Testing for Serial Correlation in Least Squares Regression: I

Author(s): J. Durbin and G. S. Watson

Source: *Biometrika*, Dec., 1950, Vol. 37, No. 3/4 (Dec., 1950), pp. 409-428

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: http://www.jstor.com/stable/2332391

TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION. I

By J. DURBIN and G. S. WATSON

Department of Applied Economics, University of Cambridge

A great deal of use has undoubtedly been made of least squares regression methods in circumstances in which they are known to be inapplicable. In particular, they have often been employed for the analysis of time series and similar data in which successive observations are serially correlated. The resulting complications are well known and have recently been studied from the standpoint of the econometrician by Cochrane & Orcutt (1949). A basic assumption underlying the application of the least squares method is that the error terms in the regression model are independent. When this assumption—among others—is satisfied the procedure is valid whether or not the observations themselves are serially correlated. The problem of testing the errors for independence forms the subject of this paper and its successor. The present paper deals mainly with the theory on which the test is based, while the second paper describes the test procedures in detail and gives tables of bounds to the significance points of the test criterion adopted. We shall not be concerned in either paper with the question of what should be done if the test gives an unfavourable result.

Since the errors in any practical case will be unknown the test must be based on the residuals from the calculated regression. Consequently the ordinary tests of independence cannot be used as they stand, since the residuals are necessarily correlated whether the errors are dependent or not. The mean and variance of an appropriate test statistic have been calculated by Moran (1950) for the case of regression on a single independent variable. The problem of constructing an exact test has been completely solved only in one special case. R. L. & T. W. Anderson (1950) have shown that for the case of regression on a short Fourier series the distribution of the circular serial correlation coefficient obtained by R. L. Anderson (1942) can be used to obtain exact significance points for the test criterion concerned. This is due to the coincidence of the regression vectors with the latent vectors of the circular serial covariance matrix. Perversely enough, this is the very case in which the test is least needed, since the least squares regression coefficients are best unbiased estimates even in the non-null case, and in addition estimates of their variance can be obtained which are at least asymptotically unbiased.

The latent vector case is in fact the only one for which an elegant solution can be obtained. It does not seem possible to find exact significance points for any other case. Nevertheless, bounds to the significance points can be obtained, and in the second paper such bounds will be tabulated. The bounds we shall give are 'best' in two senses: first they can be attained (with regression vectors of a type that will be discussed later), and secondly, when they are attained the test criterion adopted is uniformly most powerful against suitable alternative hypotheses. It is hoped that these bounds will settle the question of significance one way or the other in many cases arising in practice. For doubtful cases there does not seem to be any completely satisfactory procedure. We shall, however, indicate some approximate methods which may be useful in certain circumstances.

The bounds are applicable to all cases in which the independent variables in the regression model can be regarded as 'fixed'. They do not therefore apply to autoregressive schemes and similar models in which lagged values of the dependent variable occur as independent variables.

A further slight limitation of the tables in the form in which we shall present them is that they apply directly only to regressions in which a constant term or mean has been fitted. They cannot therefore be used as they stand for testing the residuals from a regression through the origin. In order to carry out the test in such a case it will be necessary to calculate a regression which includes a fitted mean. Once the test has been carried out the mean can be eliminated by the usual methods for eliminating an independent variable from a regression equation (e.g. Fisher, 1946).

Introduction to theoretical treatment

Any single-equation regression model can be written in the form

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon$$

in which y, the dependent variable, and x, the independent variable, are observed, the errors ϵ being unobserved. We usually require to estimate $\beta_1, \beta_2, \ldots, \beta_k$ and to make confidence statements about the estimates given only the sample

$$\begin{matrix} y_1 & x_{11} & x_{21} & \cdots & x_{k1} \\ y_2 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ y_n & x_{1n} & x_{2n} & \cdots & x_{kn} \end{matrix}$$

Estimates can be made by assuming the errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ associated with the sample to be random variables distributed with zero expectations independently of the x's. If the estimates we make are maximum likelihood estimates, and if our confidence statements are based on likelihood ratios, we can regard the x's as fixed in repeated sampling, that is, they can be treated as known constants even if they are in fact random variables. If in addition $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ can be taken to be distributed independently of each other with constant variance, then by Markoff's theorem the least squares estimates of $\beta_1, \beta_2, \ldots, \beta_k$ are best linear unbiased estimates whatever the form of distribution of the ϵ's. Unbiased estimates of the variances of the estimates can also be obtained without difficulty. These estimates of variance can then be used to make confidence statements by assuming the errors to be normally distributed.

Thus the assumptions on which the validity of the least squares method is based are as follows:

(*a*) The error is distributed independently of the independent variables with zero mean and constant variance;

(*b*) Successive errors are distributed independently of one another.

In what follows autoregressive schemes and stochastic difference equations will be excluded from further consideration, since assumption (*a*) does not hold in such cases. We shall be concerned only with assumption (*b*), that is, we shall assume that the x's can be regarded as 'fixed variables'. When (*b*) is violated the least squares procedure breaks down at three points:

(i) The estimates of the regression coefficients, though unbiased, need not have least variance.

(ii) The usual formula for the variance of an estimate is no longer applicable and is liable to give a serious underestimate of the true variance.

(iii) The t and F distributions, used for making confidence statements, lose their validity.

In stating these consequences of the violation of assumption (b) we do not overlook the fact, pointed out by Wold (1949), that the variances of the resulting estimates depend as much on the serial correlations of the independent variables as on the serial correlation of the errors. In fact, as Wold showed, when all the sample serial correlations of the x's are zero the estimates of variance given by the least squares method are strictly unbiased whether the errors are serially correlated or not. It seems to us doubtful, however, whether this result finds much application in practice. It will only rarely be the case that the independent variables are serially uncorrelated while the errors are serially correlated. Consequently, we feel that there can be little doubt of the desirability of testing the errors for independence whenever the least squares method is applied to serially correlated observations.

To find a suitable test criterion we refer to some results obtained by T. W. Anderson (1948). Anderson showed that in certain cases in which the regression vectors are latent vectors of matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ occurring in the error distribution, the statistic $\dfrac{\mathbf{z}'\boldsymbol{\Theta}\mathbf{z}}{\mathbf{z}'\boldsymbol{\Psi}\mathbf{z}}$, where $\mathbf{z}$ is the column vector of residuals from regression, provides a test that is uniformly most powerful against certain alternative hypotheses. The error distributions implied by these alternative hypotheses are given by Anderson and are such that in the cases that are likely to be useful in practice $\boldsymbol{\Psi} = \mathbf{I}$, the unit matrix. These results suggest that we should examine the distribution of the statistic $r = \dfrac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$ (changing the notation slightly) for regression on any set of fixed variables, $\mathbf{A}$ being any real symmetric matrix.

In the next section we shall consider certain formal properties of r defined in this way, and in § 3 its distribution in the null case will be examined. Expressions for its moments will be derived, and it will be shown that its distribution function lies between two distribution functions which could be determined. In § 4 we return to discuss the question of the choice of an appropriate test criterion with rather more rigour and a specific choice is made. In the final section certain special properties of this test criterion are given.

2. TRANSFORMATION OF r

We consider the linear regression of y on k independent variables $x_1, x_2, \ldots, x_k$. The model for a sample of n observations is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

or in an evident matrix notation, $\qquad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

The least squares estimate of $\boldsymbol{\beta}$ is $\mathbf{b} = \{b_1, b_2, \ldots, b_k\}$ given by $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

The vector $\mathbf{z} = \{z_1, z_2, \ldots, z_n\}$ of residuals from regression is defined by

$$\begin{aligned} \mathbf{z} &= \mathbf{y} - \mathbf{X}\mathbf{b} \\ &= \{\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\mathbf{y}, \end{aligned}$$

where $\mathbf{I}_n$ is the unit matrix of order n.

Thus
$$\begin{aligned}\mathbf{z} &= \{\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon})\\ &= \{\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\boldsymbol{\epsilon}\\ &= \mathbf{M}\boldsymbol{\epsilon} \text{ say.}\end{aligned}$$

It may be verified that $\mathbf{M} = \mathbf{M}' = \mathbf{M}^2$; that is, $\mathbf{M}$ is idempotent.*

We now examine the ratio of quadratic forms $r = \dfrac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$, where $\mathbf{A}$ is a real symmetric matrix. Transforming to the errors we have

$$r = \frac{\boldsymbol{\epsilon}'\mathbf{M}'\mathbf{A}\mathbf{M}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\epsilon}} = \frac{\boldsymbol{\epsilon}'\mathbf{M}\mathbf{A}\mathbf{M}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon}}.$$

We shall show that there exists an orthogonal transformation which simultaneously reduces the numerator and denominator of r to their canonical forms; that is, there is an orthogonal transformation $\boldsymbol{\epsilon} = \mathbf{H}\boldsymbol{\zeta}$ such that

$$r = \frac{\sum_{i=1}^{n-k} \nu_i \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2}.$$

It is well known that there is an orthogonal matrix $\mathbf{L}$ such that

$$\mathbf{L}'\mathbf{M}\mathbf{L} = \left(\begin{array}{c|c}\mathbf{I}_{n-k} & \mathbf{O}\\ \hline \mathbf{O} & \mathbf{O}\end{array}\right),$$

where $\mathbf{I}_{n-k}$ is the unit matrix of order $n-k$, and $\mathbf{O}$ stands for a zero matrix with appropriate numbers of rows and columns. This corresponds to the result that $\sum_{i=1}^{n} z_i^2$ is distributed as χ^2 with $n-k$ degrees of freedom. Thus

$$\begin{aligned}\mathbf{L}'\mathbf{M}\mathbf{A}\mathbf{M}\mathbf{L} &= \mathbf{L}'\mathbf{M}\mathbf{L}.\mathbf{L}'\mathbf{A}\mathbf{L}.\mathbf{L}'\mathbf{M}\mathbf{L}\\ &= \left(\begin{array}{c|c}\mathbf{I}_{n-k} & \mathbf{O}\\ \hline \mathbf{O} & \mathbf{O}\end{array}\right)\left(\begin{array}{c|c}\mathbf{B}_1 & \mathbf{B}_3\\ \hline \mathbf{B}_2 & \mathbf{B}_4\end{array}\right)\left(\begin{array}{c|c}\mathbf{I}_{n-k} & \mathbf{O}\\ \hline \mathbf{O} & \mathbf{O}\end{array}\right)\\ &= \left(\begin{array}{c|c}\mathbf{B}_1 & \mathbf{O}\\ \hline \mathbf{O} & \mathbf{O}\end{array}\right),\end{aligned}$$

where $\left(\begin{array}{c|c}\mathbf{B}_1 & \mathbf{B}_3\\ \hline \mathbf{B}_2 & \mathbf{B}_4\end{array}\right)$ is the appropriate partition of the real symmetric matrix $\mathbf{L}'\mathbf{A}\mathbf{L}$.

Let $\mathbf{N}_1$ be the orthogonal matrix diagonalizing $\mathbf{B}_1$, i.e.

$$\mathbf{N}_1'\mathbf{B}_1\mathbf{N}_1 = \begin{pmatrix}\nu_1 & & & \\ & \nu_2 & & \\ & & \ddots & \\ & & & \nu_{n-k}\end{pmatrix}$$

the blank spaces representing zeros. Then $\mathbf{N} = \left(\begin{array}{c|c}\mathbf{N}_1 & \mathbf{O}\\ \hline \mathbf{O} & \mathbf{I}_k\end{array}\right)$ is orthogonal, so that $\mathbf{H} = \mathbf{L}\mathbf{N}$ is orthogonal.

* This matrix treatment of the residuals is due to Aitken (1935).

Consequently
$$\mathbf{H'MH} = \mathbf{N'L'MLN}$$
$$= \mathbf{N'}\left(\begin{array}{c:c} \mathbf{I}_{n-k} & \mathbf{O} \\ \hdashline \mathbf{O} & \mathbf{O} \end{array}\right)\mathbf{N}$$
$$= \left(\begin{array}{c:c} \mathbf{I}_{n-k} & \mathbf{O} \\ \hdashline \mathbf{O} & \mathbf{O} \end{array}\right),$$

so that
$$\mathbf{H'MAMH} = \mathbf{H'MH}.\mathbf{H'AH}.\mathbf{H'MH}$$
$$= \left(\begin{array}{cccc:c} \nu_1 & & & & \\ & \nu_2 & & & \mathbf{O} \\ & & \ddots & & \\ & & & \nu_{n-k} & \\ \hdashline & & & \mathbf{O} & \mathbf{O} \end{array}\right).$$

Putting $\boldsymbol{\epsilon} = \mathbf{H}\boldsymbol{\zeta}$, we have
$$r = \frac{\sum_{i=1}^{n-k} \nu_i \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2}.$$

This result can be seen geometrically by observing that $\boldsymbol{\epsilon}'\mathbf{MAM}\boldsymbol{\epsilon}$ = constant and $\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon}$ = constant are hypercylinders with parallel generators, the cross-section of $\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon}$ constant being an $[n-k]$ hypersphere.

Determination of $\nu_1, \nu_2, \ldots, \nu_{n-k}$

By standard matrix theory $\nu_1, \nu_2, \ldots, \nu_{n-k}$ are the latent roots of **MAM** other than k zeros; that is, they are the latent roots of $\mathbf{M}^2\mathbf{A}$, since the roots of the product of two matrices are independent of the order of multiplication.* But $\mathbf{M}^2\mathbf{A} = \mathbf{MA}$ since $\mathbf{M}^2 = \mathbf{M}$. Consequently $\nu_1, \nu_2, \ldots, \nu_{n-k}$ are the latent roots of **MA** other than k zeros.

Suppose now that we make the real non-singular transformation of the x's, $\mathbf{P} = \mathbf{XG}$. Then $\mathbf{M} = \mathbf{I}_n - \mathbf{P}(\mathbf{P'P})^{-1}\mathbf{P'}$; that is, **M** is invariant under such transformations. We choose **G** so that the column vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k$ of **P** are orthogonal and are each of unit length, i.e.
$$\mathbf{p}_i'\mathbf{p}_j = \begin{cases} 1 (i = j) \\ 0 (i \neq j) \end{cases}$$

so that
$$\mathbf{P'P} = \mathbf{I}_n.$$

This amounts to saying that we can replace the original independent variables by a normalized orthogonal set without affecting the residuals.

We have, therefore,
$$\mathbf{M} = \mathbf{I}_n - (\mathbf{p}_1\mathbf{p}_1' + \mathbf{p}_2\mathbf{p}_2' + \ldots + \mathbf{p}_k\mathbf{p}_k')$$
$$= (\mathbf{I}_n - \mathbf{p}_1\mathbf{p}_1')(\mathbf{I}_n - \mathbf{p}_2\mathbf{p}_2') \ldots (\mathbf{I}_n - \mathbf{p}_k\mathbf{p}_k')$$
$$= \mathbf{M}_1\mathbf{M}_2 \ldots \mathbf{M}_k, \quad \text{say.}$$

Each factor $\mathbf{M}_i$ has the same form as **M**, the matrix **P** being replaced by the vector $\mathbf{p}_i$; it is idempotent of rank $n-1$ as can be easily verified. From the derivation it is evident that the $\mathbf{M}_i$'s commute. This is an expression in algebraic terms of the fact that we can fit regressions on orthogonal variables separately and in any order without affecting the final result.

* See, for instance, C. C. Macduffee, *The Theory of Matrices* (Chelsea Publishing Company, 1946), Theorem 16·2.

Returning to the main argument we have the result that $\nu_1, \nu_2, \ldots, \nu_{n-k}$ are the roots of $\mathbf{M}_k \ldots \mathbf{M}_2 \mathbf{M}_1 \mathbf{A}$ other than k zeros. From the form of the products we see that any result we establish about the roots of $\mathbf{M}_1 \mathbf{A}$ in terms of those of $\mathbf{A}$ will be true of the roots of $\mathbf{M}_2 \mathbf{M}_1 \mathbf{A}$ in terms of those of $\mathbf{M}_1 \mathbf{A}$. This observation suggests a method of building up a knowledge of the roots of $\mathbf{M}_k \ldots \mathbf{M}_2 \mathbf{M}_1 \mathbf{A}$ in stages starting from the roots of $\mathbf{A}$ which we assume known.

We therefore investigate the latent roots of $\mathbf{M}_1 \mathbf{A}$, say $\theta_1, \theta_2, \ldots, \theta_{n-1}, 0$. These are the roots of the determinantal equation

$$|\mathbf{I}_n \theta - \mathbf{M}_1 \mathbf{A}| = 0,$$

i.e.

$$|\mathbf{I}_n \theta - (\mathbf{I}_n - \mathbf{p}_1 \mathbf{p}_1') \mathbf{A}| = 0. \tag{1}$$

Let $\mathbf{T}$ be the orthogonal matrix diagonalizing $\mathbf{A}$, i.e.

$$\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the latent roots of $\mathbf{A}$. Pre- and post-multiplying (1) by $\mathbf{T}'$ and $\mathbf{T}$, we have

$$|\mathbf{I}_n - (\mathbf{I}_n - \mathbf{l}_1 \mathbf{l}_1') \mathbf{\Lambda}| = 0,$$

where $\mathbf{l}_1 = \{l_{11}, l_{12}, \ldots, l_{1n}\}$ is the vector of direction cosines of $\mathbf{p}_1$ referred to the latent vectors of $\mathbf{A}$ as axes. (Complications arising from multiplicities in the roots of $\mathbf{A}$ are easily overcome in the present context.) Dropping the suffix from $\mathbf{l}_1$ for the moment, we have

$$|\mathbf{I}_n \theta - (\mathbf{I}_n - \mathbf{l}\mathbf{l}') \mathbf{\Lambda}| = 0.$$

Writing out the determinant in full,

$$\begin{vmatrix} \theta - \lambda_1 + l_1^2 \lambda_1, & l_1 l_2 \lambda_2 & \cdots & l_1 l_n \lambda_n \\ l_2 l_1 \lambda_1 & \theta - \lambda_2 + l_2^2 \lambda_2 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ & & \cdots & \\ l_n l_1 \lambda_1 & \cdots & \cdots & \theta - \lambda_n + l_n^2 \lambda_n \end{vmatrix} = 0.$$

Subtracting l_2/l_1 times the first row from the second row, l_3/l_1 times the first from the third, and so on, we can expand the determinant to give the equation

$$\prod_{j=1}^{n} (\theta - \lambda_j) + \sum_{i=1}^{n} l_i^2 \lambda_i \sum_{j \neq i}^{n} (\theta - \lambda_j) = 0.$$

Reducing and taking out a factor θ corresponding to the known zero root of $\mathbf{M}_1 \mathbf{A}$ gives

$$\sum_{i=1}^{n} l_i^2 \prod_{j \neq i}^{n} (\theta - \lambda_j) = 0. \tag{2}$$

$\theta_1, \theta_2, \ldots, \theta_{n-1}$ are the roots of this equation.

We notice that when $l_r = 0$, $\theta - \lambda_r$ is a factor of (2) so that $\theta = \lambda_r$ is a solution. Thus when $\mathbf{p}_1$ coincides with a latent vector of $\mathbf{A}$, $\theta_1, \theta_2, \ldots, \theta_{n-1}$ are equal to the latent roots associated with the remaining $n-1$ latent vectors of $\mathbf{A}$. In the same way if $\mathbf{p}_2$ also coincides with a latent vector of $\mathbf{A}$, the roots of $\mathbf{M}_2 \mathbf{M}_1 \mathbf{A}$ other than two zeros are equal to the latent roots associated with the remaining $n-2$ latent vectors of $\mathbf{A}$. Thus, in general, if the k regression vectors coincide with k of the latent vectors of $\mathbf{A}$, $\nu_1, \nu_2, \ldots, \nu_{n-k}$ are equal to the roots associated with the remaining $n-k$ latent vectors of $\mathbf{A}$. This result remains true if the regression vectors are (linearly independent) linear combinations of k of the latent vectors of $\mathbf{A}$.

For other cases it would be possible to write down an equation similar to (2) giving the roots of $\mathbf{M}_2\mathbf{M}_1\mathbf{A}$ in terms of $\theta_1, \theta_2, \ldots, \theta_{n-1}$, and so on. In this way it would be theoretically possible to determine $\nu_1, \nu_2, \ldots, \nu_{n-k}$. The resulting equations would, however, be quite unmanageable except in the latent vector case just mentioned.

Inequalities on $\nu_1, \nu_2, \ldots, \nu_{n-k}$

We therefore seek inequalities on $\nu_1, \nu_2, \ldots, \nu_{n-k}$. For the sake of generality we suppose that certain of the regression vectors, say $n-k-s$ of them, coincide with latent vectors of $\mathbf{A}$ (or are linear combinations of them). We are left with s of the ν's for which we require inequalities in terms of the remaining $s+k\,\lambda$'s. We renumber them so that

$$\nu_1 \leqslant \nu_2 \leqslant \ldots \leqslant \nu_s,$$
$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_{s+k}.$$

We proceed to show that $$\lambda_i \leqslant \nu_i \leqslant \lambda_{i+k} \quad (i = 1, 2, \ldots, s). \tag{3}$$

It is convenient to establish first an analogous result for the full sets of ν's and λ's. We therefore arrange the suffixes so that

$$\nu_1 \leqslant \nu_2 \leqslant \ldots \leqslant \nu_{n-k},$$
$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n.$$

We also arrange the θ's so that $\theta_1 \leqslant \theta_2 \leqslant \ldots \leqslant \theta_{n-1}$.

It was noted above that if $l_r = 0$, λ_r is a root of (2). Also if any two of the λ's, say λ_r and λ_{r+1}, are equal, then $\lambda_r = \lambda_{r+1}$ is a root of (2). These are the only two cases in which any of the λ's is a root of (2).

For the remaining roots let $$f(\theta) = \sum_{i=1}^{n} l_i^2 \prod_{j \neq i}^{n} (\theta - \lambda_i).$$

Then $$f(\lambda_r) = l_r^2 \prod_{j \neq r}^{n} (\lambda_r - \lambda_j),$$

so that if $f(\lambda_r) > 0$ then $f(\lambda_{r+1}) \leqslant 0$, and if $f(\lambda_r) < 0$ then $f(\lambda_{r+1}) \geqslant 0$. Since $f(\theta)$ is continuous there must therefore be a root in every interval $\lambda_r \leqslant \theta \leqslant \lambda_{r+1}$. Thus

$$\lambda_i \leqslant \theta_i \leqslant \lambda_{i+1} \quad (i = 1, 2, \ldots, n-1). \tag{4}$$

To extend this result we recall that $\mathbf{M}_1\mathbf{A}$ has one zero root in addition to $\theta_1, \theta_2, \ldots, \theta_{n-1}$. Suppose $\theta_l \leqslant 0 \leqslant \theta_{l+1}$; then the roots of $\mathbf{M}_1\mathbf{A}$ can be arranged in the order

$$\theta_1 \leqslant \theta_2 \leqslant \ldots \leqslant \theta_l \leqslant 0 \leqslant \theta_{l+1} \leqslant \ldots \leqslant \theta_{n-1}.$$

Let the roots of $\mathbf{M}_2(\mathbf{M}_1\mathbf{A})$ be $\phi_1, \phi_2, \ldots, \phi_{n-1}$ together with one zero root. Then by (4)

$$\theta_1 \leqslant \phi_1 \leqslant \theta_2 \leqslant \phi_2 \leqslant \ldots \leqslant \theta_l \leqslant \phi_l \leqslant 0 \leqslant \phi_{l+1} \leqslant \ldots.$$

But $\mathbf{M}_2\mathbf{M}_1\mathbf{A}$ certainly has two zero roots, since $\mathbf{M}_2\mathbf{M}_1\mathbf{A}$ has rank at most $n-2$. Thus either ϕ_l or ϕ_{l+1} must be zero. Rejecting one of them and renumbering we have

$$\lambda_i \leqslant \phi_i \leqslant \lambda_{i+2} \quad (i = 1, 2, \ldots, n-2).$$

Applying the same argument successively we have

$$\lambda_i \leqslant \nu_i \leqslant \lambda_{i+k} \quad (i = 1, 2, \ldots, n-k).$$

Deleting cases of equality due to regression vectors coinciding with latent vectors of $\mathbf{A}$ we have (3).

The results of this section will be gathered into a lemma.

LEMMA. If $\mathbf{z}$ and $\boldsymbol{\epsilon}$ are $n \times 1$ vectors such that $\mathbf{z} = \mathbf{M}\boldsymbol{\epsilon}$, where $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and if $r = \dfrac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$, where $\mathbf{A}$ is a real symmetric matrix, then

(*a*) There is an orthogonal transformation $\boldsymbol{\epsilon} = \mathbf{H}\boldsymbol{\zeta}$, such that

$$r = \frac{\sum_{i=1}^{n-k} \nu_i \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2},$$

where $\nu_1, \nu_2, \ldots, \nu_{n-k}$ are the latent roots of $\mathbf{MA}$ other than k zeros;

(*b*) If $n-k-s$ of the columns of $\mathbf{X}$ are linear combinations of $n-k-s$ of the latent vectors of $\mathbf{A}$, then $n-k-s$ of the ν's are equal to the latent roots corresponding to these latent vectors; renumbering the remaining roots such that

$$\nu_1 \leqslant \nu_2 \leqslant \ldots \leqslant \nu_s,$$

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_{s+k},$$

then

$$\lambda_i \leqslant \nu_i \leqslant \lambda_{i+k} \quad (i = 1, 2, \ldots, s).$$

We deduce the following corollary:

COROLLARY

$$r_L \leqslant r \leqslant r_U,$$

where

$$r_L = \frac{\sum_{i=1}^{s} \lambda_i \zeta_i^2 + \sum_{i=s+1}^{n-k} \lambda_{i+k} \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2}$$

and

$$r_U = \frac{\sum_{i=1}^{n-k} \lambda_{i+k} \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2}.$$

This follows immediately by appropriate numbering of suffixes, taking $\lambda_{s+k+1} \ldots \lambda_n$ as the latent roots corresponding to the latent regression vectors and arranging the remainder so that $\lambda_i \leqslant \lambda_{i+1}$. The importance of this result is that it sets bounds upon r which do not depend upon the particular set of regression vectors. r_L and r_U are the best such bounds in that they can be attained, this being the case when the regression vectors coincide with certain of the latent vectors of $\mathbf{A}$.

3. DISTRIBUTION OF r

It has been pointed out that when the errors are distributed independently of the independent variables the latter can be regarded as fixed. There is one special case, however, in which it is more convenient to regard the x's as varying. We shall discuss this first before going on to consider the more general problem of regression on 'fixed variables'.

The case we shall consider is that of a multivariate normal system. In such a system the regressions are linear and the errors are distributed independently of the independent variables. It will be shown that if $y, x_1, \ldots, x_k$ are distributed jointly normally such that the regression of y on the x's passes through the origin, and if successive observations are independent, then r is distributed as if the residuals $z_1, \ldots, z_n$ were independent normal variables. That is, the regression effect disappears from the problem. Similarly, when the regression does not pass through the origin r is distributed as if the z's were residuals from the sample mean of n normal independent observations.

This is perhaps not a very important case in practice, since it will rarely happen that we shall wish to test the hypothesis of serial correlation in the errors when it is known that successive observations of the x's are independent. Nevertheless, it is convenient to deal with it first before going on to discuss the more important case of regression on 'fixed variables'.

To establish the result we consider the geometrical representation of the sample and observe that the sample value of r depends only on the direction in space of the residual vector z. If the x's are kept fixed and the errors are normal and independent, $\mathbf{z}$ is randomly directed in the $[n-k]$ space orthogonal to the space spanned by the x vectors. If the x's are allowed to vary $\mathbf{z}$ will be randomly directed in the $[n]$ space if and only if the x's are jointly normal and successive observations are independent (Bartlett, 1934). In this case the direction of $\mathbf{z}$ is distributed as if $z_1, z_2, \ldots, z_n$ were normal and independent with the same variance. Thus when $y, x_1, \ldots, x_k$ are multivariate normal such that the regression of y on x passes through the origin, r is distributed as if the residuals from the fitted regression through the origin were normal and independent variables.

In the same way it can be shown that if we fit a regression including a constant term, r is distributed as if the z's were residuals from a sample mean of normal independent variables whether the population regression passes through the origin or not.

Regression on 'fixed variables'

To examine the distribution of r on the null hypothesis in the 'fixed variable' case we assume that the errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent normal variables with constant variance, i.e. they are independent $N(0, \sigma^2)$ variables. Transforming as in §2 we have

$$r = \frac{\sum_{i=1}^{n-k} \nu_i \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2}.$$

Since the transformation is orthogonal, $\zeta_1, \zeta_2, \ldots, \zeta_{n-k}$ are independent $N(0, \sigma^2)$ variables. It is evident that the variation of r is limited to the range (ν_1, ν_{n-k}).

Assuming the ν's known, the exact distribution of r has been given by R. L. Anderson (1942) for two special cases: first for $n-k$ even, the ν's being equal in pairs, and second for $n-k$ odd, the ν's being equal in pairs with one value greater or less than all the others. Anderson's expressions for the distribution function are as follows:

$$P(r > r') = \sum_{i=1}^{m} \frac{(\tau_i - r')^{\frac{1}{2}(n-k)-1}}{\alpha_i} \quad (\tau_{m+1} \leqslant r' \leqslant \tau_m),$$

where

$n-k$ *even*: the ν_i's form $\frac{1}{2}(n-k)$ distinct pairs denoted by $\tau_1 > \tau_2 > \ldots > \tau_{\frac{1}{2}(n-k)}$ and $\alpha_i = \prod_{j \neq i}^{\frac{1}{2}(n-k)} (\tau_i - \tau_j)$,

$n-k$ *odd*: the ν_i's form $\frac{1}{2}(n-k-1)$ distinct pairs as above together with one isolated root τ less than all the others and $\alpha_i = \lambda \prod_{j \neq i}^{\frac{1}{2}(n-k-1)} (\tau_i - \tau_j) \sqrt{(\tau_i - \tau)}$.

The expression for $n-k$ odd, $\tau > \tau_1$ is obtained by writing $-r$ for r.

Formulae for the density function are also given by Anderson.

For the case in which the ν's are all different and $n-k$ is even the $[\frac{1}{2}(n-k)-1]$th derivative of the density function has been given by von Neumann (1941), but up to the present no elementary expression for the density function itself has been put forward. Von Neumann's expression for the derivative is as follows:

$$\frac{d^{\frac{1}{2}(n-k)-1}}{dr^{\frac{1}{2}(n-k)-1}} f(r) = 0, \quad m \text{ even}$$

$$= \frac{(-1)^{\frac{1}{2}(n-k-m-1)}\left(\dfrac{n-k}{2}-1\right)!}{\pi\sqrt{\left(-\prod\limits_{j=1}^{n-k}(r-\nu_i)\right)}}, \quad m \text{ odd}$$

for $\nu_m < r < \nu_{m+1}$, $m = 1, 2, \ldots, n-k-1$.

To use these results in any particular case the ν's would need to be known quantities, which means in practice that the regression vectors must be latent vectors of **A**. In addition, the roots associated with remaining $n-k$ latent vectors of **A** must satisfy R. L. Anderson's or von Neumann's conditions.

The results can also be applied to the distributions of r_L and r_U, the lower and upper bounds of r, provided the appropriate λ's satisfy the conditions. Using the relations

$$F_L(r) \geqslant F(r) \geqslant F_U(r), \tag{5}$$

where F_L and F_U are the distribution functions of r_L and r_U we would then have limits to the distribution function of r. The truth of the relations (5) can be seen by noting that r_L and r are in (1, 1) correspondence and $r_L \leqslant r$ always.

Approximations

R. L. Anderson's distribution becomes unwieldy to work with when $n-k$ is moderately large, and von Neumann's results can only be used to give an exact distribution when $n-k$ is very small. For practical applications, therefore, approximate methods are required.

We first mention the result, pointed out by T. W. Anderson (1948), that as $n-k$ becomes large r is asymptotically normally distributed with the mean and variance given later in this paper. For moderate values of $n-k$, however, it appears that the distributions of certain statistics of the type r are better approximated by a β-distribution, even when symmetric.* One would expect the advantage of the β over the normal approximation to be even greater when the ν's are such that the distribution of r is skew. For better approximations various expansions in terms of β-functions can be used. One such expansion was used for most of the tabulation of the distribution of von Neumann's statistic (Hart 1942). Another method is to use a series expansion in terms of Jacobi polynomials using a β-distribution expression as weight function. (See, for instance, Courant & Hilbert,† 1931, p. 76.) The first four terms of such a series will be used for calculating some of the bounds to the significance points of r tabulated in our second paper.

Moments of r

To use the above approximations we require the moments of r. First we note that since r is independent of the scale of the ζ's we can take σ^2 equal to unity. We therefore require the moments of $r = u/v$, where $u = \sum_{i=1}^{n-k} \nu_i \zeta_i^2$ and $v = \sum_{i=1}^{n-k} \zeta_i^2$, $\zeta_1, \zeta_2, \ldots, \zeta_{n-k}$ being independent $N(0, 1)$ variables.

* See, for instance, Rubin (1945), Dixon (1944), R. L. Anderson and T. W. Anderson (1950).

† Note, however, the misprint: $x^q(1-x)^{p-q}$ should read $x^{q-1}(1-x)^{p-q}$.

It is well known (Pitman, 1937; von Neumann, 1941) that r and v are distributed independently. Consequently

$$E(u^s) = E(r^s v^s) = E(r^s)\,E(v^s),$$

so that

$$E(r^s) = \frac{E(u^s)}{E(v^s)},$$

that is, the moments of the ratio are the ratios of the moments.

The moments of u are most simply obtained by noting that u is the sum of independent variables $\nu_i \zeta_i^2$, where ζ_i^2 is a χ^2 variable with one degree of freedom. Hence the sth cumulant of u is the sum of sth cumulants, that is

$$\kappa_s(u) = 2^{s-1}(s-1)!\sum_{i=1}^{n-k} \nu_i^s,$$

since

$$\kappa_s(\nu_i \zeta_i^2) = 2^{s-1}(s-1)!\,\nu_i^s.$$

In particular

$$\kappa_1(u) = \Sigma \nu_i, \quad \kappa_2(u) = 2\Sigma \nu_i^2.$$

The moments of u can then be obtained from the cumulants.

The moments of v are simply those of χ^2 with $n-k$ degrees of freedom, i.e.

$$E(v) = n-k,$$
$$E(v^2) = (n-k)(n-k+2), \text{ etc.}$$

Hence

$$E(r) = \mu_1' = \frac{1}{n-k}\sum_{i=1}^{n-k} \nu_i = \bar{\nu} \quad \text{say.} \tag{6}$$

To obtain the moments of r about the mean we have

$$r - \mu_1' = \frac{\Sigma(\nu_i - \bar{\nu})\,\zeta_i^2}{\Sigma \zeta_i^2} = \frac{u'}{v} \quad \text{say.}$$

As before the moments of $r - \mu_1'$ are the moments of u' divided by the moments of v. The moments of u' are obtained from the cumulants

$$\kappa_s(u') = 2^{s-1}(s-1)!\sum_{i=1}^{n-k} (\nu_i - \bar{\nu})^s.$$

In this way we find

$$\left.\begin{aligned}
\operatorname{var} r = \mu_2 &= \frac{2\Sigma(\nu_i - \bar{\nu})^2}{(n-k)(n-k+2)},\\
\mu_3 &= \frac{8\Sigma(\nu_i - \bar{\nu})^3,}{(n-k)(n-k+2)(n-k+4)},\\
\mu_4 &= \frac{48\Sigma(\nu_i - \bar{\nu})^4 + 12\{\Sigma(\nu_i - \bar{\nu})^2\}^2}{(n-k)(n-k+2)(n-k+4)(n-k+6)}.
\end{aligned}\right\} \tag{7}$$

It must be emphasized at this point that the moments just given refer to regression through the origin on k independent variables. If the regression model includes a constant term, that is, if the calculated regression includes a fitted mean, and if, as is usual, we wish to distinguish the remaining independent variables from the constant term, then k must be taken equal to $k'+1$ in the above expressions, k' being the number of independent variables in addition to the constant. We emphasize this point, since it is k' that is usually referred to as the number of independent variables in such a model.

The expressions given will enable the moments of r to be calculated when the ν's are known. In most cases that will arise in practice, however, the ν's will be unknown and it will be impracticable to calculate them. We therefore require means of expressing the power sums $\Sigma \nu_i^2$ in terms of known quantities, namely the matrix $\mathbf{A}$ and the independent variables.

To do this we make use of the concept of the trace of a matrix, that is, the sum of its leading diagonal elements. This is denoted for a matrix $\mathbf{S}$ by $\operatorname{tr}\mathbf{S}$, $\mathbf{S}$ being of course square. It is easy to show that the operation of taking a trace satisfies the following simple rules:

(*a*) $\operatorname{tr}(\mathbf{S}+\mathbf{T}) = \operatorname{tr}\mathbf{S}+\operatorname{tr}\mathbf{T}$,

(*b*) $\operatorname{tr}\mathbf{ST} = \operatorname{tr}\mathbf{TS}$ whether $\mathbf{S}$ and $\mathbf{T}$ are square or rectangular.

From these rules we deduce a third:

(*c*) $\operatorname{tr}(\mathbf{S}+\mathbf{T})^q = \operatorname{tr}\mathbf{S}^q + \binom{q}{1}\operatorname{tr}\mathbf{S}^{q-1}\mathbf{T} + \binom{q}{2}\operatorname{tr}\mathbf{S}^{q-2}\mathbf{T}^2 + \ldots + \operatorname{tr}\mathbf{T}^q$,

when $\mathbf{S}$ and $\mathbf{T}$ are square. In addition, we note that $\operatorname{tr}\mathbf{S} = \sum_{i=1}^{m} \sigma_i$, where $\sigma_1, \sigma_2, \ldots, \sigma_m$ are the latent roots of $\mathbf{S}$, and in general that $\operatorname{tr}\mathbf{S}^q = \sum_{i=1}^{m} \sigma_i^q$.

Thus we have immediately $$\sum_{i=1}^{n-k} \nu_i^q = \operatorname{tr}(\mathbf{MA})^q,$$

since $\nu_1, \nu_2, \ldots, \nu_{n-k}$ together with k zeros are the latent roots of $\mathbf{MA}$.

In cases in which the independent variables are known constants it is sometimes possible to construct the matrix $\mathbf{MA}$ directly and hence to obtain the mean and variance of r in a fairly straightforward way.

For models of other types in which the independent variables can take arbitrary values further reduction is needed. For the mean we require

$$\begin{aligned} \Sigma \nu_i &= \operatorname{tr}\mathbf{MA} = \operatorname{tr}\{\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\}\mathbf{A} \\ &= \operatorname{tr}\mathbf{A} - \operatorname{tr}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A} \quad \text{by rule } (a) \\ &= \operatorname{tr}\mathbf{A} - \operatorname{tr}\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad \text{by rule } (b). \end{aligned} \tag{8}$$

The calculation of this expression is not as formidable an undertaking as might at first sight appear, since $(\mathbf{X}'\mathbf{X})^{-1}$ will effectively have to be calculated in any case for the estimation of the regression coefficients. It is interesting to note incidentally that the matrix $\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ in the expression is a direct multivariate generalization of the statistic r.

For the variance we require

$$\begin{aligned} \Sigma \nu_i^2 &= \operatorname{tr}(\mathbf{MA})^2 = \operatorname{tr}\{\mathbf{A} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\}^2 \\ &= \operatorname{tr}\mathbf{A}^2 - 2\operatorname{tr}\mathbf{X}'\mathbf{A}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \operatorname{tr}\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^2, \end{aligned} \tag{9}$$

by rules (*b*) and (*c*).

Similarly

$$\begin{aligned} \Sigma \nu_i^3 = \operatorname{tr}\mathbf{A}^3 &- 3\operatorname{tr}\mathbf{X}'\mathbf{A}^3\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &+ 3\operatorname{tr}\{\mathbf{X}'\mathbf{A}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\} + \operatorname{tr}\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^3, \end{aligned} \tag{10}$$

$$\begin{aligned} \Sigma \nu_i^4 = \operatorname{tr}\mathbf{A}^4 &- 4\operatorname{tr}\mathbf{X}'\mathbf{A}^4\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + 6\operatorname{tr}\{\mathbf{X}'\mathbf{A}^3\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\} \\ &- 4\operatorname{tr}[\mathbf{X}'\mathbf{A}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^2] + \operatorname{tr}\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^4, \end{aligned} \tag{11}$$

and so on.

When the independent variables are orthogonal these expressions can be simplified somewhat since $\mathbf{X}'\mathbf{X}$ is then a diagonal matrix. Thus

$$\operatorname{tr}\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sum_{i=1}^{k} \frac{\mathbf{x}_i'\mathbf{A}\mathbf{x}_i}{\mathbf{x}_i'\mathbf{x}_i},$$

$\mathbf{x}_i$ standing for the vector of sample values of the ith independent variable. Each term in the summation has the form r in terms of one of the independent variables. Similarly for $\operatorname{tr}\mathbf{X}'\mathbf{A}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, $\operatorname{tr}\mathbf{X}'\mathbf{A}^3\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, etc. We have also

$$\operatorname{tr}\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^2 = \sum_{i=1}^{k}\left(\frac{\mathbf{x}_i'\mathbf{A}\mathbf{x}_i}{\mathbf{x}_i'\mathbf{x}_i}\right)^2 + 2\sum_{i\neq j}^{k}\frac{(\mathbf{x}_i'\mathbf{A}\mathbf{x}_j)^2}{\mathbf{x}_i'\mathbf{x}_i\,\mathbf{x}_j'\mathbf{x}_j}.$$

Thus when the regression vectors are orthogonal the following formulae enable us to calculate the mean and variance of r:

$$\left.\begin{aligned}\Sigma\nu_i &= \operatorname{tr}\mathbf{A} - \sum_{i=1}^{k}\frac{\mathbf{x}_i'\mathbf{A}\mathbf{x}_i}{\mathbf{x}_i'\mathbf{x}_i},\\ \Sigma\nu_i^2 &= \operatorname{tr}\mathbf{A}^2 - 2\sum_{i=1}^{k}\frac{\mathbf{x}_i'\mathbf{A}^2\mathbf{x}_i}{\mathbf{x}_i'\mathbf{x}_i} + \sum_{i=1}^{k}\left(\frac{\mathbf{x}_i'\mathbf{A}\mathbf{x}_i}{\mathbf{x}_i'\mathbf{x}_i}\right)^2 + 2\sum_{i<j}^{k}\frac{(\mathbf{x}_i'\mathbf{A}\mathbf{x}_j)^2}{\mathbf{x}_i'\mathbf{x}_i\mathbf{x}_j'\mathbf{x}_j}.\end{aligned}\right\} \tag{12}$$

The mean and variance are obtained by substituting these values in (6) and (7).

Similar results apply when $\mathbf{X}$ is partitioned into two or more orthogonal sets of variables. For instance, when $\mathbf{X}$ consists of the constant vector $\{c, c, \ldots, c\}$ together with the matrix $\dot{\mathbf{X}}$ of deviations from the means of the remaining $k-1$ variables, i.e.

$$\dot{\mathbf{x}}_{ij} = \mathbf{x}_{ij} - \overline{\mathbf{x}}_i (i = 2, 3, \ldots, k;\ j = 1, 2, \ldots, n),$$

then $$\operatorname{tr}\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \frac{\mathbf{i}'\mathbf{A}\mathbf{i}}{n} + \operatorname{tr}\dot{\mathbf{X}}'\mathbf{A}\dot{\mathbf{X}}(\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1},$$

$$\operatorname{tr}\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^2 = \left(\frac{\mathbf{i}'\mathbf{A}\mathbf{i}}{n}\right)^2 + \frac{2\mathbf{i}'\mathbf{A}\dot{\mathbf{X}}(\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1}\dot{\mathbf{X}}'\mathbf{A}\mathbf{i}}{n} + \operatorname{tr}\{\dot{\mathbf{X}}'\mathbf{A}\dot{\mathbf{X}}(\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1}\}^2,$$

where $\mathbf{i}$ is the equiangular vector $\{1, 1, \ldots, 1\}$. When this is a latent vector of $\mathbf{A}$ corresponding to a latent root of zero, $\mathbf{i}'\mathbf{A} = \mathbf{O}$. We then have the important result that (8)–(11) apply without change except that the original variables $\mathbf{X}$ are replaced by the deviations from their means $\dot{\mathbf{X}}$. This result holds whenever $\mathbf{x}'\mathbf{A}\mathbf{x}$ is invariant under a change of origin of $\mathbf{x}$.

Before closing this treatment of moments we should mention one difficulty in using them for obtaining approximations in terms of β-distributions and associated expansions. In constructing such approximations one usually knows the range within which the variable is distributed. In the present problem, however, the range is (ν_1, ν_{n-k}), which will often be unknown and impracticable to determine. In such cases it will accordingly be necessary to use approximations to ν_1 and ν_{n-k} before the distributions can be fitted.

Characteristic function of u and v

An alternative method of obtaining the moments of u and v is to use their joint characteristic function. This is given by

$$\begin{aligned}\phi(t_1, t_2) &= \frac{1}{(2\pi)^{\frac{1}{2}(n-k)}}\int\cdots\int\exp\left(it_1\Sigma\nu_i\zeta_i^2 + it_2\Sigma\zeta_i^2 - \tfrac{1}{2}\Sigma\zeta_i^2\right)d\zeta_1\ldots d\zeta_{n-k}\\ &= \prod_{j=1}^{n-k}(1 - 2\nu_j it_1 - 2it_2)^{-\frac{1}{2}}\\ &= (1-2it_2)^{\frac{1}{2}k}\prod_{j=1}^{n-k}(1 - 2it_2 - 2\nu_j it_1)^{-\frac{1}{2}}(1-2it_2)^{-\frac{1}{2}k}\\ &= (1-2it_2)^{\frac{1}{2}k}\,|\,\mathbf{I}_n(1-2it_2) - 2it_1\mathbf{M}\mathbf{A}\,|^{-\frac{1}{2}},\end{aligned}$$

since $\nu_1 \ldots \nu_{n-k}$ together with k zeros are the roots of the equation $|\,\mathbf{I}_n\nu - \mathbf{M}\mathbf{A}\,| = 0$.

A more manageable expression can be obtained by considering first the case of a single independent variable. The characteristic function $\phi_1(t_1 t_2)$ is then given by

$$\frac{1}{\phi_1^2} = \prod_{j=1}^{n-1} (1 - 2\theta_j it_1 - 2it_2), \tag{13}$$

where $\theta_1, \theta_2, \ldots, \theta_{n-1}$ are the roots of the equation

$$\sum_{i=1}^{n} l_i^2 \prod_{j \neq i}^{n} (\theta - \lambda_j) = 0. \tag{2 bis}$$

Consequently
$$\prod_{j=1}^{n-1} (\theta - \theta_j) = \sum_{i=1}^{n} l_i^2 \prod_{j \neq i}^{n} (\theta - \lambda_j)$$
$$= \prod_{j=1}^{n} (\theta - \lambda_j) \sum_{j=1}^{n} \frac{l_j^2}{\theta - \lambda_j}$$

for all values of θ except $\lambda_1, \lambda_2, \ldots, \lambda_n$. From (13)

$$\frac{1}{\phi_1^2} = (2it_1)^{n-1} \prod_{j=1}^{n-1} \left(\frac{1-2it_2}{2it_1} - \theta_j\right)$$
$$= (2it_1)^{n-1} \prod_{j=1}^{n} \left(\frac{1-2it_2}{2it_1} - \theta_j\right) \sum_{j=1}^{n} \frac{l_j^2}{\dfrac{1-2it_2}{2it_1} - \lambda_j}$$
$$= \prod_{j=1}^{n} (1 - 2\lambda_j it_1 - 2it_2) \sum_{j=1}^{n} \frac{l_j^2}{(1 - 2\lambda_j it_1 - 2it_2)}. \tag{14}$$

The left-hand factor of this expression is the characteristic function of u and v that would be obtained if the z's were independent normal variables, the right-hand factor giving the modification due to the fitting of a regression on a single independent variable.

To reduce the expression further we note that $\dfrac{1}{1 - 2\lambda_j it_1 - 2it_2}$ $(j = 1, 2, \ldots, n)$ are the latent roots of the matrix

$$\{(1 - 2it_2)\mathbf{I}_n - 2it_1 \mathbf{A}\}^{-1} = \mathbf{B}^{-1} \text{ say.}$$

Moreover, the latent vectors of $\mathbf{B}^{-1}$ are the same as those of $\mathbf{A}$ so that $l_1, l_2, \ldots, l_n$ are the direction cosines of the vector $\mathbf{x}$ relative to these latent vectors. Consequently

$$\sum_{j=1}^{n} \frac{l_i^2}{1 - 2\lambda_j it_1 - 2it_2} = \frac{\mathbf{x}'\mathbf{B}^{-1}\mathbf{x}}{\mathbf{x}'\mathbf{x}},$$

where $\mathbf{x}$ is the independent variable concerned. Also

$$\prod_{j=1}^{n} (1 - 2\lambda_j it_1 - 2it_2) = |\mathbf{B}|.$$

Thus
$$\frac{1}{\phi_1^2} = |\mathbf{B}| \frac{\mathbf{x}'\mathbf{B}^{-1}\mathbf{x}}{\mathbf{x}'\mathbf{x}}. \tag{15}$$

It is interesting to note that the second factor takes the general form r.

By a direct extension of this argument it can be shown that for regression on k independent variables the characteristic function is given by

$$\frac{1}{\phi^2} = |\mathbf{B}_1| \prod_{s=1}^{k} \frac{\mathbf{x}_s'\mathbf{B}_s^{-1}\mathbf{x}_s}{\mathbf{x}_s'\mathbf{x}_s},$$

where
$$\mathbf{B}_s = (1 - 2it_2)\mathbf{I}_n - 2it_1 \mathbf{M}_{s-1} \ldots \mathbf{M}_2 \mathbf{M}_1 \mathbf{A},$$

the $\mathbf{M}_i$'s being defined as in §2. This result could also be written down directly given (15) in virtue of the reproductive property of the products $\ldots \mathbf{M}_2 \mathbf{M}_1 \mathbf{A}$ mentioned in §2.

Putting $t_2 = 0$ we obtain the characteristic function of u. The cumulants and hence the moments can then be obtained by the expansion of $\log \phi$.

4. CHOICE OF TEST CRITERION*

To decide upon a suitable test criterion an important consideration is the set of alternative hypotheses against which it is desired to discriminate. The kind of alternative we have in mind in this paper is such that the correlogram of the errors diminishes approximately exponentially with increasing separation of the observations. A convenient model for such hypotheses is the stationary Markoff process

$$\epsilon_i = \rho\epsilon_{i-1} + u_i \quad (i = \ldots -1, 0, 1, \ldots), \tag{16}$$

where $|\rho| < 1$ and u_i is normal with mean zero and variance σ^2 and is independent of $\epsilon_{i-1}, \epsilon_{i-2}, \ldots$ and $u_{i-1}, u_{i-2}, \ldots$. The null hypothesis is then the hypothesis that $\rho = 0$ in (16).

It has been shown by T. W. Anderson (1948) that no test of this hypothesis exists which is uniformly most powerful against alternatives (16). Anderson also showed, however, that for certain regression systems with error distributions close to that given by (16) tests can be obtained which are uniformly most powerful against one-sided alternatives (16) and which give type B_1 regions for two-sided alternatives (16).

These regression systems include cases in which the regression vectors are constant vectors coinciding with latent vectors of a matrix $\mathbf{\Theta}$ (or with linear combinations of k of them) and in which the error distributions have density functions of the form

$$K \exp\left[-\frac{1}{2\sigma^2}\{(1+\rho^2)\,\boldsymbol{\epsilon}'\boldsymbol{\epsilon} - 2\rho\boldsymbol{\epsilon}'\mathbf{\Theta}\boldsymbol{\epsilon}\}\right]. \tag{17}$$

For such cases the uniformly most powerful test of the hypothesis $\rho = 0$ against alternatives $\rho > 0$ is given by $r > r_0$, where $r = \dfrac{\mathbf{z}'\mathbf{\Theta}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$, $\mathbf{z}$ being the vector of residuals from least squares regression, and r_0 being determined to give a critical region of appropriate size. For two-sided alternatives to $\rho = 0$ the type B_1 test is given by $r < r_2$, $r > r_3$, where r_2 and r_3 are determined so as to give a critical region of appropriate size and to satisfy the relation

$$\int_{r_2}^{r_3} rp(r)\,dr = E(r)\int_{r_2}^{r_3} p(r)\,dr,$$

$p(r)$ being the density function of r in the null case.

We recall that whatever the regression vectors,

$$r_L \leqslant r \leqslant r_U, \tag{18}$$

where r_L and r_U are defined in the Corollary, § 2. Now r_L and r_U have distributions in the null case identical with distributions of r obtained from residuals from regressions on certain latent vectors of the matrix $\mathbf{A}$. Thus if we put $\mathbf{\Theta} = \mathbf{A}$ in (17) we can say that when the lower bound r_L in (18) is attained (or the upper bound), the statistic $r = \dfrac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$ gives a test which is uniformly most powerful against one-sided alternatives (16) and which is of type B_1 against two-sided alternatives.

The error distribution for the stationary Markoff process (16) has the density function

$$K \exp\left[-\frac{1}{2\sigma^2}\left\{(1+\rho^2)\sum_{i=1}^{n}\epsilon_i^2 - \rho^2(\epsilon_1^2+\epsilon_n^2) - 2\rho\sum_{i=2}^{n}\epsilon_i\epsilon_{i-1}\right\}\right]. \tag{19}$$

* This section is based on the treatment given by T. W. Anderson (1948).

Taking $\boldsymbol{\epsilon}'\boldsymbol{\Theta}\boldsymbol{\epsilon} = \sum_{i=2}^{n} \epsilon_i \epsilon_{i-1}$ in (17) gives a density function

$$K \exp\left[-\frac{1}{2\sigma^2}\left\{(1+\rho^2)\sum_{i=1}^{n}\epsilon_i^2 - 2\rho\sum_{i=2}^{n}\epsilon_i\epsilon_{i-1}\right\}\right], \tag{20}$$

while taking $\boldsymbol{\epsilon}'\boldsymbol{\Theta}\boldsymbol{\epsilon} = \sum_{i=1}^{n} \epsilon_i^2 - \frac{1}{2}\sum_{i=2}^{n}(\epsilon_i - \epsilon_{i-1})^2$ in (17) gives a density function

$$K \exp\left[-\frac{1}{2\sigma^2}\left\{(1+\rho^2)\sum_{i=1}^{n}\epsilon_i^2 - \rho(\epsilon_1^2+\epsilon_n^2) - 2\rho\sum_{i=2}^{n}\epsilon_i\epsilon_{i-1}\right\}\right]. \tag{21}$$

These are both close to (19). Thus following Anderson we conjecture that either value of $\boldsymbol{\Theta}$ would give a good statistic r for testing against alternatives (16). Between the two statistics there is not much to choose. We ourselves have adopted a slight modification of the second, partly for reasons of computational convenience and partly because of similarity to von Neumann's statistic δ^2/s^2 (1941) already well known to research workers.

The statistic we have adopted is defined by

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

which is related to $\frac{\delta^2}{s^2}$ by $\frac{\delta^2}{s^2} = \frac{nd}{n-1}$. This is a special case of the general statistic $r = \frac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$ discussed in §2 and §3, in which

$$\mathbf{A} = \mathbf{A}_d = \tfrac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & \dots & \dots \\ 0 & -1 & 2 & -1 & \dots & \dots & \dots \\ 0 & 0 & -1 & 2 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & 2 & -1 \\ 0 & \dots & \dots & \dots & 0 & -1 & 1 \end{pmatrix}.$$

In the notation of the previous paragraph we would take $\boldsymbol{\Theta} = \mathbf{I} - \frac{1}{2}\mathbf{A}_d$ to give the density (21). Now the latent vectors of the matrices $\mathbf{A}_d$ and $\boldsymbol{\Theta}$ in this equation are the same. Thus when the regression vectors are latent vectors of $\mathbf{A}_d$ the statistic d provides a uniformly most powerful test against one-sided alternatives (21). In particular the test given by d when the bounds r_L and r_U are attained is uniformly most powerful.

The main alternative to using d or a related statistic as a test criterion would be to use one of the circular statistics such as

$$r_c = \frac{\sum_{i=1}^{n} z_i z_{i-1}}{\sum_{i=1}^{n} z_i^2}$$

or

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

where we define $z_0 \equiv z_n$ in each case. T. W. Anderson (1948) has shown that r_c and d_c give uniformly most powerful tests against one-sided alternatives in the circular population having a density function

$$K \exp\left[-\frac{1}{2\sigma^2}\left\{(1+\rho^2)\sum_{i=1}^{n}\epsilon_i^2 - 2\rho\sum_{i=1}^{n}\epsilon_i\epsilon_{i-1}\right\}\right], \tag{22}$$

where $\epsilon_0 \equiv \epsilon_n$. r_c was the statistic adopted by R. L. Anderson & T. W. Anderson (1950) for testing the residuals from regression on a Fourier series.

The disadvantage of r_c and d_c is that (22) is not so close to (19) as (20) or (21). The advantage is that since the latent roots of the associated values of $\mathbf{A}$ are equal in pairs, the results of R. L. Anderson (1942) can sometimes be used to obtain exact distributions in the null case. The roots of $\mathbf{A}_d$, on the other hand, are all distinct. We conclude that d or a related non-circular statistic would seem to be preferable whenever an approximation to the distribution is sufficient, but that a circular statistic would seem to be preferable if exact results are required at the expense of some loss of power. We mention that the computations involved in using Anderson's exact distribution become very tedious as the number of degrees of freedom increases.

The next question that arises is how good these statistics are as test criteria in cases in which the regression vectors are not latent vectors. Such cases are of course by far the more frequent in practice. It is evident that we can expect the power of the test to diminish as the regression vectors depart from the latent vectors, since the least squares regression coefficients are not then maximum likelihood estimates in the non-null case. Thus any test based on least squares residuals cannot even be a likelihood ratio test. Against this three points can be made. The first is that we still have a valid test, though possibly of reduced power. Secondly, it is desirable on grounds of convenience to have a test based on least squares residuals even though it is not an optimal test. Thirdly, the statistic r necessarily lies between the bounds r_L and r_U and when these bounds are attained the test is optimal. We note also that it is only for the latent vector case that the distribution problems have been approached with any success.

5. SOME SPECIAL RESULTS

To obtain the moments of d we need the powers of $\mathbf{A}_d$. Because of the symmetry of these matrices they are completely specified by the top left-hand triangle. Thus we can write

$$\mathbf{A}_d \doteq \begin{array}{rrr} 1 & -1 & 0 \\ & 2 & -1 \\ & & 2 \end{array}$$

We find

$$\mathbf{A}_d^2 \doteq \begin{array}{rrrr} 2 & -3 & 1 & 0 \\ & 6 & -4 & 1 \\ & & 6 & -4 \\ & & & 6 \end{array}$$

$$\mathbf{A}_d^3 \doteq \begin{array}{rrrrrr} 5 & -9 & 5 & -1 & 0 & 0 \\ & 19 & -15 & 6 & -1 & 0 \\ & & 20 & -15 & 6 & -1 \\ & & & 20 & -15 & 6 \end{array}$$

$$\mathbf{A}_d^4 \doteq \begin{array}{rrrrrrr} 14 & -28 & 20 & -7 & 1 & 0 & 0 \\ & 62 & -55 & 28 & -8 & 1 & 0 \\ & & 70 & -56 & 28 & -8 & 1 \\ & & & 70 & -56 & 28 & 1 \end{array}$$

Rather than use these matrices as they stand, however, it will probably be more convenient to proceed by finding the sums of squares of the successive differences of the z's. Denoting the sth differences by $\Delta^s z$ we have

$$\left.\begin{aligned} z'\mathbf{A}_d z &= \sum_{i=1}^{n-1} (\Delta z_i)^2, \\ z'\mathbf{A}_d^2 z &= \Sigma(\Delta^2 z_i)^2 + (z_1 - z_2)^2 + (z_{n-1} - z_n)^2, \\ z'\mathbf{A}_d^3 z &= \Sigma(\Delta^3 z_i)^2 + 4z_1^2 + 9z_2^2 + z_3^2 - 12z_1z_2 - 6z_2z_3 + 4z_1z_3 + \text{a similar expression in } z_n, z_{n-1}, z_{n-2}, \\ z'\mathbf{A}_d^4 z &= \Sigma(\Delta^4 z_i)^2 + 13z_1^2 + 45z_2^2 + 17z_3^2 + z_4^2 - 48z_1z_2 - 54z_2z_3 - 8z_3z_4 + 28z_1z_3 + 12z_2z_4 \\ &\quad - 6z_1z_4 + \text{a similar expression in } z_n, z_{n-1}, z_{n-2}, z_{n-3}. \end{aligned}\right\} \tag{23}$$

For the circular definition of d, i.e.

$$d_c = \frac{\sum_{i=1}^{n} (z_i - z_{i-1})^2}{\sum_{i=1}^{n} z_i^2} = \frac{\mathbf{z}'\mathbf{A}_{dc}\mathbf{z}}{\mathbf{z}'\mathbf{z}} \quad \text{with } z_0 \equiv z_n,$$

the correction terms disappear, giving

$$z'\mathbf{A}_{dc}^s z = \sum_{i=1}^{n} (\Delta^s z_i)^2, \quad \text{where} \quad z_{-i} \equiv z_{n-i}.$$

The latent roots of $\mathbf{A}_d$ are given by

$$\lambda_j = 2\left\{1 - \cos\frac{\pi(j-1)}{n}\right\} \quad (j = 1, 2, \ldots, n) \tag{24}$$

(von Neumann 1941). The first four power sums are:

$$\left.\begin{aligned} \sum_{j=1}^{n} \lambda_j &= 2(n-1), \\ \Sigma\lambda_j^2 &= 2(3n-4), \\ \Sigma\lambda_j^3 &= 4(5n-8), \\ \Sigma\lambda_j^4 &= 2(35n-64). \end{aligned}\right\} \tag{25}$$

The latent vector corresponding to the zero root λ_1 is $\{1, 1, \ldots, 1\}$, which is the regression vector corresponding to a constant term in the regression model. For regressions with a fitted mean, therefore, we need only consider the remaining $n-1$ λ's which we renumber accordingly so that

$$\lambda j = 2\left(1 - \cos\frac{\pi j}{n}\right) \quad (j = 1, 2, \ldots, n-1).$$

With these λ's we have from the Corollary, §2,

$$d_L \leqslant d \leqslant d_U, \tag{26}$$

where

$$d_L = \frac{\sum_{i=1}^{n-k'-1} \lambda_i \zeta_i^2}{\sum_{i=1}^{n-k'-1} \zeta_i^2}, \tag{27}$$

$$d_U = \frac{\sum_{i=1}^{n-k'-1} \lambda_{i+k'} \zeta_i^2}{\sum_{i=1}^{n=k'-1} \zeta_i^2}, \tag{28}$$

k' being the number of independent variables in the model in addition to the constant term.

With the error distribution assumed in §3 the limits of the mean of d are given by

$$E(d) \leqslant E(d_U) = 2 - \frac{2}{n-k'-1} \sum_{j=k'+1}^{n-1} \cos \frac{\pi j}{n}$$

$$\geqslant E(d_L) = 2 - \frac{2}{n-k'-1} \sum_{j=1}^{n-k'-1} \cos \frac{\pi j}{n}.$$

We state without proof the limits of the variance of d:

$$\operatorname{var}(d) \leqslant \frac{16}{(n-k'-1)(n-k'+1)} \sum_{j=1}^{\frac{1}{2}(n-k'-1)} \cos^2 \frac{\pi j}{n} \quad (n-k' \text{ odd}),$$

$$\leqslant \frac{16}{(n-k'-1)(n-k'+1)} \sum_{j=1}^{\frac{1}{2}(n-k')-1} \cos^2 \frac{\pi j}{n} + \frac{8(n-k'-2)}{(n-k'-1)^2(n-k'+1)} \cos^2 \frac{(n-k')\pi}{2n} \quad (n-k' \text{ even}),$$

$$\geqslant \frac{16}{(n-k'-1)(n-k'+1)} \sum_{i=k'+1}^{\frac{1}{2}(n-1)} \cos^2 \frac{\pi j}{n} \quad (n \text{ odd}),$$

$$\geqslant \frac{16}{(n-k'-1)(n-k'+1)} \sum_{i=k'+1}^{\frac{1}{2}(n-2)} \cos^2 \frac{\pi j}{n} \quad (n \text{ even}).$$

To give some idea of how the distribution of d can vary for different regression vectors we give a short table of the limiting means and variances.

		$k'=1$		$k'=3$		$k'=5$	
		Mean	Variance	Mean	Variance	Mean	Variance
$n=20$	Lower	1·89	0·157	1·65	0·101	1·38	0·048
	Upper	2·11	0·200	2·35	0·249	2·62	0·313
$n=40$	Lower	1·95	0·090	1·84	0·077	1·72	0·063
	Upper	2·05	0·100	2·16	0·111	2·28	0·124
$n=60$	Lower	1·97	0·062	1·89	0·057	1·82	0·051
	Upper	2·03	0·067	2·11	0·071	2·18	0·077

We wish to record our indebtedness to Prof. R. L. Anderson for suggesting this problem to one of us.

REFERENCES

AITKEN, A. C. (1935). *Proc. Roy. Soc. Edinb.* **55**, 42.
ANDERSON, R. L. (1942). *Ann. Math. Statist.* **13**, 1.
ANDERSON, R. L. & ANDERSON, T. W. (1950). *Ann. Math. Statist.* **21**, 59.
ANDERSON, T. W. (1948). *Skand. AktuarTidskr.* **31**, 88.
BARTLETT, M. S. (1934). *Proc. Camb. Phil. Soc.* **30**, 327.
COCHRANE, D. & ORCUTT, G. H. (1949). *J. Amer. Statist. Soc.* **44**, 32.

Courant, R. & Hilbert, D. (1931). *Methoden der Mathematischen Physik.* Julius Springer.
Dixon, W. J. (1944). *Ann. Math. Statist.* **15**, 119.
Fisher, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed. Oliver and Boyd.
Hart, B. I. (1942). *Ann. Math. Statist.* **13**, 207.
Moran, P. A. P. (1950). *Biometrika* **37**, 178.
von Neumann, J. (1941). *Ann. Math. Statist.* **12**, 367.
Pitman, E. J. G. (1937). *Proc. Camb. Phil. Soc.* **33**, 212.
Rubin, H. (1945). *Ann. Math. Statist.* **16**, 211.
Wold, H. (1949). 'On least squares regression with auto-correlated variables and residuals.' (Paper read at the 1949 Conference of the International Statistical Institute.)