COVARIANCES OF LEAST-SQUARES ESTIMATES WHEN RESIDUALS ARE CORRELATED[1]

By M. M. Siddiqui[2]

University of North Carolina

1. Summary. In this paper we will study the effects on the covariance matrix of the least-squares estimates of regression coefficients and on the estimate of the residual variance when the usual condition of independence of residuals is violated. The cases of linear trend and of regression on trigonometric functions will be considered in some detail.

2. Introduction. Several authors have studied the problem of estimating regression coefficients when residuals are autocorrelated. We refer here only to the work of Grenander and Rosenblatt [2, 3, 4]. Grenander [2] gives conditions on the regression variables for the existence of consistent estimates of the regression coefficients. He also gives conditions on the residual process under which the least-squares (L.S.) estimate of a regression coefficient is asymptotically efficient with respect to the Markov estimate. The covariances of the L.S. estimates as summarized in a matrix form are well known and are given at the end of section 3. The exact expression for an individual covariance or variance in the general case is easily extracted from this matrix and is given in section 4. The variance of the L.S. estimate in the general case is also given by Grenander [2, (8) p. 258]. Asymptotic expressions for the covariances of these estimates are also available [2, 4]. However, it seemed desirable to present here, in some detail, exact expressions or high order approximations to them for the individual variances and covariances of the L.S. estimates of regression coefficients and for the expectation of the estimate of residual variance, particularly for the cases of general interest, in readily usable form, and derived in an elementary fashion. The first term of each of our expressions coincide with the asymptotic expression given in [2, 4], when the regression coefficients are made comparable.

Bounds on the covariances of L.S. estimates are also provided in (7).

3. The L.S. estimates. Let $y = x'\beta + \Delta$ be the regression equation, where y and Δ are $N \times 1$ column vectors, β is a $p \times 1$ column vector, x is a $p \times N$ matrix and a prime is used to denote the transpose of a matrix or a vector. It is assumed that $N > p$, x is non-stochastic and of rank p, and Δ is a $N(0, \sigma^2 P)$ vector variate, where 0 is a zero vector and P is a positive definite correlation matrix.

Introducing $c = (xx')^{-1}$, $b = cxy$, $v = y - x'b$, $n = N - p$, and writing δq

Received December 3, 1957; revised June 13, 1958.

[1] Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at the Boulder Laboratories, National Bureau of Standards.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

for $q - Eq$, where q is a variate with expected value Eq, it is known that b and $s^2 = v'v/n$ are the least-squares estimates of β and σ^2 respectively. It is also known that $Eb = \beta$, and

$$B = E\delta b\delta b' = \sigma^2 cxPx'c.$$

In case $P = I_N$, where I_N is the $N \times N$ identity matrix, $Es^2 = \sigma^2$ and $B = \sigma^2 c$.

4. The covariance matrix B. We propose to study the effects on B and Es^2 when P is given by

$$P = I_N + \sum_{k=1}^{N-1} \rho_k(C^k + C'^k), \tag{1}$$

where

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{2}$$

i.e. when

$$E\Delta_t\Delta_{t\pm k} = \sigma^2\rho_k, \qquad k = 0, 1, \cdots, N-1, \rho_0 = 1. \tag{3}$$

We have

$$v = y - x'b = x'(\beta - b) + \Delta = (I_N - x'cx)\Delta \tag{4}$$

as $b = \beta + cx\Delta$. Writing $m = x'cx$, we have $m' = m$ and $m^2 = m$. Hence if λ is a characteristic root of m, $\lambda = 0$ or 1. Writing "tr" for the trace of a matrix we obtain tr $m = p$. Now, by simple evaluation

$$Es^2 = \frac{\sigma^2}{n}[N - \operatorname{tr} Pm] = \sigma^2\left[1 - \frac{2}{n}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k} \rho_k\, m_{t,t+k}\right]. \tag{5}$$

Here, if e is a matrix, e_{ij} or $e_{i,j}$ refers to its element in the ith row and the jth column.

If we write $d = cx$, we find that

$$B_{ij} = E\delta b_i\,\delta b_j = \sigma^2\left[\sum_{t=1}^{N} d_{it}\,d_{jt} + \sum_{k=1}^{N-1}\sum_{t=1}^{N-k}\rho_k(d_{jt}\,d_{i,t+k} + d_{it}\,d_{j,t+k})\right].$$

If, by a proper choice of x or with a suitable transformation on x, we make $xx' = c^{-1} = I_p$, we have $d = x$. Writing x_i' for the row vector in the ith row of x, we find

$$B_{ij} - \sigma^2\delta_{ij} = \sigma^2\sum_{k=1}^{N-1}\rho_k\, x_i'(C^k + C'^k)x_j, \qquad i, j = 1, \cdots, p; \tag{6}$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$.

It has been shown [1, p. 130] that if A is an $N \times N$ real symmetric matrix with characteristic roots $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_N$, and u and v are $N \times 1$ real vectors, then, under the conditions $u'u = v'v = 1$, $u'v = 0$, the bilinear form $u'Ay$ has a maximum $(\alpha_N - \alpha_1)/2$ and a minimum $(\alpha_1 - \alpha_N)/2$. Also the quadratic form $u'Au \leqq \alpha_N$. Now the maximum characteristic root of $C^k + C'^k$, where k is a positive integer,

$$\alpha_N = 2\cos\left\{\frac{\pi}{\left[\frac{N+k-1}{k}\right]+1}\right\} \leqq 2\cos\left\{\frac{k\pi}{N+2k-1}\right\},$$

where $[q]$ denotes the largest integer $\leqq q$, and the minimum characteristic root $\alpha_1 = -\alpha_N$, [1, p. 101]. Hence, we obtain

$$(7) \qquad | B_{ij} - \sigma^2\delta_{ij} | < 2\sigma^2 \sum_{k=1}^{N-1} \left| \rho_k \cos\frac{k\pi}{N+2k-1} \right| < 2\sigma^2 \sum_{k=1}^{N-1} | \rho_k | .$$

In the case $\rho_k = \rho^k$, $k = 0, 1, \cdots$, where $\alpha = |\rho| < 1$, we have

$$(8) \qquad | B_{ij} | < \frac{2\alpha\sigma^2}{1-\alpha} \quad \text{if } i \neq j, \qquad B_{ii} < \sigma^2\left(\frac{1+\alpha}{1-\alpha}\right).$$

5. Linear trend. Let $N = 2r + 1$ where r is a positive integer and consider the linear trend in the form

$$(9) \qquad y_t = \beta_1(2r+1)^{-1/2} + \beta_2(t-r-1)/a + \Delta_t, \qquad t = 1, \cdots, N,$$

where

$$(10) \qquad a^2 = r(r+1)(2r+1)/3 = N(N^2-1)/12.$$

In the notation of section 3

$$(11) \qquad \begin{aligned} x_{1t} &= (2r+1)^{-1/2}, \qquad x_{2t} = (t-r-1)/a, \qquad t = 1, \cdots, N, \\ b_1 &= \sqrt{N}\bar{y}, \qquad b_2 = \left[\sum_{t=1}^{N} ty_t - (r+1)\sum_{t=1}^{N} y_t\right] \Big/ a. \end{aligned}$$

Furthermore

$$c = I_p, \qquad p = 2, \qquad n = N - 2,$$

$$m_{ij} = (x'x)_{ij} = \frac{1}{N} + \frac{3(2i-N-1)(2j-N-1)}{N(N^2-1)}$$

$$ns^2 = \sum y_t^2 - b_1^2 - b_2^2,$$

$$(12) \qquad B_{11} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right], \qquad B_{12} = 0,$$

$$B_{22} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\rho_k - \frac{2}{N}\left(3 + \frac{2}{N^2-1}\right)\sum_{k=1}^{N-1} k\rho_k + \frac{4}{N(N^2-1)}\sum_{k=1}^{N-1} k^3\rho_k\right]$$

and

$$Es^2 = \sigma^2\left[1 - \frac{4}{n}\sum_{k=1}^{N-1}\rho_k + \frac{2}{nN}\left(4 + \frac{2}{N^2-1}\right)\sum_{k=1}^{N-1}k\rho_k - \frac{4}{nN(N^2-1)}\sum_{k=1}^{N-1}k^3\rho_k\right]$$

For the case when $\rho_k = \rho^k$, we can evaluate the summations $\sum \rho_k$, $\sum k\rho_k$, etc. If N is moderately large we may neglect ρ^N and thus find

$$E\frac{s^2}{\sigma^2} \cong 1 - \frac{4\rho}{n(1-\rho)} + \frac{8\rho}{nN(1-\rho)^2} + \frac{4(\rho + 4\rho^2 + \rho^3)}{nN(N^2-1)(1-\rho)^4},$$

(13)
$$\frac{B_{11}}{\sigma^2} = 1 - \frac{2}{N}\frac{\rho - N\rho^N + (N-1)\rho^{N+1}}{(1-\rho)^2} + \frac{2(\rho - \rho^N)}{1-\rho} \cong \frac{1+\rho}{1-\rho} - \frac{2\rho}{N(1-\rho)^2},$$

$$B_{12} = 0, \qquad \frac{B_{22}}{\sigma^2} \cong \frac{1+\rho}{1-\rho} - \frac{6\rho}{N(1-\rho)^2}.$$

We note that b_i are independently distributed $N(\beta_i, B_{ii})$, $i = 1, 2$, variates. If we set

$$b_1' = N^{-1/2}b_1 = \bar{y}, \qquad b_2' = \frac{\sqrt{12}b_2}{\sqrt{N(N^2-1)}},$$

the estimate of Ey_t is given by

(14) $$Y_t = \bar{y} + b_2'(t - r - 1)$$

and under the first order autoregressive scheme for Δ's,

(15) $$\sigma_{\bar{y}}^2 \cong \frac{\sigma^2}{N}\left(\frac{1+\rho}{1-\rho}\right), \qquad \sigma_{b_2'}^2 \cong \frac{12\sigma^2}{N(N^2-1)}\left(\frac{1+\rho}{1-\rho}\right), \qquad \text{cov}(\bar{y}, b_2') = 0.$$

Thus

$$\sigma_{Y_t}^2 \cong \frac{\sigma^2}{N}\left(\frac{1+\rho}{1-\rho}\right)\left[1 + \frac{12(t-r-1)^2}{N^2-1}\right].$$

6. Regression on trigonometric functions. Consider

(16)
$$y_t = \beta_1/\sqrt{N} + \sqrt{2/N}\sum_{i=1}^{q}\beta_{2i}\cos\lambda_i t + \sqrt{2/N}\sum_{i=1}^{q}\beta_{2i+1}\sin\lambda_i t + \Delta_t, \qquad t = 1, \cdots, N,$$

where $\lambda_i = 2\pi\omega_i/N$ and ω_i is a positive integer less than N for $i = 1, 2, \cdots, q$ and $\omega_i \neq \omega_j$ if $i \neq j$.

In the notation of section 2

$$x_{1t} = 1/\sqrt{N}, \qquad x_{2i,t} = \sqrt{2/N}\cos\lambda_i t,$$
$$x_{2i+1,t} = \sqrt{2/N}\sin\lambda_i t, i = 1, 2, \cdots, q; t = 1, 2, \cdots, N,$$
$$xx' = c^{-1} = I_{2q+1}, \qquad n = N - 2q - 1,$$
$$b_1 = \sqrt{N}\bar{y}, \qquad b_{2i} = \sqrt{2/N}\sum_t y_t \cos\lambda_i t,$$
$$b_{2i+1} = \sqrt{2/N}\sum_t y_t \sin\lambda_i t, i = 1, \cdots, q, \tag{17}$$
$$m_{ts} = 1/N + 2/N\sum_{i=1}^{q}\cos(t - s)\lambda_i, \qquad t, s = 1, \cdots, N,$$
$$s^2 = \left(\sum_t y_t^2 - \sum_{i=1}^{2q+1} b_i^2\right)\Big/ n.$$

We find

$$E\frac{s^2}{\sigma^2} = 1 - \frac{2}{n}\sum_{k=1}^{N-1}\rho_k + \frac{2}{nN}\sum_{k=1}^{N-1}k\rho_k - \frac{4}{n}\sum_{k=1}^{N-1}\sum_{i=1}^{q}\left(1 - \frac{k}{N}\right)\rho_k\cos k\lambda_i. \tag{18}$$

For the covariances of b_i and b_j we obtain

$$B_{11} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right],$$
$$B_{1,2i} = \frac{\sqrt{2}\,\sigma^2}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k}\rho_k\{\cos(t + k)\lambda_i + \cos t\lambda_i\}, \qquad i = 1, \cdots, q,$$
$$B_{2i,2i} = \sigma^2\left[1 + \frac{4}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k}\rho_k\cos(k + t)\lambda_i\cos t\lambda_i\right], \qquad i = 1, \cdots, q. \tag{19}$$
$$B_{2i,2j+1} = \frac{2\sigma^2}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k}\rho_k\{\cos t\lambda_i\sin(k + t)\lambda_j + \cos(k + t)\lambda_i\sin t\lambda_j\}, \qquad i,j = 1, \cdots, q.$$

$B_{1,2i+1}$ and $B_{2i+1,2i+1}$ are obtainable from the expressions for $B_{1,2i}$ and $B_{2i,2i}$ respectively by replacing cosine by sine.

If $\rho_k = \rho^k$, and ρ^N is negligible, we find for the variances, after some reduction,

$$\frac{B_{11}}{\sigma^2} \cong \frac{1 + \rho}{1 - \rho} - \frac{2\rho}{N(1 - \rho)^2},$$
$$\frac{B_{2i,2i}}{\sigma^2} \cong \frac{1 - \rho^2}{1 - 2\rho\cos\lambda_i + \rho^2} + \frac{\rho\cos\lambda_i}{N(1 - 2\rho\cos\lambda_i + \rho^2)} - \frac{\rho(1 + \rho^2)\cos\lambda_i - 2\rho^2}{N(1 - 2\rho\cos\lambda_i + \rho^2)^2}, \tag{20}$$
$$\frac{B_{2i+1,2i+1}}{\sigma^2} \cong \frac{1 - \rho^2}{1 - 2\rho\cos\lambda_i + \rho^2} - \frac{\rho\cos\lambda_i}{N(1 - 2\rho\cos\lambda_i + \rho^2)} - \frac{\rho(1 + \rho^2)\cos\lambda_i - 2\rho^2}{N(1 - 2\rho\cos\lambda_i + \rho^2)^2}, \qquad i = 1, \cdots, q.$$

Also

$$E\frac{s^2}{\sigma^2} \cong 1 - \frac{2\rho}{n(1-\rho)} - \frac{4\rho}{n}\sum_{i=1}^{q}\frac{\cos\lambda_i - \rho}{1 - 2\rho\cos\lambda_i + \rho^2} + O\left(\frac{1}{N^2}\right). \tag{21}$$

7. Concluding remarks. We conclude with the remarks that in most practical cases the correlation matrix for Δ's will not be known. However, if Δ's may be represented as a stationary autoregressive process of some small order—in many cases first or second order scheme gives a reasonably good fit—we would be required to estimate a few parameters $\rho_1, \rho_2, \cdots, \rho_k$. We, however, note that these quantities do not appear in b and s^2, only in B and Es^2.

We further observe that the estimates, $\hat{\beta}$ and $\hat{\sigma}^2$, of β and σ^2 obtained from maximizing the likelihood function will depend on the parameters of P, i.e. on $\rho_1, \rho_2, \cdots, \rho_{N-1}$, which will mean using sample serial correlation coefficients to estimate ρ's in the expression for $\hat{\beta}$ and $\hat{\sigma}^2$. These estimates will obviously be non-linear. Thus it seems more desirable to stick to the least-squares estimates b and s^2 rather than to attempt to develop maximum-likelihood (or minimum χ^2) estimates.

8. Acknowledgement. The writer wishes to express his indebtedness to Professor Harold Hotelling for drawing his attention to this problem.

REFERENCES

[1] M. M. SIDDIQUI, *Distributions of Some Serial Correlation Coefficients*, University of North Carolina Institute of Mathematical Statistics Mimeograph Series No. 164, Chapel Hill (1957).

[2] ULF GRENANDER, "On the estimation of regression coefficients in the case of an autocorrelated disturbance," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 252–272.

[3] MURRAY ROSENBLATT, "On the estimation of regression coefficients of a vector-valued time series with a stationary residual," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 99–121.

[4] ULF GRENANDER AND MURRAY ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957, Chapter 7.