TESTS FOR REGRESSION COEFFICIENTS WHEN ERRORS ARE CORRELATED

By M. M. Siddiqui

Boulder Laboratories, National Bureau of Standards

1. Summary. In a previous paper [6] the covariances of least-squares estimates of regression coefficients and the expected value of the estimate of residual variance were investigated when the errors are assumed to be correlated. In this paper we will investigate the distribution of the usual test statistics for regression coefficients under the same assumptions. Applications of the theory to the cases of testing a single sample mean, the difference between the means of two samples, the coefficients in a linear trend and in regression on trigonometric functions will be discussed in some detail under an assumed covariance matrix for errors.

2. Introduction. Several authors have studied the effects on common tests of significance when one or another of the ideal conditions is not satisfied. The effect of correlation between errors on t and z tests for means has been investigated by Daniels [3]. Box, in a series of excellent papers, including [1] and [2], has studied the problem of unequal variances of errors and correlations between them in analysis of variance situations. In continuation of these investigations, it seemed desirable to study in some detail the distributions of common tests of significance, or their variations, for regression coefficients in the usual cases of interest. The results contained in this paper may be considered an extension of Daniels', Box's and Welch's [7] work.

3. Test statistics for regression coefficients. Let $y = x'\beta + \Delta$ be the observation 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 $N(\mathbf{0}, \sigma^2 P)$ vector variate, where $\mathbf{0}$ is a zero vector and P is a positive definite correlation matrix. The notation $N(a, D)$ is used for the normal distribution with mean vector a and covariance matrix D. P will be assumed to have a specified structure, given by (3.8), and to be known. Although, in principle, a non-singular transformation on y exists which takes one back to the standard case, from practical considerations it seemed worthwhile to study the effects on the usual test statistics when β is estimated by minimizing $\Delta'\Delta$ instead of $\Delta' P \Delta$. Some of the reasons for doing this are given in the last section. We further assume that x is so chosen that $xx' = I_p$, the $p \times p$ identity matrix, so that the elements x_{ij} of x are of the order $N^{-\frac{1}{2}}$. This assumption is no restriction in principle, and, even in practice, a simple modification in x may be sufficient. For example, in the case of a linear or a polynomial trend, orthogonal polynomials may be used; in the case of regression on the mean or on trigonometric functions, a normalizing factor may be introduced.

Received February 22, 1960.

Writing $b \equiv xy$, $v \equiv y - x'b$, $n \equiv N - p$, we recall that b and $S^2 \equiv v'v$ are the least-squares estimates of β and $n\sigma^2$ respectively. Now

$$b = x(x'\beta + \Delta) = \beta + x\Delta; \tag{3.1}$$

hence

$$Eb = \beta, \qquad \sigma^2 B \equiv E(b - \beta)(b' - \beta') = \sigma^2 xPx'. \tag{3.2}$$

Also,

$$v \equiv y - x'b = (I_N - x'x)\Delta \equiv M\Delta, \tag{3.3}$$

where $M \equiv I_N - x'x$. Since $M' = M = M^2$, the characteristic roots of M are zeros or ones, and since the trace of $M = n$, M is of rank n. Further,

$$M_{ij} = \delta_{ij} - \sum_{s=1}^{p} x_{si}x_{sj} = \delta_{ij} + O(N^{-1}), \tag{3.4}$$

and

$$Ev = 0, \sigma^2 V \equiv Evv' = \sigma^2 MPM, \sigma^2 R \equiv Ev(b' - \beta') = \sigma^2 MPx'. \tag{3.5}$$

If $P = I_N$, then $B = I_p$, $V = M$, $R = Mx' = 0$, so that v is independent of b. Also $ES^2 = n\sigma^2$ and S^2/σ^2 is independently distributed of b as a χ^2 variate with n degrees of freedom. The usual statistic test to the hypothesis concerning the value of β_j is

$$u_j = (b_j - \beta_j)n^{\frac{1}{2}}/S, \tag{3.6}$$

which is distributed as a Student variate with n degrees of freedom. In general, when $P = I_N$,

$$u = n^{\frac{1}{2}}a'(b - \beta)/S(a'a)^{\frac{1}{2}} \tag{3.7}$$

is a Student variate with n degrees of freedom for any non-null vector a.

When $P \neq I_N$, neither $ES^2 = n\sigma^2$ nor the distribution of S^2/σ^2 is that of a χ^2 variate. Furthermore, v and b are correlated so that S^2 is not independent of b.

We now consider the special case

$$P_{ij} = \rho_{|i-j|}, \qquad \rho_0 = 1. \tag{3.8}$$

It will be assumed that ρ_1 is small and $\sum_{k=2}^{N-1} \rho_k$ negligible so that the departure from the ideal conditions is not very great. For example, $\rho_k = e^{-k^2}$, or $1/(a^2k^2 + 1)$, $a \geqq 2$. At first glance, these assumptions may seem somewhat restrictive, but a little reflection will show that they are quite reasonable. If one or more autocorrelations are high, it would be desirable to modify the initial model by introducing additional regression variables, presumably stochastic in nature. For instance, we may introduce a small order autoregressive scheme for Δ. In the applications which will follow after the general discussion we will actually set $\rho_k = 0$ for $k > 1$. In this case P will be written as $P^{(1)}$ and for the positive-definiteness of $P^{(1)}$ we will require $| \rho_1 | < (\frac{1}{2}) \sec [\pi/N + 1)]$.

Evaluating B_{jj}, V_{ii}, and R_{ij}, where B is defined in (3.2), and V and R in (3.5), we find

$$B_{jj} = 1 + 2\sum_{k=1}^{N-1} \rho_k \sum_{s=1}^{N-k} x_{js}x_{j,s+k} \tag{3.9}$$
$$\cong 1 + 2\rho_1 \sum_{s=1}^{N-1} x_{js}x_{j,s+1},$$

$$V_{ii} = M_{ii} + 2\sum_{k=1}^{N-1} \rho_k \sum_{s=1}^{N-k} M_{is} M_{i,s+k}$$
$$\cong 1 - \sum_{r=1}^{p} x_{ri}^2 - 2\rho_1 \left[\sum_{t=1}^{p} x_{ti}\, x_{t,i+1} + \sum_{t=1}^{p} x_{ti}\, x_{t,i-1} + \sum_{s=1}^{N-1} \sum_{r=1}^{p} \sum_{t=1}^{p} x_{ri}\, x_{rs}\, x_{ti}\, x_{t,s+1} \right], \tag{3.10}$$

$$R_{ij} = \sum_{k=1}^{N-1} \rho_k \sum_{s=1}^{N-k} (M_{is}\, x_{j,s+k} + M_{i,s+k}\, x_{js})$$
$$\cong \rho_1 \left[x_{j,i-1} + x_{j,i+1} - \sum_{j=1}^{N-1} \sum_{r=1}^{p} (x_{ri}\, x_{rs}\, x_{j,s+1} + x_{ri}\, x_{r,s+1}\, x_{js}) \right], \tag{3.11}$$

where $x_{j0} = x_{j,N+1} = 0$, and $\cong$ is to be replaced by $=$ when $P = P^{(1)}$. We notice that B_{jj} and V_{ii} are of the order of unity while R_{ij} is of the order of $\rho_1 N^{-\frac{1}{2}}$. The correlation coefficient between v_i and b_j is given by $R_{ij}(V_{ii}B_{jj})^{-\frac{1}{2}} = O(\rho_1 N^{-\frac{1}{2}})$. Hence, if either N is large or if ρ_1 is small, the vector v is almost independent of b. As a first approximation, therefore, we will derive the distributions of the test statistics as though S^2 were distributed independently of b. Now,

$$S^2 \equiv v'v = \Delta' M \Delta$$

is a non-negative definite quadratic form of rank n in Δ. If $\lambda_1, \cdots, \lambda_n$ are the non-zero characteristic roots of $A \equiv MP$, the distribution of S^2/σ^2 is that of $Q = \sum_{i=1}^{n} \lambda_i \chi_i^2(1)$. Here $\chi_i^2(\nu)$ denotes a χ^2 variate with ν degrees of freedom and all such variates appearing in a linear combination are independent. We approximate the probability density function (pdf) of Q by the pdf of a $g\chi^2(h)$ variate, i.e.,

$$k(z; g, h,) = \frac{e^{-z/2g}\, z^{h/2-1}}{(2g)^{h/2}\Gamma(h/2)} \quad \text{for } z > 0, \qquad 0 \text{ for } z \leq 0, \tag{3.12}$$

where (see, for example, [1])

$$g = \sum \lambda^2 / \sum \lambda, \qquad h = \left(\sum \lambda\right)^2 / \sum \lambda^2, \tag{3.13}$$

so that the first two moments of Q are equal to the first two moments of $k(z; g, h)$. g will be called the *scaling factor* and h the *effective number of degrees of freedom* associated with Q. Box [1] has shown that $h \leqq n$. If greater accuracy is desired this approximation can be improved in several ways [5]. One way is to write the pdf of Q, $p(z)$, as

$$(3.14) \qquad p(z) = k(z; g, h) \sum_{m=0}^{\infty} \frac{m!\,\Gamma(h/2)}{\Gamma(m + h/2)} \frac{d_m}{(2g)^m} L_m^{(h/2-1)}\left(\frac{z}{2g}\right),$$

where

$$L_m^{(c)}(x) = \sum_{j=0}^{m} \binom{m + c}{m - j} \frac{(-x)^j}{j!}$$

is the Laguerre polynomial of degree m, and the d_m are given by

$$(3.15) \qquad \begin{aligned} d_m &= (2g)^m \int_0^\infty p(z) L_m^{(h/2-1)}(z/2g)\, dz \\ &= \sum_{j=0}^{m} \binom{m - 1 + h/2}{m - j} (-1)^j (2g)^{m-j} \gamma_j / j!, \end{aligned}$$

where $\gamma_j = EQ^j$. In particular, $d_0 = 1$, $d_1 = d_2 = 0$, and

$$(3.16) \qquad d_3 = -\tfrac{4}{3}[\sum \lambda^3 - hg^3].$$

The convergence of such series as (3.14) in the case of a linear combination of χ^2 variates with all coefficients positive has been proved by Gurland [4].

If Q_1 and Q_2 are two independent quadratic forms in normal variates with zero means, the distribution of their ratio can be obtained easily from their joint distribution, where each distribution is developed in the form (3.14) In particular if Z is a $N(0, 1)$ variate and Q is a $\sum \lambda_i \chi_i^2(1)$ variate independent of Z, then the distribution of $t = Z(gh)^{\frac{1}{2}}/Q^{\frac{1}{2}}$ is given by

$$(3.17) \qquad \begin{aligned} \Pr(|t| \geqq t_0) &= I_{x_0}(h/2, 1/2) \\ &\quad + \sum_{m=3}^{\infty} (2g)^{-m} d_m \sum_{j=0}^{m} (-1)^j \binom{m}{j} I_x^0(j + h/2, 1/2), \end{aligned}$$

where $x_0 = (1 + t_0^2/h)^{-1}$, and, for $p > 0$, $q > 0$,

$$\begin{aligned} I_z(p, q) &= [B(p, q)]^{-1} \int_0^z x^{p-1}(1 - x)^{q-1}\, dx, \qquad \text{for } 0 < z < 1, \\ &= 0 \quad \text{for} \quad z \leqq 0, \qquad\qquad 1 \quad \text{for} \quad z \geqq 1. \end{aligned}$$

The leading term of (3.17) indicates that t is approximately a Student variate with h degrees of freedom.

Now,

$$z = a'(b - \beta)/\sigma(a'Ba)^{\frac{1}{2}}$$

is a $N(0, 1)$ variate for any non-zero vector a, and is approximately independent of S^2/σ^2. Hence

$$t = a'(b - \beta)(gh)^{\frac{1}{2}}/S(a'Ba)^{\frac{1}{2}}$$

is approximately a Student variate with h degrees of freedom. The alternative test statistics,

$$w = a'(b - \beta n^{\frac{1}{2}}/S(a'Ba)^{\frac{1}{2}},$$
$$u = a'(b - \beta)n^{\frac{1}{2}}/S,$$

are related to t by the relations

$$w = \alpha_1 t, \qquad u = \alpha_1\alpha_2 t,$$
$$\alpha_1 = (n/gh)^{\frac{1}{2}}, \qquad \alpha_2 = (a'Ba/a'a)^{\frac{1}{2}}.$$

Since α_1 does not depend on the vector a we observe that the distribution of w does not depend on the choice of a, so that $w_1, w_2, \cdots, w_p$ will be identically distributed as $\alpha_1 t$. It will be found in many cases that $\alpha_1 = 1 + O(1/n)$, so that considering w as a Student variate with n degrees of freedom involves mainly an error in degrees of freedom, which is not very serious if n and h both are moderately large. However, in general α_2 will not be close to unity and considering u as a Student variate with n degrees of freedom will lead to serious errors in probability statements. Since α_2 depends on the choice of a, the distribution of u will change with a change in a. In particular $u_1, \cdots, u_p$, in general, will have different distributions.

Let t_α denote a number such that $\Pr(|t| \geqq t_\alpha) = \alpha$. This number can be approximately determined through interpolation in existing tables of Student distributions. The 100 $(1 - \alpha)$ per cent confidence interval on β_j is approximately given by

$$b_j - t_\alpha S B_{jj}^{\frac{1}{2}}/(gh)^{\frac{1}{2}} \leqq \beta_j \leqq b_j + t_\alpha S B_{jj}^{\frac{1}{2}}/(gh)^{\frac{1}{2}}. \tag{3.18}$$

In many cases it will be difficult to determine the characteristic roots $\lambda_1, \cdots, \lambda_n$ of the matrix A. We only require, however, the sums of powers of these roots to determine the values of g, h, d_3, d_4 etc. These may be found by the relations

$$\sum_{j=1}^{n} \lambda_j^r = \operatorname{tr} A^r, \qquad r = 1, 2, \cdots, \tag{3.19}$$

where "tr" stands for the trace of a matrix.

In the following applications we will confine attention to the case when $P = P^{(1)}$, i.e., when

$$P_{ij}^{(1)} = \rho_{|i-j|}, \qquad \rho_j = 0, \qquad \text{for } j > 1.$$

In the case of testing a single sample mean we will also consider $P = P^{(2)}$, where

$$P_{ij}^{(2)} = \rho^{|i-j|}, \qquad |\rho| < 1.$$

It is believed that applications of the theory presented in this section will be found mostly in the analysis of time series. If we have a record on a time series, which we believe to be stationary, we may wish to test the hypothesis that the process mean is zero. If we have several samples we may wish to compare their means. In some other cases we may wish to test the existence of a linear trend or of cycles. In all such cases we may assume that the errors form a stationary process with an autocorrelation function ρ_k. The theory then provides adequate test statistics, when N is large, noting that $P_{ij} = \rho_{|i-j|}$.

4. Single sample mean. Let

$$y_t = \beta/N^{\frac{1}{2}} + \Delta_t, \qquad t = 1, 2, \cdots, N.$$

Since there is only one regression coefficient, we omit the suffixes from β and B. Now

$$b = (N)^{\frac{1}{2}}\bar{y} = N^{-\frac{1}{2}} \sum_{t=1}^{N} y_t, \qquad S^2 = \sum_{t=1}^{N} t_t^2 - N\bar{y}^2.$$

The elements of the matrix M are obviously $\delta_{ij} - 1/N$, where $\delta_{ij} = 0$ if $i \neq j$, 1 if $i = j$. The usual test statistic concerning β is

$$u = (\bar{y} - \beta/N^{\frac{1}{2}})(nN)^{\frac{1}{2}}/S, \qquad n = N - 1.$$

(a) In case $P = P^{(1)}$, we have $B = 1 + 2\rho_1 - 2\rho_1/N$, and, evaluating $\sum \lambda$ and $\sum \lambda^2$ from the relations (3.19), we obtain

$$\sum \lambda = \text{tr } MP^{(1)} = n(1 - 2\rho_1/N)$$

$$\sum \lambda^2 = \text{tr } (MP^{(1)})^2 = n(1 - 4\rho_1/N) + 2(n-2)\rho_1^2 + 4(N+1)\rho_1^2/N^2.$$

From these, g and h are easily determined for any given value of ρ_1, and then

$$\alpha_1 = (n/\sum \lambda)^{\frac{1}{2}}, \qquad \alpha_2 = (B)^{\frac{1}{2}}$$

$$w = \alpha_1 t, \qquad u = \alpha_1\alpha_2 t.$$

(b) In case $P = P^{(2)}$, we have, neglecting ρ^N,

$$B = (1+\rho)/(1-\rho) - 2\rho/\{N(1-\rho)^2\},$$

$$\sum \lambda = N - (1+\rho)/(1-\rho) + 2\rho/N(1-\rho)^2,$$

$$\sum \lambda^2 = N(1+\rho^2)/(1-\rho^2) - (1+\rho)^2/(1-\rho)^2 - 2\rho^2/(1-\rho^2)^2 + 4\rho(1-\rho^3)/\{N(1-\rho)^4(1+\rho)\} + 4\rho^2/\{N^2(1-\rho)^4\}.$$

As an illustration, values of g, h and approximate 5% points of t, w and u for $\rho_1 = -.2, 0, +.2$ when $N = 10$ are given in the following table. The top value in each column corresponds to $P = P^{(1)}$ and the bottom value to $P = P^{(2)}$.

ρ_1	h	g	α_1	α_2	$t_{.05}$	$w_{.05}$	$u_{.05}$
−.2	8.51	1.100	.981	.800	2.282	2.239	1.791
	8.44	1.102	.983	.833	2.285	2.246	1.871
0	9.00	1.000	1.000	1.000	2.262	2.262	2.262
	9.00	1.000	1.000	1.000	2.262	2.262	2.262
+.2	8.43	1.025	1.021	1.166	2.285	2.333	2.720
	8.46	1.012	1.025	1.200	2.284	2.341	2.809

5. Two samples. We distinguish the sample and associated quantities by a subscript or an additional subscript, e.g.,

$$y_{it} = \beta_i/N^{\frac{1}{2}} + \Delta_{it}, \qquad t = 1, 2, \cdots, N_i, \qquad i = 1, 2,$$

where Δ_i, $i = 1, 2$, are independent $N(0, \sigma_i^2 P_i)$ vector variates. The case $P_i = I_{N_i}$, $i = 1, 2$ and $\sigma_1^2 \neq \sigma_2^2$ has been studied by Welch [7]. We will treat here the general case when σ_1^2, σ_2^2, P_1 and P_2 are arbitrary. We will assume that P_1 and P_2 and $\theta = \sigma_1^2/\sigma_2^2$ are known and that N_1 and N_2 are large. The variate

$$Z = [\bar{y}_1 - \bar{y}_2 - \beta_1/(N_1)^{\frac{1}{2}} + \beta_2/(N_2)^{\frac{1}{2}}]/\sigma_2(\theta B_1/N_1 + B_2/N_2)$$

is a $N(0, 1)$ variate and

$$S_i^2 = \sum_{t=1}^{N_i} t_{it}^2 - N_i\bar{y}_i^2 = \sum_{i=1}^{N_i} \Delta_{it}^2 - N_i\bar{\Delta}_i^2$$
$$= \Delta_i' M_i \Delta_i, \qquad i = 1, 2,$$

are distributed, independently of each other and approximately independently of Z, as $\sigma_i^2 \sum_{j=1}^{n_i} \lambda_{ij}\chi_i^2(1)$, where λ_{ij}, $j = 1, \cdots, n_i$ are non-zero characteristic roots of $A_i = M_i P_i$ and $n_i = N_i - 1$. Hence $Q = S_1^2/\sigma_1^2 + S_2^2/\sigma_2^2$ is distributed as $\sum_{i=1}^{2} \sum_{j=1}^{n_i} \lambda_{ij}\chi^2(1)$. Let g and h be the scaling factor and the effective degrees of freedom associated with Q, i.e.,

$$h = (\sum \sum \lambda_{ij})^2/\sum \sum \lambda_{ij}^2, \qquad g = \sum \sum \lambda_{ij}^2/\sum \sum \lambda_{ij},$$

where the summations over j are from 1 to n_i and over i from 1 to 2. Then

$$t = [\bar{y}_1 - \bar{y}_2 - \beta_1/(N_1)^{\frac{1}{2}} + \beta_2/(N_2)^{\frac{1}{2}}](gh)^{\frac{1}{2}}/[\theta B_1/N_1 + B_2/N_2)(S_1^2/\theta + S_2^2)]^{\frac{1}{2}}$$

is approximately a Student variate with h degrees of freedom.

6. Linear trend. We take N to be an odd integer and consider the linear trend in the form

$$y_t = N^{-\frac{1}{2}}\beta_1 + [N(N^2 - 1]^{-\frac{1}{2}}(12)^{\frac{1}{2}}\beta_2\{t - (N + 1)/2\} + \Delta_t,$$
$$t = 1, 2, \cdots, N.$$

From [6], we have

$$b_1 = N^{\frac{1}{2}}\bar{y}, \qquad b_2 = (12)^{\frac{1}{2}}N^{-\frac{1}{2}}(N^2 - 1)^{-\frac{1}{2}}\sum_{t=1}^{N} ty_t - (3)^{\frac{1}{2}}N^{\frac{1}{2}}(N + 1)^{\frac{1}{2}}(N - 1)^{-\frac{1}{2}}\bar{y},$$

$$S^2 = \sum_{t=1}^{N} y_t^2 - b_1^2 - b_2^2.$$

The elements of the matrix M are given by

$$M_{ij} = \delta_{ij} - 1/N - 3[N(N^2 - 1)]^{-1}(2i - N - 1)(2j - N - 1).$$

If $P = P^{(1)}$, we have

$$B_{11} = 1 + 2\rho_1 - 2\rho_1/N, \qquad B_{12} = 0,$$
$$B_{22} = 1 + 2\rho_1 - 6\rho_1/N - 4\rho_1/\{N(N^2 - 1)\}.$$

Evaluating $\sum \lambda$ and $\sum \lambda^2$, we find, writing $n = N - 2$,

$$\sum \lambda = n(1 - 4\rho_1/N),$$

$$\sum \lambda^2 = n(1 - 8\rho_1/N) + 2(n - 3)\rho_1^2 + 16\rho_1^2/N + 16\rho_1^2/N^2 - 24\rho_1^2/\{N^2(N - 1)\},$$

from which h and g are determined. Finally

$$t = S^{-1}[a_1(b_1 - \beta_1) + a_2(b_2 - \beta_2)](gh)^{\frac{1}{2}}/[a_1^2B_{11} + a_2^2B_{22}]^{\frac{1}{2}}$$

is approximately a Student variate with h degrees of freedom for arbitrary constants a_1 and a_2 not both equal to zero.

7. Regression on trigonometric functions. Consider

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

where $\mu_i = 2\pi\omega_i/N$, $i = 1, \cdots, q$, and ω_i are positive integers less than N and different from each other. Again, from [6] we obtain

$$b_1 = N^{\frac{1}{2}}\bar{y}, \qquad b_{2i} = (2/N)^{\frac{1}{2}}\sum_{t=1}^{N} y_t \cos \mu_i t, \qquad b_{2i+1} = (2/N)^{\frac{1}{2}}\sum_{t=1}^{N} y_t \sin \mu_i t, \qquad i = 1, \cdots, q,$$

$$S^2 = \sum_{t=1}^{N} y_t^2 - b_1^2 - \sum_{i=2}^{2q+1} b_i^2,$$

$$n = N - 2q - 1,$$

$$M_{st} = \delta_{st} - N^{-1} - 2N^{-1}\sum_{i=1}^{q} \cos \mu_i(s - t).$$

Assuming $P = P^{(1)}$, we also have

$$B_{11} = 1 + 2\rho_1 - 2\rho_1/N$$

$$B_{2i,2i} = 1 + 2\rho_1 \cos \mu_i - 4N^{-1}\rho_1 \cos \mu_i$$

$$B_{2i+1,2i+1} = 1 + 2\rho_1 \cos \mu_i, \qquad i = 1, \cdots, q.$$

Evaluating $\sum \lambda$ and $\sum \lambda^2$, we find

$$\sum \lambda = n - 2\rho_1 - 4\rho_1\sum_{i=1}^{q} \cos \mu_i + 2N^{-1}\rho_1 + 4N^{-1}\rho_1\sum_{i=1}^{q} \cos \mu_i,$$

$$\sum \lambda^2 = n - 2(n - 2)\rho_1^2 - 4\rho_1 - 8\rho_1\sum_{i=1}^{q} \cos \mu_i - 4\rho_1^2\sum_{i=1}^{q} \cos 2\mu_i + 4N^{-1}\rho_1\left(1 + 2\sum_{i=1}^{q} \cos \mu_i\right) + 4N^{-1}\rho_1^2\left(1 + 2\sum_{i=1}^{q} \cos 2\mu_i\right)$$

$$+ 4N^{-2}\rho_1^2 \left(2q^2 + 3q + 1 + 2 \sum_{i=1}^{q} \cos \mu_i + \sum_{i=1}^{q} \cos 2\, \mu_i \right.$$

$$\left. + 2 \sum_{i \neq j}^{q}\sum \cos \mu_i \cos \mu_j \right).$$

The remaining steps for testing any one of the regression coefficients are straightforward.

8. Concluding remarks. In the preceding discussion we have assumed that

(i) the elements $P_{ij} = \rho_{|i-j|}$, $\rho_0 = 1$,
(ii) ρ_1 is small and ρ_2, ρ_3, $\cdots$ are negligible, and ρ_1 is known, or
(ii′) N is large and ρ_1 is known.

As was remarked earlier, if P is known, it is possible to find a non-singular matrix, D, such that $DPD' = I_N$. The transformation $y^* = Dy$, then, takes us back to the ideal situation as the covariance matrix of $\Delta^* = D\Delta$ is $\sigma^2 I_N$. Since, on theoretical grounds, such a transformation is desirable before applying the least-squares method, or, equivalently, the minimum variance unbiased linear estimate of β, obtained by minimizing $\Delta' P \Delta$, must be preferred over the least-squares estimate, b, obtained by minimizing $\Delta'\Delta$, the reasons for using the latter procedure must be sought in practical considerations. Some of the reasons may be enumerated as the following. Firstly, if N is moderatley large, it may become quite laborious to evaluate D or to work with the transformed variable y^*. Secondly, one may be dealing with several regression problems, the covariance matrices of errors in different problems being different, and one may wish to streamline the calculations. Thirdly, and this is the most important reason, in almost all practical situations, P will be unknown. In this case, if we estimate β and the elements of P (under some assumed structure other than I_N) simultaneously, say, by the maximum likelihood method, the estimate, $\hat{\beta}$, of β will become non-linear in y. The problem of finding the distribution of $\hat{\beta}$, and of obtaining suitable statistics for testing hypotheses concerning β, will become extremely complicated. The only suitable procedure seems to be to proceed as if $P = I_N$, and to obtain the least-squares estimates, b, which are linear, unbiased and asymptotically efficient. The autocorrelations, appearing in the statistic t, will have to be replaced by the serial correlations calculated from the residual, v. Although this point needs further investigation, it is the feeling of the writer that, for large N, the significance level of t will not be affected seriously, at least under the assumption that only the first autocorrelation will be estimated. The error involved in using the sample serial correlation in place of the unknown autocorrelation will, presumably, be of order $N^{-\frac{1}{2}}$ in probability.

We further observe that, as a first approximation, the distribution of t was obtained as if S^2 were independent of b. It would be of interest to improve this approximation by taking into consideration the correlation between b and S^2.

REFERENCES

[1] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, I.," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 290–302.

[2] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, II.," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 484–498.

[3] H. E. Daniels, "The effect of departures from ideal conditions other than non-normality on the t and z tests of significance." *Proc. Cambridge Philos. Soc.*, Vol. 34 (1938), pp. 321–328.

[4] John Gurland, "Distribution of quadratic forms and ratios of quadratic forms," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 416–427.

[5] P. B. Patnaik, "The non-central χ^2 and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[6] M. M. Siddiqui, "Covariances of least-squares estimates when residuals are correlated," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1251–1256.

[7] B. L. Welch, "The significance of the difference between two means when the population variances are unequal," *Biometrika*, Vol. 29 (1937), pp. 350–362.