A New Test for Autocorrelated Errors in the Linear Regression Model
Author(s): I. I. Berenblut and G. I. Webb
Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 35, No. 1 (1973), pp. 33-50
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/2985124
Accessed: 04/09/2014 20:24

A New Test for Autocorrelated Errors in the Linear Regression Model

By I. I. BERENBLUT and G. I. WEBB
The City University, London

[Received June 1971. Final revision June 1972]

SUMMARY

A test for autocorrelated errors in the linear model is introduced and shown to have, in general, greater power than the Durbin and Watson test for high values of autocorrelation.

Keywords: LINEAR REGRESSION; TESTS FOR AUTOCORRELATED ERRORS

1. INTRODUCTION

LEAST squares estimation of parameters in the general linear model may be highly inefficient in the presence of autocorrelated errors: among others, Cochrane and Orcutt (1949) and Watson (1955) have given examples drawn from economic models. Consequently the problem of testing for autocorrelation has received much attention from statisticians concerned with the analysis of economic time series. In these applications the correlation has been generally assumed to take the form of a stationary first-order autoregressive process in the errors.

There are also many well-known systematic experimental designs in which, because of the limited randomization possible, the efficiency of estimating treatment contrasts depends heavily on the assumption of independence of errors—e.g. E. J. Williams's designs for estimating residual effects (1949), D. R. Cox's designs for estimating treatment effects in the presence of a polynomial trend with time (1952). R. M. Williams (1952), on the other hand, has presented designs for long sequences of treatments applied to one experimental unit so that a postulated stationary autocorrelation can be allowed for in the analysis. The new methods we develop in this paper arose from consideration of the problem of testing the hypothesis of zero autocorrelation in results from experiments designed to be efficient for independent errors; typically these experiments involve relatively short series of observations. In particular, in these circumstances it does not seem realistic to assume stationarity for the autoregressive process.

Anderson (1948) showed that for neither the stationary nor the non-stationary first order autoregressive model does a UMP test exist, even for one-sided alternatives. However, with slight modifications of the respective density functions it is possible to devise tests which are UMP against one-sided alternatives for certain restricted classes of design matrices. Durbin and Watson (1950) considered a close approximation to the density function of the stationary first-order autoregressive process and on this basis derived a suitable test statistic (their d statistic). Unfortunately the distribution of d depends upon the design matrix used; however, they were further able to show that the significance points of the distribution have upper and lower bounds, and later published tables of these bounds (Durbin and Watson, 1951). Koerts and Abrahamse (1968), using results due to Imhof (1961), gave a method for calculating the exact significance points of statistics of the type proposed by Anderson and adopted by Durbin and Watson (i.e. ratios of quadratic forms in normal variables) for any particular design matrix. The amount of calculation involved is not trivial and requires the use of a computer.

Various other tests against the stationary alternative have been suggested, but none of these has been shown to be as powerful in general as the exact Durbin and Watson test. Two of such proposed statistics have the advantage, which the Durbin and Watson statistic lacks, of having distributions independent of the design matrix, so that significance points may be tabulated once and for all. These tests are due to (a) Durbin (1970), where the residuals used in the test are based on estimates of parameters obtained from a derived set of regressors; and (b) Abrahamse and Louter (1971), in which the statistic is based upon a new class of estimators for the disturbance vector. The calculation of the values of these two statistics is much more complicated than the d statistic of Durbin and Watson, and if a computer is to be employed at all it would seem better practice to use it for the exact Durbin and Watson test.

From considerations of the non-stationary model and of the likelihood ratio test we have developed a statistic suitable for testing the hypothesis of zero error autocorrelation. It is shown to possess good power properties—including precisely the optimum property of the Durbin and Watson test—which suggest it might be used effectively for testing against stationary and non-stationary alternatives. The amounts of computation involved in the use of the new test and the exact Durbin and Watson test are about the same. Some practical examples are provided; it will be seen that the test given by the new statistic is more powerful than the Durbin and Watson test for high values of autocorrelation when applied to some standard designs and to economic data treated in the literature. Furthermore, the tables in Durbin and Watson (1951) can be utilized when making a bounds test using the new statistic.

2. Stationary and Non-stationary Error Processes

The general linear model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{2.1}$$

where

$\mathbf{y}$ is the $n \times 1$ vector of observations;

$\mathbf{X}$ is the $n \times k$ design matrix, assumed without loss of generality to be of full rank (k);

$\boldsymbol{\beta}$ is the $k \times 1$ vector of parameters; and

$\mathbf{u}$ is the $n \times 1$ vector of errors and $E(\mathbf{u}) = \mathbf{0}$.

The dispersion matrix is given by

$$D(\mathbf{y}) = E(\mathbf{u}\mathbf{u}') = \sigma^2 \mathbf{V}.$$

Specifically we consider the case where $\mathbf{u}$ has a multivariate normal p.d.f. given by

$$\{(2\pi\sigma^2)^{n/2} \sqrt{|\mathbf{V}|}\}^{-1} \exp\left(-\frac{\mathbf{u}'\mathbf{V}^{-1}\mathbf{u}}{2\sigma^2}\right). \tag{2.2}$$

In the present model $\mathbf{V}$ is chosen to represent the first-order autoregressive process:

$$u_i = \rho u_{i-1} + \epsilon_i \quad (i = 2, 3, \ldots, n),$$

where $u_1 = \epsilon_1$ and ϵ_i is the independent component of error present in the ith observation, i.e. $\mathrm{cov}(\epsilon_i, \epsilon_j) = 0$ $(i \neq j)$.

In the stationary model already mentioned the assumption is made that the variance of each u_i is constant; this amounts to saying that $\mathrm{var}(\epsilon_i) = \sigma^2$ for $i \neq 1$ and $\mathrm{var}(\epsilon_1) = \sigma^2/(1-\rho^2)$. Hence $|\rho| < 1$ is a requirement of this model and a special role is allotted to the first observation, which seems artificial in the case of short sequences. We prefer to make the more reasonable assumption for finite sequences that the independent component of error ϵ_i has constant variance. Hildreth and Lu (1960) also make the latter assumption but, since they are dealing with economic processes, they impose the restriction that $|\rho| < 1$. We do not think there is a valid physical basis for such a restriction and in the discussion below ρ can take any value.

Thus

$$\mathrm{var}(\epsilon_i) = \sigma^2 \quad \text{for all } i \tag{2.3}$$

implies that

$$\mathbf{V} = \mathbf{V}(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1+\rho^2 & \rho+\rho^3 & & \vdots \\ \vdots & & & & \vdots \\ \rho^{n-1} & \cdots & \cdots & \cdots & (1+\rho^2+\ldots+\rho^{2n-2}) \end{bmatrix}$$

and

$$\mathbf{V}^{-1} = \mathbf{V}^{-1}(\rho) = \begin{bmatrix} 1+\rho^2 & -\rho & 0 & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & & & \vdots \\ \vdots & & & & & 0 \\ \vdots & & & & 1+\rho^2 & -\rho \\ 0 & \cdots & \cdots & \cdots & -\rho & 1 \end{bmatrix},$$

where $\mathbf{V}(\rho)$ and $\mathbf{V}^{-1}(\rho)$ are symmetric and positive definite for all ρ.

Representing $\mathbf{V}^{-1}$ by $\mathbf{C}'\mathbf{C}$, where

$$\mathbf{C} = \mathbf{C}(\rho) = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots \\ -\rho & 1 & 0 & & & \vdots \\ 0 & -\rho & 1 & & & \vdots \\ \vdots & & & \cdot & & \vdots \\ \vdots & & & & \cdot & \vdots \\ \vdots & & & & \cdot & \vdots \\ \cdots & \cdots & \cdots & 0 & -\rho & 1 \end{bmatrix},$$

we see that $|\mathbf{V}^{-1}| = |\mathbf{C}|^2 = 1$, and so $|\mathbf{V}| = 1$.

If ρ is known it is possible to transform (2.1) to $\mathbf{Cy} = \mathbf{CX\beta}+\boldsymbol{\epsilon}$, and proceed with the usual least squares analysis.

The maximum likelihood estimate of ρ, denoted by $\hat{\rho}$, may be found by minimization of $(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})$ with respect to ρ, subject to the m.l.e. of $\boldsymbol{\beta}$ given by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{2.4}$$

Note:

$$(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{u}'(\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\,\mathbf{u}. \tag{2.5}$$

When minimized with respect to ρ the expression given in (2.5) is defined as the Generalized Residual Sum of Squares, GRSS. The residual sum of squares after fitting least squares estimates of the parameters $\boldsymbol{\beta}$ is

$$\text{RSS} = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{u}'\mathbf{M}\mathbf{u}, \tag{2.6}$$

where $\mathbf{M} = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

It is of interest to note, in view of well-known practice in economic time series analysis, that for $\rho = 1$ premultiplying (2.1) by $\mathbf{C}$ as indicated above amounts to taking first differenced variables in both the observations and the regression vectors, except by this method the first rows remain unchanged.

3. UMP Tests Based on Least Squares Residuals

Anderson (1948) showed that, for a design matrix in which the columns are linear combinations of k latent vectors of the matrix $\boldsymbol{\Theta}$, the statistic

$$r = \frac{\mathbf{z}'\boldsymbol{\Theta}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$$

($\mathbf{z}$ being the vector of least squares residuals, $\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}$) provides a UMP test against one-sided alternatives when the error structure has density

$$K\exp\left[-(2\sigma^2)^{-1}\{(1+\rho^2)\,\mathbf{u}'\mathbf{u}-2\rho\mathbf{u}'\boldsymbol{\Theta}\mathbf{u}\}\right]. \tag{3.1}$$

The density function associated with the stationary first-order autoregressive process is proportional to

$$\exp\left[-(2\sigma^2)^{-1}\left\{(1+\rho^2)\sum_1^n u_i^2-\rho^2(u_1^2+u_n^2)-2\rho\sum_2^n u_i u_{i-1}\right\}\right], \tag{3.2}$$

for which no UMP test exists (Anderson, 1948). By replacing (3.2) by a close approximation, namely,

$$K_1\exp\left[-(2\sigma^2)^{-1}\left\{(1+\rho^2)\sum_1^n u_i^2-\rho(u_1^2+u_n^2)-2\rho\sum_2^n u_i u_{i-1}\right\}\right] \tag{3.3}$$

and applying Anderson's result, Durbin and Watson obtained their statistic (subscripted for clarity in this context)

$$d_A = \frac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}},$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 2 & -1 & & & & \vdots \\ \vdots & & \cdot & & & & \vdots \\ \vdots & & & \cdot & & & \vdots \\ \vdots & & & & \cdot & & \vdots \\ \vdots & & & & & 2 & -1 \\ 0 & \cdots & \cdots & \cdots & \cdots & -1 & 1 \end{bmatrix}.$$

The test for zero autocorrelation provided by the d_A statistic is UMP against one-sided alternatives defined by (3.3) if the k regression vectors (columns of $\mathbf{X}$) are linear combinations of k latent vectors of $\mathbf{A}$.

The latent vector associated with the zero latent root of $\mathbf{A}$ is the unit vector $\mathbf{1}$, and this must be included in the model as a regression vector (Durbin and Watson, 1971, p. 10).

If we adopt the hypothesis of non-stationary first-order autoregressive errors, we have the density function

$$\{(2\pi\sigma^2)^{n/2}\}^{-1} \exp\left[-(2\sigma^2)^{-1}\left\{(1+\rho^2)\sum_1^n u_i^2 - \rho^2 u_n^2 - 2\rho\sum_2^n u_i u_{i-1}\right\}\right], \tag{3.4}$$

and again no UMP test exists. A close approximation to (3.4) is

$$K_2 \exp\left[-(2\sigma^2)^{-1}\left\{(1+\rho^2)\sum_1^n u_i^2 - \rho u_n^2 - 2\rho\sum_2^n u_i u_{i-1}\right\}\right]. \tag{3.5}$$

Using Anderson's result, a test for the null hypothesis of $\rho = 0$ using the statistic

$$d_B = \frac{\mathbf{z}'\mathbf{B}\mathbf{z}}{\mathbf{z}'\mathbf{z}},$$

where

$$\mathbf{B} = \begin{bmatrix} 2 & -1 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 2 & -1 & & & & \vdots \\ \vdots & & \cdot & & & & \vdots \\ \vdots & & & & & & \vdots \\ \vdots & & & & \cdot & & \vdots \\ \vdots & & & & & 2 & -1 \\ 0 & \cdots & \cdots & \cdots & \cdots & -1 & 1 \end{bmatrix},$$

is UMP against one-sided alternatives given by (3.5) when the k regression vectors are linear combinations of k latent vectors of $\mathbf{B}$. The latent roots of B are given by Bellman (1960, p. 66) as

$$\theta_i = 2 - 2\cos\{(2i-1)\pi/(2n+1)\} \quad (i = 1, 2, \ldots, n). \tag{3.6}$$

It is important to note that $\mathbf{B}$ and $\mathbf{V}^{-1}(1)$ are identical.

4. A TEST FOR AUTOCORRELATION BASED ON MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS

We now introduce an improved alternative to the d_A test which will be seen to possess both the optimum power properties of d_A and d_B.

The likelihood ratio test for the hypothesis $\rho = 0$ against the alternative hypothesis $\rho \neq 0$ is given by considering the distribution of a statistic of the form

$$\frac{(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\hat{\rho})(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})}{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}, \tag{4.1}$$

i.e. the ratio of GRSS to RSS.

This statistic takes values between zero and one and the critical region is in the left-hand tail. Since this is a very awkward expression to deal with, we propose, as a test statistic for positive autocorrelation, the following modification of (4.1) obtained by replacing $\hat{\rho}$ by 1 and hence $\mathbf{V}^{-1}(\hat{\rho})$ by $\mathbf{B}$:

$$g = \frac{(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{B}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})}{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})} = \frac{\mathbf{y}'\{\mathbf{B}-\mathbf{B}\mathbf{X}(\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\}\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}}, \tag{4.2}$$

where

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\mathbf{y}.$$

The test between the hypothesis H_0: $\rho = 0$ and H_1: $\rho = \rho^* > 0$ (density function given by (3.4)) is defined by

$$\omega = (g \leqslant v),$$

where ω is the critical region and v is determined so that $P(g \leqslant v \mid H_0) = \alpha$, the size of the test. For completeness, we state, without proof, the inequality $0 < g < 4$, which is also true for d_A and d_B.

If restrictions on the form of the design matrix are made as in Section 3 for the tests given by d_A and d_B, we find that the power properties of both the test statistics d_B and Durbin and Watson's d_A are inherent in the test given by the g statistic. This is proved by establishing equality of g to the statistics d_B and d_A when the design matrix takes the appropriate form, that is,

(1) $g = d_B$ when the k regression vectors are linear combinations of k latent vectors of $\mathbf{B}$. This follows from the fact (Durbin and Watson, 1971) that in this case the minimum variance estimates coincide with the least squares estimates;

(2) $g = d_A$ when the k regression vectors are linear combinations of k latent vectors of $\mathbf{A}$, including the vector $\mathbf{1}$ associated with the zero latent root (i.e. a constant term is fitted). The proof of this statement is given in Appendix B.

5. CONDITIONS FOR MOST POWERFUL TESTS

5.1 *Special Case of Kadiyala's Statistic*

We are using a simplified likelihood ratio statistic to test the hypothesis H_0: $\rho = 0$ against H_1: $\rho = 1$, i.e.

$$\left.\begin{aligned} H_0&: \mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n), \\ H_1&: \mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{B}^{-1}). \end{aligned}\right\} \tag{5.1}$$

In this form we do not have a test between simple hypotheses, the parameters $\boldsymbol{\beta}$ being unknown.

It will be shown that there exist simple hypotheses (5.4) obtained by transformation of (5.1) for which the g test is most powerful. The general result is due to Kadiyala (1970), although our method of proof is different.

We may choose a $n\times(n-k)$ matrix $\mathbf{Q}$ whose columns are orthonormal (Appendix A) such that, where $\boldsymbol{\xi}=\mathbf{Q}'\mathbf{y}$ and $\mathbf{W}=\mathbf{B}-\mathbf{BX}(\mathbf{X}'\mathbf{BX})^{-1}\mathbf{X}'\mathbf{B}$,

$$E(\boldsymbol{\xi})=\mathbf{0}, \tag{5.2}$$

$$g=\frac{\mathbf{y}'\mathbf{W}\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}}=\frac{\boldsymbol{\xi}'\mathbf{J}\boldsymbol{\xi}}{\boldsymbol{\xi}'\boldsymbol{\xi}}, \tag{5.3}$$

($\mathbf{J}$ being the diagonal matrix of the $n-k$ non-zero latent roots of $\mathbf{W}$) and

$$\left.\begin{aligned} H_0'&:\ \boldsymbol{\xi}\sim N(\mathbf{0},\sigma^2\mathbf{I}_{n-k})\\ H_1'&:\ \boldsymbol{\xi}\sim N(\mathbf{0},\sigma^2\mathbf{J}^{-1}).\end{aligned}\right\} \tag{5.4}$$

Thus H_0 implies H_0' and H_1 implies H_0' but the converses are not true as the transformation given by $\mathbf{Q}$ is singular. Kadiyala (1970) considered an approximation of the type (5.4) for (5.1) and we make use of his results to state that the test given by the statistic g for the hypotheses defined in (5.4) is most powerful (by the Lehmann–Stein result, 1948) and unbiased. Moreover, no restrictions about the form of the design matrix have been made in this approximation. These results do not hold for the d_A and d_E tests, which are not particular cases of Kadiyala's statistic.

5.2. *Invariance Theory Approach*

(*i*) *Non-stationary model*

Making use of equation (7) in Durbin and Watson (1971), it follows that the test given by g for the hypotheses defined in (5.1) is the most powerful invariant test, as the problem remains invariant under transformations of the type

$$\mathbf{y}^*=\gamma_0\mathbf{y}+\mathbf{X}\boldsymbol{\gamma},$$

where $\boldsymbol{\gamma}'=(\gamma_1,\ldots,\gamma_k)$, $-\infty<\gamma_i<\infty$ $(i=1,\ldots,k)$, and $0<|\gamma_0|<\infty$. Furthermore, in the region of $\rho=+1$, the g statistic gives the locally most powerful invariant test for the hypotheses defined by

$$\begin{aligned} H_0&:\ \mathbf{y}\sim N(\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I}_n),\\ H_1&:\ \mathbf{y}\sim N(\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{V}(\rho)).\end{aligned} \tag{5.5}$$

These results, which are true for any design matrix, follow directly from Durbin and Watson's work by considering the density function for the non-stationary process given in (3.4). The algebraic development in this case is to be found in Appendix C.

(*ii*) *Stationary model*

Pursuing the remark made in Section 2 that if we followed the procedure of taking first differences and use as a statistic the ratio of the residual sums of squares with first differenced and original variables, we would obtain $\{\mathbf{y}'(\mathbf{A}\text{-}\mathbf{AXGX}'\mathbf{A})\mathbf{y}\}/\mathbf{y}'\mathbf{My}$ where $\mathbf{G}$ is a generalized inverse of $\mathbf{X}'\mathbf{AX}$. It is worthy of note that this ratio is the same as the g statistic with $\mathbf{B}$ replaced by $\mathbf{A}$. Supposing that an overall mean i included in the regression, it can easily be shown that this statistic gives the locally most powerful invariant test for $\rho \to 1$, in the case of the stationary autoregressive process (3.2). More remarkably, for regression involving an overall mean, this statistic can be shown to be identical with g.

6. PRACTICAL EXAMPLES

In this section we illustrate the power functions of the g and d_A statistics in the cases of some randomly selected standard types of design: three from economic sources (Hildreth and Lu, 1960), two 4×4 block designs and a 5×5 latin square. Figs. 1–6 show plots of the powers of the g test (unbroken line) and the d_A test (broken line) against values of ρ between 0 and 1 in the non-stationary alternative (3.4) (Figs. a); and between 0 and 0.95 in the stationary alternative (3.2) (Figs. b). It will be seen that in all these cases the g test is more powerful than the d_A test for high values of autocorrelation—i.e. when it is most important to reject the null hypothesis if it is false. On the other hand, as claimed by Durbin and Watson (1971), the d_A statistic is seen to have greater power as $\rho \to 0$. There is also an indeterminate region in which no definite advantage is apparent for one test over the other. The six examples are as follows.

Fig. 1: Pears data from Hildreth and Lu (1960, p. 65). Secondary source Henshaw (1966, Table 1). Here $n = 16$ and $k = 5$.

Fig. 2: Plums data from Hildreth and Lu (1960, p. 61). Secondary source is Kadiyala (1970, example 2). Here $n = 23$ and $k = 6$.

Fig. 3: Wheat data from Hildreth and Lu (1960, p. 59). Secondary source is Kadiyala (1970, example 1). Here $n = 18$ and $k = 5$. However, the value we obtain for d_A (1.91) does not agree with the value of 0·96 given by Hildreth and Lu and Kadiyala.

Fig. 4: The 5×5 latin square given by

		Subjects				
		1	2	3	4	5
	1	A	B	C	D	E
	2	B	A	E	C	D
Periods	3	C	D	B	E	A
	4	D	E	A	B	C
	5	E	C	D	A	B.

The parameters fitted in the model are subjects, periods, treatments and an overall mean. Autocorrelation is assumed only within subjects. Here $n = 25$ and $k = 13$.

Fig. 5: The 4×4 block design given by

		Subjects			
		1	2	3	4
	1	A	B	C	D
Periods	2	B	C	D	A
	3	C	D	A	B
	4	D	A	B	C,

in which the parameters to be fitted refer to subjects treatments and overall mean. Autocorrelation is assumed only within subjects. Here $n = 16$ and $k = 7$.

Fig. 6: As Fig. 5 but with the design

$$\begin{array}{cccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} \\ \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{A} \\ \mathrm{D} & \mathrm{A} & \mathrm{B} & \mathrm{C} \\ \mathrm{C} & \mathrm{D} & \mathrm{A} & \mathrm{B}. \end{array}$$

In every case given the size of the test was 0·05. Details of the method used for calculating the significance points and powers are given in Koerts and Abrahamse (1968).

7. BOUNDS FOR g

The statistic g is defined by $\mathbf{y}'\mathbf{W}\mathbf{y}/\mathbf{y}'\mathbf{M}\mathbf{y}$, where $\mathbf{W} = \mathbf{B} - \mathbf{B}\mathbf{X}(\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}$ and $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. In the class of design models including an overall mean, we may partition $\mathbf{X}$ in the form $(\mathbf{1} \vdots \mathbf{X}_0)$, where $\mathbf{1}$ denotes the normalized unit vector. Let the latent vectors of the symmetric matrix $\mathbf{A}$ be represented by the columns of $\mathbf{L} = (\mathbf{1} \vdots \mathbf{L}_0)$ so that $\mathbf{L}'\mathbf{L} = \mathbf{L}\mathbf{L}' = \mathbf{I}_n$ and

$$\mathbf{L}'\mathbf{A}\mathbf{L} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \lambda_1 & & & & & \vdots \\ \vdots & & \lambda_2 & & & & \vdots \\ \vdots & & & \cdot & & & \vdots \\ \vdots & & & & \cdot & & \vdots \\ \vdots & & & & & \cdot & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \lambda_{n-1} \end{bmatrix} = \mathbf{\Lambda} = \left[\begin{array}{c:c} 0 & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{\Lambda}_{n-1} \end{array}\right];$$

$\lambda_i = 2 - 2\cos(i\pi/n)$, as proved by von Neumann (1941). Conversely

$$\mathbf{A} = \mathbf{L}\mathbf{\Lambda}\mathbf{L}' = \mathbf{L}_0\mathbf{\Lambda}_{n-1}\mathbf{L}_0'.$$

Writing $\mathbf{X} = \mathbf{S}\mathbf{G}$, where $\mathbf{S} = (\mathbf{1} \vdots \mathbf{S}_0)$ is an orthonormal matrix and $\mathbf{G}$ is a $k \times k$ non-singular matrix, then $\mathbf{W} = \mathbf{B} - \mathbf{B}\mathbf{S}(\mathbf{S}'\mathbf{B}\mathbf{S})^{-1}\mathbf{S}'\mathbf{B}$ and $\mathbf{M} = \mathbf{I} - \mathbf{S}\mathbf{S}'$. Let $\mathbf{U} = (\mathbf{S} \vdots \mathbf{T})$ where $\mathbf{T}$ is the orthonormal complement of $\mathbf{S}$ so that $\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}_n$. Partitioning $(\mathbf{S}'\mathbf{B}\mathbf{S})^{-1}$ as

$$\left[\begin{array}{c:c} \mathbf{1}'\mathbf{B}\mathbf{1} & \mathbf{1}'\mathbf{B}\mathbf{S}_0 \\ \hdashline \mathbf{S}_0'\mathbf{B}\mathbf{1} & \mathbf{S}_0'\mathbf{B}\mathbf{S}_0 \end{array}\right]^{-1},$$

expressing $\mathbf{B}$ in terms of $\mathbf{A}$ as in (2.1), and writing

$$\left[\begin{array}{c:c} \mathbf{A}_1 & \mathbf{B}_1 \\ \hdashline \mathbf{B}_1' & \mathbf{C}_1 \end{array}\right]^{-1} \text{ in the form } \left[\begin{array}{c:c} \mathbf{A}_2 & \mathbf{B}_2 \\ \hdashline \mathbf{B}_2' & \mathbf{C}_2 \end{array}\right],$$

where $\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2$, for example, are given in Rao (1965a, p. 29), it may be verified that $\mathbf{W} = \mathbf{A} - \mathbf{A}\mathbf{S}_0(\mathbf{S}_0'\mathbf{A}\mathbf{S}_0)^{-1}\mathbf{S}_0'\mathbf{A}$.

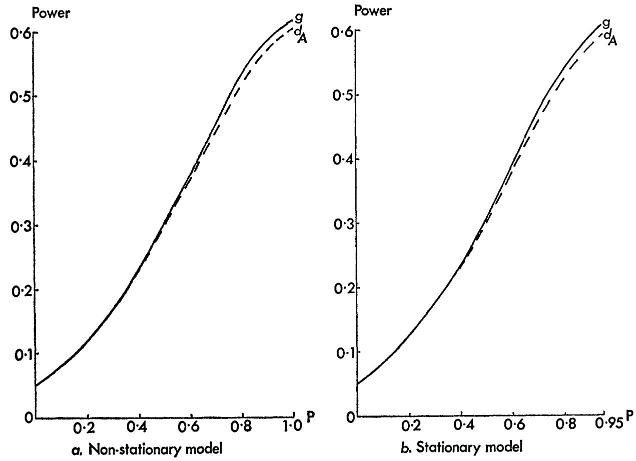

FIG. 1. Pears data (Hildreth and Lu).

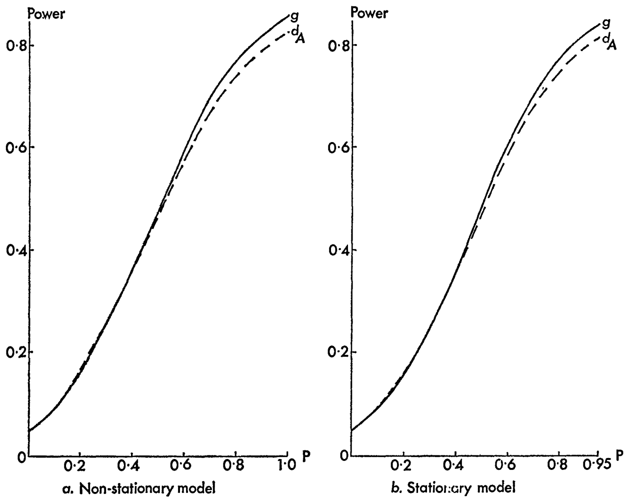

FIG. 2. Plums data (Hildreth and Lu).

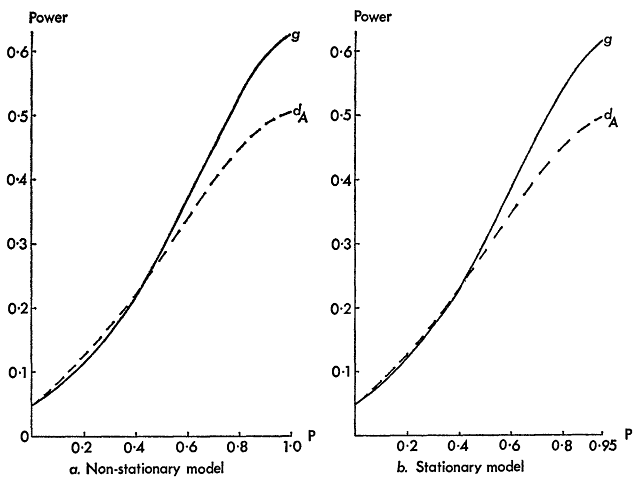

FIG. 3. Wheat data (Hildreth and Lu).

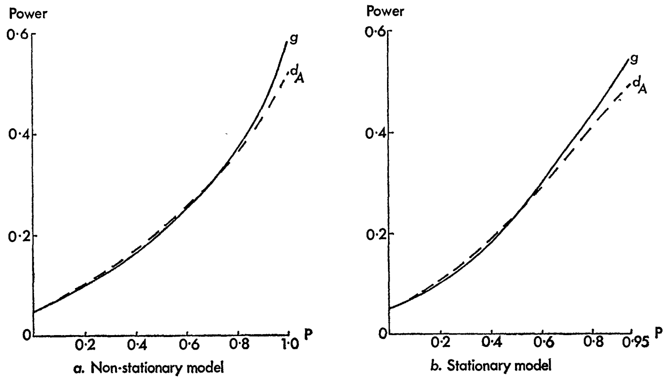

FIG. 4. 5 × 5 Latin square.

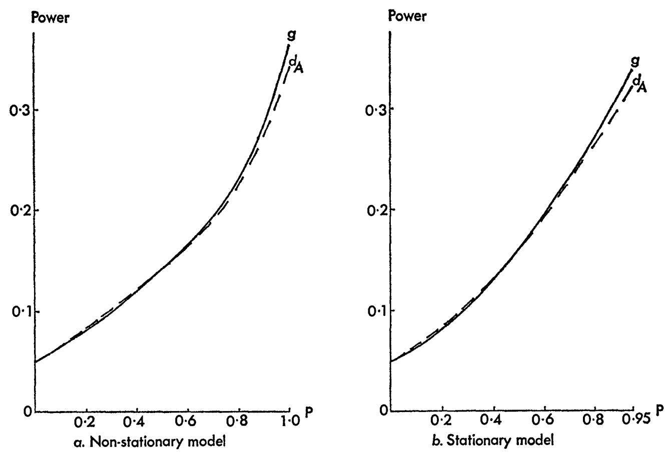

FIG. 5. First 4 × 4 block design.

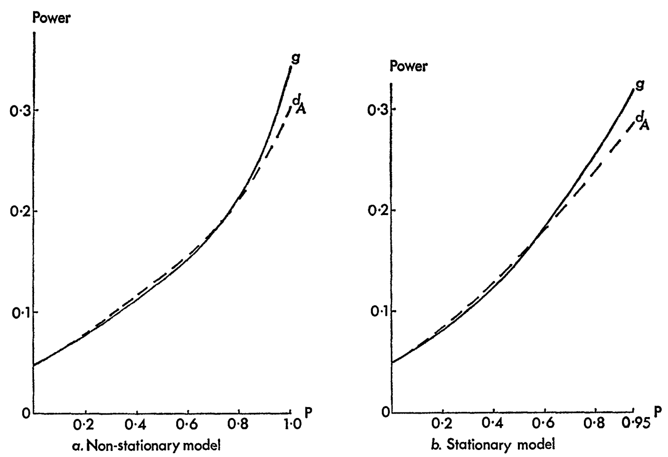

FIG. 6. Second 4 × 4 block design.

Under the null hypothesis of zero autocorrelation the statistic g can be reduced to the form

$$\sum_{i=1}^{n-k} \nu_i \xi_i^2 \Big/ \sum_{i=1}^{n-k} \xi_i^2 \quad \text{(see Appendix A)},$$

where $\nu_i\ (i = 1, 2, \ldots, n-k)$ are the latent roots of $\mathbf{T'WT}$ arranged in order of ascending magnitude and the ξ_i's are distributed independently $N(0, 1)$. Since

$$d_L = \sum_{i=1}^{n-k} \lambda_i \xi_i^2 \Big/ \sum_{i=1}^{n-k} \xi_i^2 \quad \text{and} \quad d_U = \sum_{i=1}^{n-k} \lambda_{i+k-1} \xi_i^2 \Big/ \sum_{i=1}^{n-k} \xi_i^2,$$

where the λ_i are the positive latent roots of $\mathbf{A}$ arranged in order of ascending magnitude, given the bounds of the d statistic, it is necessary and sufficient to prove that

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

for g to have those same bounds.

First, replace $\mathbf{A}$ by $\mathbf{L}_0 \mathbf{\Lambda}_{n-1} \mathbf{L}_0'$; it follows that $\mathbf{T'WT}$ can be written as

$$\mathbf{T'L}_0 \mathbf{\Lambda}_{n-1} \mathbf{L}_0' \mathbf{T} - \mathbf{T'L}_0 \mathbf{\Lambda}_{n-1} \mathbf{L}_0' \mathbf{S}_0 (\mathbf{S}_0' \mathbf{L}_0 \mathbf{\Lambda}_{n-1} \mathbf{L}_0' \mathbf{S}_0)^{-1} \mathbf{S}_0' \mathbf{L}_0 \mathbf{\Lambda}_{n-1} \mathbf{L}_0' \mathbf{T}.$$

Now substitute $\mathbf{S}_0' \mathbf{L}_0 - \mathbf{S}_0' \mathbf{L}_0 \mathbf{\Lambda}_{n-1}^{-1} \mathbf{L}_0' \mathbf{T} (\mathbf{T'L}_0 \mathbf{\Lambda}_{n-1}^{-1} \mathbf{L}_0' \mathbf{T})^{-1} \mathbf{T'L}_0$ for $(\mathbf{S}_0' \mathbf{L}_0 \mathbf{\Lambda}_{n-1} \mathbf{L}_0' \mathbf{S}_0)^{-1} \mathbf{S}_0' \mathbf{L}_0 \mathbf{\Lambda}_{n-1}$, by use of Lemma 2b in Rao (1965b, p. 358), noting that $\mathbf{S}_0' \mathbf{L}_0 \mathbf{L}_0' \mathbf{T} = \mathbf{S}_0'(\mathbf{I} - \mathbf{11}')\mathbf{T} = \mathbf{0}, \mathbf{S}_0' \mathbf{L}_0 \mathbf{L}_0' \mathbf{S}_0 = \mathbf{I}_k$, $\mathbf{T'L}_0 \mathbf{L}_0' \mathbf{T} = \mathbf{I}_{n-k}$, $\mathbf{L}_0' \mathbf{T}$ is of full rank $n-k$, $\mathbf{L}_0' \mathbf{S}_0$ is of full rank k, and finally $\mathbf{L}_0' \mathbf{S}_0 \mathbf{S}_0' \mathbf{L}_0 = \mathbf{I}_{n-1} - \mathbf{L}_0' \mathbf{TT'L}_0$. We thus find that $\mathbf{T'WT} = (\mathbf{T'L}_0 \mathbf{\Lambda}_{n-1}^{-1} \mathbf{L}_0' \mathbf{T})^{-1}$. The relationship between the latent roots of $\mathbf{T'L}_0 \mathbf{\Lambda}_{n-1}^{-1} \mathbf{L}_0' \mathbf{T}$ and those of $\mathbf{\Lambda}_{n-1}^{-1}$ can be inferred from the inequalities found by Durbin and Watson (1950) concerning the latent roots of $\mathbf{MA}$ and $\mathbf{A}$. Since $\mathbf{M} = \mathbf{TT}'$ the positive latent roots of $\mathbf{MA}$ are the latent roots of $\mathbf{T'AT}$. Therefore, substituting for $\mathbf{A}$ by $\mathbf{\Lambda}_{n-1}^{-1}$ and for $\mathbf{T}$ by $\mathbf{L}_0' \mathbf{T}$ their result implies that $1/\lambda_i \geqslant 1/\nu_i \geqslant 1/\lambda_{i+k-1}$, $(i = 1, 2, \ldots, n-k)$. Thus the upper and lower bounds for the significance points of g are precisely those of the d_A statistic, tables of which have been published in Durbin and Watson (1951). These tables are therefore directly applicable to bounds tests employing the g statistic in place of the d_A statistic.

When an overall mean is not included as a parameter in the regression scheme it can be shown along similar lines that the bounding inequality of the g statistic under the null hypothesis is $g_L \leqslant g \leqslant g_u$, where

$$g_L = \sum_{i=1}^{n-k} \theta_i \xi_i^2 \Big/ \sum_{i=1}^{n-k} \xi_i^2, \quad g_U = \sum_{i=1}^{n-k} \theta_{i+k} \xi_i^2 \Big/ \sum_{i=1}^{n-k} \xi_i^2,$$

and the θ_i are the latent roots of $\mathbf{B}$, as given in (3.6). Clearly $g_L < d_L < d_U < g_U$.

8. Testing for Negative Autocorrelation and Two-tail Tests

Hildreth and Lu (1960) have shown that even in economic data negative autocorrelation is not an uncommon occurrence. If a test against the one-sided alternative of negative autocorrelation is required a statistic g_n can be developed along analogous lines to g and is found to possess similar properties. We define

$$g_n = \frac{\mathbf{y}'(\mathbf{B}_1 - \mathbf{B}_1 \mathbf{X}(\mathbf{X}'\mathbf{B}_1 \mathbf{X})^{-1} \mathbf{X}'\mathbf{B}_1)\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}},$$

where $\mathbf{B}_1 = \mathbf{V}^{-1}(-1)$. In particular the test is locally most powerful invariant in the region of $\rho = -1$.

However, it is likely that a composite (i.e. two-tail) test against the null hypothesis of zero autocorrelation will have more application in practice. A suitable choice of test statistic then would be g, where the critical region is defined by

$$\omega_1 = \{g \leqslant V_1\}, \quad \omega_2 = \{g \geqslant V_2\}$$

and $P(g \leqslant V_1 | H_0) = P(g \geqslant V_2 | H_0) = \alpha/2$ (i.e. equal tails). A rejection in ω_1 indicates positive autocorrelation and in ω_2 negative autocorrelation. Durbin and Watson similarly extended their d_A test. In general no special power properties can be claimed for either two-sided test.

9. Test Procedures using the g-Statistic

In general the numerator of g may be evaluated easily as the residual sum of squares of transformed dependent variables, $\mathbf{Cy}$, regressed on the correspondingly transformed design matrix, $\mathbf{CX}$, where

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 1 & 0 & & & \vdots \\ 0 & -1 & 1 & & & \vdots \\ \vdots & & \cdot & & & \vdots \\ \vdots & & & \cdot & & \vdots \\ \vdots & & & & \cdot & 0 \\ 0 & & & & -1 & 1 \end{bmatrix};$$

i.e. the residual sum of squares of first differenced variables with y_1 and X_{1j} $(j = 1, \ldots, k)$ remaining unchanged.

In the particular case where $\mathbf{X}$ contains the column $\mathbf{1}$ the calculation can be simplified further by ignoring y_1 and X_{1j} $(j = 1, \ldots, k)$ and computing the ratio of the residual sum of squares using the $n-1$ first differenced variables and $k-1$ parameters to the residual sum of squares for the original variables.

The significance points for the statistics d_L and d_U are tabulated in Durbin and Watson (1951). When an overall mean is included as a parameter in the regression scheme we have shown that the corresponding significance points of g must lie between these tabulated values. Thus when the calculated value of g is less than the critical value of d_L the null hypothesis of independent errors is rejected in favour of the alternative of positive autocorrelation. If g is greater than the critical value d_U the alternative of positively autocorrelated errors is rejected. An observed g lying between the tabulated significance points of d_L and d_U results in an inconclusive test. In this case the critical value of g itself may be evaluated using the methods reviewed in Durbin and Watson (1971).

If a test against negative autocorrelation or a two-sided test is required the bounds for the significance points in the right-hand tail are found by subtracting from 4 the tabulated values for the left-hand tail. For example, if we obtain a value of g greater than 4-d_L we should reject the null hypothesis in favour of the alternative that the errors are negatively autocorrelated.

10. CONCLUSION

Listing the power properties of the g test derived in this paper we have:

(1) the g statistic provides a UMP test of the hypothesis H_1: $\rho = \rho^* > 0$ against H_0: $\rho = 0$ (density function (3.5)) when the k regression vectors are linear combinations of k latent vectors of $\mathbf{B}$;

(2) the g statistic provides a UMP test of the hypothesis H_1: $\rho = \rho^* > 0$ against H_0: $\rho = 0$, with density function (3.3), when the k regression vectors are linear combinations of k latent vectors of A (including the latent vector $\mathbf{1}$);

(3) the g statistic provides a MP test of the transformed hypotheses (5.4), independently of the design matrix;

(4) the g statistic provides the locally most powerful invariant test in the region of $\rho = +1$ for the hypotheses (5.5), independently of the design matrix;

(5) the g statistic provides the locally most powerful invariant test against the stationary alternative as $\rho \to 1$, provided that an overall mean is included as a parameter in the regression.

The theory has been illustrated with the use of some practical examples and it is seen that the power function characteristics of g (and d_A) are similar for stationary and non-stationary alternatives, and that g has greater power than d_A for higher values of autocorrelation in both alternatives, i.e. when rejection of the null hypothesis if it is false is most important.

Since the testing for autocorrelation is a precaution against obtaining least squares estimates of parameters which have lower than optimum efficiency, the most important contingency against which we have to guard is the acceptance of the hypothesis of independent errors when in fact they are strongly autocorrelated. Thus we conclude, from power considerations, the g test is preferable to other tests of autocorrelation for both types of autoregressive model (stationary and non-stationary) which we have consideıed in this paper.

ACKNOWLEDGEMENT

The authors are grateful for the suggestions made by the referee which improved the presentation of this paper. G. I. Webb is a recipient of a Science Research Council Research Studentship.

REFERENCES

ABRAHAMSE, A. P. J. and LOUTER, A. S. (1971). On a new test for autocorrelation in least squares regression. *Biometrika*, **58**, 53–60.

ANDERSON, T. W. (1948). On the theory of testing serial correlation. *Skand. Akt.*, **31**, 88–116.

BELLMAN, R. (1960). *Introduction to Matrix Analysis*. New York: McGraw-Hill.

COCHRANE, D., and ORCUTT, G. H. (1949). Applications of least squares regressions to relations containing autocorrelated error terms. *J. Amer. Statist. Ass.*, **44**, 32–00.

COX, D. R. (1952). Some recent work on systematic experimental designs. *J. R. Statist. Soc.* B, **14**, 211–219.

DURBIN, J. (1970). An alternative to the bounds test for testing for serial correlation in least squares regression. *Econometrics*, **38**, 422–429.

DURBIN, J. and WATSON, G. B. (1950). Testing for serial correlation in least squares regression, I. *Biometrika*, **37**, 409–428.

—— (1951). Testing for serial correlation in least squares regression, II. *Biometrika*, **38**, 159–178.

—— (1971). Testing for serial correlation in least squares regression, III. *Biometrika*, **58**, 1–19.

HENSHAW, R. C. (1966). Testing single-equation least squares regression models for autocorrelated disturbances. *Econometrica*, **34**, 646–660.

HILDRETH, C. and LU, J. Y. (1960). Demand relations with autocorrelated disturbances. Technical Bulletin No. 276, Agricultural Experiment Station, Michigan State University.
IMHOF, P. J. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 419–426.
KADIYALA, K. R. (1970). Testing for the independence of regression disturbances. *Econometrica*, **38**, 97–117.
KOERTS, J. and ABRAHAMSE, A. P. J. (1968). On the power of the BLUS procedure. *J. Amer. Statist. Ass.*, **63**, 1226–1236.
LEHMANN, E. L. and STEIN, C. (1948). Most powerful tests of composite hypotheses. I. Normal distributions. *Ann. Math. Statist.*, **19**, 495–516.
VON NEUMANN, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **12**, 367–395.
RAO, C. R. (1965a). *Linear Statistical Inference and its Applications*. New York: Wiley.
—— (1965b). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Fifth Berkeley Symposium.
WATSON, G. S. (1955). Serial correlation in regression analysis, I. *Biometrika*, **42**, 327–341.
WILLIAMS, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Austral. J. Sci. Res.* A, **2**, 149–168.
WILLIAMS, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika*, **39**, 151–167.

APPENDIX A

Following Durbin and Watson (1950, p. 413), we may write the design matrix $\mathbf{X}$ in the form $\mathbf{SG}$ where $\mathbf{S}$ is an $n \times k$ orthonormal matrix and $\mathbf{G}$ is a $k \times k$ non-singular matrix. Defining $\mathbf{U} = (\mathbf{S} \vdots \mathbf{T})$, where

$$\left.\begin{aligned} \mathbf{U'U} = \mathbf{UU'} = \mathbf{I}_n \\ \mathbf{S'T} = \mathbf{0} \end{aligned}\right\} \tag{A.1}$$

then

$$\mathbf{W} = \mathbf{B} - \mathbf{BX}(\mathbf{X'BX})^{-1}\mathbf{X'B} = \mathbf{B} - \mathbf{BS}(\mathbf{S'BS})^{-1}\mathbf{S'B} \tag{A.2}$$

and

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'} = \mathbf{I} - \mathbf{SS'} = \mathbf{TT'}. \tag{A.3}$$

Now, from Rao (1965, p. 358)

$$(\mathbf{S'BS})^{-1}\mathbf{S'B} = (\mathbf{S'S})^{-1}\mathbf{S'} - (\mathbf{S'S})^{-1}\mathbf{S'B}^{-1}\mathbf{T}(\mathbf{T'B}^{-1}\mathbf{T})^{-1}\mathbf{T'},$$

that is,

$$(\mathbf{S'BS})^{-1}\mathbf{S'B} = \mathbf{S'} - \mathbf{S'B}^{-1}\mathbf{T}(\mathbf{T'B}^{-1}\mathbf{T})^{-1}\mathbf{T'}. \tag{A.4}$$

It will be shown possible (A.6) to diagonalize $\mathbf{W}$ and $\mathbf{M}$ simultaneously using an orthonormal transformation. First consider

$$\mathbf{U'WU} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{T'WT} \end{array}\right]$$

and

$$\mathbf{U'MU} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{n-k} \end{array}\right].$$

From (A.2) and (A.4)

$$\begin{aligned}\mathbf{T'WT} &= \mathbf{T'BT}+\mathbf{T'BSS'B^{-1}T(T'B^{-1}T)^{-1}}\\ &= \mathbf{T'BT}+\mathbf{T'BB^{-1}T(T'B^{-1}T)^{-1}}-\mathbf{T'BTT'B^{-1}T(T'B^{-1}T)^{-1}}\\ &= \mathbf{(T'B^{-1}T)^{-1}}.\end{aligned} \tag{A.5}$$

Now let

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{I}_k & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P}_1 \end{array}\right],$$

where $\mathbf{P}_1$ is the orthonormal matrix of latent vectors of the symmetric matrix $(\mathbf{T'B^{-1}T})^{-1}$. Thus $\mathbf{P}_1'\mathbf{P}_1 = \mathbf{I}_{n-k}$ and $\mathbf{P}_1'(\mathbf{T'B^{-1}T})^{-1}\mathbf{P}_1 = \mathbf{J}$ say, the diagonal matrix of latent roots of $(\mathbf{T'B^{-1}T})^{-1}$, which are also the non-zero latent roots of $\mathbf{W}$. Using the orthonormal matrix $\mathbf{UP}$ then

$$\left.\begin{aligned}\mathbf{P'U'WUP} &= \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{J} \end{array}\right],\\ \mathbf{P'U'MUP} &= \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{n-k} \end{array}\right]\end{aligned}\right\} \tag{A.6}$$

and

$$g = \mathbf{y'Wy/y'My} = \mathbf{y'UP}\left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{J} \end{array}\right]\mathbf{P'U'y} \Big/ \mathbf{y'UP}\left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{n-k} \end{array}\right]\mathbf{P'U'y}.$$

Now

$$\mathbf{UP} = (\mathbf{S}\,\vdots\,\mathbf{T})\left[\begin{array}{c|c} \mathbf{I}_k & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P}_1 \end{array}\right] = (\mathbf{S}\,\vdots\,\mathbf{TP}_1).$$

Therefore, where $\mathbf{Q} = \mathbf{TP}_1$, g becomes $\mathbf{y'QJQ'y}/(\mathbf{y'QQ'y})$. The mean and dispersion matrix of $\mathbf{Q'y}$ are given by

$$E(\mathbf{Q'y}) = \mathbf{P'T'SG}\boldsymbol{\beta} = \mathbf{0}$$

and

$$D(\mathbf{Q'y}) = \sigma^2\mathbf{Q'B^{-1}Q} = \sigma^2\mathbf{P}_1'\mathbf{T'B^{-1}TP}_1 = \sigma^2\mathbf{J}^{-1}.$$

APPENDIX B

Choosing $\mathbf{U}, \mathbf{S}, \mathbf{T}$ as defined generally in Appendix A so that the columns of $\mathbf{U}$ are the normalized latent vectors of $\mathbf{A}$, then $\mathbf{U'AU} = \boldsymbol{\Lambda}_n$, the diagonal matrix of latent roots of $\mathbf{A}$, and

$$\mathbf{AT} = \mathbf{T}\boldsymbol{\Lambda}_{n-k}, \tag{B.1}$$

where $\boldsymbol{\Lambda}_{n-k}$ is the diagonal matrix of those latent roots of $\mathbf{A}$ associated with the latent vectors $\mathbf{T}$. It is easily verified that $\mathbf{B^{-1}A} = \mathbf{I}-(\mathbf{1}\,\vdots\,\mathbf{0})$.

Now $g = \mathbf{y}'\mathbf{W}\mathbf{y}/(\mathbf{y}'\mathbf{M}\mathbf{y})$ and $d_A = \mathbf{y}'\mathbf{MAM}\mathbf{y}/(\mathbf{y}'\mathbf{M}\mathbf{y})$. Therefore, to prove equality it is sufficient to prove that $\mathbf{W} = \mathbf{MAM}$ under the given conditions. From (A.3)

$$\mathbf{MAM} = \mathbf{TT}'\mathbf{ATT}' = \mathbf{T}\mathbf{\Lambda}_{n-k}\mathbf{T}',$$

and so

$$\mathbf{U}'\mathbf{MAMU} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Lambda}_{n-k} \end{array}\right].$$

Thus from (A.5) we are left to prove that

$$(\mathbf{T}'\mathbf{B}^{-1}\mathbf{T})^{-1} = \mathbf{\Lambda}_{n-k}.$$

The matrix $\mathbf{\Lambda}_{n-k}$ does not contain the zero latent root and is therefore non-singular. From (B.1)

$$\mathbf{T} = \mathbf{AT}\mathbf{\Lambda}_{n-k}^{-1}$$

and

$$\begin{aligned} \mathbf{T}'\mathbf{B}^{-1}\mathbf{T} &= \mathbf{T}'\mathbf{B}^{-1}\mathbf{AT}\mathbf{\Lambda}_{n-k}^{-1} \\ &= \{\mathbf{T}'\mathbf{T} - \mathbf{T}'(\mathbf{1} \vdots \mathbf{0})\mathbf{T}\}\mathbf{\Lambda}_{n-k}^{-1} \\ &= \mathbf{\Lambda}_{n-k}^{-1}, \end{aligned}$$

as $\mathbf{T}'\mathbf{1} = \mathbf{0}$, $\mathbf{1}$ being a column of $\mathbf{S}$. Thus

$$(\mathbf{T}'\mathbf{B}^{-1}\mathbf{T})^{-1} = \mathbf{\Lambda}_{n-k}$$

and equality of g and d_A is proved when the regression vectors are linear combinations of k latent vectors of $\mathbf{A}$, including the vector $\mathbf{1}$.

APPENDIX C

Here we consider the invariance theory approach used by Durbin and Watson (1971) and apply it to the non-stationary process (3.4) and the test of the hypotheses defined by

$$H_0: \mathbf{y} \sim N(\mathbf{X\beta}, \sigma^2\mathbf{I}_n)$$

and

$$H_1: \mathbf{y} \sim N(\mathbf{X\beta}, \sigma^2\mathbf{V}(\rho^*)),$$

that is,

$$H_0: \rho = 0 \quad \text{and} \quad H_1: \rho = \rho^* > 0.$$

Proceeding along similar lines to Durbin and Watson (1971, pp. 9 and 10) (but remembering that, under H_0, $\mathbf{B}$ is replaced by $\mathbf{I}$ and $\hat{\mathbf{u}}$ by $\mathbf{z}$) it may be easily verified that the most powerful invariant test is given by the rejection region

$$g(\rho^*) = \frac{\hat{\mathbf{u}}'\mathbf{V}^{-1}(\rho^*)\hat{\mathbf{u}}}{\mathbf{z}'\mathbf{z}} < C_1,$$

where

$$\hat{\mathbf{u}} = [\mathbf{I} - \mathbf{X}\{\mathbf{X}'\mathbf{V}^{-1}(\rho^*)\mathbf{X}\}^{-1}\mathbf{X}'\mathbf{V}^{-1}(\rho^*)]\mathbf{y}.$$

Thus for the hypothesis (5.1), i.e. $\rho^* = 1$, g provides the most powerful invariant test.

Examining the behaviour of $g(\rho^*)$ as $\rho^* \to 1$ we see that $\mathbf{V}^{-1}(\rho^*) \to \mathbf{B}$, so that $g(\rho^*) \to g$, and hence the g test is the most powerful invariant test in the region of $\rho = 1$.

For completeness a study has been made into the power functions of tests such as that given by g (0.7). The test given by g (0·7) is (in the practical cases considered) slightly more powerful than g in the region of $\rho = 0{\cdot}7$ (as would be expected), but this is offset by lower power in the region of $\rho = 1$. If a test is required against an *a priori* value of ρ, say ρ^* (or even approximately ρ^*), then $g(\rho^*)$ has the property of providing the most powerful invariant test in the region of ρ^*. However, if there is no obvious value of ρ^*, i.e. when testing against the composite hypothesis of positive autocorrelation, $\rho = \rho^* > 0$, $g\,[= g(1)]$ has the advantageous properties (for certain classes of design matrices) indicated in Section 4.

Corrigendum

Sampling Moments of Moments Associated with Univariate Distributions

By L. R. SHENTON, K. O. BOWMAN and D. SHEEHAN

J. R. Statist. Soc. B, **33**, 444–457

On p. 450, for (22) read

$$\mu_s = \sum_{x=0}^{n} (x-np)^s \binom{n}{x} p^x (1-p)^{n-x}.$$

On p. 453, for the coefficients of N^{-8} and N^{-9} in (31b) read respectively $-1{,}972{,}877{,}770{,}080$ and $272{,}866{,}864{,}838{,}\overline{360}$.