Estimation of the Regressive and Covariance Parameters in Linear Regression

by John Gurland

1. Introduction.

If the method of unweighted least squares is applied to estimate the coefficients in linear regression, the estimators, in general, will not be efficient within the whole class of linear unbiased estimators unless the covariance matrix Σ of the disturbances is the identity matrix multiplied by a scalar σ^2, say, which need not be known a priori. This method, which does not require any particular distribution of the disturbances (except that the second moments exist), also yields an unbiased estimate of σ^2. (cf. David and Neyman [1]).

If the most general Σ is considered, in which other unknown parameters besides σ^2 are present (and the disturbances may be correlated), the problem of estimating the regressive parameters becomes more complicated. A more general form of least-squares will produce B.L.U. (best linear unbiased) estimators if all the parameters in Σ, except σ^2, are necessarily known a priori. (cf. Aitken [2]). The present article was initiated by an attempt to estimate the autoregressive parameter in a first order Markoff process when such a process generates the disturbances of a linear regression equation. In this instance there is only one other unknown parameter in Σ besides σ^2.

It was subsequently discovered, in the formulation of a method of estimation, that the method is general, and may be applied in estimating the unknown parameters of Σ, whatever be the form of Σ, with the sole restriction that it be non-singular. For this reason, the article will deal with the general case, and refer to the auto-correlation problem as a particular example. The methods set forth below are tentatively suggested, and require further investigation. It should be remarked that these methods are distribution free in the sense that the form of the distribution of the disturbances is not specified. If the assumption of normality of the disturbances were made, the maximization of the likelihood function would, under certain conditions, yield estimators with asymptotically optimal properties; but how well this method compares with those proposed here is not yet known.

2. Least-squares estimators.

Since most of the sections in this paper lean heavily on least-squares estimation, it seems advisable, in order to have the required formulae available, to develop the theory for the general case. This development is essentially that of Aitken [2], although the following exposition is somewhat shorter and simpler. It should be noted that no assumption of normality is made.

Let the random variables

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

have the covariance matrix

$$\Sigma = \sigma^2 \| w_{ij} \| = \sigma^2 \Omega . \tag{2}$$

The elements w_{ij} $(i = 1,2,\ldots,n)$ $(j = 1,2,\ldots,n)$ of Ω are assumed known. Let the expected value of y be given by

$$Ey = \xi = \theta A \tag{3}$$

where

$$\theta = (\theta_1, \theta_2, \ldots, \theta_k) \qquad (k \leq n) \tag{4}$$

is a vector of k unknown parameters, and

$$A = \| a_{ij} \| \tag{5}$$

is a matrix of rank k consisting of elements a_{ij} ($i = 1,2,\ldots,k$; $j = 1,2,\ldots,n$) which are known. To find the B.L.U. estimators of θ we proceed as follows: Let

$$\chi = \theta\, b' \tag{6}$$

where

$$b = (b_1, b_2, \ldots, b_k) \tag{7}$$

is a vector of k given constants. Suppose $\hat{\chi}$ is the B.L.U. estimator of χ, and

$$\hat{\chi} = y\, c' \tag{8}$$

The problem is to find c in terms of the known constants A, Ω, and b. The unbiasedness of $\hat{\chi}$ yields following condition

$$A\, c' = b'. \tag{9}$$

The variance of $\hat{\chi}$ is given by

$$c\, \Sigma\, c' = \sigma^2\, c\, \Omega\, c'. \tag{10}$$

The minimization of (10) subject to (9), whatever be the value of σ^2, requires

$$c = b(A\, \Omega^{-1} A')^{-1} A\, \Omega^{-1} \tag{11}$$

Hence, the B.L.U. estimator $\hat{\theta}$ of θ is given by

$$\hat{\theta} = y(A\, \Omega^{-1})'(A\, \Omega^{-1} A')^{-1}. \tag{12}$$

In particular, if $\Omega = I$, the identity, this gives

$$\hat{\theta} = y\, A'(A\, A')^{-1}. \tag{13}$$

Let

$$S^2 = (y - \xi)\, \Omega^{-1} (y - \xi)'. \tag{14}$$

The minimization of S^2 with respect to θ yields precisely the estimate $\hat{\theta}$ given (12). This constitutes the main result of least-squares theory.

3. Unbiased estimator of σ^2.

It is rather well known that when $\Omega = I$, an unbiassed estimator of σ^2 is given by

$$\frac{S^2_{min}}{n-k} = \frac{1}{n-k}(y - \hat{\theta} A)(y - \hat{\theta} A)' \tag{15}$$

where $\hat{\theta}$ satisfies (13). The corresponding result holds for a general Ω, and could be conjectured from maximum likelihood estimation on the assumption of normality, although the author has never seen it stated for the general case. As the general result will be referred to later, it is advisable to prove it here briefly.

Replacing $\xi = \theta A$ in (14) by $\hat{\theta} A$ from (12), we obtain

$$S^2_{min.} = y D y' \tag{16}$$

where

$$D = \Omega^{-1} - \Omega^{-1} A'(A \Omega^{-1} A')^{-1} A \Omega^{-1}. \tag{17}$$

Since

$$\xi D \xi' = 0 \tag{18}$$

it follows that

$$E\, S^2_{min.} = E(y - \xi) D(y - \xi)' = tr\, D \Sigma = \sigma^2\, tr\, D \Omega. \tag{19}$$

It merely remains to show

$$tr\, D\Omega = n - k\ . \tag{20}$$

Expanding $D\Omega$, we obtain

$$tr\, D\Omega = tr\, I - tr\, \Omega^{-1} A'(A \Omega^{-1} A')^{-1} A$$

$$= n - tr\, (A \Omega^{-1} A')^{-1} A \Omega^{-1} A' = n - k.$$

Q.E.D.

4. Transformation of the covariance matrix.

If $\Omega \neq I$, it is possible to simplify the general least-squares procedure by means of a linear transformation. Suppose

$$\det \Omega \neq 0\ . \tag{21}$$

Since Ω is positive definite, a matrix G exists, such that

$$G' \Omega G = I \tag{22}$$

and hence

$$G G' = \Omega^{-1}\ . \tag{23}$$

Let

$$Y = y G \tag{24}$$

and

$$\eta = \xi\, G = \theta\, A\, G. \qquad (25)$$

Then the minimization of

$$(Y - \eta)\,(Y - \eta)' \qquad (26)$$

with respect to θ yields the same solution $\hat{\theta}$ as in (12). To prove this we observe that the minimum value of (26) occurs when

$$\hat{\theta} = Y\,(A\,G)'(A\,G\,G'\,A')^{-1} \qquad (27)$$

as can be seen from (13). That is

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

which is the same as (12).

Consider as an example the case of autocorrelated disturbances generated by a Markoff process. Denote the disturbance by u

$$u = y - \xi \qquad (28)$$

and suppose

$$u_t - \rho\, u_{t-1} = v_t \qquad t = 1,2,\ldots,n \qquad (29)$$

where $v_1, v_2, \ldots, v_n$ are independent, with common mean 0 and variance σ^2. Taking this process to be stationary, with $|\rho| < 1$ and covariance matrix

$$\Sigma = \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \rho^{n-3} \\ & & & \ddots & \rho \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{pmatrix} \qquad (30)$$

it is easily verified that

$$\Omega^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \cdots & & & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots & & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ \vdots & & & & & & 0 \\ \vdots & & & \cdots & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & \cdots & \cdots & 0 & \cdots & -\rho & 1 \end{pmatrix} \tag{31}$$

In this case,

$$G = \begin{pmatrix} \sqrt{1-\rho^2} & -\rho & 0 & \cdots & & 0 \\ 0 & 1 & -\rho & 0 & \cdot & 0 \\ 0 & 0 & 1 & -\rho & \cdots & 0 \\ \vdots & & & & & \\ \vdots & & \cdots & 0 & 1 & -\rho \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix} \tag{32}$$

Hence

$$Y = (\sqrt{1-\rho^2}\, y_1,\ y_2 - \rho\, y_1,\ y_3 - \rho\, y_2, \ldots\ldots,\ y_n - \rho\, y_{n-1}). \tag{33}$$

This is the transformation suggested by Cochrane and Orcutt [3].

5. Families of unbiased estimators of θ.

It should be pointed out that any non-singular matrix Λ, used in place of Ω in (12), yields an unbiased estimator of θ. Let

$$\hat{\hat{\theta}} = y\,(A\Lambda^{-1})'\,(A\Lambda^{-1}A')^{-1} \tag{34}$$

which minimizes

$$(y - \xi)\Lambda^{-1}(y - \xi)' \tag{35}$$

with respect to θ. Although the covariance matrix of y is $\Sigma = \sigma^2\Omega$, $\hat{\hat{\theta}}$ is unbiased, since

$$E\,\hat{\hat{\theta}} = \theta A\,(A\Lambda^{-1})'\,(A\Lambda^{-1}A')^{-1} = \theta. \tag{36}$$

This fact is useful when considering the parameters of Ω as unknown. Suppose the elements w_{ij} of Ω are functions of r parameters $\gamma_1, \gamma_2, \ldots, \gamma_r$.

$$w_{ij} = w_{ij}(\gamma_1, \gamma_2, \ldots, \gamma_r) \tag{37}$$

Write

$$\Omega = \Omega_\gamma \tag{38}$$

as a symbolic notation of (37), where

$$\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r) \tag{39}$$

and

$$\hat{\theta} = \hat{\theta}(\gamma) \tag{40}$$

when in (12), $\Omega = \Omega_\gamma$. Let

$$E\{\hat{\theta}(\gamma) | \overset{\circ}{\gamma}\} \tag{41}$$

be the expectation of $\hat{\theta}(\gamma)$ on the assumption that $\overset{\circ}{\gamma}$ is the true value of γ. Then, in virtue of (36)

$$E\{\hat{\theta}(\gamma) | \overset{\circ}{\gamma}\} = \theta \tag{42}$$

no matter what value γ assumes. Hence $\hat{\theta}(\gamma)$ may be regarded as an r-parameter family of unbiased estimates of θ.

Let

$$E[\{\hat{\theta}(\gamma) - \theta\} | \gamma_o]^2 = V_{\overset{\circ}{\gamma}} \hat{\theta}(\gamma) \tag{43}$$

be the variance of $\hat{\theta}(\gamma)$ when $\overset{\circ}{\gamma}$ is the correct value. The implication of the least-squares theory of §2 is

$$\min_\gamma V_{\overset{\circ}{\gamma}} \hat{\theta}(\gamma) = V_{\overset{\circ}{\gamma}} \hat{\theta}(\overset{\circ}{\gamma}). \tag{44}$$

If the transformation of §4 is used, where

$$G \Omega_\gamma G' = I \tag{45}$$

the same estimate $\hat{\theta}(\gamma)$ as above is obtained, with the same properties.

6. The concentration ellipsoid of Cramér.

It is advisable, at this point, to state briefly the results of Cramér ([4], [5]). Suppose the random variables $z_1, z_2, \ldots, z_n$ have a joint distribution involving k unknown parameters

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \tag{46}$$

where $k < n$. Let

$$\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_k) \tag{47}$$

be a vector of k functions of $z_1, z_2, \ldots, z_n$, not involving α, which are unbiassed estimators of α.

That is

$$E\,\hat{\alpha} = \alpha \tag{48}$$

whatever be the values of α. Assume the joint distribution of $\hat{\alpha}$ is non-singular, and that its covariance matrix Λ is non-singular. The equation

$$(w - \alpha)\Lambda^{-1}(w - \alpha)' = k + 2 \tag{49}$$

defines the concentration ellipsoid of the estimators $\hat{\alpha}$. Its volume is given by

$$\frac{(k+2)^{k+2}}{\Gamma(\frac{k}{2}+1)} \qquad \sqrt{\det \Lambda} \tag{50}$$

The quantity $\det \Lambda$ is sometimes called the generalized variance. If a unit mass were uniformly distributed over the interior of the ellipsoid (49), it would have the same first moments α and the same second order moments Λ as the distribution of $\hat{\alpha}$.

It is proven by Cramér that subject to certain conditions of regularity depending on the distribution of $z_1, z_2, \ldots, z_n$, there is a fixed ellipsoid which lies entirely within the concentration ellipsoid of any set of unbiassed estimators $\hat{\alpha}$. This ellipsoid has the equation

$$(w - \alpha)\,L\,(w - \alpha)' = k + 2 \tag{51}$$

where

$$L = \left\| E\,\frac{\partial \log f}{\partial \alpha_i}\;\frac{\partial \log f}{\partial \alpha_j} \right\| \tag{52}$$

and where f is the probability density of $z_1, z_2, \ldots, z_n$. (If the z's are discrete, f is taken as the probability of the z's.)

If a set of estimators $\hat{\alpha}$ has a concentration ellipsoid which coincides with (51), it is called a set of joint efficient estimators. In such a case, the variance of these estimators is minimal within the respective classes of unbiased estimators.

7. Proposed methods of estimating γ.

First, let us find the covariance matrix of $\hat{\theta}(\overset{o}{\gamma})$, on the assumption $\overset{o}{\gamma}$ is the correct value. Let χ be as defined in (6), and m be defined by

$$m = \theta d' \tag{53}$$

where

$$d = (d_1, d_2, \ldots, d_k). \tag{54}$$

is a vector of k given constants. Let $\hat{\chi}_{\overset{o}{\gamma}}$, $\hat{m}_{\overset{o}{\gamma}}$ be the B.L.U. estimates of χ, m respectively. Then

$$\hat{\chi}_{\overset{o}{\gamma}} = y c' \tag{55}$$

where

$$c = b (A \Omega_{\overset{o}{\gamma}}^{-1} A')^{-1} A \Omega_{\overset{o}{\gamma}}^{-1} \tag{56}$$

and

$$\hat{m}_{\overset{o}{\gamma}} = y e' \tag{57}$$

where

$$e = d (A \Omega_{\overset{o}{\gamma}}^{-1} A')^{-1} A \Omega_{\overset{o}{\gamma}}^{-1}. \tag{58}$$

The covariance of $\hat{\chi}_{\overset{o}{\gamma}}$, $\hat{m}_{\overset{o}{\gamma}}$ is given by

$$E \left\{ (\hat{\chi}_{\overset{o}{\gamma}} - \chi)(\hat{m}_{\overset{o}{\gamma}} - m) \mid \overset{o}{\gamma} \right\} = c \Sigma_{\overset{o}{\gamma}} e' \tag{59}$$

which, by (56, (58) reduces to

$$\sigma^2 b (A \Omega_{\overset{o}{\gamma}}^{-1} A')^{-1} d' . \tag{60}$$

It follows that

$$E \left\{ (\hat{\theta}(\overset{o}{\gamma}) - \theta)'(\hat{\theta}(\overset{o}{\gamma}) - \theta) \mid \overset{o}{\gamma} \right\} = \sigma^2 (A \Omega_{\overset{o}{\gamma}}^{-1} A')^{-1}. \tag{61}$$

Hence, in Cramér's terminology, the concentration ellipsoid of the estimators $\hat{\theta}(\gamma)$ is given by

$$\sigma^2 (w - \theta) A \Omega_{\gamma}^{-1} A' (w - \theta)' = k + 2. \tag{62}$$

Let $\hat{\sigma}^2(\gamma)$ be the unbiased estimator of σ^2 given in § 3. That is

$$\hat{\sigma}^2(\gamma) = \frac{1}{n-k}(y - \hat{\theta}(\gamma)A)\Omega_{\gamma}^{-1}(y - \hat{\theta}(\gamma)A)'. \quad (63).$$

The first suggested method of finding an estimator $\hat{\gamma}$ of γ is as follows:

Method I:

Minimize

$$\left\{\hat{\sigma}^2(\gamma)\right\}^k \det(A\Omega_{\gamma}^{-1} A')^{-1} \quad (64)$$

with respect to γ. That is

$$\min_{\gamma}\left\{\hat{\sigma}^2(\gamma)\right\}^k \det(A\Omega_{\gamma}^{-1} A')^{-1} = \left\{\hat{\sigma}^2(\hat{\gamma})\right\}^k \det(A\Omega_{\hat{\gamma}}^{-1} A')^{-1}. \quad (65)$$

The rationale behind this method of estimation is as follows. As stated in § 6, Cramér has shown that unbiased estimators with minimal variance have the smallest generalized variance. Our suggested method minimizes the generalized variance for a concentration ellipsoid which differs from Cramer's ellipsoid in two respects. First, the estimators $\hat{\theta}(\gamma)$ involve unknown parameters γ; second, the matrix of coefficients

$$\hat{\sigma}^2(\gamma) A\Omega_{\gamma}^{-1} A' \quad (66)$$

is random, due to the introduction of $\hat{\sigma}^2(\gamma)$ in place of σ^2. Hsu [6], has shown, for the case $\Omega = I$, that in a large class of situations the estimator $\hat{\sigma}^2$ of §3 is the best quadratic unbiased estimator of σ^2; that is, the variance of $\hat{\sigma}^2$ is minimal within the set of quadratic unbiassed estimators of σ^2. By the same methods it can be shown that $\hat{\sigma}^2$ is best quadratic unbiased when $\Omega \neq I$. Hence, the choice of $\hat{\sigma}^2(\gamma)$ to replace σ^2 seems reasonable. The stochastization of the concentration ellipsoid in this manner, and the minimization of its volume suggests a method of obtaining estimators the properties of which should come close to those given by Cramér's theory.

The estimators $\hat{\theta}(\hat{\gamma})$ are, of course, unbiased, in virtue of (42), but the author has not yet succeeded in deriving all the properties of the estimators $\hat{\theta}(\hat{\gamma})$ and $\hat{\gamma}$, the method of obtaining which has been formulated above. A somewhat heuristic justification of the procedure may be given as follows: The covariance of $\hat{l}_\gamma$, $\hat{m}_\gamma$, on the assumption $\overset{\circ}{\gamma}$ is the correct value, is given by

$$\sigma^2 b(A\,\Omega_\gamma^{-1}\,A')^{-1}\,A\,\Omega_\gamma^{-1}\,\Omega_{\overset{\circ}{\gamma}}\,\Omega_\gamma^{-1}\,A'(A\,\Omega_\gamma^{-1}\,A')^{-1}\,d'. \qquad (67)$$

Hence, by (44)

$$\min_\gamma \det (A\,\Omega_\gamma^{-1}\,A')^{-1}\,A\,\Omega_\gamma^{-1}\,\Omega_{\overset{\circ}{\gamma}}\,\Omega_\gamma^{-1}\,A'(A\,\Omega_\gamma^{-1}\,A')^{-1} = \det (A\,\Omega_{\overset{\circ}{\gamma}}^{-1}\,A')^{-1}. \qquad (68)$$

In a small neighborhood of $\overset{\circ}{\gamma}$, the determinant in the left member of (68) is approximately

$$\det (A\,\Omega_\gamma^{-1}\,A')^{-1}. \qquad (69)$$

Now consider the estimator $\hat{\sigma}^2(\gamma)$ in a small neighborhood of $\overset{\circ}{\gamma}$.

$$E\left\{\sigma^2(\gamma)\,\middle|\,\gamma_0\right\} = E\left\{(y\,D\,y')\,\middle|\,\gamma^0\right\} \qquad (70)$$

where D is given by (17). Hence

$$E\left\{\hat{\sigma}^2(\gamma)\,\middle|\,\overset{\circ}{\gamma}\right\} = \frac{\sigma^2}{n-k}\,\mathrm{tr}\,D\,\Omega_{\overset{\circ}{\gamma}} \qquad (71)$$

$$= \sigma^2 + \frac{\sigma^2}{n-k}\,\mathrm{tr}\left\{\Omega_\gamma^{-1} - \Omega_\gamma^{-1}\,A'(A\,\Omega_\gamma^{-1}\,A')^{-1}\,A\,\Omega_\gamma^{-1}\right\}\left\{\Omega_{\overset{\circ}{\gamma}} - \Omega_\gamma\right\}.$$

and the bias will be insignificant in a small neighborhood about $\overset{\circ}{\gamma}$. Thus, the minimization of (64), should yield an estimator $\hat{\gamma}$ which, with high probability, clusters about the true value $\overset{\circ}{\gamma}$; consequently the variance of $\hat{e}(\hat{\gamma})$ should be close to the lower bound given in (61).

A second method is suggested which considers only the diagonal elements of the covariance matrix

$$\sigma^2(A\,\Omega_\gamma^{-1}\,A')^{-1} \qquad (72)$$

Let the diagonal elements be

$$\sigma^2\,\tau_1(\gamma),\ \sigma^2\,\tau_2(\gamma),\ldots\ldots,\sigma^2\,\tau_k(\gamma). \qquad (73)$$

The following method is suggested for obtaining an estimator $\hat{\gamma}$ of γ.

<u>Method II.</u>

Minimize

$$\left\{\sigma^2(\gamma)\right\}^k \prod_{i=1}^{k} \tau_i(\gamma), \qquad (74)$$

with respect to γ. That is

$$\min_{\gamma} \left\{ \hat{\sigma}^2(\gamma) \right\}^k \prod_{i=1}^{k} \tau_i(\gamma) = \left\{ \hat{\sigma}^2(\hat{\gamma}) \right\}^k \prod_{i=1}^{k} \tau_i(\hat{\gamma}). \qquad (75)$$

The justification of this method is suggested by (44) and the properties of $\hat{\sigma}^2(\gamma)$ discussed under method I. Since the true variances of all k components of $\hat{\theta}(\gamma)$ are minimum when $\gamma = \overset{o}{\gamma}$, it seems at least plausible that the value $\hat{\gamma}$ which minimizes the product of estimators of these variances should cluster around $\overset{o}{\gamma}$ with high probability. This should imply a relatively small variance of $\hat{\theta}(\hat{\gamma})$.

As an example, linear regression with autocorrelated disturbances will be discussed in § 7.

7. <u>Example: Linear regression with autocorrelated disturbances.</u>

The above methods will now be applied to the example mentioned in § 4. Let

$$A = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & & x_{2n} \\ \vdots & & & \\ x_{k-1,1} & x_{k-1,2} & \cdots & x_{k-1,n} \\ 1 & 1 & & 1 \end{pmatrix} \qquad (76)$$

Let

$$\left. \begin{aligned} Y_j &= y_j - \rho y_{j-1} \\ X_{ij} &= x_{ij} - \rho x_{i,j-1} \end{aligned} \right\} \qquad \begin{aligned} i &= 1,2,\ldots,k \\ j &= 2,3,\ldots,n \end{aligned} \qquad (77)$$

and

$$\left. \begin{aligned} Y_1 &\quad \sqrt{1-\rho^2}\, y_1 \\ X_{i1} &\quad \sqrt{1-\rho^2}\, x_{i1} \end{aligned} \right\} \qquad i = 1,2,\ldots,k \qquad (78)$$

Then

$$A\,G = \| X_{ij} \| \qquad i = 1,2,\ldots,k \quad j = 1,2,\ldots\ldots,n. \tag{79}$$

where G is defined by (32). Let

$$\left.\begin{array}{l} S_{ih} = \sum_{t=1}^{n} X_{it} X_{jt} \\ S_{Yh} = \sum_{t=1}^{n} Y_t X_{ht} \end{array}\right\} \qquad i = 1,2,\ldots,k \quad h = 1,2,\ldots,k. \tag{80}$$

and

$$S_{YY} = \sum_{t=1}^{n} Y_t^{\,2} \tag{81}$$

Then

$$A\,\Omega_{\rho}^{-1} A' = A\,G_{\rho}\,G_{\rho}'A' = \| S_{ih} \| = S_{\rho} \text{ , say.} \tag{82}$$

and

$$\hat{\sigma}^2(\rho) = \frac{\det T}{\det S} \tag{83}$$

where

$$T_{\rho} = \begin{pmatrix} S_{11} & \ldots & S_{1k} & S_{Y1} \\ \vdots & & & \vdots \\ S_{k1} & \ldots & S_{kk} & S_{Yk} \\ S_{Y1} & \ldots & S_{Yk} & S_{YY} \end{pmatrix} \tag{84}$$

Hence, $\hat{\rho}$ is obtained, according to Method I, by minimizing

$$\frac{(\det T_{\rho})^{k}}{(\det S_{\rho})^{k+1}} \tag{85}$$

with respect to ρ . When k = 2, (85) becomes

$$\frac{\begin{vmatrix} S_{11} & S_{12} & S_{Y1} \\ S_{21} & S_{22} & S_{Y2} \\ S_{Y1} & S_{Y2} & S_{YY} \end{vmatrix}^{2}}{\begin{vmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}^{3}} \tag{86}$$

whilst, according to Method II, an estimate of ρ is obtained by minimizing

$$\frac{s_{11} \quad s_{22}}{\begin{vmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{vmatrix}^4} \begin{vmatrix} s_{11} & s_{12} & s_{Y1} \\ s_{21} & s_{22} & s_{Y2} \\ s_{Y1} & s_{Y2} & s_{YY} \end{vmatrix}^2 \qquad (87)$$

An approximate solution, in which the computations are considerably simplified, is obtained by neglecting the terms in (78). For instance, when $k = 2$, $\hat{\rho}$ would be obtained, according to Method II, by solving the following sixth degree equation in ρ.

$$2(1-\rho)\, s'_{11} \left\{ s'_{11}(r_{221}+2r_{222}\rho) + s'_{22}(r_{111}+2r_{112}\rho) - 2s'_{12}(r_{121}+2r_{122}\rho) \right\}$$
$$- 3(1-\rho)(s'_{11}\, s'_{22} - s'^2_{12})(r_{111} + r_{112}\rho) + 2s'_{11}(s'_{11}\, s'_{22} - s'^2_{12}) = 0 \qquad (88)$$

where

$$s'_{ij} = r_{ijo} + r_{ij,1}\rho + r_{ij,2}\rho^2 \qquad i = 1,2 \quad j = 1,2 \qquad (89)$$

$$\left.\begin{aligned} r_{ij,o} &= \sum_2^n (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j) \\ - r_{ij,1} &= \sum_2^n (z_{i,t-1} - \bar{z}_{i,-1})(z_{j,t-1} - \bar{z}_{j,-1}) + \sum_2^n (z_{jt} - \bar{z}_j)(z_{i,t-1} - \bar{z}_{i,-1}) \\ r_{ij,2} &\quad \sum_2^n (z_{i,t-1} - \bar{z}_{i,-1})(z_{j,t-1} - \bar{z}_{j,-1}) \end{aligned}\right\} \qquad (90)$$

$$\left.\begin{aligned} z_{1t} &= x_t \\ z_{2t} &= y_t \end{aligned}\right\} \qquad t = 1,2,\ldots,n. \qquad (91)$$

and

$$\left.\begin{aligned} \bar{z}_i &= \frac{1}{n-1} \sum_2^n z_{it} \\ \bar{z}_{i,-1} &= \frac{1}{n-1} \sum_2^n z_{i,t-1} \end{aligned}\right\} \qquad i = 1,2. \qquad (92)$$

8. Conclusion.

It appears from the example above that the computations required are no less involved than those required in maximum likelihood estimation. For certain forms of Ω the computations could be considerably simplified. There are certain advantages in the suggested methods, however, as stated earlier. First, the principles on which the estimation is based are stated in terms of finite size samples. Second, no assumption regarding the form of the distribution of the disturbances is required. The properties of these estimates is now under study.

REFERENCES

[1] F. N. David and J. Neyman, "Extension of the Markoff Theorem on Least Squares," Statistical Research Memoirs, Vol. II, London, 1938, pp. 105-116.

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

[3] D. Cochrane and G.H. Orcutt, "Application of least Squares Regression to Relationship containing Autocorrelated Error Terms," Journal of the American Statistical Association, Vol. 44, 1949, pp. 32-61.

[4] H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, 1946.

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

[6] P. L. Hsu, "On the Best Unbiased Quadratic Estimate of the Variance," Vol. II, London, 1938, pp. 91-104.