Cowles Foundation Paper 85
Reprinted from Econometrica, 22(2), April 1954

# AN EXAMPLE OF AUTOCORRELATED DISTURBANCES IN LINEAR REGRESSION[1]

By John Gurland

In linear regression models in which the disturbances are autocorrelated, it is often assumed that these are given by a Markov process. This article investigates the loss of efficiency of estimators of the regression parameters when there are certain types of specification bias concerning the disturbances. Specifically, it points out that the loss of efficiency can be serious if the initial conditions of the process which are assumed to be true are in fact not true. In particular, if the process is incorrectly assumed to be stationary, then although the estimation procedure which is presumed to yield the best linear unbiased estimators will produce unbiased estimators, their joint efficiency may nevertheless be close to zero.

1. *Introduction*. Attention has recently been given to a particular instance of the incorrect specification of the covariance matrix of the disturbances in linear regression. This is the situation in which the disturbances are assumed to be independent, with common variance, when in fact they are generated by a first order Markoff process. In a recent empirical study, Cochrane and Orcutt [1] investigated the loss of efficiency for a certain linear regression model. As has been pointed out by Wold [2] however, many of the series considered by Cochrane and Orcutt are not stationary, but evolutionary; consequently the loss of efficiency which Cochrane and Orcutt attribute to the presence of autocorrelation may be due in great part to the evolutionary nature of the series.

The motivation of this paper is somewhat similar to that of Wold [2], but it does differ in two main respects. First, the loss of efficiency in estimating the regression parameters is attributed, here, to another source, namely, to the incorrect specification of the initial conditions in the series of disturbances constructed by Cochrane and Orcutt. This could be responsible for considerable loss even if evolutionary series are permitted. Second, Wold considers the asymptotic behavior of the estimates for large sample size, whereas here the loss of efficiency is considered for any fixed sample size.

In the following section the relation between approach $(A)$ of minimizing the residual variance and $(B)$ of minimizing the variances of the estimates of the regression parameters is recalled. It is seen that, in approach $(B)$, a particular quadratic form involving the inverse of the covariance matrix is minimized to yield the best linear unbiased estimates of the regression parameters. If the residual variance under approach $(A)$ is minimized, however, the resulting estimates, although linear unbiased, will not generally be best in the sense of being of minimum variance. It is evident, of course, that approach $(A)$ is covariance-free in the sense that the covariance matrix need not be specified in order to arrive at the regression estimates, whereas approach $(B)$ is not.

[1] This paper will be reprinted as Cowles Commission Paper, New Series, No. 85 and as Iowa State College, Statistical Laboratory, Reprint Series, No. 1.

2. *Linear regression with autocorrelated disturbances.* Let

$$y_i = \xi_i + u_i, \qquad (i = 1, 2, \cdots, n)$$

where $\xi_i$, the expected value of $y_i$ is a linear combination

$$\xi_i = \theta_1 x_{1i} + \theta_2 x_{2i} + \cdots + \theta_{k-1} x_{k-1,i} + \theta_k$$

of the unknown parameters $\theta$, and

$$Eu_i = 0.$$

The covariance matrix of the disturbances $u_i$ consists of elements

$$Eu_i u_j = \sigma^2 \omega_{ij}.$$

It is assumed here that the $x_{hi}$ $(h = 1, 2, \cdots, k; i = 1, 2, \cdots, n)$ are known "fixed" variates, and that the elements $\omega_{ij}$ are also known. Let $\Omega$ designate the matrix of elements $\omega_{ij}$, and $\Omega^{-1}$ the inverse matrix with elements $\omega^{ij}$. As shown by Aitken [3], the best linear unbiased estimates of $\theta$ are obtained by solving the following $k$ linear equations for $\theta_1, \theta_2, \cdots, \theta_k$.

$$\frac{\partial}{\partial \theta_h} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega^{ij}(y_i - \xi_i)(y_j - \xi_j) = 0; \quad (h = 1, 2, \cdots, k).$$

Denote these estimates by $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$.

By means of a linear transformation it is possible[2] to obtain the same estimates $\hat{\theta}$ by minimizing a quadratic form which corresponds to the unit matrix. Thus, there exists a linear transformation defined by

$$v_i = \sum_{j=1}^{n} u_j \gamma_{ij} \qquad (i = 1, 2, \cdots, n)$$

such that

$$Ev_i v_j = \begin{cases} \sigma^2 & (\text{for } i = j) \\ 0 & (\text{for } i \neq j). \end{cases}$$

Let

$$s_i = \sum_{j=1}^{n} y_j \gamma_{ij}$$

$$\eta_i = \sum_{j=1}^{n} \xi_j \gamma_{ij}.$$

Then the best linear unbiased estimates $\hat{\theta}$ of $\theta$ are also obtained by solving the following $k$ linear equations for $\theta_1, \theta_2, \cdots, \theta_k$.

$$\frac{\partial}{\partial \theta_h} \sum_{i=1}^{n} (s_i - \eta_i)^2 = 0, \qquad (h = 1, 2, \cdots, k).$$

[2] Refer to Appendix.

As mentioned in the introduction, estimates obtained under approach $(A)$ will always be unbiased. This is a special case of the following theorem, the proof of which is given in the appendix.

THEOREM 1. Suppose $\Omega$ is the correct covariance matrix of the disturbances $u_1, u_2, \cdots, u_n$; but it is assumed that the covariance matrix is $\Omega^*$, with elements $\omega^*_{ij}$. Then the linear estimates of $\theta_1, \theta_2, \cdots, \theta_k$ obtained by solving the $k$ linear equations

$$\frac{\partial}{\partial\theta_h} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega^{*ij}(y_i - \xi_i)(y_j - \xi_j) = 0 \qquad (h = 1, 2, \cdots, k)$$

are unbiased.

3. *Incorrectly specified covariance matrix.* The following result which is important in the subsequent sections is stated as a theorem. The proof of this theorem will be found in the appendix.

THEOREM 2. Suppose it is assumed that

$$Eu_iu_j = \sigma^2\omega^*_{ij} \qquad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, n)$$

when, in fact, the correct covariances and variances are given by

$$Eu_iu_j = \sigma^2\omega^0_{ij}.$$

Let the estimates obtained by solving

$$\frac{\partial}{\partial\theta_h} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega^{0ij}(y_i - \xi_i)(y_j - \xi_j) = 0 \qquad (h = 1, 2, \cdots, k)$$

be denoted by $\hat\theta^0_1, \hat\theta^0_2, \cdots, \hat\theta^0_K$. Let the estimates obtained by solving the corresponding equations with $\omega^{*ij}$ be denoted by $\hat\theta^*_1, \hat\theta^*_2, \cdots, \hat\theta^*_k$. Then the covariance matrices $M^0$ of $\hat\theta^0_1, \hat\theta^0_2, \cdots, \hat\theta^0_k$ and $M^*$ of $\hat\theta^*_1, \hat\theta^*_2, \cdots, \hat\theta^*_k$ are

$$M^0 = \sigma^2(X\Omega^0X')^{-1},$$

and

$$M^* = \sigma^2(X\Omega^{*^{-1}}X')^{-1}X\Omega^{*^{-1}}\Omega^0\Omega^{*^{-1}}X'(X\Omega^{*^{-1}}X')^{-1}.$$

The joint efficiency, $\det M^0/\det M^* = F(\hat\theta^*_1, \hat\theta^*_2, \cdots, \hat\theta^*_k)$, say, becomes

$$F(\hat\theta^*_1, \hat\theta^*_2, \cdots, \hat\theta^*_k) = \det{}^2(ZZ')/\det(ZYZ')\det(ZY^{-1}Z')$$

where $Z = X\Pi$, $Y = \Pi'\Omega^0\Pi$, and $\Pi$ is such a matrix that $\Pi'\Omega^*\Pi = I$.

4. *Disturbances generated by first-order Markoff process.* The following first order Markoff process $u_t$ will be considered:

$$u_t - \rho u_{t-1} = v_t \quad \text{where } Ev_tu_{t-1} = 0 \quad (t = -N + 1, \cdots -1, 0, 1, 2, \cdots, n),$$

and

$$\left.\begin{aligned} Ev_t &= 0 \\ Ev_tv_{t'} &= 0 \qquad (t \neq t') \\ Ev_t^2 &= \sigma^2 \end{aligned}\right\} \quad (t = -N, -N + 1, \cdots -1, 0, 1, 2, \cdots, n).$$

The "initial value" $u_{-N}$ is defined[3] by

$$u_{-N} = \delta v_{-N}$$

where the value of the constant $\delta$ is selected arbitrarily. Define

$$g_N^2 = 1 + \rho^2 + \rho^4 + \cdots + \rho^{2N+2}\delta^2.$$

By direct calculation it is easily verified that the variances of the disturbances are given by

$$Eu_1^2 = \sigma^2 g_N^2$$

$$Eu_j^2 = \sigma^2(1 + \rho^2 + \rho^4 + \cdots + \rho^{2j-2}g_N^2), \qquad (j = 2, 3, \cdots, n)$$

and the covariances are given by

$$Eu_iu_j = \sigma^2\rho^{|j-i|}Eu_i^2 \qquad (j \neq i).$$

Setting

$$g_\infty^2 = \lim_{N=\infty} g_N^2$$

it is clear that

$$g_\infty^2 = 1/(1 - \rho^2) \qquad (\text{for } \rho^2 < 1)$$

and the value of $\delta$ is irrelevant in this case. It should also be noted that for

$$\text{(1)} \qquad \begin{array}{l} \rho^2 < 1 \text{ and } N \text{ infinite} \\ \text{or } \rho^2 < 1,\ N \text{ finite, and } \delta^2 = 1/(1 - \rho^2) \end{array}$$

the variances and covariances are given by the expression:

$$Eu_i\, u_j = \frac{\sigma^2}{1 - \rho^2}\rho^{|i-j|}, \qquad (i = 1, 2, \cdots, n;\, j = 1, 2, \cdots, n).$$

As this expression in $i$ and $j$ depends only on the difference $i - j$, and since $Eu_t = 0$, the $u_t$ process, in this case, is stationary to the second order. (Cf. Cramér [4].)

If the true values of $\rho$ and $\delta$ are known, the best linear unbiased estimates of the parameters $\theta$ may be obtained by means of the following transformation[4]:

$$\left.\begin{aligned} s_t &= y_t - \rho y_{t-1} \\ \eta_t &= \xi_t - \rho\xi_{t-1} \\ z_{ht} &= x_{ht} - \rho x_{ht-1} \end{aligned}\right\} \qquad (t = 2, 3, \cdots, n;\, h = 1, 2, \cdots, k)$$

with $s_1 = y_1/g_N$, $\eta_1 = \xi_1/g_N$, and $z_{h1} = x_{h1}/g_N$.
Solving the $k$ linear equations

$$\text{(2)} \qquad \frac{\partial}{\partial\theta_h}\left[\frac{1}{g_N^2}(y_1 - \xi_1)^2 + \sum_2^n (s_t - \eta_t)^2\right] = 0$$

[3] A referee has pointed out that $u_{-N}$ could be defined more generally by requiring only that $Eu_{-N}^2 = \delta^2\sigma^2$.

[4] Refer to Appendix.

will yield the best linear unbiased estimates. If, however, incorrect values of $\rho$ or $\delta$ or both are used, the estimates, although unbiased, will no longer be "best" in general.

5. *Incorrect specification of the initial condition.* Since the parameter $\delta$ occurs only in the term $(1/g_N^2)(y_1 - \xi_1)^2$ in (2), one might be tempted to conclude that it plays a minor role and that incorrect specification of its value would be rather inocuous. Cochrane and Orcutt recommend neglecting this term and solving the system of equations

$$\frac{\partial}{\partial \theta_h} \sum_2^n (s_t - \eta_t)^2 = 0.$$

They claim this is justified if the true value of $\rho$ is very close to 1. It will be shown, however, that it is precisely in such a situation that the loss of efficiency can be extreme if the initial conditions are assigned incorrectly.

To show how serious the loss could be solely for incorrect specification of the initial conditions, it will be assumed that the true value of the autoregressive parameter $\rho$ is known. The question of the justifiability of omitting the term $(1/g_N^2)(y_1 - \xi_1)^2$ is closely related to the following question: "What is the loss of efficiency in estimating (by the method of §2) the parameters $\theta$, with the value of $g_N$ assumed to be extremely large when, as a matter of fact, it is not?" If this loss is serious, then the omission of this term will, of course, be unjustifiable.

For simplicity, consider the case $k = 2$, with

$$y_t = \theta_1 x_t + \theta_2 + u_t$$

and $u_t$ a first-order Markoff process as described in §4. Let $\hat{\theta}_1^*$, $\hat{\theta}_2^*$ be the estimates obtained with $g^*$ as the assumed value of $g_N$. Suppose $g^0$ is the correct value of $g_N$. In order to apply the formula for $F(\hat{\theta}_1^*, \hat{\theta}_2^*)$ of §3, it is convenient to have the following expressions and determinants represented by symbols. Thus, let

$$s_1^{(\rho)} = \sum_2^n (x_t - \rho x_{t-1}), \qquad s_2^{(\rho)} = \sum_2^n (x_t - \rho x_{t-1})^2,$$

$$A_0 = \det \begin{bmatrix} s_2^{(\rho)} & (1-\rho)s_1^{(\rho)} \\ (1-\rho)s_1^{(\rho)} & (n-1)(1-\rho)^2 \end{bmatrix}, \quad A_1 = \det \begin{bmatrix} x_1 & (1-\rho)s_1^{(\rho)} \\ 1 & (n-1)(1-\rho)^2 \end{bmatrix},$$

$$\text{and} \quad A_2 = \det \begin{bmatrix} s_2^{(\rho)} & x_1 \\ (1-\rho)s_1^{(\rho)} & 1 \end{bmatrix}.$$

Then

$$F(\hat{\theta}_1^*, \hat{\theta}_2^*) = \frac{\left[A_0 + \dfrac{x_1 A_1 + x_2}{g^{*2}}\right]^2}{\left[A_0 + \dfrac{g^{02}(x_1 A_1 + A_2)}{g^{*4}}\right]\left[A_0 + \dfrac{x_1 A_1 + A_2}{g^{02}}\right]}.$$

It easily follows that

$$\lim_{g^* = \infty} F(\hat{\theta}_1^*, \hat{\theta}_2^*) = \frac{1}{1 + \dfrac{x_1 A_1 + A_2}{A_0 g^{02}}},$$

$$\lim_{x_1 = \infty} \lim_{g^* = \infty} F(\hat{\theta}_1^*, \hat{\theta}_2^*) = 0, \tag{3}$$

$$\lim_{\rho = 1} \lim_{g^* = \infty} F(\hat{\theta}_1^*, \hat{\theta}_2^*) = 0. \tag{4}$$

Although the limits in (3) and (4) are taken in a particular order, there exist values of $x_1$ and $g^*$ in the case of (3), and values of $\rho$ and $g^*$ in the case of (4) for which the efficiency is arbitrarily close to zero. This shows that the loss of efficiency can be serious precisely in those situations where the term $1/g_N^2(y_1 - \xi_1)^2$ is omitted in obtaining estimates of $\theta_1$, $\theta_2$ according to the method of §2. It should also be remarked that (3) and (4) hold also in the case of evolutionary series for $u_t$, since $\rho^2$ need not be less than unity.[5] If, on the other hand, the estimates $\hat{\theta}_1^*$, $\hat{\theta}_2^*$ are obtained on the assumption that $u_t$ is a stationary process (note (1)) then the limit in (4) must be taken for values $\rho < 1$.

The assumption that $g^*$ is very large can also be interpreted in several ways, of which the following appear to be relevant.

| | | I ($\rho^2 < 1$) | II ($\rho^2 \geq 1$) | III ($\rho^2 \geq 1$) |
|---|---|---|---|---|
| Assumed | $g^*$ large | $\delta^*$ large | $\delta^*$ large | $N$ large and $\delta^*$ small |
| True | $g^0$ (fixed) | $\delta^0$ (fixed) | $\delta^0$ (fixed) | $N^0$ (fixed) and $\delta^0$ (fixed) |

Under case I, since $\rho^2 < 1$ and $N$ is finite,[6] $\delta^*$ must be large to make $g^*$ large. Under case II, $N$ is also finite, and has as well the same value for both the assumed and true categories. Under case III, however, the assumed value $N^*$ of $N$ must come into play (cf. (1)) to make $g$ large, and the true value $N^0$ of $N$ to make $g^0$ fixed. Any of the above three cases may be proposed as an explanation of how the term $(1/g_N^2)(y_1 - \xi_1)^2$ comes to be omitted. Some possible consequences of such an omission have been indicated above by the limits in (3) and (4).

In the evaluation of the limits in (3) and (4) it has been tacitly assumed that $\rho$ and $\delta$ are not functionally related; nevertheless the same limits might be found for particular relationships that are chosen. As a special instance of a functional relationship, consider $\delta^2 = 1/(1 - \rho^2)$ where $\rho^2 < 1$. This is as pointed out on page (221) equivalent to the assumption that the process $u_t$ is stationary to the second order. It is interesting to ascertain what damage results in assuming $u_t$ to be a stationary process, when in fact it has an initial fixed value $u_{-N} = 0$. This case is of interest here, because Cochrane and Orcutt establish

[5] Refer to the definition of $u_t$ in §4.

[6] If $N$ were infinite in this case, then $g^0 = g^*$, and $g^*$ would not be permitted to grow large, since $g^0$ is fixed.

the validity of their transformation and employ it on the assumption of stationarity, whereas in their artificially constructed series[7] they take $u_0 = 0$.

Since $u_0 = 0$ implies $\delta = 0$, it is possible to consider the situation with slightly greater generality by assuming $u_{-N} = 0$, where $N$ is an arbitrary (but fixed) nonnegative integer. This allows the process to be started up at time $t = -N$. For this situation

$$g^0 = 1 + \rho^2 + \rho^4 \cdots + \rho^{2N} \qquad \text{(since } \delta^0 = 0\text{)}$$

and

$$g^{*^2} = 1/(1 - \rho^2).$$

The expression for efficiency becomes

$$F(\hat{\theta}_1^*, \hat{\theta}_2^*) = \frac{[A_0 + x_1(1 - \rho^2)A_1 + (1 - \rho^2)A_2]^2}{[A_0 + x_1 g^{0^2}(1 - \rho^2)^2 A_1 + g^{0^2}(1 - \rho^2)^2 A_2]\left[A_0 + \frac{x_1}{g^{0^2}} A_1 + \frac{1}{g^{0^2}} A_2\right]}.$$

On substituting for $A_0$, $A_1$, $A_2$ as defined on page (222), it may easily be verified that

$$\lim_{\rho \to 1} F(\hat{\theta}_1^*, \hat{\theta}_2^*) = \frac{1}{1 + \frac{n - 1}{4(N + 1)}\left(1 - \frac{s_1^{(1)^2}}{(n - 1)s_2^{(1)}}\right)}.$$

The greatest lower bound of this expression, for all possible values of $x_1, x_2, \cdots, x_n$ is attained when $s_1^{(1)} = 0$ (i.e. when $x_1 = x_n$). In this case it reduces to $4(N + 1)/(n + 3 + 4N)$ which is close to 1 for $N$ large and $n/N$ small, and close to zero for $n$ and $n/N$ large. In Cochrane and Orcutt's situation, $N = 0$, $u_0 = 0$, $\rho = 1$, $n = 20$; so that for series with $x_1 = x_n$ the efficiency would be .17. This could be a source of the considerable loss of efficiency of the estimates obtained from their series designated by $(B)$.

6. *Conclusion.* Although the estimates of the regression parameters are unbiased regardless of whether the covariance matrix of disturbances is correctly specified, the corresponding loss of efficiency may be far from negligible, as shown in the above investigation. Although this discussion was prompted by the type of incorrect specification of the Cochrane and Orcutt series mentioned above, it should be remarked here that, in general, the assumption of stationarity is open to question when considering linear regression models such as the above, having disturbances generated by a Markoff process. If one is so fortunate as to know that the absolute value of the autoregressive parameter $\rho$ is less than one, is he also so fortunate as to know that the process started up at time $t = -\infty$? If not, then in making statistical inferences he should consider, along with the autoregressive parameters of the Markoff process, the parameters specifying the initial conditions.

[7] In the Cochrane and Orcutt series the values of $x_t$ are selected at random. As the present investigation refers to the fixed variate case, it shows what could happen for some particular selections of the values of $x_t$.

## APPENDIX

Let

$$y = (y_1, y_2, \cdots, y_n)$$

$$\theta = (\theta_1, \theta_2, \cdots, \theta_k)$$

$$\xi = Ey = \theta X$$

where

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{k-1,1} & x_{k-1,2} & & x_{k-1,n} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \qquad (\text{rank } X = k < n).$$

Let

$$u = y - \xi \qquad (\text{with } Eu'u = \sigma^2\Omega, \text{ and } \det \Omega \neq 0.)$$

The best linear unbiased estimate of $\theta$,

$$\hat{\theta} = y(X\Omega^{-1})' \, (X\Omega^{-1}X')^{-1}$$

may be obtained by minimizing

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

with respect to $\theta$. The covariance matrix of $\hat{\theta}$ is given by

$$E(\hat{\theta} - \theta)' \, (\hat{\theta} - \theta) = \sigma^2(X\Omega^{-1}X')^{-1}.$$

Let $G$ be a matrix with elements $\gamma_{ij}$ such that

$$G'\Omega G = I$$

where $I$ is the $n \times n$ unit matrix.

Let

$$s = yG,$$

$$\eta = \xi G.$$

Then $\hat{\theta}$ may also be obtained by minimizing

$$(s - \eta)(s - \eta)'$$

with respect to $\theta$. Suppose it is assumed that

$$Eu'u = \sigma^2\Omega^*$$

when in fact

$$Eu'u = \sigma^2\Omega^0.$$

The covariance matrix of the estimates $\hat{\theta}$ obtained by minimizing

$$(y - \xi)\Omega^{*^{-1}}(y - \xi)'$$

with respect to $\theta$ is

$$E(\hat{\theta}^* - \theta)'(\hat{\theta}^* - \theta) = \sigma^2(X\Omega^{*^{-1}}X')^{-1}X\Omega^{*^{-1}}\Omega\,\Omega^{*^{-1}}X'(X\Omega^{*^{-1}}X')^{-1}.$$

Note that $\hat{\theta}^*$ is unbiased because

$$\hat{\theta}^* = y(X\Omega^{*^{-1}})'(X\Omega^{*^{-1}}X')^{-1}$$

and

$$Ey = \theta X.$$

Defining, as usual, the joint efficiency of the estimates by the ratio of the determinants of their covariance matrices, it follows that

$$F(\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_k^*) = \frac{\det (X\Omega^0 X')}{\det^{-2}(X\Omega^* X') \det (X\Omega^{*^{-1}}\Omega\,\Omega^{*^{-1}}X')}.$$

Suppose $G'\Omega^* G = I$ (i.e. $GG' = \Omega^{*^{-1}}$); $G$, here, then corresponds to $\Pi$ in §3. Let

$$XG = Z,$$

$$G'\Omega^0 G = U.$$

Then, the above expression for joint efficiency becomes

$$\frac{\det^2 (ZZ')}{\det (ZUZ') \det (ZU^{-1}Z')}.$$

For the Markoff process defined in §4

$$\Omega^* = \begin{bmatrix} \hat{g}^* & \rho g^{*2} & \cdots & \rho^{n-1}g^{*2} \\ \rho g^{*2} & 1 + \rho^2 g^{*2} & \cdots & \rho^{n-2}(1 + \rho^2 g^{*2}) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \rho^{n-1}g^{*2} & \rho^{n-2}(1 + \rho^2 g^{*2}) & \cdots & 1 + \rho^2 + \rho^4 + \cdots + \rho^{2n-2}g^{*2} \end{bmatrix}.$$

The determinants of $ZZ'$, $ZUZ'$, $ZU^{-1}Z'$ may be computed by noting that in this case

$$G = \begin{bmatrix} \frac{1}{g^*} & -\rho & 0 & \cdots & 0 \\ 0 & 1 & -\rho & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & -\rho \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

$$U = \begin{bmatrix} \frac{g^{0^2}}{g^{*2}} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$Z = \begin{bmatrix} \frac{x_1}{g^*} & x_2 - \rho x_1 & \cdots & x_n - \rho x_{n-1} \\ \frac{1}{g^*} & 1 - \rho & \cdots & 1 - \rho \end{bmatrix}$$

*University of Chicago and Iowa State College*

REFERENCES

[1] Cochrane, D. and G. H. Orcutt: "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms," *Journal of the American Statistical Association*, Vol. 44, 1949.

[2] Wold, H.: "On Least Squares Regression with Auto-correlated Variables and Residuals," *Transactions of the International Institute of Statistics*, Session at Berne, Switzerland, September 1949.

[3] Aitken, A. C.: "On Least Squares and Linear Combinations of Observations," *Proceedings of the Royal Society of Edinburgh*, Vol. 55, 1935.

[4] Cramér, H.: "Problems in Probability Theory," *Annals of Mathematical Statistics*, Vol. 18, 1947.