The Philippine Statistician
Vol. xxii, Nos. 3-4, 1973

TESTING THE INDEPENDENCE OF REGRESSION ERRORS*

FE R. BERNARDO**

I. *The Linear Regression Model*

Consider the equation

$$L = HA$$

where $L = (l_1, l_2, \ldots, l_n)'$ is a $(n,1)$ matrix of unknown real numbers, $A = (a_1, a_2, \ldots, a_k)'$ is a $(K, 1)$ matrix of unknown real numbers and $H = (hij)$ $(i=1, \ldots, n\ ;\ j=1,2,\ldots,k)$ is a (n,k) matrix of known real numbers.

The quantities l_i $(i=1,2, \ldots, n)$ are not directly observed. However they are supposed to differ from the observed quantities Z_i $(i = 1,2,\ldots,n)$ by unknown random variables w_i so that for every i,

$$Z_i = l_i + w_i \qquad (1.1)$$

If $W = (w_1, w_2, \ldots, w_n)'$ and $Z = z_1, z_2, \ldots z_n)'$ then (1.1) maybe written in matrix form as

$$Z = L + W$$

$$\text{or} \quad Z = HA + W \qquad (1.2)$$

The random variable W is assumed to be multivariate normal with zero mean and variance-covariance matrix

$$V(W) = \left\{ \begin{matrix} s_1^2 & 0 & \ldots & 0 \\ o & s_2^2 & \ldots & 0 \\ . & . & . & . \\ o & 0 & \ldots & s^2_n \end{matrix} \right\}$$

where s_i^2 is the variance of w_i. It is also assumed that each of the s_i^2 is a multiple of an unknown quantity s^2,

that is

$$s^2_i = \frac{s_2}{d_i}$$

where the d_i 's are known "weights"

Let $D = (\sqrt{d_i}\ \delta_{ij})$, where δ_{ij} is the Kronecher symbol. Applying the transformation

$$Y = DZ$$

equation (1.2) becomes

$$Y = XA + \bar{U} \qquad (1.3)$$

where $X = DH$, $U = DW$. Therefore the first two moments of U are

$$E(\bar{U}) = DE(W) = 0$$
$$V(\bar{U}) = D^2V(W) = (d_i s_i^2 \delta_{ij}) = s^2\ I.$$

Since $\bar{U} = DW$ is a linear transformation, then $\bar{U}$ is a multinormal vector (with mean zero and variance-covariance matrix s^2 I).

Equation (1.3) is known as the linear regression model.

Since (1.3) has more unknowns than equations, it has no unique solution. However, A can be estimated so that U'U is minimum. This estimator of A, called the least squares estimator, is

$$\hat{A} = (X'X)^{-1}\ X'Y.$$

Under the assumptions on X and U mentioned above, we have

$$\begin{aligned} E(\hat{A}) &= E\ (X'X)^{-1}\ X'\ (XA + U) \\ &= A + (X'X)^{-1}\ X'\ E(\bar{U}) \\ &= A \end{aligned}$$

and

$$\begin{aligned} V(\hat{A}) &= (X'X)^{-1} V(Y) (X'X)^{-1} X' \\ &= (X'X)^{-1} X' (s^2) X(X'X)^{-1} \\ &= s^2 (X'X)^{-1} \end{aligned}$$

The likelihood function of $\bar{U}$ is

$$L = \frac{1}{(2\pi)^{n/2} s^n} \exp\left[- \frac{U'U}{2s^2} \right]$$

Hence:

$$\ln L = - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln s^2 \frac{1}{2s^2} (Y\text{-}XA)' (Y - XA)$$

$$\text{and } \ln \hat{L} = - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln s^2 \frac{-1}{2s^2} (Y - X\hat{A})' (Y - X\hat{A}),$$

$$\text{so that} \quad \frac{\partial^2 \ln L}{\partial \hat{A}^2} = \frac{-X'X}{s^2}.$$

Thus, Fisher's information matrix I is

$$I = - E\left[\frac{\partial^2 \ln L}{\partial \hat{A}^2} \right].$$

$$= - E\left[- \frac{X'X}{s^2} \right] = \frac{X'X}{s^2} = \left[V(\hat{A}) \right]^{-1}.$$

We therefore conclude $\hat{A}$ is an unbiased estimator of A, and that in the class of all unbiased estimators of A, $\hat{A}$ has the least variance.

However, when the components of $\bar{U}$ are not independent of each other, then $\hat{A}$ is no longer the best estimator of A. This is proven in section 3.1 of text.

It is therefore necessary to test for the independence of the $\bar{U}_i$'s whenever least squares regression methods are used, that is test the hypothesis

$$H_o : \bar{U} \sim N(O, s^2 I) \tag{1.4}$$

against the hypothesis

$$H_a : \bar{U} \sim N(O, s^2 \Sigma) \tag{1.5}$$

when Σ is a given matrix different from I.

Since the errors U_i 's are unknown the test must be based on the residuals from the estimated regression line $\hat{U}$, (also called the least squares estimator of U) which is defined as

$$\begin{aligned} \hat{U} &= Y - X\hat{A} \\ &= [I - X(X'X)^{-1} X'] Y \\ &= M Y = M U \end{aligned}$$

Now,

$$\begin{aligned} V(\hat{U}) &= M\, V(\bar{U})\, M' \\ &= M\, (s^2 I)\, M' \\ &= s^2 M \end{aligned}$$

under the null hypothesis (1.4). This implies that the residuals are correlated. Consequently, the ordinary tests of independence can not be used to test (1.4) against (1.5).

II. *Tests of the Null Hypothesis*

Von Neuman, reviving a procedure introduced in the 19th century by a German scientist, tackles the problem of testing for independence of regression errors by using the ratio of the mean square successive difference to the variance, that is:

$$Q = \frac{\dfrac{1}{n-1} \sum_{i=2}^{n} (\hat{u}_i - \hat{u}_{i-1})^2}{\dfrac{1}{n} \sum_{i=1}^{n} (\hat{u}_i - \bar{\hat{u}})^2}. \tag{2.1}$$

The null hypothesis (1.4) is rejected at significance level $\mathcal{L}$ whenever an observed $Q < Q_{\mathcal{L}}$ where $Pr\ (Q < Q_{\mathcal{L}}/H_0) = \mathcal{L}$.

T.W. Anderson and R. L. Anderson studied the model

$$y_i - \mu_i = (y_{i-1} - \mu_{i-1}) + \epsilon_i \qquad (i=1, \ldots, n) \quad (2.2)$$

where the y_i 's are observed values, the ϵ_i 's are random errors which are normally and independently distributed with mean zero and variance δ^2; while the μ_i 's are linear combinations of Fourier terms. They then defined the circular serial correlation coefficient of this model as

$$R = \frac{\sum_{i=1}^{n} (y_i - m_i)\ (y_{i-1} - m_{i\ 1})}{\sum_{i=1}^{n} (_i - m_i)^2}$$

where $m_o = m_n$ and m_i is an estimate of u_i. This statistic can be used to test the null hypothesis

$$H_o : \rho = 0$$

against the alternative hypothesis

$$H_a : \rho > 0$$

or $$H_a' : \rho < 0$$

where the respective critical regions are

$$\begin{aligned} \beta_1 &= \{R \mid R \geq R_1 > 0\} \\ \beta_2 &= \{R \mid R \leq R_2 < 0\} \end{aligned} \qquad (2.3)$$

with R_1 and R_2 being the critical values corresponding to a given significance level.

Suppose that the alternative to the null hypothesis (1.4) is that the u_1 's follow a stationary Markoff scheme, i.e.

$$u_i = \rho(u_{i-1}) + \epsilon_i \quad (i=\ldots-1, 0, 1, \ldots) \qquad (2.4)$$

where $|\beta| \leq 1$ and $\epsilon_i \ \underline{/}\overline{\ }\ N(0, s^2)$. Then the hypotheses (1.4) and (1.5) are equivalent to the hypotheses

$$H_o^* : \rho = 0 \qquad (2.5)$$
$$H_a^* : \rho > 0 \qquad (2.6)$$

respectively.

Now, the regression equation (1.3) can be written as

$$y_i = \sum_{j=1}^{k} a_j X_{ij} + u_i \qquad (i=1, \ldots, n).$$

Then (2.4) becomes

$$y_i - \sum_{j=1}^{k} a_j X_{i-j} = \rho(y_{i-1} - \sum_{j=1}^{k} a_j X_{i-1j}) + \epsilon_i \quad (i=1, \ldots, n)$$

which falls under the model (2.2) with $\mu_i = \sum_{i=1}^{n} a_j X_{ij}$.

Therefore, when the regression vectors, that is the culomns of the matrix X in the regression model (1.3) coincide with vectors $(\cos \frac{2\pi i}{n}, \cos \frac{4\pi i}{n}, \ldots, \cos \frac{2n\pi i}{n})'$ and

$(\sin \frac{2\pi i}{n}, \sin \frac{4\pi i}{n}, \ldots, \sin \frac{2n\pi i}{n})'$, then

$$R = \frac{\sum_{i=1}^{n} (y_i - m_i)(y_{j-1} - m_{i-1})}{\sum_{i=1}^{n} (y_i - m_i)^2} = \frac{\sum_{i=1}^{n} \hat{u}_i \hat{u}_{i-1}}{\sum_{i=1}^{n} \hat{u}_i^2}$$

can be used to test hypothesis (2.5) against hypothesis (2.6) with critical regions defined in (2.3).

The most commonly used test procedure is the one introduced by Durbin and Watson. The statistic used us

$$d = \frac{\hat{U} A_d \hat{U}}{\hat{U}' \hat{U}}$$

$$A_d = \begin{Bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & \dots & -1 & 1 \end{Bmatrix}$$

The exact distribution of d was found by Durbin-Watson using the Imhof method. But because of the fact that $\hat{U} = [I - X(X'X)^{-1} XI] U$, the distribution of d depends on X. Hence its significance points can be tabulated only for a given X, implying that these significance points will have to be computed everytime this test procedure is used.

Durbin and Watson showed that when S of the K regression vectors coincide with S of the latent vectors of A_d, then bounds of d can be found equal to

$$r_l = \frac{\sum_{i=1}^{n-k} \lambda_i z_i^2}{\sum_{i=1}^{n-k} z_i^2}$$

$$r_u = \frac{\sum_{i=1}^{n-k} \lambda_i + {}_{k-s} z_i^2}{\sum_{i=1}^{n-k} z_i^2}$$

where $\lambda_1, \dots, \lambda_{n-s}$ are the eigen values associated with the remaining n-s eigenvectors of A_d. Now, when a constant is fitted in the regression model, the first column of the X matrix consists of all ones. Hence it coincides with the eigen-vector of A_d corresponding to the zero eigenvalue. Thus there exists bounds of d, that is,

$$d_L < d < d_U$$

where

$$d_L = \frac{\sum_{i=1}^{n-k} \lambda_i z_i^2}{\sum_{i=1}^{n-k} z_i^2}$$

$$d_U = \frac{\sum_{i=1}^{n-k} \lambda_{i+k-1} z_i^2}{\sum_{i=1}^{n-k} z_i^2}$$

Instead of looking for the significance points of d, the following procedure called the "bounds test" maybe used:

reject H_0 whenever computed $d < d_L$
do not reject H_0 whenever computed $d > d_U$
test inconclusive if computed d is between d_L and d_U.

This dependence on X of the distribution of test statistics based on the least squares estimator $\hat{U}$ of U led to the discovery of other estimators of U.

Theil is the first to introduce an estimator of U which is independent of X, the BLUS estimator U*. He constructed U* with variance-covariance matrix s^2 I and such that it is also best linear unbiased estimator of U. Because of this additional restriction on U*, it can estimate only n-k of the errors $u_1, u_2, \ldots, u_n$. He therefore partitioned the matrices in the regression equation (1.3) into

$$\begin{Bmatrix} Y_0 \\ \hline Y_1 \end{Bmatrix} = \begin{Bmatrix} X_0 \\ \hline X_1 \end{Bmatrix} A + \begin{Bmatrix} U_0 \\ \hline U_1 \end{Bmatrix}$$

where U_0 consists of the K components of U which are not represented in U*. The matrix M is also partitioned into

$$M = \left\{ \begin{array}{c|c} M_{00} & M_{01} \\ \hline M_{1}0 & M_{11} \end{array} \right\}$$

where M_0 is (k, k) and M_{11} is a principal (n-k, n-k) minor. Theil then derived U* in terms of U as

$$U^* = \hat{U}_1 + \sum_{i=1}^{k} \left(\frac{1}{\sqrt{g_i}} - 1 \right) P_i P_i' \hat{U}_i$$

where $\hat{U} = \left\{ \frac{\hat{U}_0}{\hat{U}_1} \right\}$, and g_i $(i=1,\ldots,k)$ are the K roots of M_{11}

which are less than one, and P_i are its corresponding eigen-vectors. The von Neuman ratio of U*

$$Q^* = \frac{\sum_{i=2}^{n-k} (u_i^* - u^*_{i-1})^2}{\sum_{i=1}^{n-r} (u^*_i - \bar{u}^*)^2} \qquad (2.7)$$

can be used to test (1.4) versus (1.5) rejecting the null hypothesis when an observed $Q^* < Q^*_0$ where Q^*_0 is a constant such that $Pr(Q^* < Q^*_0 \mid H_0) = \alpha$, the pre-assigned size of the test

Abrahamse and Koerst derived another estimator W* of U which is best in the class of all linear unbiased estimators of U. To make W* independent of X, the authors imposed the condition that the covariance matrix of W* is a fixed matrix F chosen a priori to be independent of X. The expression for W* is

$$W^* = [K'MK]^{-\frac{1}{2}} K' \hat{U} \qquad (2.8)$$

where $M = I - X(X'X)^{-1}X'$ and $K'K = F$.

Equation (2.8) shows that given K, or equivalently, specifying F, a corresponding W* can be formed. The authors proved that if K is chosen to be the matrix consisting of the eigenvectors corresponding to the n-k largest roots of A_d, then

$$Q' = \frac{\underset{***}{W^{*\prime} \quad A_d \quad W^*}}{W^{*\prime} \quad W^*} \qquad o$$

has the same distribution as Durbin-Watson's upper bound d_U, and hence is independent of X.

Durbin proposes the following procedure as an alternative to the bounds test when the regression vectors do not coincide with the eigenvectors of A_d. Let L be the matrix whose columns are the k-1 eigenvectors of A_d corresponding to the k-1 smallest non-zero eigenvalue. Then instead of (1.3), he considered the model

$$Y = (E \mid X \mid L) \left\{ \begin{array}{c} a_1 \\ \hline A_1 \\ \hline A_2 \end{array} \right\} + U$$

where E is the vector with unit elements. Suppose $\hat{a}_1$, $\hat{A}_2$, A_3 are the least squares estimates of a_1, A_2, A_3 with

$$V(\hat{A}_1) = s^2 \; G_1 \;= s^2 \; P_1P_1'$$
$$V(\hat{A}_1) = s^2 \; G^2{}_2 = s^2 \; P_2P_2'$$

Durbin then defined the vector

$$\Sigma = Y - \hat{a}_1 \; E - X\hat{A}_2 - LA_3 + X_2A_4$$

with $A_4 = P_1P_2^{-1}\hat{A}_2$; $X_t + X - L(L'L)^{-1} L'X$. Then he showed that the statistic

$$d' = \frac{\sum_{i=2}^{n} (z_i - z_{i-1})^2}{\sum_{i=1}^{n} z_i^2}$$

has the same distribution as d_U whose significance points have already been tabulated since distribution is independent of X.

Koteswara Rao Kadiyala suggests three test criteria based on the estimator W of U which is defined as

$$W = P\hat{U}$$

Here P is a set of eigenvectors of $M = I - X(X'X)^{-1}X$ which simultaneously diagonalizes Σ, the variance-covariance matrix U under the alternative hypothesis. Like Theil's BLUS estimator, W estimates only n-k of the components of the error vector U.

Rao's first test statistic is

$$S_1 = \frac{W' D^{-1} W}{W' W}$$

where $D = P \Sigma P'$ is a diagonal matrix but whose diagonal elements need not be the eigenvalues of Σ. H_0 is rejected when an observed $S_1 \leq S_{1\alpha}$, where $S_{1\alpha}$ is the significance point of S_1 corresponding to α, the size of the rest.

The second test procedure proposed by Rao is based on von Neuman ratio

$$S_2 = \frac{W' \Lambda W}{W' W}$$

where is the (n-k n-k) diagonal matrix with the non-zero characteristic roots of A_d arranged in decreasing order of magnitude along the diagonal. The critical region is

$$\beta_4 = \{W \mid s_2 \leq S_{2\alpha}\}$$

where $S_{2\alpha}$ is a constant such that $Pr(S_2 \leq S_{2\alpha}) = \alpha$, the preassigned size of the test.

For his third test criterion, Rao considers two (n-k, 1) vector, L_1 and L_2 such that

$$L_1L_2' = 0; \quad L_1'L_1 = 1 = L_2'L_2.$$

$$\beta_5 = \{ W \mid |s_3| = \frac{|L_2' W|}{|L_1' W|} \geq S_3 \}$$

where $S_{3\alpha}$ is determined from the equation

$$Pr\ (S_3 \leq S_{3\alpha} | H_o) = \alpha,$$

the size of the test given in advance. S_3 under H_0 and $b + dS_3)$ under H_a where

$$b = \frac{L_1'\ DL_2}{(L_1'\ DL_1)^{1/2} \left[L_2'\ DL_2 - \frac{(L_1'\ DL_2)'}{(L_1'\ DL_1)'} \right]^{1/2}}$$

and

$$d = \frac{-\ (L_1'\ DL_1)^{1/2}}{L_2'\ DL_2 - \frac{(L_1'\ DL_2)^2}{L_1'\ DL_1} \quad ^{1/2}}$$

both follow a cauchy distribution. Hence significance points and power of S_3 can easily be obtained.

III. *Derivation of the Distribution. of Rao's Test Criterion S_1*

Imhof, in his paper "Computing the Distributions of Quadratic forms in Normal Variables" [20] proved the following theorem:

Theorem 3.1. Let $Z = (z_1, \ldots, z_m)'$ be a random vector which is normally distributed with mean 0 and covariance matrix F. Let $\mu = (\mu_i, \ldots, \mu_m)'$ be a constant vector and consider the quadratic form:

$$Q = (Z + \mu)'\ \triangle\ (Z + \mu)$$

where $\triangle$ is a given matrix. If F is non-singular, Q can be written as

$$Q = \sum_{i=1}^{m} \delta_i\ X^2_{hi}\ ;\ \lambda^2 i^2\ ,$$

where

δ_i is a non-zero root of F;

λ_i is a linear combination of $_1, \ldots, _m$;
$X^2_{hi}; \lambda^2_i$ is an independent chi-squarevariable with h_i degrees of freedom and non-centrality parameter λ_i^2

Then

$$\Pr(Q \geq Z) = \tfrac{1}{2} + \int_0 \frac{\sin \theta (r)}{r \phi (r)} dr, \qquad (3.1)$$

where

$$\theta (r) = \tfrac{1}{2} \sum_{i=1}^{m} [h_i \tan^{-1} (\delta_i r) + \lambda^2_i \delta_i^r (1 + \delta_j^2 r^2)^{-1}] \quad -\tfrac{1}{2}zr$$

$$\phi (r) = \prod_{i=1}^{m} (1 + \delta_i^2 r^2)^{\tfrac{1}{2}hi} \exp \tfrac{1}{2} \sum_{i=1}^{m} [\lambda_i \delta_j r)^2 / (1 + \delta_i^2 r^2)].$$

$$\lim_{r \to 0} \frac{\sin \theta (r)}{r \phi (r)} = \tfrac{1}{2} \sum_{i=1}^{m} \delta_i (h_i + \lambda_j)^2 - \tfrac{1}{2}z$$

$$\lim_{r \to \infty} \left\{ \phi (r) = \right\} \begin{cases} - \infty & , \text{if } z > 0 \\ \infty & , \text{if } z < 0 \\ \frac{\pi}{4} \sum_{i=1}^{m} h_i \delta_i \mid \delta_i \mid^{-1} & \text{if } z = 0 \end{cases}$$

Let us use above theorem to determine the cumulative distribution of Rao's S_1 (henceforth to be denoted by S), under the null hypothesis $H_0 : \bar{U} \not\sim N(O, I)$. We have

$$\Pr (S \leq S) = \Pr \left[\frac{W' D^{-1} W}{W' W} \leq S \right]$$

$$= \Pr[(P\bar{U})' D^{-1} - SI) (PU) \leq 0].$$

Since P' $(D^{-1} - SI)P$ is symmetric, then there exists an orthogonal (n,n) matrix H such that

$$H'P' (D^{-1} - SI) PH = \Delta$$

or

$$P'(D^{-1} - SI) P = H \Delta H',$$

where Δ is the diagonal matrix whose elements are the eigenvalues of $P'(D^{-1} - SI)P$. To determine these diagonal elements of Δ, note that the characteristic roots of $P'(D^{-1} - SI)P$ are equal to the roots of $(D^{-1} - SI) PP' = (D^{-1} - SI)$. Relation $| D^{-1} - SI - \mu I | = 0$ implies

$$\begin{vmatrix} d_1^{-1} - (S + \mu) & 0 & \dots & 0 \\ 0 & d_2^{-1} - (S + \mu) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_{n-k}^{-1} - (S + \mu) \end{vmatrix} = 0$$

which can be written as

$$\prod_{i=1}^{n-k} [d_i^{-1} - (S + \mu)] = 0,$$

and therefore

$$u_i = d_i^{-1} - S \qquad (i=, \dots, n\text{-}k).$$

Let $Z = H'U$. Then the first two moments of Z are

$$\begin{aligned} E(Z) &= H'E(\bar{U}) = 0 \\ &= H'V(\bar{U})H \\ &= H'IH \\ &= I \end{aligned}$$

Moreover, $Z = H'U$ is a linear transformation from U to Z.

Hence $Z \sim N(O,I)$. Consequently

$$z_i^2 \sim X^2(1) \text{ and } \sum_{i=1}^{n-k} z_i^2 \sim X^2(n\text{-}k).$$

We can therefore write

$$\Pr(S \leq S) = PR(\bar{U}'H\Delta H'\bar{U} \leq 0)$$

$$= \Pr(Z'\Delta Z \leq 0)$$

$$= \Pr \left[\sum_{i=1}^{n-k} (d_i^{-1} - S)\, z_i^2 > 0 \right].$$

The assumptions of theorem 3.1 are satisfied by $\sum_{i=1}^{i-1} (d_i^{-1} - S) z_i^2$ playing the role of Q with $m = n\text{-}k$, $F = I$, $\sigma_i = (d_i^{-1} - S)$ $h_i = 1$ and $\lambda_i = 0$. Therefore by using (3.1) we obtain the following expression for the cumulative distribution of S under H_o, i.e.

$$\Pr(S \leq S) = 1 - \left(\tfrac{1}{2} + \frac{1}{\pi} \int_0 \frac{\sin \theta\,(r)}{r\, \theta\,(r)}\, dr \right) \qquad (3.2)$$

where

$$\theta\,(r) = \tfrac{1}{2} \sum_{i=1}^{n-k} \tan^{-1} [(d_i^{-1} - S)\, r].$$

$$\phi\,(r) = \Pi_{i=1}^{n-k} [1 + (d_i^{-1} - S)^2 r^2]^{1/4},$$

$$\lim_{r \to 0} \frac{\sin \theta\,(r)}{r\, \phi\,(r)} = \tfrac{1}{2} \sum_{i=1}^{n-k} (d_i^{-1} - S),$$

$$\lim_{r \to \infty} [\theta\,(r)] = \frac{\pi}{4} \sum_{i=1}^{n-k} (d_i^{-1} - S)\, |d_i^{-1} - S|^{-1}.$$

Under the alternative hypothesis H_a, we have

$$E(\bar{U}) = 0$$

$$V(\bar{U}) = \Sigma.$$

Since Σ is positive definite, there exists a non-singular matrix C such that

$$\Sigma = CC' \qquad (3.3)$$

Let us now derive the cumulative distribution function of S under H_a. We have

$$Pr(S \le S) = Pr\left[\frac{W' D^{-1} W}{W' W} \le S\right]$$

$$= Pr[W'(D^{-1} - SI) \; W \le 0]$$

$$= Pr[\bar{U}'P' \; (D^{-1} - SI) \; P\bar{U} \le 0]$$

Now, let $V = C^{-1}U$, where C is defined as in (3.3). Then $U = CV$ so that when H_a is true, we obtain

$$\begin{aligned} E(V) &= C^{-1} \; E(\bar{U}) = 0 \\ V(V) &= C^{-1} \; V(\bar{U}) \; (C')^{-1} \\ &= C^{-1} \; \Sigma \; (C')^{-1} \\ &= C^{-1}CC' \; (C')^{-1} \\ &= I. \end{aligned} \tag{3.4}$$

Therefore

$$Pr(S \le S) = Pr(V'C'P' \; (D^{-1} - SI) \; PCV \le 0).$$

It can easily be seen that $C'P' \; (D^{-1} - SI) \; PC$ is a symmetric matrix. Moreover, its eigenvalues are the characteristic roots of

$$\begin{aligned} (D^{-1} - SI)PCC'P' &= (D^{-1} - SI) \; P \; \Sigma \; P' \\ &= (D^{-1} - SI) \; D \\ &= (I - SD). \end{aligned}$$

But $|I - SD - \lambda \; I| = 0$ implies $|(1 - \lambda) \; I - SD| = 0$, which imples $\prod_{i=1}^{n-k} (1 - \lambda - Sd_i) = 0$. Consequently, the characteristic roots of $I - SD$ are

$$\lambda_i = 1 - Sd_i \qquad (i=1 \ldots, n\text{-}k) \tag{3.5}$$

Above statements imply that there is an orthogonal matrix G such that $G[C'P' \; (D^{-1} - SI) PC] \; G' = \Lambda$ Where Λ is a diagonal matrix whose diagonal elements are the λ_i's defined in (3.5). Thus, we have

$$Pr(S \le S) = Pr \; (V'G' \; \Lambda GV \le 0.$$

$$\theta(r) = \prod_{i=1}^{n-k} [1 + (1-Sd_i)^2 r^2]^{1/4},$$

$$\lim_{r\to 0} \frac{\sin \theta (r)}{r \phi (r)} = \tfrac{1}{2} \sum_{i=1}^{n-k} (1-Sd_i).$$

Note that the distributions of S both under the null and the null and the alternative hypothesis depend on d_i, the elements of $D = P \Sigma P'$, which in turn is dependent on X through P. Hence the significance points and power of S can be determined only for a given matrix X.

IV. *Application*

Let us consider the example of the consumption of textile in the Netherlands. The data is tabulated below

TABLE I

TIME SERIES FOR THE TEXTILE EXAMPLE

Year	y	x_2	x_3
1923	1.99651	1.98543	2.00432
1924	1.99564	1.99167	2.00043
1925	2.00000	2.00000	2.00000
1926	2.04766	2.02078	1.95713
1927	2.08797	2.02078	1.93702
1928	2.07041	2.03941	1.95279
1929	2.08314	2.04454	1.95713
1930	2.13354	2.05038	1.91803
1931	2.18808	2.03862	1.84572
1932	2.18639	2.02243	1.81558
1933	2.20003	2.00732	1.78746
1934	2.14799	1.97955	1.79588
1935	2.13418	1.98408	1.80346
1936	2.22531	1.98945	1.72099
1937	2.18837	2.01030	1.77597
1938	2.17319	2.00689	1.77452
1939	2.21880	2.01620	1.78746

To simplify further above expression, let

$$Z = GV.$$

Then from relation (3.4) we have

$$\begin{aligned} E(Z) &= GE(V) = 0 \\ V(Z) &= GV(V)G' \\ &= G\,I\,G' \\ &= I. \end{aligned}$$

This means that $\sum_{i=1}^{n-k} z_i^2$ is a central chi-square variable with n-k degrees of freedom. Therefore

$$\begin{aligned} Pr(S \leq S) &= Pr(Z'\Omega Z \leq 0) \\ &= Pr(\sum_{i=1}^{n-k} \lambda_i z_i^2 \leq 0) \\ &= 1 - Pr \sum_{i=1}^{n-k} (1 - Sd_i) z_i^2 > 0) \end{aligned}$$

Since the assumptions of theorem 3.1 are satisfied by $\sum_{i=1}^{n-k} (1-Sd_i) z_i^2$, playing the role of Q, then the cumulative density function of S under H_a is of the form:

$$\begin{aligned} Pr(S \leq S) &= 1 - (\tfrac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \theta\ (r)}{r\ \phi\ (r)} dr) \\ &= \tfrac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \theta\ (r)}{r\ \phi\ (r)} dr, \quad (3.6) \end{aligned}$$

where

$$\theta(r) = \tfrac{1}{2} \sum_{i=1}^{n-k} \tan^{-1} [(1 - Sd_i) r],$$

y = logarithm of per capita consumption of textile, obtained by dividing the money value of textile consumption by family households by PN.

p = retail price index of clothing for the city of Amsterdam

N = population of the Netherlands

X_1 = logarithm of real per capita income, obtained by dividing the money value of income of family households by N

π = general retail price index

X_2 = logarithm of the deflated price index of clothing, i.e., of the ratio P/π.

Rao's test criterion S_1 was applied to this example. Here we have

$$Y = \begin{Bmatrix} 1.99651 \\ 1.99564 \\ 2.00000 \\ 2.04766 \\ 2.08707 \\ 2.07041 \\ 2.08314 \\ 2.13354 \\ 2.18808 \\ 2.18639 \\ 2.20003 \\ 2.14799 \\ 2.13418 \\ 2.22531 \\ 2.18837 \\ 2.17319 \\ 2.21880 \end{Bmatrix}, X = \begin{Bmatrix} 1.00000 & 1.98543 & 2.00432 \\ 1.00000 & 1.99167 & 2.00043 \\ 1.00000 & 2.00000 & 2.00000 \\ 1.00000 & 2.02078 & 1.95713 \\ 1.00000 & 2.02078 & 1.93702 \\ 1.00000 & 2.03941 & 1.95279 \\ 1.00000 & 2.04454 & 1.95713 \\ 1.00000 & 2.05038 & 1.91803 \\ 1.00000 & 2.03862 & 1.84572 \\ 1.00000 & 2.02243 & 1.81558 \\ 1.00000 & 2.00732 & 1.78746 \\ 1.00000 & 1.97955 & 1.79588 \\ 1.00000 & 1.98408 & 1.80346 \\ 1.00000 & 1.98945 & 1.72099 \\ 1.00000 & 2.01030 & 1.77597 \\ 1.00000 & 2.00689 & 1.77452 \\ 1.00000 & 2.01620 & 1.78746 \end{Bmatrix} \qquad (4.1)$$

We would like to test the null hypothesis

$$H_0 : \bar{U} \sim N(O, I) \qquad (4.2)$$

against the alternative hypothesis that U_i follow a stationary Markoff scheme (see equation (2.4)), with correlation coefficient $\rho = .8$. That is

$$H_a : \bar{U} \sim N(0, \Sigma)$$

where

$$\Sigma = \frac{1}{1-(.8)^2} \begin{Bmatrix} 1 & .8 & (.8)^2 & \dots & (.8)^{16} \\ .8 & 1 & .8 & \dots & (.8)^{15} \\ \dots & \dots & \dots & \dots & \dots \\ (.8)^{15} & (.8)^{14} & (.8)^{13} & \dots & (.8) \\ (.8)^{16} & (.8)^{15} & (.8)^{14} & \dots & 1 \end{Bmatrix} \qquad (4.3)$$

First, let us find the matrix P of the transformation $W = P\bar{U}$. In section II, P was defined as a (n-k, n) row-orthogonal matrix whose rows form a set of eigenvectors of $M = I - X(X'X)^{-1}X'$ corresponding to the eigenvalue one; and which simultaneously diagonilizes Σ, the variance-covariance matrix of U under H_a.

Since for any X of rank k, M is a symmetric, idempotent matrix of rank n-k, there exists an orthogonal matrix R such that

$$RMR' = \left\{ \begin{array}{c|c} I_{n-k} & 0 \\ \hline 0 & 0_k \end{array} \right\} \qquad (4.4)$$

Such an R was found using the SSP program EIGEN. Let us partition R into

$$R = \left\{ \begin{array}{c} R_1 \\ \hline R_2 \end{array} \right\}$$

where R_1 is (14, 17). Then (4.4) implies that R_1 form a set set of eigenvectors of M corresponding to the eigenvalue one.

Next we define $K = R_1 \Sigma R_1'$. Since K is symmetric, there exists an orthogonal (14, 14) matrix T (such a T can be found using SSP program EIGEN) such that

$$TKT' = TR_1\Sigma R_1'T' = D$$

where D is a diagonal matrix whose non-zero elements are the eigenvalues of K. We have

$$D = \left\{\begin{array}{cccccccccccccc}
.312 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & .348 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & .355 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & .383 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & .403 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & .468 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .515 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .565 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .697 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .892 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.192 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.654 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.225 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.832
\end{array}\right\} \tag{4.5}$$

Let

$$P = TR_1. \quad \text{Then}$$

$$PMP' = TR_1MR_1'T'$$

$$= T_{15}$$

Hence P satisfies the conditions imposed on the matrix of the transformation $W = PU$. Carrying out the matrix multiplication, we obtain

P =

0.0207	-0.0724	0.1190	-0.1802	0.2468	-0.2982	0.3366	-0.3746	0.3914
	-0.3741	0.3414	-0.2755	0.1862	-0.1167	0.0826	-0.0525	0.0217
0.1454	-0.4362	0.5124	-0.4619	0.3056	-0.0481	-0.1890	0.2906	-0.2218
	0.1221	0.0356	-0.0837	-0.0389	0.0393	0.0701	-0.1086	0.0663
0.0755	-0.1027	0.0886	0.0498	-0.2516	0.3676	-0.3352	0.1508	0.0625
	-0.2906	0.3637	-0.3150	0.2227	0.0480	-0.3201	0.3646	-0.1801
0.1837	-0.3684	0.2599	0.0718	-0.3466	0.3145	-0.0186	-0.2648	0.3707
	-0.2006	-0.1241	0.3405	-0.3147	0.0883	0.1078	-0.1750	0.0817
0.1005	-0.2435	0.1323	0.0533	-0.0175	-0.2430	0.4310	-0.1342	-0.0622
	0.1957	-0.1532	0.0261	-0.1861	0.2676	-0.4397	0.4857	-0.2121
-0.2167	0.3533	-0.0085	-0.2798	-0.2435	-0.0519	-0.0906	0.2462	0.1378
	-0.4185	0.1071	0.3077	-0.4797	0.2179	-0.1357	0.1345	-0.0652
0.2256	-0.1603	-0.1472	0.3599	-0.1433	-0.2541	0.1703	0.2528	-0.3296
	-0.2319	0.4076	-0.0594	-0.1995	-0.2737	-0.0320	-0.3276	0.1965
-0.2649	0.3127	0.2195	0.2703	0.1276	0.2716	0.1831	0.2342	0.1088
	0.1849	0.0707	-0.2641	-0.0089	0.3904	-0.2515	-0.2934	0.2801
0.1584	-0.0929	-0.1912	0.0708	0.2433	-0.1150	-0.2104	0.1300	0.3506
	-0.0899	-0.3874	-0.0338	0.2662	0.2171	-0.4763	-0.1942	0.3549
-0.4714	0.0217	0.4851	0.2044	-0.3256	-0.3318	0.0725	0.2812	0.0257
	-0.1745	-0.1239	0.1730	0.2710	-0.1729	-0.0649	0.0091	0.1241
-0.0105	-0.1164	-0.1239	-0.1306	0.0938	0.2072	0.1406	-0.1382	-0.2665
	-0.0447	0.2248	0.3732	0.0310	-0.4842	-0.3568	0.1313	0.4693
-0.1686	0.0343	0.2282	0.4122	0.2133	-0.1835	-0.4443	-0.2688	0.1469
	0.3323	0.2696	-0.1164	-0.3584	-0.1173	0.1386	0.0209	0.1415
-0.3970	-0.2379	0.0448	0.3296	0.4719	0.3165	0.0447	-0.2017	-0.2740
	-0.2292	-0.0672	0.1202	0.1887	0.2379	0.0135	-0.1246	-0.2338
0.0366	0.0548	0.0674	0.1042	0.0442	0.0316	-0.0429	-0.1161	-0.2351
	-0.3206	-0.3564	-0.3557	-0.1725	0.0487	0.3115	0.4236	0.4860

The relation $W = P\,\hat{U}$ can be simplified into

$$= P\,[Y - X(X'X)^{-1}X'\,Y]$$

$$= P\,M\,Y = P\,Y$$

Using the above equation the obtain the following form for W by multiplying the matrix P given by (4.5) and the vector Y given by (4.1). We have

$$W = \begin{Bmatrix} .02430 \\ -.29779 \\ -.17819 \\ -.39676 \\ -.22469 \\ .47677 \\ -.45960 \\ .57347 \\ .31691 \\ 1.00539 \\ .02746 \\ .37536 \\ .83127 \\ -.08035 \end{Bmatrix}$$

We are now ready to compute the test statistic S. We have

$$S = \frac{W'\,D^{-1}\,W}{W'\,W} = \frac{\sum_{i=1}^{15} d_i^{-1}\,W_i^2}{\sum_{i=1}^{15} W_i^2} = .329$$

Table II, next page, gives the Pr $(S \leq S \mid H_o)$ and $Pr(S \leq S \mid H_1)$ for selected values of S. Table II was obtained using formulas (3.2) and 3.6) where the integral

$$\int_0^{\infty} \frac{\sin\theta\,(r)}{r\phi(r)}\,dr$$ was approximated using the SSP QSF.

From this table, we can see that Ho is rejected for $\alpha = .055$.

TABLE II

S	Pr $(S \leq SH_0)$	Pr $(S \leq S H_1)$
.1	.000008	.000006
.2	.000010	.000013
.3	.000011	.000455
.4	.000013	.009166
.5	.000027	.037201
.6	.000110	.087500
.7	.000410	.159454
.8	.001800	.249752
.9	.005290	.352139
1.0	.013152	.458964
1.1	.028482	.563044
1.2	.055004	.658782
1.3	.096448	.742552
1.4	.155667	.812595
1.5	.233693	.868682
1.6	.329014	.911720
1.7	.437210	.943309
1.8	.551321	.965419
1.9	.662868	.980093
2.0	.763482	.989279
2.1	.846733	.994656
2.2	.909473	.997578
2.3	.952071	.999029
2.4	.977718	.999682
2.5	.991116	.999936
2.6	.997033	.999968
2.7	.999187	
2.8	.999819	
2.9	.999960	
3.0	.999984	

The null hypothesis (4.2) was tested against the alternative (4.3) for the same textile example, using the BLUS procedure with the last three components of U not estimated. Computing for the von Neuman of U* which was defined as

$$Q^* = \frac{\sum_{i=2}^{14} (u^*_i - u^*_{i-1})^2}{\sum_{i=1}^{14} (u^*_i - \bar{u}^*)^2}$$

we obtain $Q^* = .85072$.

For a 5% level of significance, the critical region is $Q^* < 1.1276$. This is taken from a table of significance points of Q^* tabulated by Abrahamse and Koerts [1]. Hence Ho is rejected.

Theil and Nagar [26] computed for the value of Durbin-Watson's test statistic d for the textile example and found it to be equal to d = 1926. We were able to determine the $\Pr (d \leq d \mid Ho)$ and $\Pr (d \leq \mid H_1)$ for the textile example, using the Imhof theorem, theorem 3.1. Using these significance points, H_0 is not rejected at 4.1% level of significance.

The following table summarizes the above results.

TABLE III

Testing the Independence of the Regression Errors for the Textile Example Using Test Statistics S_1 Q^2 and d.

Rao's S		Theil's Q*		Durbin-Watson d	
α	Conclusion	α	Conclusion	α	
.055	H_0 Not Rejected	.05	H_0 Rejected	.041	H_0 Rejected

The power of a test is defined as the probability that the alternative hypothesis is accepted when it is true. Using this definition we were able to compute the powers of S and d for the textile example from the tabulated values of $\Pr(S \leq S \mid H_1)$ and $\Pr(d \leq d \mid H_1)$. The power of the BLUS test was computed by Abrahamse and Koerst [1].

Below is a comparison of the power of S with the powers of d and Q*

TABLE IV

Powers of S, Q, d for the Textile Example

Rao's S		Theil's Q*		Durbin-Watson's d	
α	Power	α	Power	α	Power
.055	.659	.05	.36	.041	.676

To check the cumulative distribution of S derived in Section III, the emperical distributions of this statistic under both the null and the alternative hypothesis were constructed using Monte Carlo procedure. The observed frequencies were then compared with the corresponding theoretical frequencies as computed from table II, by means of the chi-square test. For a 5% level of significance, we conclude that the fit is good.

REFERENCES

1) ABRAHAMSE, A. P. J. and Koerts J., "On the Power of the BLUS Test", Journal of the American Statistical Association, Vol. 63 (1968), pp. 1227-1236.

2) ABRAHAMSE, A. P. J. and KOERTS J., "A Comparison Between the Power of the Durbin Watson Test and the Power of the BLUS Test", Journal of the American Statistical Association, Vol. 64 (1969), pp. 938-948.

3) ABRAHAMSE. A. P. and KOERTS, J., "New Estimators of Disturbances in Regression Analysis", Journal of the American Statistical Association, Vol. 66 (1971), 71-74.

4) ANDERSON, R. L., "Distribution of the Serial Correlation Coefficient", Annals of Mathematical Statistics, Vol. 13 (1942), pp. 1-13.

5) ANDERSON, R. L. and ANDERSON, T. W., "Distribution of the Circular Serial Correlation Coefficient for Residuals From a Fitted Fourier Series, Annals of Mathematical Statistics, Vol. 21 (1950), pp. 59-81.

6) ANDERSON T. W., "An Introduction to Multivariate Statistical Analysis", John Wiley and Sons. Inc. 1958.

7) ANDERSON, R. L. and BANCROFT, T. A., "Statistical Theory in Research", McGraw-Hill Book Company, Inc., 1952.

8) BELMAN, Richard, "Introduction to Matrix Analysis", MacGraw-Hill. Book Company, Inc., 1960.

9) COCHRANE, D. and ORCUTT, G. H., "Applications of Least Squares Regression to Relations Containing Autocorrelated Error Terms", Journal of the American Statistical Association. Vol. 44 (9149), pp. 32-61.

10) DURBIN, John, "An Alternative to the Bounds Test for Testing Serial Correlation in Least Squares Regression", Econometrika, Vol. 38 (1970), 422-429.

11) DURBIN, J. and Watson, G. S., "Testing for Serial Correlation in Least Squares Regression I," Biometrika, Vol. 37 (1950) pp. 409-428.

12) DURBIN, J. and WATSON, G. S., "Testing for Serial Correlation in Least Squares Regression II", Biometrika, Vol. 38 (1951), pp. 159-178.

13) DURBIN, J. and WATSON, G. S., "Testing for Serial Correlation in Least Squares Regression III", Biometrika, Vol. 58 (1971), pp. 1-19.

14) FINKBINER, Daniel, "Introduction to Matrices and Linear Transformations," W. H. Freeman and Co., San Francisco, 1960.

15) GANTMACKER, F. R., "The Theory of Matrices," Vol. I, Chelsea Publishing Company, New York, 1959.

16) HANNAN, E. J., "Testing for Serial Correlation in Least Squares Regression", Biometrica, Vol. 44, pp. 57-66.

17) HANNAN, E. J., "Times Series Analysis", Metheen, London, 1960.

18) HANNAN, E. J. and TERRELL, R. D., "Testing for Serial Correlation After Least Squares Regression", Econometrika, Vol. 36(1968), pp. 133-150.

19) HENSHAW, R., "Testing Single Equation Least Squares Regression Models for Antocorrelated Disturbances", Econometrica, Vol. 34 (1966), pp. 646-660.

20) IMHOF, P. J., "Computing the Distribution of Quadratic Forms in Normal Variables", Biometrika, Vol. 48 (1961), pp. 419-426.

21) KENDALL, M. G. and STUART, A., "The Advanced Theory of Statistics", Vol. III, Charles Griffin and Company, London, 1961.

22) LEHMAN and STEIN, "Most Powerful Tests of Composite Hypothesis I, Annals of Mathematical Statistics, Vol. 19 (1948), pp. 495-516.

23) PRESS, J., "On Serial Correlation", Annals of Mathematical Statistics Vol. 40 (1969), pp. 188-196.

24) RAO, C. R., Linear Statistical Inference and Its Applications, John Wiley and Sons, Inc., New York, 1965.

25) RAO, K. K., "Testing the Independence of Regression Disturbances", Econometrika, Vol. 38 (1970), pp. 97-113.

26) THEIL, H., "The Analysis of Disturbances in Regression Analysis", Journal of the American Statistical Association, Vol. 60 (1965), pp. 1067-1079.

27) THEIL, H. and NAGAR, A. L., "Testing the Independence of Regression Disturbances", Journal of the American Statistical Association", Vol. 56 (1961), pp. 793-806.

28) THEIL. H., "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances", Journal of the American Statistical Association, Vol. 63 (1968), pp. 242-251.

29) VON NEUMANN, J., KENT, R. H., BELLINSON, H. R. and HART, B. I., "The Mean Square Successive Difference", Annals of Mathematical Statistics, Vol. 12 (1941), 153-162.

30) VON NEUMANN, J., "Distribution of the Mean Square Successive Difference ot the Variance", Annals of Mathematical Statistics, Vol. 12 (1941), 367-395.

31) YULE, G. U., "ON the Time Correlation Problem". Journal of the Royal Statistical Society 84 (1921), pp. 496-537.