The estimates of the regressive parameters in linear regression obtained by least squares are unbiassed whether or not the covariance matrix of the disturbances is correctly specified. On the other hand, the corresponding loss of efficiency may be far from negligible. A particular instance of incorrect specification of this covariance matrix which has received some attention recently is the situation where the disturbances are assumed to be independent, with common variance, when in fact they are autocorrelated by a Markoff process. By means of artificially constructed samples, Cochrane and Orcutt ([1], [2]) have investigated the loss of efficiency for a special linear regression equation, and a special system of two linear equations. A recent discussion on testing for autocorrelation in the disturbance of a linear regression equation is given by Durbin and Watson ([3], [4]).

The present paper is also concerned with the particular instance where the disturbances are distributed according to a Markoff process. Two situations are investigated. First, the disturbance is assumed to be stationary when in reality it is not. Second, in a two-equation system (reduced form), some consequences of falsely assuming the same process (i.e., same autoregressive parameters in the Markoff process) for the disturbances of both equations are exhibited. Finally, in the last section of the paper, a few remarks are made regarding the aforementioned papers of Cochrane and Orcutt.

2. Least-squares procedure under linear transformations.

Let the random variables

$$y = (y_1, y_2, \ldots, y_n) \tag{1}$$

have the expected values

$$\xi = Ey = \theta A \tag{2}$$

where

$$\theta = (\theta_1, \theta_2, \ldots, \theta_k), \qquad k < n \tag{3}$$

are k unknown parameters, and

$$A = [a_{ij}] \qquad \begin{array}{l} i = 1, 2, \ldots, k \\ j = 1, 2, \ldots, n \end{array} \tag{4}$$

is a matrix of rank k the elements of which are known and fixed.* Suppose the disturbances

$$u = y - \xi \tag{5}$$

have the covariance matrix

$$Eu'u = \sigma^2 \Omega \tag{6}$$

where Ω is nonsingular. Then, as Aitken [5] has shown, the best linear unbiassed estimators $\hat{\theta}$ of θ are obtained by minimizing

$$(y - \xi)\, \Omega^{-1}\, (y - \xi)' \tag{7}$$

---

* This is the fixed variate case. If the elements of A are regarded as random, the expectation (2) is conditional, and the disturbances $y - \xi$ are independent of the elements of A.

with respect to θ, where we assume, of course, that the elements of Ω are known.*

Now let G be such a matrix that

$$G' \Omega G = I_n \tag{8}$$

where I_n is the unit nxn matrix; and set

$$\begin{aligned} Y &= y\,G \\ \eta &= \xi\,G \end{aligned} \tag{9}$$

Then it can be shown (cf. Gurland [6]) that the minimization of

$$(Y - \eta)(Y - \eta)' \tag{10}$$

with respect to θ yields the same estimators as obtained by minimizing (7).

An expression for the joint efficiency of unbiassed estimators $\hat{\theta}^{(\Lambda)}$ which are obtained by minimizing the quadratic form

$$(y - \xi)\Lambda.(y - \xi)' \tag{11}$$

will be used in the sequel. This expression can be derived by referring to formula (68) in Gurland [6]. The joint efficiency of $\hat{\theta}^{(\Lambda)}$ is given by**

$$\begin{aligned} \text{Eff.}(\hat{\theta}^{(\Lambda)}) &= \frac{\det(A\,\Omega^{-1}A')^{-1}}{\det[(A\Lambda A')^{-1}\,A\Lambda\,\Omega\,\Lambda' A'(A\Lambda A')^{-1}]} \\ &= \frac{\det^2(A\Lambda A')}{\det(A\Lambda\,\Omega\,\Lambda\,A')\,\det(A\,\Omega^{-1}\,A')} . \end{aligned} \tag{12}$$

Applying the transformation (9) with G given by

$$G'\Lambda^{-1}G = I_n \qquad , GG' = \Lambda$$

---

* For the case some or all of the elements of Ω are not known, the author [6] has proposed a method of estimation which does not rely on the assumption of normality of u.

** Considering only the class of linear unbiassed estimators, this expression for joint efficiency is in agreement with that of Cramer [8]; since the ellipsoid $b(A\,\Omega^{-1}A')^{-1}b' = k + 2$ is contained in the ellipsoid $b[(A\Lambda\,A')^{-1}\,A\Lambda\Omega\Lambda'A'(A\Lambda A')^{-1}]b' = k + 2$.

then

$$\text{Eff.}(\hat{\theta}^{(\Lambda)}) = \frac{\det^2(BB')}{\det(BMB')\det(BM^{-1}B')} \tag{13}$$

where

$$G'\Omega G = I$$
$$AG = B \tag{14}$$

3. Linear transformation in two special cases.

Consider the first-order Markoff process

$$u_t - \rho u_{t-1} = v_t \qquad t = 1, 2, \ldots, n \tag{15}$$

where $v_1, v_2, \ldots, v_n$ are independent with common mean zero, and variance σ^2, and

$$E\, v_t u_{t-1} = 0 \qquad t = 1, 2, \ldots, n. \tag{16}$$

Supposing u_0 has* mean zero and variance $\lambda^2\sigma^2$, the covariance matrix of u becomes

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} g^2 & \rho g^2 & \rho^2 g^2 & \cdots & \rho^{n-1}g^2 \\ \rho g^2 & 1+\rho^2 g^2 & \rho(1+\rho^2 g^2 & & \rho^{n-2}(1+\rho^2 g^2) \\ \rho^2 g^2 & \rho(1+\rho^2 g^2) & 1+\rho^2+\rho^4 g^2 & & \rho^{n-2}(1+\rho^2+\rho^4 g^2) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1}g^2 & \rho^{n-2}(1+\rho^2 g^2) & \rho^{n-3}(1+\rho^2+\rho^4 g^2) & & 1+\rho^2+\rho^4+\ldots\rho^{2n-2}g^2 \end{bmatrix} \tag{17}$$

where

$$g^2 = 1+\rho^2\lambda^2 \tag{18}$$

---

* It would be equivalent here to take $u_0 = \lambda v_0$ and define v_t for $t = 0, 1, 2, \ldots, n$.

It can be verified that the following matrix G satisfies (8).

$$G = \begin{bmatrix} \frac{1}{g} & -\rho & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\rho & & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 \\ 0 & 0 & 0 & & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & -\rho & 0 \\ 0 & 0 & 0 & & 1 & -\rho \\ 0 & 0 & 0 & & 0 & 1 \end{bmatrix} \qquad (19)$$

Note that for $\lambda = 0$, the process (15) has a fixed initial value $u_o = 0$, whereas if $\lambda^2 = \frac{1}{1-\rho^2}$ $\quad |\rho| < 1$ the process is stationary.*

The second case for which a matrix G will be found satisfying (8) is the following, to which we shall refer later in the discussion of a two equation system. Consider the two Markoff processes

$$u_{1,t} - \rho_1 u_{1,t-1} = v_{1t} \qquad (20)$$

$$u_{2,t} - \rho_2 u_{2,t-1} = v_{2t} \qquad (21)$$

each of which has the same general specification as (15) with the additional** conditions

$$\begin{aligned} u_{10} &= \lambda_1 v_{10} \\ u_{20} &= \lambda_2 v_{20} \end{aligned} \qquad (22)$$

---

* Here we mean stationary in the wide sense. (Cf. Doob [7]).

** See Footnote on page 4.

$$E\, v_{1t}\, v_{2\tau} = K\,\sigma^2 \quad , \text{ when } t = \tau = 0, 1, 2, \ldots, n. \qquad\qquad (23)$$
$$= 0 \quad , \text{ when } t \neq \tau .$$

In this case, the covariance matrix of

$$u = (u_{11}, \ldots, u_{1n}, u_{21}, \ldots, u_{2n}) \qquad (24)$$

is given by

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{bmatrix} \qquad (25)$$

where

$$\Omega_{ii} = \begin{bmatrix} g_i^2 & \rho_i g_i^2 & \cdots\cdots\cdots & \rho_i^{n-1} g_i^2 \\ \rho_i g_i^2 & 1+\rho_i^2 g_i^2 & & \rho_i^{n-2}(1+\rho_i^2 g_i^2) \\ \rho_i^2 g_i^2 & \rho_i(1+\rho_i^2 g_i^2) & & \rho_i^{n-3}(1+\rho_i^2+\rho_i^4 g_i^2) \\ \vdots & \vdots & & \vdots \\ \rho_i^{n-1} g_i^2 & \rho_i^{n-2}(1+\rho_i^2 g_i^2) & & 1+\rho_i^2+\rho_i^4+\cdots+\rho_i^{2n-2} g_i^2 \end{bmatrix} \qquad (26)$$

$i = 1, 2.$

$$\Omega_{12} = K \begin{bmatrix} h & \rho_2 h & \rho_2^2 h & \cdots\cdots & \rho_2^{n-1} h \\ \rho_1 h & 1+\rho_1\rho_2 h & \rho_2(1+\rho_1\rho_2 h) & & \rho_2^{n-2}(1+\rho_1\rho_2 h) \\ \rho_1^2 h & \rho_1(1+\rho_1\rho_2 h) & 1+\rho_1\rho_2+\rho_1^2\rho_2^2 h & & \rho_2^{n-3}(1+\rho_1\rho_2+\rho_1^2\rho_2^2 h) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_1^{n-1} h & \rho_1^{n-2}(1+\rho_1\rho_2 h) & \rho_1^{n-3}(1+\rho_1\rho_2+\rho_1^2\rho_2^2 h) & & 1+\rho_1\rho_2+\cdots+(\rho_1\rho_2)^{2n-2}(1+\rho_1\rho_2 h) \end{bmatrix} \tag{27}$$

$$g_i^2 = 1 + \rho_i^2 \lambda_i^2 \qquad i = 1,\ 2. \tag{28}$$

$$h = 1 + \rho_1 \rho_2 \lambda_1 \lambda_2 \quad . \tag{29}$$

Taking

$$Q_1 = \left[\begin{array}{ccccc|ccccc} \frac{1}{g_1} & -\rho_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & & 0 \\ 0 & 1 & -\rho_1 & & 0 & \vdots & \vdots & & & \vdots \\ 0 & 0 & 1 & & 0 & & & & & \\ 0 & 0 & 0 & & 0 & & & & & \\ \vdots & \vdots & \vdots & & \vdots & & & & & \\ 0 & 0 & 0 & & -\rho_1 & & & & & \\ 0 & 0 & 0 & & 1 & 0 & 0 & \cdots & & 0 \\ \hline 0 & 0 & 0 & \cdots & 0 & \frac{1}{g_2} & -\rho_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & 0 & 1 & -\rho_2 & \cdots & 0 \\ & & & & & 0 & 0 & 1 & \cdots & 0 \\ & & & & & 0 & 0 & 0 & \cdots & 0 \\ & & & & & \vdots & \vdots & \vdots & & \\ & & & & & 0 & 0 & 0 & & -\rho_1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & & 1 \end{array}\right] \tag{30}$$

it can be verified that

$$G_1' \Omega G_1 = \left[\begin{array}{ccc|ccc} 1 & & 0 & K & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & K \\ \hline K & & 0 & 1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & K & 0 & & 1 \end{array}\right] \qquad (31)$$

Letting

$$H = G_1' \Omega G_1 \qquad (32)$$

and defining G_2 by

$$G_2 = \frac{1}{\sqrt{1-K^2}} \left[\begin{array}{ccc|ccc} \sqrt{1-K^2} & & & -K & & \\ & \ddots & & & \ddots & \\ & & \sqrt{1-K^2} & & & -K \\ \hline 0 & \cdots & 0 & 1 & & \\ \vdots & & \vdots & & \ddots & \\ 0 & \cdots & 0 & & & 1 \\ \vdots & & \vdots & & & \ddots \\ 0 & & 0 & & & 1 \end{array}\right] \qquad (33)$$

it may also be verified that

$$G_2' \, H \, G_2 = I_{2n} \qquad (34)$$

Setting

$$k = \sqrt{1-K^2} \qquad (35)$$

note that

$$G_1 G_2 = \frac{1}{k} \left[\begin{array}{cccccc:cccccc}
\frac{k}{g_1} & -\rho_1 k & 0 & \cdots & \cdots & 0 & \frac{-K}{g_1} & \rho_1 K & 0 & \cdots & \cdots & 0 \\
0 & k & -\rho_1 k & & & \vdots & 0 & -K & \rho_1 K & & & 0 \\
0 & 0 & k & & & \vdots & 0 & 0 & -K & & & 0 \\
0 & 0 & 0 & & & \vdots & 0 & 0 & 0 & & & \vdots \\
\vdots & \vdots & \vdots & & & -\rho_1 k & \vdots & \vdots & \vdots & & & \vdots \\
0 & 0 & 0 & & & & 0 & 0 & 0 & & & \rho_1 K \\
0 & 0 & 0 & & & k & 0 & 0 & 0 & & & -K \\
\hdashline
0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{1}{g_2} & -\rho_2 & 0 & & & 0 \\
\vdots & & & & & \vdots & 0 & 1 & -\rho_2 & & & 0 \\
\vdots & & & & & \vdots & 0 & 0 & 1 & & & 0 \\
\vdots & & & & & \vdots & 0 & 0 & 0 & & & 0 \\
\vdots & & & & & \vdots & \vdots & \vdots & \vdots & & & \vdots \\
\vdots & & & & & \vdots & 0 & 0 & 0 & & & -\rho_2 \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 & 0 & 0 & & & 1
\end{array}\right] \qquad (36)$$

4. Consequences of incorrect specification of initial conditions.

Suppose the disturbance u_t is distributed according to the process (15) with covariance matrix (17). We shall now obtain an expression for the joint efficiency of the least-squares estimators $\hat{\theta}^{(\lambda)}$, where* λ is arbitrary (> 0), when in fact the true value of λ is $\mathring{\lambda}$. The true covariance matrix** of the disturbances will, of course, be given by (17), with $\lambda = \mathring{\lambda}$. Denote this true covariance matrix by $\mathring{\Omega}$.

Let

$$A = \begin{pmatrix} x_1 & x_2 & \cdots\cdots & x_n \\ 1 & 1 & \cdots\cdots & 1 \end{pmatrix} \qquad (37)$$

---

* The notation $\hat{\theta}^{(\lambda)}$ used here is more suggestive than the notation $\hat{\theta}^{(\Lambda)}$ for the estimators obtained from (11). More precisely, $\hat{\theta}^{(\lambda)}$ will now be used to represent the estimators obtained by minimising $(y - \xi)\Omega^{-1}(y - \xi)'$, where Ω is given by (17) with λ arbitrary.

** Here it is assumed the parameter ρ is known; hence the loss of efficiency will be due solely to incorrect specification of λ.

so that the regression equation becomes

$$y_t = x_t\,\theta_1 + \theta_2 + u_t \quad . \tag{38}$$

Applying the transformation

$$Y = y\,G$$

with G given by (19) and arbitrary λ, we obtain

$$M = G'\mathring{\Omega}\,G = \begin{bmatrix} \frac{\mathring{g}^2}{g^2} & 0 & 0 & \cdots\cdots\cdots & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & 0 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & 1 \end{bmatrix} \tag{39}$$

where

$$\mathring{g}^2 = 1 + \rho^2\,\mathring{\lambda}^2 \; . \tag{40}$$

Note that, for the same λ in G and Ω,

$$G'\Omega\,G = I_n$$

so that formula (13) for the joint efficiency of $\hat{\theta}_1^{(\lambda)}$, $\hat{\theta}_2^{(\lambda)}$ is applicable. It now remains to find, explicitly, $\det(BB')$, $\det(BMB')$, $\det BM^{-1}B'$.

$$B = AG = \begin{bmatrix} \frac{x_1}{g}, & x_2 - \rho x_1, & \cdots\cdots\cdots\cdots, & x_n - \rho x_{n-1} \\ \frac{1}{g}, & 1-\rho, & \cdots\cdots\cdots\cdots, & 1-\rho \end{bmatrix} \; . \tag{41}$$

Let $\quad s_1^{(\rho)} = \sum_{t=2}^{n} (x_t - \rho x_{t-1})$ (42)

$$s_2^{(\rho)} = \sum_{t=2}^{n} (x_t - \rho x_{t-1})^2 \; . \tag{43}$$

Then

$$BB' = \begin{bmatrix} \frac{x_1}{g^2} + s_2^{(\rho)} & \frac{x_1}{g^2} + (1-\ )s_1^{(\rho)} \\ \frac{x_1}{g^2} + (1-\rho)s_1^{(\rho)} & \frac{1}{g^2} + (n-1)(1-\rho)^2 \end{bmatrix} \qquad (44)$$

$$BMB' = \begin{bmatrix} \frac{x_1}{g^4}\overset{\circ}{g}{}^2 + s_2^{(\rho)} & \frac{x_1}{g^4}\overset{\circ}{g}{}^2 + (1-\ )s_1^{(\rho)} \\ \frac{x_1}{g^4}\overset{\circ}{g}{}^2 + (1-\rho)s_1^{(\rho)} & \frac{\overset{\circ}{g}{}^2}{g^4} + (n-1)(1-\rho)^2 \end{bmatrix} \qquad (45)$$

$$BM^{-1}B' = \begin{bmatrix} \frac{x_1}{\overset{\circ}{g}{}^2} + s_2^{(\rho)} & \frac{x_1}{\overset{\circ}{g}{}^2} + (1-\rho)s_1^{(\rho)} \\ \frac{x_1}{\overset{\circ}{g}{}^2} + (1-\rho)s_1^{(\rho)} & \frac{1}{\overset{\circ}{g}{}^2} + (n-1)(1-\rho)^2 \end{bmatrix} \qquad (46)$$

Let

$$A_o = \det \begin{bmatrix} s_2^{(\rho)} & (1-\rho)s_1^{(\rho)} \\ (1-\rho)s_1^{(\rho)} & (n-1)(1-\rho)^2 \end{bmatrix} \qquad (47)$$

$$A_1 = \det \begin{bmatrix} x_1 & (1-\rho)s_2^{(\rho)} \\ 1 & (n-1)(1-\rho)^2 \end{bmatrix} \qquad (48)$$

$$A_2 = \det \begin{bmatrix} s_2^{(\rho)} & x_1 \\ (1-\rho)s_1^{(\rho)} & 1 \end{bmatrix} \qquad (49)$$

Then

$$\text{Eff.}(\hat{\theta}_1^{(\lambda)}, \hat{\beta}_2^{(\lambda)}) = \frac{\left[A_o + \frac{x_1}{g^2}A_1 + \frac{1}{g^2}A_2\right]^2}{\left[A_o + x_1\frac{\overset{\circ}{g}{}^2}{g^4}A_1 + \frac{\overset{\circ}{g}{}^2}{g^4}A_2\right]\left[A_o + \frac{x_1}{\overset{\circ}{g}{}^2}A_1 + \frac{1}{\overset{\circ}{g}{}^2}A_2\right]} \qquad (50)$$

It is apparent from the expression (50) that the efficiency depends not only on the discrepancy between λ and $\overset{\circ}{\lambda}$ but also on the values of ρ and the elements of A. This would, of course, be expected, but the above expression indicates precisely how all these quantities are involved. The following values and limits of (50) are informative

$$\text{Eff.}(\hat{\theta}^{(\lambda)}) = 1 \quad \text{when } x_t - \rho x_{t-1} = 0, \quad x_1 \neq 0, \quad \rho \neq 1. \tag{51}$$

$$\lim_{x_1 \to \infty} \text{Eff.}(\hat{\theta}^{(\lambda)}) = 1 \tag{52}$$

$$\lim_{\lambda \to \infty} \text{Eff.}(\hat{\theta}^{(\lambda)}) = \frac{1}{1 + \frac{x_1}{\overset{\circ}{g}^2} \frac{A_1}{A_o} + \frac{A_2}{A_o} \frac{1}{\overset{\circ}{g}^2}} \tag{53}$$

$$\lim_{x_1 \to \infty} \lim_{\lambda \to \infty} \text{Eff.}(\hat{\theta}^{(\lambda)}) = 0 \tag{54}$$

$$\lim_{\rho \to 1} \lim_{\lambda \to \infty} \text{Eff.}(\hat{\theta}^{(\lambda)}) = 0 \tag{55}.$$

The above limits indicate that the values of x_1 and ρ are important in evaluating the efficiency for different values of λ when $\overset{\circ}{\lambda}$ is the true value. It is clear the efficiency can be made arbitrarily close to zero by using the properties signified by (54) and (55), and to one by (51) and (52).

It should be remarked that in deriving (54) and (55) we assume ρ and λ are not functionally related, although the same limits might hold for particular relationships chosen.

As a special instance of a functional relationship, consider

$$\lambda^2 = \frac{1}{1 - \rho^2}, \qquad |\rho| < 1. \tag{56}$$

This is equivalent to assuming the process u_t is stationary in the wide*

---

* If the process is normal, stationarity in the wide sense is equivalent to stationarity in the strict sense.

sense. It is interesting to see what damage results in assuming u_t to be stationary when in fact it has an initial* fixed value $u_o = 0$. This case is interesting because Cochrane and Orcutt [1] (appendix, p.57) establish the validity of their transformation on the assumption of stationarity, whereas in their artificially constructed series they take** $u_o = 0$. The damage when u_o has an initial stochastic value such that $\dot{\lambda} \neq 0$ can also be exhibited by means of (50), but, for brevity, we consider here only the case $\dot{\lambda} = 0$, $\lambda^2 = \frac{1}{1-\rho^2}$.

Referring to (50) and setting $\dot{\lambda} = 0$, $\lambda^2 = \frac{1}{1-\rho^2}$, we obtain

$$\text{Eff.}(\hat{\theta}_1^{(\lambda)}, \hat{\theta}_2^{(\lambda)}) = \frac{[A_o + x_1(1-\rho^2)A_1 + (1-\rho^2)A_2]^2}{[A_o + x_1(1-\rho^2)^2 A_1 + (1-\rho^2)^2 A_2][A_o + x_1 A_1 + A_2]} \tag{57}$$

Note that for $\rho = 0$, the efficiency is unity, as would obviously be expected. Also,

$$\lim_{\rho \to 1} \text{Eff.}(\hat{\theta}_1^{(\lambda)}, \hat{\theta}_2^{(\lambda)}) = \frac{1}{1 + \frac{n}{4}\left(1 - \frac{s_1^{(1)2}}{n\, s_2^{(1)}}\right)} \tag{58}$$

Now

$$0 \leq \frac{s_1^{(1)2}}{n\, s_2^{(1)}} \leq 1 \tag{59}$$

---

* Investigating the bias of least-squares estimators in a stochastic difference equation, Hurwicz [9] considers separately these two initial conditions of the process.

** In their empirical series, the values $(x_1, x_2, \ldots, x_n)$ are also random. Our investigation here refers to the fixed variate case.

The lower bound is attained when $a_1^{(1)} = 0$, i.e., when $x_n = x_1$. The upper bound is attained when $x_2 - x_1 = x_3 - x_2 = \ldots\ldots = x_n - x_{n-1}$. Hence the loss of efficiency for falsely assuming stationarity when $\lambda = 0$, and ρ is close to 1 is expressed by the inequalities

$$\frac{4}{n+4} \leq \lim_{\rho \to 1} E(\hat{\theta}_1^{(\lambda)} \hat{\theta}_2^{(\lambda)}) \leq 1 \tag{60}$$

Note that this lower bound decreases as n increases. This is not incompatible with variances that decrease as n increases.

5. Incorrect specification of autoregressive parameters in a two equation system.

A simple two equation system will be considered as an extension of the single regression equation of § 4, wherein each equation contains two regressive parameters (and one autoregressive parameter for the disturbance process). The method herein described works more generally; this particular model is chosen for simplicity.

The following notation extends naturally from the one-equation case.

$$y_1 = (y_{11}, y_{12}, \ldots, y_{1n}) \qquad y_2 = (y_{21}, y_{22}, \ldots, y_{2n}) \tag{61}$$

$$\xi_1 = (\xi_{11}\ \xi_{12}, \ldots, \xi_{1n}) \qquad \xi_2 = (\xi_{21}\ \xi_{22}, \ldots, \xi_{2n}) \tag{62}$$

$$u_1 = (u_{11}, u_{12}, \ldots, u_{1n}) \qquad u_2 = (u_{21}, u_{22}, \ldots, u_{2n}) \tag{63}$$

$$y = (y_1, y_2) \qquad \xi = (\xi_1, \xi_2) \qquad u = (u_1, u_2) \tag{64}$$

$$y_1 - \xi_1 = u_1 \qquad y_2 - \xi_2 = u_2 \tag{65}$$

u_1 and u_2 are defined by (20) and (21), with the additional conditions (22) and (23).

Writing A as

$$A = \begin{bmatrix} x_1 & x_2 & \cdots\cdots & x_n & 0 & 0 & \cdots\cdots & 0 \\ 1 & 1 & \cdots\cdots & 1 & 0 & 0 & \cdots\cdots & 0 \\ 0 & 0 & & 0 & x_1 & x_2 & \cdots\cdots & x_n \\ 0 & 0 & \cdots\cdots & 0 & 1 & 1 & \cdots\cdots & 1 \end{bmatrix} \tag{66}$$

the regression equations become

$$\begin{aligned} y_{1t} &= \theta_1 x_t + \theta_2 + u_{1t} \\ y_{2t} &= \theta_3 x_t + \theta_4 + u_{2t} \end{aligned} \tag{67}$$

We are interested here in examining the damage caused by falsely assuming the processes u_{1t} and u_{2t} to have the same autoregressive parameter, i.e., $\rho_1 = \rho_2$. It will therefore be assumed that the values of λ_1 and λ_2 are known correctly, and for convenience we take

$$\lambda_1 = \lambda_2 = \mathring{\lambda}_1 = \mathring{\lambda}_2 = 0. \tag{68}$$

The expression for the joint efficiency will be further simplified by taking

$$\begin{aligned} \rho_1 &= \mathring{\rho}_1 \\ \rho_2 &= \mathring{\rho}_1 \end{aligned} \quad . \tag{69}$$

This is equivalent to the assumption that the value of ρ_1 is accurately known to be $\mathring{\rho}_1$, and ρ_2 is falsely assumed to be equal to $\mathring{\rho}_1$ when in fact $\mathring{\rho}_2$ is the true value of ρ_2, and $\mathring{\rho}_2 \neq \mathring{\rho}_1$.

Now, the covariance matrix of $u = (u_1, u_2)$ is given by (25), with ρ_1, ρ_2, λ_1, λ_2, K all arbitrary. The true covariance matrix $\sigma^2 \Omega$ may therefore be written as follows

$$\sigma^2 \mathring{\Omega} = \sigma^2 \begin{bmatrix} \mathring{\Omega}_{11} & \mathring{\Omega}_{12} \\ \mathring{\Omega}'_{12} & \mathring{\Omega}_{22} \end{bmatrix} \tag{70}$$

where

$$\mathring{\Omega}_{ii} = \begin{bmatrix} 1 & \mathring{\rho}_i & \mathring{\rho}_i^2 & \cdots\cdots & \mathring{\rho}_i^{n-1} \\ \mathring{\rho}_i & 1+\mathring{\rho}_i^2 & \mathring{\rho}_i(1+\mathring{\rho}_i^2) & & \mathring{\rho}_i^{n-2}(1+\mathring{\rho}_i^2) \\ \mathring{\rho}_i^2 & \mathring{\rho}_i(1+\mathring{\rho}_i^2) & 1+\mathring{\rho}_i^2+\mathring{\rho}_i^4 & \cdots\cdots & \mathring{\rho}_i^{n-3}(1+\mathring{\rho}_i^2+\mathring{\rho}_i^4) \\ \vdots & \vdots & \vdots & & \vdots \\ \mathring{\rho}_i^{n-1} & \mathring{\rho}_i^{n-2}(1+\mathring{\rho}_i^2) & \mathring{\rho}_i^{n-3}(1+\mathring{\rho}_i^2+\mathring{\rho}_i^4) & \cdots & 1+\mathring{\rho}_i^2+\cdots+\mathring{\rho}_i^{2n-2} \end{bmatrix} \qquad (71)$$

$i = 1, 2$

$$\mathring{\Omega}_{12} = K \begin{bmatrix} 1 & \mathring{\rho}_2 & \mathring{\rho}_2^2 & \cdots\cdots & \mathring{\rho}_2^{n-1} \\ \mathring{\rho}_1 & 1+\mathring{\rho}_1\mathring{\rho}_2 & \mathring{\rho}_2(1+\mathring{\rho}_1\mathring{\rho}_2) & & \mathring{\rho}_2^{n-2}(1+\mathring{\rho}_1\mathring{\rho}_2) \\ \mathring{\rho}_1^2 & \mathring{\rho}_1(1+\mathring{\rho}_1\mathring{\rho}_2) & 1+\mathring{\rho}_1\mathring{\rho}_2+\mathring{\rho}_1^2\mathring{\rho}_2^2 & & \mathring{\rho}_2^{n-3}(1+\mathring{\rho}_1\mathring{\rho}_2+\mathring{\rho}_1^2\mathring{\rho}_1^2) \\ \vdots & \vdots & \vdots & & \vdots \\ \mathring{\rho}_1^{n-1} & \mathring{\rho}_1^{n-2}(1+\mathring{\rho}_1\mathring{\rho}_2) & \mathring{\rho}_1^{n-3}(1+\mathring{\rho}_1\mathring{\rho}_2+\mathring{\rho}_1^2\mathring{\rho}_2^2) & \cdots & 1+\mathring{\rho}_1\mathring{\rho}_2+\cdots+(\mathring{\rho}_1\mathring{\rho}_2)^{n-1} \end{bmatrix} \qquad (72)$$

By means of the transformation $Y = yG$, with $G = G_1 G_2$ given by (36), viz.

$$G = \frac{1}{k}\left[\begin{array}{cccccc|cccccc}
k & -\mathring{\rho}_1 k & 0 & \cdots & \cdots & 0 & -K & \rho_1 K & 0 & \cdots & \cdots & 0 \\
0 & k & -\mathring{\rho}_1 k & & & \vdots & 0 & -K & \mathring{\rho}_1 K & & & 0 \\
0 & 0 & k & & & \vdots & 0 & 0 & -K & & & \vdots \\
\vdots & \vdots & \vdots & & & \vdots & 0 & 0 & 0 & & & \vdots \\
0 & 0 & 0 & & & -\mathring{\rho}_1 k & \vdots & \vdots & \vdots & & & \mathring{\rho}_1 K \\
0 & 0 & 0 & & & k & 0 & 0 & 0 & & & -K \\
\hline
0 & 0 & 0 & \cdots & \cdots & 0 & 1 & -\rho_2 & 0 & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & & & \vdots & 0 & 1 & -\rho_2 & & & \vdots \\
\vdots & \vdots & \vdots & & & \vdots & 0 & 0 & 1 & & & \vdots \\
\vdots & \vdots & \vdots & & & \vdots & 0 & 0 & 0 & & & \vdots \\
\vdots & \vdots & \vdots & & & \vdots & 0 & 0 & 0 & & & -\rho_2 \\
0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 & & & 1
\end{array}\right] \qquad (73)$$

the expression (13) for the joint efficiency is now applicable.

Let

$$M = G'\mathring{\Omega}G = \frac{1}{k^2} = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}' & M_{22} \end{bmatrix} \qquad (74)$$

where

$$M_{11} = \begin{bmatrix}
k^2 & 0 & 0 & \cdots & 0 \\
0 & k^2 & 0 & \cdots & 0 \\
0 & 0 & k^2 & \cdots & 0 \\
\vdots & & & & \\
0 & 0 & 0 & \cdots & k^2
\end{bmatrix} \qquad (75)$$

$$M_{12} = \delta kK \begin{bmatrix} 0 & 1 & \mathring{\rho}_2 & \mathring{\rho}_2^2 & \mathring{\rho}_2^3 \cdots\cdots \mathring{\rho}_2^{n-2} \\ 0 & 0 & 1 & \mathring{\rho}_2 & \mathring{\rho}_2^2 \cdots\cdots \mathring{\rho}_2^{n-3} \\ 0 & 0 & 0 & 1 & \mathring{\rho}_2 \cdots\cdots \mathring{\rho}_2^{n-4} \\ \vdots & & & & \\ 0 & 0 & 0 & 0 & \mathring{\rho}_2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \cdots\cdots\cdots\cdots & 0 \end{bmatrix} \qquad (76)$$

$$M_{22} = \begin{bmatrix} k^2 & \delta k^2 & \delta \mathring{\rho}_2 k^2 \cdots\cdots\cdots & \delta \mathring{\rho}_2^{n-2} k^2 \\ \delta k^2 & k^2 + \delta^2 & \delta(k^2 + \mathring{\rho}_2 \delta) \cdots\cdots & \delta \mathring{\rho}_2^{n-3}[k^2 + \delta_2 \mathring{\rho}_2] \\ \delta \mathring{\rho}_2 k^2 & \delta(1^2 + \mathring{\rho}_2 \delta_2) & k^2 + \delta^2(1 + \mathring{\rho}_2^2) \cdots & \delta \mathring{\rho}_2^{n-4}[k^2 + \delta_2 \mathring{\rho}_2(1 + \mathring{\rho}_2^2)] \\ \delta \mathring{\rho}_2^2 k^2 & \delta \mathring{\rho}_2(k^2 + \mathring{\rho}_2 \delta_2) & \delta[k^2 + \delta \mathring{\rho}_2(1 + \mathring{\rho}_2^2)] \cdots & \delta \mathring{\rho}_2^{n-5}[k^2 + \delta_2 \mathring{\rho}_2(1 + \mathring{\rho}_2^2 + \mathring{\rho}_2^4] \\ \vdots & \vdots & \vdots & \vdots \\ \delta \mathring{\rho}_2^{n-2} & \delta \mathring{\rho}_2^{n-3}(k^2 + \mathring{\rho}_2 \delta_2) & \delta \mathring{\rho}_2^{n-4}[k^2 + \delta \mathring{\rho}_2(1 + \mathring{\rho}_2^2)] & k^2 + \delta^2[1 + \mathring{\rho}_2^2 + \cdots \mathring{\rho}_2^{2n-4}] \end{bmatrix} \qquad (77)$$

where

$$\delta = \rho_2 - \mathring{\rho}_2 \qquad (78)$$

To facilitate the computation of (13), and at the same time to examine its values when δ is small, and K close to 1 numerically, we assume δ^2 is negligible and that k^2, δ, $\mathring{\rho}_2$ are of the same order of magnitude. The factor $\frac{1}{k^2}$ appearing with M_1 and the factor $\frac{1}{k}$ appearing with G, will cancel out in

(13). For our purpose, then, it is convenient to write

$$M_{12} = \begin{bmatrix} 0 & \delta kK & 0 & 0 & \cdots & 0 \\ 0 & 0 & \delta kK & 0 & & 0 \\ 0 & 0 & 0 & \delta kK & & 0 \\ 0 & 0 & 0 & 0 & & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ & & & & & 0 \\ & & & & & \delta kK \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \qquad (79)$$

$$M_{11} = M_{22} = \begin{bmatrix} k^2 & 0 & 0 & \cdots & 0 \\ 0 & k^2 & 0 & & 0 \\ 0 & 0 & k^2 & & 0 \\ 0 & 0 & 0 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & k^2 \end{bmatrix} \qquad (80)$$

The inverse of M is given as follows.

$$M^{-1} = k^2 \begin{bmatrix} M^{11} & M^{12} \\ M^{12'} & M^{22} \end{bmatrix} \qquad (81)$$

$$M^{11} = M^{22} = \begin{bmatrix} rk^2 & 0 & 0 & \cdots & 0 \\ 0 & rk^2 & 0 & & 0 \\ 0 & 0 & rk^2 & & 0 \\ 0 & 0 & 0 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & rk^2 \end{bmatrix} \tag{82}$$

$$M^{12} = \begin{bmatrix} 0 & -rkK\delta & 0 & 0 & \cdots & 0 \\ 0 & 0 & -rkK\delta & 0 & \cdots & 0 \\ 0 & 0 & 0 & -rkK\delta & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -rkK\delta \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{83}$$

where

$$r = \frac{1}{k^4 - k^2 K^2 \delta^2}. \tag{84}$$

Referring to (66) and (73) we obtain

$$B = AG = \frac{1}{k} \begin{bmatrix} kx_1 & kX_2 & kX_3 \cdots & kX_n & -Kx_1 & -KX_2 & -KX_3 \cdots & -KX_n \\ k & k(1-\dot{\rho}_1) & k(1-\dot{\rho}_1) \cdots & k(1-\dot{\rho}_1) & -K & -K(1-\dot{\rho}_1) & -K(1-\ddot{\rho}_1) & -K(1-\dot{\rho}_1) \\ 0 & 0 & 0 \cdots & 0 & x_1 & X_2 & X_3 & X_n \\ 0 & 0 & 0 \cdots & 0 & 0 & 1-\dot{\rho}_1 & 1-\dot{\rho}_1 & 1-\dot{\rho}_1 \end{bmatrix} \tag{85}$$

where

$$X_t = x_t - \hat{\rho}_1 x_{t-1}, \qquad t = 2, 3, \ldots, n. \tag{86}$$

Note that we are taking $\rho_2 = \hat{\rho}_1$ in G.

A fairly simple case to examine is the following:* Suppose

$$X_2 = X_3 = \ldots\ldots = X_n \quad . \tag{87}$$

Then the only nonzero minor of (85) will be

$$\det \begin{vmatrix} kx_1 & kX_2 & -Kx_1 & -KX_2 \\ k & k(1-\hat{\rho}_1) & -K & -K(1-\hat{\rho}_1) \\ 0 & 0 & x_1 & X_2 \\ 0 & 0 & 1 & 1- \quad 1 \end{vmatrix} = k^2 \det{}^2 \begin{vmatrix} x_1 & X_2 \\ 1 & 1-\hat{\rho}_1 \end{vmatrix} \tag{88}$$

Employing Laplacian** expansions for BMB', $BM^{-1}B'$, we obtain

$$\text{Eff.}(\hat{\theta}_1^{(\rho_2)}, \hat{\theta}_2^{(\rho_2)}, \hat{\theta}_3^{(\rho_2)}, \hat{\theta}_4^{(\rho_2)}) = \frac{\det M}{\det \begin{vmatrix} k^2 & 0 & 0 & \delta kK \\ 0 & k^2 & 0 & 0 \\ 0 & 0 & k^2 & 0 \\ \delta kK & 0 & 0 & k^2 \end{vmatrix} \det M_{2n-4}} \tag{89}$$

where M_{2n-4} is the matrix M of (74) with dimension (2n-4)x(2n-4) instead of 2nx2n. Now, it is easy to prove by induction that a 2nx2n matrix of the form

---

* Other cases will also be considered presently.

** The first factor in the denominator of (89) corresponds to the orientation of the minor (88) in (85). The second factor, divided by det M, gives the corresponding cofactor minor in M^{-1}. Note that $\det^2(BB')$ cancels out from numerator and denominator.

$$\left[\begin{array}{cccccc|cccc}
a & 0 & 0 & 0 \cdots & 0 & 0 & 0 & b & 0 \cdots & 0 \\
0 & a & 0 & 0 \cdots & 0 & 0 & 0 & 0 & b & 0 \\
0 & 0 & a & 0 \cdots & 0 & 0 & 0 & 0 & 0 & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 & 0 & 0 \cdots & b \\
0 & 0 & 0 & 0 \cdots & 0 & a & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & a & 0 & 0 & 0 \\
b & 0 & 0 & 0 & 0 & 0 & 0 & a & 0 & 0 \\
0 & b & 0 & 0 & 0 & 0 & 0 & 0 & a & 0 \\
\vdots & & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & b & 0 & 0 & 0 & 0 & a
\end{array}\right] \tag{90}$$

has for its determinant

$$a^2(a^2 - b^2)^{n-1}. \tag{91}$$

Thus, taking

$$a = k^2 \; ; \qquad b = \delta kK \tag{92}$$

the joint efficiency reduces to

$$\frac{a^2(a^2 - b^2)^{n-1}}{a^2(a^2 - b^2)\, a^2(a^2 - b^2)^{n-3}} = \frac{a^2 - b^2}{a^2} = 1 - \frac{\delta^2 K^2}{k^2} \tag{93}$$

Since k^2 has the same order of magnitude as δ, it is clear the loss of efficiency in this case is negligible if δ is sufficiently* small. This

---

* If $\delta = .3$, the efficiency is 73 percent, which suggests the loss might not always be neglected.

might have been expected, intuitively, since $\overset{\circ}{\rho}_1, \overset{\circ}{\rho}_2$ are both very small in this case; however, the situation may have been saved by imposing the condition (87) that the values of x_t are in geometric progression.* Since $\overset{\circ}{\rho}_1$ is small in this situation, this would imply the x's are "close," especially for large values of n.

It will be seen presently, in considering other restrictions on the nature of the sequence $x_1, \ldots, x_n$, that the efficiency may be diminished appreciably even for $\overset{\circ}{\rho}_1, \overset{\circ}{\rho}_2$ very close, and small. This would, indeed, serve as a grim warning about the possible unreliability of our intuition in such instances of incorrect specification of the autoregressive parameters of the disturbance processes.

In the following situation the same order of approximation will be maintained as above, but the following restrictions will be imposed on the sequence $x_1, x_2, \ldots, x_n$. Let**

$$x_1 = x_2 = x_1 - \overset{\circ}{\rho}_1 x_{n-1} = s_1^{(\overset{\circ}{\rho}_1)} = \sum_{t=2}^{n-1} x_t x_{t+1} = 0. \tag{94}$$

Then the matrices BB', BMB', $BM^{-1}B'$ are given as follows

$$BB' = \frac{1}{k^2} \begin{bmatrix} s_2^{(\overset{\circ}{\rho}_1)} & 0 & -Ks_2^{(\overset{\circ}{\rho}_1)} & 0 \\ 0 & \beta & 0 & -K\beta \\ -Ks_2^{(\overset{\circ}{\rho}_1)} & 0 & s_2 & 0 \\ 0 & -K\beta & 0 & \beta \end{bmatrix} \tag{95}$$

---

* Compare the situation in (59), where the upper bound is attained for a geometric series for the x's.

** For definition of $s_1^{(\overset{\circ}{\rho}_1)}$, $s_2^{(\overset{\circ}{\rho}_1)}$, see (42), (43).

$$BMB' = \frac{1}{k^2} \begin{bmatrix} k^2 s_2^{(\mathring{\rho}_1)} & 0 & -k^2 K s_2^{(\mathring{\rho}_1)} & 0 \\ 0 & k^2[\beta - 2\delta K^2 \alpha] & 0 & k^2[-K\beta + \delta K \alpha] \\ -k^2 K s_2^{(\mathring{\rho}_1)} & 0 & k^2 s_2^{(\mathring{\rho}_1)} & 0 \\ 0 & k^2[-K\beta + \delta K \alpha] & 0 & k^2 \beta \end{bmatrix} \tag{96}$$

$$BM^{-1}B' = \frac{r}{k^2} \begin{bmatrix} k^2 s_2^{(\mathring{\rho}_1)} & 0 & -k^2 K s_2^{(\mathring{\rho}_1)} & 0 \\ 0 & k^2[\beta + 2\delta K^2 \alpha] & 0 & k^2[-K\beta - \delta K \alpha] \\ -Kk^2 s_2 & 0 & k^2 s_2^{(\mathring{\rho}_1)} & 0 \\ 0 & k^2[-K\beta - 2\delta K \alpha] & 0 & k^2 \beta \end{bmatrix} \tag{97}$$

where

$$\alpha = (1-\mathring{\rho}_1)\ [1+(n-1)(1-\mathring{\rho}_1)] \tag{98}$$

$$\beta = 1+(n-1)(1-\mathring{\rho}_1)^2 \tag{99}$$

Hence, by the Laplacian expansion of determinants, using 2x2 minors,

$$\det(k^2BB') = \beta^2 k^4 (s_2^{(\mathring{\rho}_1)})^2$$

$$\det(k^2BMB') = k^6 (s_2^{(\mathring{\rho}_1)})^2 \cdot k^4 \begin{vmatrix} \beta - 2\delta K^2\alpha & -K\beta + \delta K\alpha \\ -K\beta + \delta K \alpha & \beta \end{vmatrix} \tag{100}$$

$$\det(k^2BM^{-1}B') = k^6 (s_2^{(\mathring{\rho}_1)})^2 \cdot k^4 \begin{vmatrix} \beta + 2\delta K^2\alpha & -K\beta - \delta K \alpha \\ -K\beta - 2\delta K \alpha & \beta \end{vmatrix} \tag{101}$$

Hence, the joint efficiency is given by

$$\underset{(\rho_2=\overset{\circ}{\rho}_1)}{\text{Eff.}}(\hat{\theta}_1^{(\rho_2)}, \hat{\theta}_2^{(\rho_2)}, \hat{\theta}_3^{(\rho_2)}, \hat{\theta}_4^{(\rho_2)}) = \frac{k^8 \beta^4}{k^{20} r^4 [k^2\beta^2 - \delta^2 K^2 \alpha^2]^2} = \left[\frac{\left(1 - \frac{\delta^2 K^2}{k^2}\right)^2}{1 - \frac{\delta^2 K^2}{k^2} \cdot \frac{\alpha^2}{\beta^2}}\right]^2 \tag{102}$$

A few comments will now be made about this interesting formula. First, the square of the denominator will be greater than the numerator provided $\frac{\alpha^2}{\beta^2} < 1$. Now

$$\frac{\alpha^2}{\beta^2} = \left[1 - \frac{\overset{\circ}{\rho}_1}{1+(n-1)(1-\overset{\circ}{\rho}_1)^2}\right]^2 \tag{103}$$

which is less than 1 provided $\overset{\circ}{\rho}_1 > 0$. However, the numerator has a factor which is less than one, but all squared, so the efficiency still remains ≤ 1, as it should, for $\overset{\circ}{\rho}_1$ sufficiently small numerically. This means the loss of efficiency is greater for $\overset{\circ}{\rho}_1 > 0$. Secondly, (103) shows the efficiency depends on the value of n; moreover, as in the case of (58), the efficiency decreases as n increases, if $\overset{\circ}{\rho}_1 > 0$. For $\overset{\circ}{\rho}_1 < 0$, the efficiency increases as n increases.

For n = 11, $\rho_1 = \delta = k^2 = .2$, the efficiency given by (102) is 80 percent, whilst for n = 21, the efficiency is 75 percent. Note also, that the limit of (102) as $n \longrightarrow \infty$ is $\left(1 - \frac{\delta^2 K^2}{k^2}\right)^2$, which is the same as (106) given below and has the value 64 percent for $\rho_1 = \delta = k^2 = .2$.

Now consider the characteristic roots of the matrix M given in (74), (79), (80). These are given by the values of v satisfying

$$(a-v)^2 \, [(a-v)^2 - b^2]^{n-1} = 0 \tag{104}$$

where a and b have been defined by (92). This yields as possible values for v a, a-b, a + b.

Thus, in decreasing order of magnitude, the characteristic roots are

$$\underbrace{k^2+\delta kK,\ k^2+\delta kK,\ \ldots\ldots,\ k^2+\delta kK}_{n-1},\ k^2,\ k^2,\ \underbrace{k^2-\delta kK,\ k^2-\delta kK,\ \ldots\ldots,\ k^2-\delta kK}_{n-1} \qquad (105)$$

Watson [10], has given a lower bound for efficiency of least-squares estimators in terms of the characteristic roots of the covariance matrix by an inequality due to Cassel [11]. Applying this method here, we obtain, as a lower bound for the efficiency

$$\frac{4^2[(k^2+\delta kK)(k^2-\delta kK)]^2}{4^2\cdot k^4\cdot k^4} = \left(1-\frac{\delta^2K^2}{k^2}\right)^2 \qquad (106)$$

Note that this is the square of the expression given by (93). Compare it also with (102). If in (106), we set $\delta = .2 = k^2$, the lower bound of efficiency is 64 percent, whilst if $\delta = .3 = k^2$, the lower bound is 53 percent. The approximation may not be as accurate for $\delta = .3$, but it is evident, at least, that the loss* of efficiency may not be negligible, even for small values of $\mathring{\rho}_1$ and $\mathring{\rho}_2$.

6. Comments on the papers by Cochrane and Orcutt [1], [2].

These papers focus attention on the difficulties encountered in regression analysis when the underlying assumptions do not hold. Many interesting points are brought out in the results of their extensive artificial sampling experiment.

---

* As Watson's thesis is not yet available, it is not clear to the author whether the bound (106) is always attainable, and if it is, how realistic the required "fixed" values would be. Note, for instance, that (106) is the limit of (102) as $n \to \infty$; this would require the fixed values to be extremely close.

Particular instances are exhibited [2] in which the small sample bias of maximum likelihood estimators is serious. A case is also given [2] where the maximum-likelihood estimators obtained by the method of limited information [14] are, empirically, more efficient than maximum-likelihood estimators which are obtained by utilising all the information.

Apart from the sampling aspect of such investigations, however, great caution must be exercised in how particular assumptions which are violated, are isolated from other violations, and in ascribing loss of unbiassedness or of efficiency to particular causes. For instance,* C.O., actually use quasi-maximum likelihood estimators (cf.[14]), since the disturbances have a discrete rectangular distribution, not a continuous normal distribution. Thus, the curious case mentioned above, in which loss of information yields an apparently more efficient estimator is very similar to a phenomenon described by Hotelling [13], where the lack of normality brings about great changes in the tail of the distribution of the estimator.

It has also been pointed out by Wold [12], that the empirical loss of efficiency obtained by C. O. [1] for the least-squares estimators when the disturbances and the explanatory variables are autocorrelated may well be due to the fact that the underlying stochastic processes selected by C. O. are, in some cases, evolutive and not stationary.

Another violation which may well be responsible for some loss of efficiency of estimators in [2] is the possible lack of identifiability (cf. [14]). Little is known of identifiability in the case of autocorrelation of the disturbances in a system of equations, and it seems appropriate to settle

---

* Hereafter, the letters C. O. are used as an abbreviation of Cochrane and Orcutt.

this question before attempting to estimate parameters for which, under certain conditions, there may not even exist nontrivial estimators.

The investigation of § 4, above, points out another violation in [1] not mentioned by Wold [12], since he considers the problem from the large sample viewpoint. Reference to [1] will show that the initial values of the artificial series constructed are taken to be zero. As pointed out above, this could lead to a serious loss of efficiency, especially when the autoregressive parameter is close to unity in value; and this value is actually unity for series B in C. O. [1].

As for the investigation in § 6, above, it points out a danger rather than a violation in the methods of C. O. [2]. Although the two equation system we have selected is much simpler, it still exhibits the appreciable loss of efficiency which may result in taking the values of the autoregressive parameters to be the same in the disturbance processes of both equations, when the processes are related in a particularly simple way. What is perhaps more dangerous, even in these simple situations, is that the true values of the parameters could be very close and small. This would indicate the need for sensitive tests of equality of autoregressive parameters before selecting particular models of the disturbance processes in linear regression equations.

REFERENCES

[1] Guy H. Orcutt and Donald Cochrane, "A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regressive Analysis," Journal of the American Statistical Association, Vol. 44, 1949, pp. 356-372.

[2] D. Cochrane and G. H. Orcutt, "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, Vol. 44, 1949, pp. 32-61.

[3] J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression, I," Biometrika, Vol. 37, 1950, p. 409.

[4] J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression, II," Biometrika, Vol. 38, 1951.

[5] A. C. Aitken, "On Least Squares and Linear Combinations of Observations," Proceedings of the Royal Society of Edinburgh, Vol. 55, 1935, pp. 42-48.

[6] J. Gurland, "Estimation of the Regressive and Covariance Parameters," Cowles Commission Discussion Paper, Statistics 349.

[7] J. L. Doob, "Time Series and Harmonic Analysis," Proceedings of the First Berkeley Symposium, 1946.

[8] H. Cramer, "A Contribution to the Theory of Statistical Estimation," Skand. Aktuartids, Vol. 29, 1946, pp. 85-94.

[9] Leonid Hurwicz, "Least Squares Bias in Time Series," Cowles Commission Monograph 10, John Wiley and Sons, Inc., New York, 1950.

[10] G. S. Watson, Unpublished thesis, University of Cambridge.

[11] Cassel, Result stated by Watson at Cowles Commission staff meeting, Oct. 1, 1951.

[12] H. Wold, "On Least Square Regression with Autocorrelated Variables and Residuals," Transactions of the International Institute of Statistics, Session at Berne, Switzerland, September, 1949, 13 pp.

[13] H. Hotelling Rietz Memorial Lecture at Minneapolis Meeting of Institute of Mathematical Statistics, September, 1951.

[14] T. C. Koopmans (Editor), Statistical Inference In Dynamic Economic Problems, Cowles Commission Monograph no. 10, John Wiley & Sons, New York, 421 pp.