THESES SIS/LIBRARY
R.G. MENZIES LIBRARY BUILDING NO:2
THE AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200 AUSTRALIA

TELEPHONE: +61 2 6125 4631
FACSIMILE: +61 2 6125 4063
EMAIL: library.theses@anu.edu.au

USE OF THESES

THE THEORY AND APPLICATION OF STOCHASTIC PROCESSES

by E.J. HANNAN

A thesis submitted to
The Australian National University
for the degree of
Doctor of Philosophy in Statistics

1955

PREFACE

The work presented in this thesis was done during the period October, 1953 to July, 1955.

The work is original in the sense that it was wholly done by myself. However, in the course of presenting the results it has been necessary to refer to the work of other men. Where this has been done it has always been made clear by means of a reference.

Chapter I contains no original research but is rather a review of the existing state of the subject.

In # 2.2. a theorem due to Pitman is proved so that the only original research in Chapter 2 is contained in # 2.3. where Pitman's theorem is extended.

Chapter 3 is based on a paper by Ogawara, but most of the chapter is original and consists either of extensions of Ogawara's results or of an examination of their efficiency.

Chapters 4, 5 and 6 present original results obtained by the author, while Chapter 7, where a theory due to Dr. E.G. Bowen is tested, is also wholly original.

Most of the contents of Chapters 3 and 4 were published by the author in a paper in Biometrika, Volume 42, Parts 1 and 2. Most of the contents of

Chapter 5 will appear under the author's name in Biometrika, Volume 42, Parts 3 and 4. The material making up Chapter 7 appeared as an article by the author in The Australian Journal of Physics, Volume 8, No. 2. Most of the subject matter of Chapter 6 has been communicated (by Professor P.A.P. Moran) for publication under the author's name in The Proceedings of the Cambridge Philosophical Society.

I wish to give my most sincere thanks to Professor P.A.P. Moran for suggesting the subject of this thesis to me and for his advice in the research. I have also received helpful advice from the referees of my papers, who are, however, unknown to me.

A great part of the computations required for the work presented in Chapter 7 were done for me on the C.S.&.I.R.O. electronic computer, through the help of Mr. Pearcey and Mr. Beard.

SUMMARY

THEORY AND APPLICATION OF STOCHASTIC PROCESSES

PART I

EXACT TESTS FOR CORRELATION IN TIME SERIES

Chapter I. Testing for correlation in time series.

The second section of this chapter reviews the work which has been done, mainly in recent years, on the problem of testing for serial correlation in a stationary time series. In # 1.3. the particular problem of testing for serial correlation in the residuals from a least squares regression is discussed. Finally, in # 1.4. the correlation of time series is considered.

Chapter 2. The asymptotic efficiency of tests.

If t_1 and t_2 are statistics computed from samples of size n, their asymptotic relative efficiency as test statistics of an hypothesis specified by a parameter value θ_0 is defined as:

$$E(t_1, t_2 \mid \theta_0) = \lim_{n\to\infty} \left\{ \frac{[\partial/\partial\theta\, \mathcal{E}(t_1)]^2}{Var(t_1)} \cdot \frac{Var(t_2)}{[\partial/\partial\theta\, \mathcal{E}(t_2)]^2} \right\}_{\theta=\theta_0}$$

Here $\mathcal{E}(t_j)$ and $Var(t_j)$ are respectively the mean and variance of the t_j. The justification for the use of this criterion rests upon a theorem due to Pitman which

is presented in # 2.2. Pitman shows that under certain regularity conditions for t_1 and t_2 with limiting normal distributions and variances or order n^{-1} the ratio of the sample sizes required for t_1 and t_2 to have the same power against alternative values of θ which differ from θ_0 by quantities of order $n^{-\frac{1}{2}}$ is in the limit given by $E(t_1, t_2 / \theta_0)$

Pitman's theorem is extended in # 2.3. to enable the asymptotic powers ~~to~~ of two variates distributed as χ^2 with p degrees of freedom to be compared. This covers also the case where it is required to compare the asymptotic powers of two tests based on multiple correlations computed from regressions on the same number of regressors.

Chapter 3. An exact test for serial correlation.

An exact test for serial correlation was derived by Ogawara by considering the conditional distribution of the x_{2t} for fixed x_{2t+1} when x_t comes from a stationary simple Markoff process. The asymptotic efficiency of this test is shown to be unity. When the hypothetical serial correlation ρ is not zero the procedure can be modified so as to obtain another exact test whose asymptotic efficiency is $(1-\rho^2)$. The procedure was also extended by Ogawara to the case of a stationary multiple Markoff process but here it is shown that the asymptotic efficiency of the estimates of the autoregression coefficients never rises above $\frac{2}{h+1}$, where h is the order of the autoregressive process (multiple Markoff process).

Chapter 4. Testing for serial correlation in the Residuals from a regression.

Ogawara's method is further extended to obtain an exact test for serial correlation in the residuals from a least squares regression. The test is shown to be asymptotically fully efficient.

Chapter 5. Exact tests for correlation between time series.

Two principal models are considered in this chapter. In the first, two processes are considered as correlated by means of a regression relation, the residual being a simple Markoff process. In the second, two linear processes, of which at least one is Markovian, are correlated by means of a correlation between the shocks, of which they are (infinite) moving averages. In each case the procedure of Chapter 4 leads to an exact test and in each case this test is for a range of circumstances of some interest in practice, asymptotically more efficient than the ordinary correlation coefficient. However, a statistic due to Quenouille is shown to be asymptotically fully efficient when the alternatives are prescribed by the second model and of fairly high efficiency when the first model is appropriate. This statistic is the partial correlation between the two series with the effects of lagged values of both removed. Though the exact distribution of Quenouille's statistic is not known it has been shown, by Quenouille, by sampling experiments, that its large sample theoretical mean and variance on the null hypothesis are also appropriate for small samples so that its treatment as an ordinary correlation will probably give a good approximation to the true significance point.

Chapter 6. Exact tests for correlation in vector processes.

The considerations of Chapters 3, 4 and 5 are extended to processes whose values at each point of time constitute a vector. The methods of correlation analysis are now replaced by those of canonical correlations. It is shown that when the order of the vector is greater than unity the test for serial independence of a vector process (including the case of a vector of residuals) obtained by Ogawara's method is no longer as powerful (asymptotically) as that which would be obtained by a straightforward canonical analysis of the relation between the vector and its lagged values. The loss of power arises from the necessity of introducing additional parameters, which follows from the (general) lack of commutativity of the matrix of regression coefficients and the covariance matrix of the residuals. Of course, in the case of the straightforward canonical analysis the distribution of the test statistic would only be known asymptotically.

A particular case of a vector process can be obtained by a redefinition of an autoregressive process. The approach of the present chapter gives no improvement (with respect to the power of the test) on that of Chapter 3 and, in general, the case of a simple Markoff process appears to stand in a unique position in relation to Ogawara's method.

PART II

APPLIED RESEARCH

Chapter 7. Testing for singularities in rainfall data.

A test of the homogeneity of daily mean rainfalls in Sydney over the period 1859 to 1952 is made. This test was suggested by the theory proposed by E.G. Bowen that showers of meteors passing into the atmosphere provided the nuclei without which certain cloud formations may not rain and, thus, since some such meteor showers are annually recurring, lead to singularities (peaks) in the seasonal rainfall pattern.

The test is made difficult by the complicated stochastic nature of the process generating the observations so that the actual significance point used may be appropriate to a more stringent test than that proposed. As the sample point does not fall in the critical region, however, there can be no doubt that the test gives a "not significant" result at the 5% level so that the test does not support Bowen's theory.

CONTENTS

- - - - -

PART 1

EXACT TESTS FOR CORRELATION IN TIME SERIES

Chapter 1

Testing for correlation in time series

1.1 Introduction

The first part of this thesis is concerned with the problem of testing for correlation when the set of observations from which the test statistic is being constructed is generated by a (possibly multidimensional) stationary discrete stochastic process.

By a discrete, strictly stationary (vector) process is meant a family of random vectors $\underset{\sim}{x}_t$ (where t is any integer), of k elements, such that the joint distribution of

(1.1.1.) $\underset{\sim}{x}_{t_1}, \underset{\sim}{x}_{t_2}, \ldots, \underset{\sim}{x}_{t_j}$

is the same as the joint distribution of

(1.1.2.) $\underset{\sim}{x}_{t_1+h}, \underset{\sim}{x}_{t_2+h}, \ldots, \underset{\sim}{x}_{t_j+h}$

for all $j, t_1, \ldots, t_j$ and h.

Throughout most of the thesis it will also be assumed that the joint distribution of the elements of the vector $\underset{\sim}{x}_t$ is normal. For the stationarity of such processes it is sufficient that $\mathcal{E}(\underset{\sim}{x}_t)$ should be constant and $\mathcal{E}(\underset{\sim}{x}_t \underset{\sim}{x}'_{t+s})$ should depend only s. Following Doob (1953) any process for which these last two conditions are fulfilled may be called a stationary process in the wide sense.

1.2. <u>Testing for serial correlation in a scalar process.</u>

In this section we will be concerned with a series of observations

(1.2.1.) $x_1, x_2, \ldots, x_n$

from a stationary scalar process.

Very many tests have been constructed for the serial independence of such a series. However, we shall not in this thesis be concerned with the most general situation, where the null hypothesis is specified no further while the alternatives to independence are only very generally specified, so that non parametric and distribution free tests must be used. In general it will be assumed that the x_i, on the null hypothesis, come from a distribution which is sufficiently near normal for the results suggested by normal theory to be reasonable approximations for medium sized samples.

The alternative hypotheses to randomness which have received the greatest attention have been those which have involved the simplest linear processes; low order autoregressions or moving averages.* These are processes whose realisations satisfy relations of the form

---

* The process of concealed harmonics will be considered in Part II of this thesis.

(1.2.2.) $(x_t-\mu)+a_1(x_{t-1}-\mu)+\cdots\cdots+a_k(x_{t-k}-\mu)=\varepsilon_t$ (Autoregression)

(1.2.3.) $(x_t-\mu)=\varepsilon_t+b_1\varepsilon_{t-1}+\cdots\cdots+b_k\varepsilon_{t-k}$ (Moving Average)

In both cases the realisation ε_t is generated by a process of independent random variates (strict or weak sense) with mean value zero and variance σ^2. In the case of the process (1.2.2.) it is necessary further, in order to ensure stationarity, to add the condition that all the roots of

(1.2.2.)$'$ $z^k+a_1z^{k-1}+\cdots\cdots+a_k=0$

should lie inside the unit circle.

In a fundamental paper T.W. Anderson (1948) considered the problem of finding most powerful tests, independent of the nuisance parameters in $\underset{\sim}{\mu}$ and σ^2, in case the probability density for x_t is of the form,

(1.2.4.) $K\exp\left\{-\frac{1}{2\sigma^2}\left[(1+\rho^2)(\underset{\sim}{x}-\underset{\sim}{\mu})'(\underset{\sim}{x}-\underset{\sim}{\mu})-2\rho(\underset{\sim}{x}-\underset{\sim}{\mu})'\underset{\sim}{\Theta}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]\right\}$

and $|\rho|<1$.

Here $\underset{\sim}{x}$ and $\underset{\sim}{\mu}$ are vectors of n elements and $\underset{\sim}{\Theta}$ is an $n\times n$ symmetric matrix.

Anderson showed that in the case where the vector $\underset{\sim}{\mu}$ can be expressed linearly in terms of characteristic vectors of the matrix Θ a uniformly most powerful (one sided) test of the hypothesis $\rho=\rho_0$ is given by the region, $r<r_0$ where

(1.2.5.)
$$r=\frac{(\underset{\sim}{x}-\hat{\underset{\sim}{\mu}})'\underset{\sim}{\Theta}(\underset{\sim}{x}-\hat{\underset{\sim}{\mu}})}{(\underset{\sim}{x}-\hat{\underset{\sim}{\mu}})'(\underset{\sim}{x}-\hat{\underset{\sim}{\mu}})}$$

and $\hat{\underset{\sim}{\mu}}$ is the least squares estimate of $\underset{\sim}{\mu}$. The constant r_0 is, of course, chosen so as to give the region $r < r_0$ the appropriate content on the null hypothesis. The distribution of r is independent of the parameter σ^2 and those parameters involved in $\underset{\sim}{\mu}$. The same statistic can be used to define a critical region of type B_1, i.e. one which is uniformly most powerful in the class of regions "similar to the sample space" (with respect to the same nuisance parameters) and which are unbiassed at least in a neighbourhood of the null hypothesis.

If the alternative hypothesis is the process (1.2.2.) with $h=1$ and ϵ_t normal (i.e. a Gaussian Markoff process) Anderson showed that no uniformly most powerful (one sided) test exists. The matrix $\underset{\sim}{\Theta}$ for this alternative is

(1.2.6.)
$$\frac{1}{2}\begin{bmatrix} \rho & 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 1 & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & & 1 & \rho \end{bmatrix}$$

When the two corner elements of (1.2.6.) are replaced by zeros and the vector $\underset{\sim}{\mu}$ is the null vector the statistic (1.2.5.) becomes

(1.2.7.)
$$r_B = \frac{\sum_{2}^{n} x_t x_{t-1}}{\sum_{1}^{n} x_t^2}$$

and this statistic satisfies the conditions of Anderson's theorem when the alternative hypothesis is prescribed by (1.2.4.) with Θ replaced by the modified form of the matrix (1.2.6.) and μ null.

Similarly when (1.2.6.) is modified by replacing ρ by unity the test statistic

$$(1.2.8.) \quad r_c = \frac{\frac{1}{2}(x_1-\bar{x})^2 + \frac{1}{2}(x_n-\bar{x})^2 + \sum_2^n (x_t-\bar{x})(x_{t-1}-\bar{x})}{\sum_1^n (x_t-\bar{x})^2}$$

also satisfies Anderson's theorem for (1.2.4.) containing this modified Θ as the density function specifying the alternative hypothesis.

Moreover, the distribution of (1.2.8.) for $n = n_c$ is the same as that of (1.2.7.) for $n = n_c - 1$. The 5% and 1% significance points of these two statistics (for the hypothesis $\rho = 0$) were tabulated by Anderson in his paper, having been obtained from the (accurate) approximation to the significance points for Von Neumann's ratio tabulated by Hart (1942).

In practice, of the alternative hypotheses mentioned above, the one which is most likely to be in view is that specified by (1.2.2.), with h equalling unity.

It is clear that for n reasonably large the statistic (1.2.8.) will provide a test against this hypothesis whose power will be very near to the power of the most powerful test for all values of ρ. As pointed out by Watson (1955) the tabulated significance points given by Anderson show that even for n quite small (say 15) the distribution of r_c is very near to that of the ordinary correlation coefficient computed from $n+3$ observations.

Thus when the null hypothesis is independence and the alternative hypothesis is a Markoff process, and normality is maintained, a more or less complete solution to the problem of a test (within the context of the Neyman Pearson theory) has been given.

In practice r_c is not the statistic which has most frequently been used. The most commonly used statistics have probably been the ratio $\frac{\delta^2}{s^2} = \frac{2n}{n-1}(1-r_c)$, the statistic

$$(1.2.9.) \qquad r_1 = \frac{\sum_2^n (x_t - \bar{x})(x_{t-1} - \bar{x})}{\left\{\sum_2^n (x_t - \bar{x})^2 \sum_1^{n-1} (x_t - \bar{x})^2\right\}^{1/2}} \quad *$$

and the statistic

$$(1.2.10.) \qquad r_1' = \frac{\sum_1^n (x_t - \bar{x})(x_{t-1} - \bar{x})}{\sum_1^n (x_t - \bar{x})^2}$$

where $x_0 = x_n$.

As mentioned above the distribution of δ^2/s^2, on

* Quenouille (1949b) has investigated a number of modifications of this statistic, by means of sampling experiments, from the point of view of bias and variance on the null hypothesis, for small samples.

the null hypothesis, was investigated by Von Neumann (1941) and its significance points were tabulated by Hart (1942).

The exact distribution of r_1' was determined by R.L. Anderson (1942) (see also Watson (1951a)) who also tabulated the significance points.

On the null hypothesis the joint characteristic function of the numerator and denominator of the ratio (1.2.10.) is

$$(1.2.11.) \qquad \phi(\theta_1, \theta_2) = \prod_{j=1}^{n-1} \left(1 - 2i\theta_2 - 2i\lambda_j\theta_1\right)^{-\frac{1}{2}}$$

where $\lambda_j = \cos\frac{2\pi j}{n}$ (see for example Whittle (1951)). This may be written (Koopmans (1942)).

$$(1.2.12.) \qquad \phi(\theta_1, \theta_2) = \left\{\exp -\tfrac{1}{2}\sum_{1}^{n} \log\left(1 - 2i\theta_2 - 2i\theta_1\cos\tfrac{2\pi j}{n}\right)\right\}\left(1 - 2i\theta_2 - 2i\theta_1\right.$$
$$\approx \left\{\exp -\tfrac{1}{2}\int_0^{2\pi} \log\left(1 - 2i\theta_2 - 2i\theta_1\cos t\right)dt\right\}\left(1 - 2i\theta_2 - 2i\theta_1\right)^{\frac{1}{2}}$$

From this expression Dixon (1944) obtained approximations to the moments of the ratio (1.2.10) and showed that these approximations were in fact exact for the moments of order not greater than $\frac{n}{2}$. By identifying these moments with those of the distribution with density function

$$(1.2.13.) \qquad K\left(1 - x^2\right)^{\frac{n-2}{2}}$$

he showed that the statistic (1.2.10) was approximately distributed as the ordinary correlation coefficient based on $n+2$ pairs of observations.

This result was checked by Rubin (1945) who inverted the integral approximation to the characteristic function (1.2.11.) for the case where the population mean was zero.

In practice it is clear that it will be sufficient to use the statistic r_1 and refer it to the tabulated significance points of the distribution (1.2.13.), for n reasonably large.

For the case in which the alternative hypothesis to independence is prescribed by (1.2.2.) with $h > 1$ T.W. Anderson (1948) showed that no uniformly most powerful test exists, as was to be expected. The test statistic which is intuitively suggested is the multiple correlation coefficient of x_t and $x_{t-1}, x_{t-2}, \ldots, x_{t-h}$. Similarly for the test, of the hypothesis that h = m against the hypothesis that $h = k+m$, the test statistic which suggests itself is the, partial, multiple correlation of x_t and $x_{t-m-1}, x_{t-m-2}, \ldots, x_{t-m-k}$ with the effects of $x_{t-1}, x_{t-2}, \ldots, x_{t-m}$ removed. (The use of such statistics was, in fact, suggested originally by Yule (1927)). By analogy with the results for h = 1 it might be hoped that these statistics could be treated in the same way as their counterparts when the relation between sets of variates, independent at successive points of time, is being considered.

This possibility was investigated by Quenouille (1949) who obtained the exact joint distribution of the first m, circularly defined, serial correlations when the number of observations, n, is odd. The circular serial correlations will be indicated by r_j' where

$$(1.2.14.)\qquad r_j' = \frac{\sum_1^n x_t x_{t+j} - \frac{1}{n}\left(\sum_1^n x_t\right)^2}{\sum_1^n x_t^2 - \frac{1}{n}\left(\sum_1^n x_t\right)^2}$$

and x_{n+i} is defined as equal to x_i.

A simpler and more general derivation of Quenouille's result was given by Watson (1955) who extended the distribution to even n.

From the computational point of view the exact distribution is extremely difficult to use, the density function being for n even

$$(1.2.15.)\qquad \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-2m}{2})}\sum_{j_1,\ldots,j_m \in R}\cdots\sum \frac{\begin{vmatrix} 1 & r_1' & r_2' & \cdots & r_m' \\ 1 & \cos\frac{2\pi j_1}{n} & \cos\frac{4\pi j_1}{n} & \cdots & \cos\frac{2m\pi j_1}{n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \cos\frac{2\pi j_m}{n} & \cos\frac{4\pi j_m}{n} & \cdots & \cos\frac{2m\pi j_m}{n} \end{vmatrix}^{\frac{n}{2}-m-1} \mathrm{Sgn} \begin{vmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & \cos\frac{2\pi j_1}{n} & \cos\frac{4\pi j_1}{n} & \cdots & \cos\frac{2m\pi j_1}{n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \cos\frac{2\pi j_m}{n} & \cos\frac{4\pi j_m}{n} & \cdots & \cos\frac{2m\pi j_m}{n} \end{vmatrix}}{\begin{vmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & \cos\frac{2\pi j_1}{n} & \cos\frac{4\pi j_1}{n} & \cdots & \cos\frac{2m\pi j_1}{n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \cos\frac{2\pi j_m}{n} & \cos\frac{4\pi j_m}{n} & \cdots & \cos\frac{2m\pi j_m}{n} \end{vmatrix}^{1/2} \prod_{\substack{j=1 \\ j \neq (j_1,\ldots,j_m)}}^{\frac{n}{2}-1} \begin{vmatrix} 1 & \cos\frac{2\pi j}{n} & \cdots & \cos\frac{2m\pi j}{n} \\ 1 & \cos\frac{2\pi j_1}{n} & \cdots & \cos\frac{2m\pi j_1}{n} \\ \cdot & \cdot & \cdots & \cdot \\ 1 & \cos\frac{2\pi j_m}{n} & \cdots & \cos\frac{2m\pi j_m}{n} \end{vmatrix}}$$

The summation is over all suffixes j_i $(i = 1 \cdots m)$ for which the point, $r_1' \cdots r_m'$ in euclidian m -space, falls in the least convex polyhedron with vertices

$$(-1, -1 \cdots -1), \quad \left(\cos\frac{2\pi j_1}{n}, \cos\frac{4\pi j_1}{n}, \cdots\cdots, \cos\frac{2m\pi j_1}{n}\right),$$

$$\cdots\cdots\cdots$$

$$, \left(\cos\frac{2\pi j_m}{n}, \cos\frac{4\pi j_m}{n}, \cdots\cdots, \cos\frac{2m\pi j_m}{n}\right).$$

For n odd the constant term becomes $\Gamma(\frac{n-1}{2}) / \Gamma(\frac{n-2m-1}{2})$ while the terms, -1, in the determinants are replaced by zeros.

The distribution is clearly of little practical use because of the complicated summation involved combined with the fact that density function has a different analytic form for different ranges of values of the correlations.

Quenouille conjectured that a smoothing operation analogous to that suggested by Koopmans for r_1' would give an approximation to the joint density of the r_j' (uncorrected for mean) with n odd as

$$(1.2.16.) \quad \pi^{-m/2} \frac{\Gamma(\frac{n}{2}+1) \cdots \Gamma(\frac{n}{2}-m+2)}{\Gamma(\frac{n}{2}+\frac{1}{2}) \cdots \Gamma(\frac{n}{2}-m+\frac{3}{2})} \left(\frac{R_m}{R_{m-1}}\right)^{\frac{1}{2}(n-2m+1)}$$

where

$$(1.2.17.) \quad R_m = \begin{vmatrix} 1 & r_1' & \cdots & r_m' \\ r_1' & 1 & & r_{m-1}' \\ & & & \\ r_m' & r_{m-1}' & & 1 \end{vmatrix}$$

This conjecture was supported by the following arguments:

(1) The joint distribution of the correlations between a variate and m sets of other independent variates was shown to be of this form by Garding so that this was the natural generalisation of Dixon's result for r_1'.

(2) Smoothed outer limits to the values of $r_1', \ldots, r_m'$ will be provided by the equations $R_p = 0 \ (p = 1, \ldots, m)$ so that these determinants could be expected to play a part in the frequency function.

(3) Quenouille showed that if $r_m', r_{m-1}', \ldots r_2'$ are integrated out of the function (1.2.16.) the result is Dixon's approximation to the distribution of r_1'.

(4) If the joint characteristic function of the r_j' is smoothed in a manner analogous to that suggested by Koopmans the result suggests that the smoothed frequency function is a polynomial in the r_i' which vanishes for $R_m = 0$.

Quenouille's conjecture was substantially verified by Watson who observed that for n odd the joint density of $r_1', \ldots, r_{n-1}'$ is constant. The true joint domain of these variates is polyhedral but the vertices of the polyhedron can easily be seen to lie on the boundary of the region for which R_{n-1} is positive definite. Using this region for the density the variates $r_{m+1}', \ldots r_{n-1}'$ can be integrated out to give a density of the same form as (1.2.16.) but with the power, to which the term R_m/R_{m-1} is raised, reduced by $\frac{3}{2}$. As Watson pointed out the integration out of all of the remaining r_j', save r_1', suggests that Quenouille's conjectured approximation is the better approximation.

A further check on Quenouille's approximation was carried out by Watson (1951a) and Jenkins (1954b) who took the case $m=2$ and integrated out r_1' to get the approximation to the distribution of r_2'. Dixon had shown that on the null hypothesis the distribution of r_2' should be the same as that of r_1' but the approximation resulting from Quenouille's form differs from this; though the difference will be small for large n.

This case was considered in more detail by Jenkins (1954b) who, by presuming that r_1' and the first partial serial correlation

$$(1.2.18.)\qquad r_{1.2}' = \frac{r_2' - r_1'^2}{1 - r_1'^2}$$

were independent obtained an approximation to the joint distribution of r_1' and r_2' as

$$(1.2.19.)\qquad \frac{n}{2\pi}\left(\frac{R_2}{R_1}\right)^{\frac{1}{2}(n-2)} \frac{1-r_2}{R_1^{3/2}} \qquad \text{(corrected for mean)}$$

compared with

$$\frac{1}{\pi}\;\frac{\Gamma(\frac{n}{2}+1)\,\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2}+\frac{1}{2})\,\Gamma(\frac{n}{2}-\frac{1}{2})}\left(\frac{R_2}{R_1}\right)^{\frac{n-3}{2}}$$

from Quenouille's form.

It appears, however, that for n reasonably large, and m small relative to n, the distribution of the first m circular serial correlation coefficients (without mean corrections) will be reasonably well approximated by the distributions arising in classical least squares analysis treated as having come from samples of n+ 3 observations. In practice the non circular definition of the serial correlations (with mean corrections) will be used and in this case the number of observations should probably be treated as n+ 2. *

Finally, as pointed out by Watson (1955), the hypothesis which will be under test will often be that the observations are generated by a process of the form (1.2.2.) of order m against the hypothesis that the order is m+ k. When the process (1.2.2.) is circularly defined, so that $x_{n+i} = x_i$, Madow's (1945) result for r_1' can be generalised (Quenouille (1949, Watson (1955)) to show that the density is obtained from (1.2.15.) by adjoining a factor

$$(1.2.20.) \qquad \left\{ \sum_{0}^{m} a_j^2 + 2 \sum_{j=1}^{m} \left(\sum_{i=0}^{m-j} a_i a_{i+j} \right) r_j' \right\}^{-\left(\frac{n-1}{2}\right)}$$

---

* As pointed out on page 6 the distribution of r_B is close to that of the ordinary correlation from n+ 4 observations. As compared with r_1, however, r_B has no mean correction and has an additional term in the numerator.

It is then evident that the distribution of the multiple correlation $r'_{m+1(1\cdots m)}$ and the partial multiple correlation $r'_{1(m+1,\cdots,m+k)\cdot 23\cdots m}$ will be independent of the a_j so that, to the extent that the tests based on the distribution (1.2.16.) are a satisfactory approximation, and the assumption of a circular universe does not distort the situation, these statistics can be treated as if they were ordinary correlations based on three additional observations.

Less study seems to have been devoted to the model (1.2.3.) than to (1.2.2.). In most applications, on prior grounds, the process (1.2.2.) seems more acceptable than (1.2.3.). This is so, for example, in econometrics.

No uniformly most powerful test against an hypothesis of the form (1.2.3.) is available. Consider, for example, the process (1.2.3.) with $k=1$. Then a test against the hypothesis $b_1=-\alpha$ will be provided by the statistic

(1.2.20.) $$t=\sum_{1}^{n-1}\alpha^{j-1}r_j$$

where r_j is the j^{th} serial correlation coefficient (non-circular). This test, for fixed α, will be near to a most powerful test, for the covariance matrix of the process is

$$(1.2.21.)\quad \sigma^2\begin{bmatrix} 1+\alpha^2 & -\alpha & 0 & \cdot & \cdot & \cdot & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \cdot & \cdot & \cdot & 0 \\ 0 & -\alpha & 1+\alpha^2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1+\alpha^2 \end{bmatrix}$$

and if the elements $1+\alpha^2$ in the two corners are altered to unity the inverse becomes

$$(1.2.22.)\quad \frac{1}{\sigma^2(1-\alpha^2)}\begin{bmatrix} 1 & \alpha & \cdot & \cdot & \cdot & \cdot & \alpha^{n-1} \\ \alpha & 1 & \cdot & \cdot & \cdot & \cdot & \alpha^{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha^{n-1} & \alpha^{n-2} & \cdot & \cdot & \cdot & & 1 \end{bmatrix}$$

Then for $\mu=0$ the statistic (1.2.20.) (computed without mean corrections) is most powerful against this fixed value of α , when the alternative hypothesis is prescribed by the matrix (1.2.22.), as the matrix of the quadratic form in the exponent of the density function of the x_t . (T.W. Anderson (1948)).

Clearly the best critical region (of fixed "content") will depend on b_1 so that no uniformly most powerful test will exist.

If the range of alternative held in mind suggest that $|b_1|$ is small then only the first few terms of (1.2.20.) need be included and a reasonable test could be found by choosing suitable coefficients for the r_j on prior grounds. (See Whittle (1951)).

So far the null hypothesis considered has been the hypothesis that the x_t are independent. It will sometimes be necessary to test the hypothesis that the process generating the x_t is of the form (1.2.2.) with prescribed a_j, not all equal to zero. The only case which has so far received anything approaching an exact treatment is the case $h=1$. In this case the null density is multiplied by the factor (see 1.2.20.)

$$(1.2.23.)\qquad \left\{1 + a_1^2 + 2a_1 r_1'\right\}^{-\left(\frac{n-1}{2}\right)}$$

when the universe is circularly defined, so that $x_i = x_{n+i}$.

Putting $a_1 = -\rho$ (as is usual) an approximation to the density is obtained by adjoining (1.2.23.) to (1.2.13.) to obtain

$$(1.2.24.)\qquad \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\,\Gamma\left(\frac{n}{2}\right)}\left(1-r_1'\right)^{\frac{n-2}{2}}\left(1+\rho^2-2\rho r_1'\right)^{-\frac{n-1}{2}}$$

This approximation was also obtained by Leipnik (1947) directly from the smoothed form of the characteristic function for the case where the true mean is known to be zero, so that r_1' is computed without the use of a mean correction. Leipnik showed that the mean and variance of the variate in this distribution are

(1.2.25.) $\mu_{r_1'} = \dfrac{\rho(n-1)}{n+1}$, $\sigma^2_{r_1'} = \dfrac{1}{n+1}\left\{1 - \dfrac{\rho^2(n-1)n}{(n+1)(n+3)}\right\}$

In practice, of course, the universe will not be circular and the mean will not be known to be zero. In this case r_1 will be used and it is known that this statistic has a large bias when ρ is not zero and n is small. In fact the mean and variance are (Bartlett (1946) Marriott and Pope (1954));

(1.2.26.)
$$\mu_{r_1} = \rho - \frac{1}{n}\{(1+\rho)(1+2\rho)\} + O(n^{-2})$$
$$\sigma^2_{r_1} = \frac{1}{n}(1-\rho^2) + O(n^{-2})$$

Though the two variances agree reasonably well the bias in μ_{r_1} is much larger (for ρ positive) than in $\mu_{r_1'}$,for n small.

In practice, therefore, the distribution (1.2.24.) will not be satisfactory for n small and ρ large positive (near unity). Even though, if the exact distribution of a statistic is known, bias is not important in relation to tests of significance it can still be important if it is desired to average a number of, otherwise non homogeneous, estimates of the same parameter, since the resulting statistic will not tend to the population value in probability as the number of estimates averaged is increased. (An example is mentioned in part II of this thesis).

The distribution (1.2.24.) is not in a usable form as it stands. Quenouille (1948) suggested the use of the transformation $r_1' = \tanh z$, which reduces the distribution to approximate normality. However, the variance is not independent of ρ (to order n^{-1}) as is the case with the ordinary correlation coefficient so that other devices have to be used to obtain a confidence interval. A preferable procedure was suggested by Jenkins (1954a) who observed that $\sigma^2_{r_1'} \approx \frac{1}{n+1}(1-\mu^2_{r_1'})$ so that the transformation $r_1' = \sin z$ makes the variance approximately independent of the mean. The mean and variance now become

(1.2.27.)
$$\mu_z = \sin^{-1}\rho - \frac{3\rho}{2(n-1)(1-\rho^2)^{1/2}} + O(n^{-2}).$$
$$\sigma_z^2 = \frac{1}{n-1} + O(n^{-2}).$$

The variate z can then be approximately treated as normal with mean $\sin^{-1}\rho$ and variance $\frac{1}{n-1}$ * so that a confidence interval can be obtained directly. The transformation is not satisfactory for $\rho > 0.9$.

The tests considered in this section have all related to a specified range of alternatives. Often it will be necessary to test the appropriateness, of a prescribed stochastic model, to an observed series without

* An examination of the term in n^{-2} in σ_z^2 suggests that, for ρ small, n^{-1} is a better approximation to the variance.

specifying alternatives. In such cases a test of goodness of fit is required. For a prescribed process of the form (1.2.2.) a test was given by Quenouille (1947b) who used Bartlett's (1946) results for the covariance matrix of the observed serial correlations to obtain linear forms, R_j , in the serial correlations, the j^{th} form depending only on the serial correlations of order j and under. The R_j are asymptotically normally and independently distributed, while for $j > h$ the R_j are (asymptotically) distributed independently of the r_k $(k \leq h)$. Thus a goodness of fit test is obtained based on the χ^2 distribution. For the case of a simple Markoff process, the parameter ρ being estimated by r_1 the forms are

(1.2.28.) $$R_j = r_j - 2r_1 r_{j-1} + r_1^2 r_{j-2}$$

which have, asymptotically, the variance $\frac{(1-\rho^2)^2}{n}$

Quenouille's test was extended to other schemes by Walker (1950) and Wold (1949), (see also Bartlett and Diananda (1950)), but the results will not be discussed here as they are not used in this thesis.

1.3. Testing for Correlation in the Residuals from a Regression.

We will be concerned, in this section, with the problem of estimating the regression coefficients in a relation

(1.3.1.) $$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_p x_{pt} + \varepsilon_t$$

where, (1) the ε_t are wholly independent of the x_{jt}

(2) the residuals ε_t are generated by a stationary stochastic process. (At least to the second order).

The relation (1.3.1.) may be written

(1.3.2.) $$\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

where $\underset{\sim}{y}$ and $\underset{\sim}{\varepsilon}$ are vectors of n elements y_t and ε_t ; $\underset{\sim}{\beta}$ is a vector of the $p+1$ coefficients β_j and $\underset{\sim}{X}$ is an $n \times (p+1)$ matrix whose first column consists of n units.

It is well known (Aitken (1934/35)) that if the covariance matrix of the ε_t is known to be $\underset{\sim}{\Gamma}$ the best linear unbiassed estimate of the coefficient vector $\underset{\sim}{\beta}$ is obtained by minimising the quadratic form

(1.3.3.) $$(\underset{\sim}{y} - \underset{\sim}{X}\,\underset{\sim}{\beta})'\,\underset{\sim}{\Gamma}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\,\underset{\sim}{\beta})$$

and that this estimate will also be the maximum likelihood estimate when the ε_t are jointly normal.

In general the matrix $\underset{\sim}{\Gamma}$ will not be known. If it is wrongly prescribed the estimation procedure can become very inefficient (Watson (1951a)). In particular the use of straightforward least squares (which would correspond to $\underset{\sim}{\Gamma} = \underset{\sim}{I}$) will in general be inefficient if the ε_t are not in fact uncorrelated. Since the straightforward least squares procedure will often be used it is, therefore, important to be able to test the resulting (observed) residuals for serial correlation.

The analysis of #1.2 can be carried over to the present case. The vector $\underset{\sim}{\mu}$ in (1.2.4.) will now be the vector $\underset{\sim}{X}\underset{\sim}{\beta}$ and, if this vector can be expressed linearly in terms of the characteristic vectors of $\underset{\sim}{Ⓗ}$, then Anderson's theorem shows that the test criterion giving a uniformly most powerful (one sided) test (or a uniformly most powerful locally unbiassed test) will be

$$(1.3.4.) \qquad \frac{(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})' \, \underset{\sim}{Ⓗ} \, (\underset{\sim}{y} - \underset{\sim}{X}\hat{\beta})}{(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})}$$

In case $\mathcal{E}_t$ is generated by a circular simple Markoff process the matrix $\underset{\sim}{Ⓗ}$ becomes $\frac{1}{2}\underset{\sim}{W}$ where

$$(1.3.5.) \qquad \underset{\sim}{W} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 0 \\ & & & & \\ \cdots & & & & \cdot \\ & & & & \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

which has characteristic vectors

$$(1.3.6.) \qquad \begin{Bmatrix} \cos\frac{2\pi j}{n} \\ \cos\frac{4\pi j}{n} \\ \vdots \\ \cos\frac{2n\pi j}{n} \end{Bmatrix}, \quad \begin{Bmatrix} \sin\frac{2\pi k}{n} \\ \sin\frac{4\pi k}{n} \\ \vdots \\ \sin\frac{2n\pi k}{n} \end{Bmatrix}, \quad \begin{array}{l} n \text{ even} \\ \begin{cases} j = 0 \cdots\cdots \frac{n}{2} \\ k = 1 \cdots\cdots \frac{n-2}{2} \end{cases} \\ n \text{ odd} \\ \begin{cases} j = 0 \cdots\cdots \frac{n-1}{2} \\ k = 1 \cdots\cdots \frac{n-1}{2} \end{cases} \end{array}$$

It is, therefore, evident that when the alternative to independence of residuals is the hypothesis that the residuals are generated by a simple Markoff process, and the regressands are the terms of a Fourier series, the first serial correlation of the residuals will provide a test which will, for all values of the parameter ρ, be near to the most powerful (one sided) test. This case is also to some extent unique in that the distribution of the test statistic can be obtained exactly (R.L. & T.W. Anderson (1950)). Since any reasonably smooth function can be well approximated by a relatively few terms of a Fourier series it follows that in any case of "analytical" regression* the first serial correlation coefficient of the residuals will provide a good approximation to a most powerful test.

It is in just this case that the test is least important, however since the straightforward least squares procedure will be, asymptotically at least, fully efficient. For the estimates of $\underset{\sim}{\beta}$ obtained by minimising the quadratic form (1.3.3.), is (Aitken (1934/35))

(1.3.7.) $$(\underset{\sim}{X}'\underset{\sim}{\Gamma}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{\Gamma}^{-1}\underset{\sim}{y}$$

and if the column vectors of $\underset{\sim}{X}$ are characteristic vectors of $\underset{\sim}{H}$ they are characteristic vectors of $\underset{\sim}{\Gamma}^{-1}$ so that

* That is, regression upon the "variates" which are the values, at equidistant points, of an analytic function of t.

(1.3.8.) $\underset{\sim}{\Gamma}^{-1}\underset{\sim}{X} = \underset{\sim}{X}\underset{\sim}{\Lambda}$

where $\underset{\sim}{\Lambda}$ is the $(p+1)\times(p+1)$ matrix of the characteristic roots of $\underset{\sim}{\Gamma}^{-1}$ corresponding to the characteristic vectors in $\underset{\sim}{X}$.

The expression (1.3.7.) then becomes

(1.3.7.)' $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$

This is the same as straightforward least squares estimate of $\underset{\sim}{\beta}$.

The covariance matrix of the estimates is, however, not given by the usual formula. For this covariance matrix is (Aitken (1934/35))

(1.3.9.) $(\underset{\sim}{X}'\underset{\sim}{\Gamma}^{-1}\underset{\sim}{X})^{-1} = \underset{\sim}{\Lambda}^{-1}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$

This is, of course, important in connection with tests of significance.

Since in practice the serial correlations ρ_j of the residuals ϵ_t will converge to zero fairly quickly, it it evident that for n large the matrix $\underset{\sim}{\Gamma}$ will be well approximated by a sum of the form (Whittle (1951)).

(1.3.10.) $\underset{\sim}{\Gamma} = \sigma^2 \sum_{-(n-1)}^{n-1} \rho_j \underset{\sim}{W}^j$

for which the vectors (1.3.6.) are still characteristic vectors. It follows, therefore, that in almost any case of analytical regression the straightforward method of least squares will be, asymptotically at least, fully efficient. This result has been checked by a rigorous investigation by Grenander (1954).

The matrix $\underset{\sim}{X}$ will not usually be expressible as a linear combination of characteristic vectors of $\underset{\sim}{\Theta}$. However, Durbin and Watson (1950) showed that as $\underset{\sim}{X}$ varies the statistic (1.3.4.) will still lie between two bounds, the bounds being attained for regression matrices $\underset{\sim}{X}$ which are matrices composed of characteristic vectors of $\underset{\sim}{\Theta}$ so that, when the bounds are attained, (1.3.4.) is an optimal test statistic. These bounds were shown to be of the form

$$r_{lower} = \frac{\sum_{i=1}^{n-p-1} \lambda_i \zeta_i^2}{\sum_{i=1}^{n-p-1} \zeta_i^2} \tag{1.3.11.}$$

$$r_{upper} = \frac{\sum_{i=1}^{n-p-1} \lambda_{i+p+1} \zeta_i^2}{\sum_{i=1}^{n-p-1} \zeta_i^2}$$

where the ζ_i are (orthogonal) linear combinations of the true residuals and the λ_i are the characteristic roots of $\underset{\sim}{\Theta}$ arranged in order of magnitude. When $\underset{\sim}{\Theta}$ is of the form (1.3.5.) the λ_i are the values of $\cos\theta$ for θ at equal intervals in the range π to zero (n even) and it is evident that (r (upper) - r (lower)) will tend to zero in probability as n increases, (for fixed p).

Thus the first serial correlation coefficient of the residuals will have considerable appeal as a test statistic, against the hypothesis of independence of the ϵ_t for the cases where the alternative hypothesis is the hypothesis of a simple Markoff process or some process which is reasonably well approximated by such a scheme.

Durbin and Watson showed that the statistic (1.3.4.) could be reduced to the canonical form

(1.3.12.) $$\frac{\sum_{1}^{n-p-1} \nu_i \zeta_i^2}{\sum_{i=1}^{n-p-1} \zeta_i^2}$$

where the ν_i are the non zero roots of the matrix

(1.3.13.) $$[\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'] \underset{\sim}{\Theta}$$

The distribution of (1.3.12.) has not beeh given in a closed form except in cases where the ν_i have special properties (e.g. equal in pairs). Moreover, the distribution would depend upon the ν_i and it would be computationally impracticable to determine these in every case. The only case for which an exact treatment has been given is the case of residuals from a fitted Fourier series (R.L. & T.W. Anderson (1950)).

For the case in which $p = 1$ Moran (1950) obtained the mean and variance of the circular serial correlation of the residuals as

(1.3.14.) $$\mathcal{E}(r_1') = -\frac{1 + r_{1x}}{n-2}$$
$$Var(r_1') = \frac{n+1}{n^2} + O(n^{-2}).$$

where r_{1x} is the observed first circular serial correlation of the regressand series.

These formulae were extended to the general case by Durbin and Watson (1950) (pages 420, 421.)

Durbin and Watson (1951) also tabulated bounds to the 5%, 2.5% and 1% significance points of (1.3.4.) when

$$(1.3.15.) \qquad \Theta = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

The test statistic is then $d = \frac{n-1}{n}\frac{\delta^2}{s^2}$

The bounds whose significance points were tabulated will be given by (1.3.11.) with the λ_i as latent roots of the matrix (1.3.15.).

In cases for which the bounds test is inconclusive (that is, in which d falls between the bounds) Durbin and Watson recommended the use of $\frac{1}{4}d$ * as a Beta variate with the appropriate mean and variance, provided the number of observations is reasonably large.

The bounds for small samples are fairly far apart. For example for $n=30$ and $p=5$ the bounds for a (one sided) 5% test are 0.98 and 1.73, corresponding (roughly) to a serial correlation between .13 and .49.

---

* $d \approx 2(1-r_1)$ which varies between 0 and 4.

For small samples the situation is not, therefore, entirely satisfactory. One suspects that the treatment of the serial correlation of the residuals as an ordinary correlation from $n+s-p$ observations will (in most cases) lead to a significance point, even for samples of around 30 observations, not far from the true significance point, but this is merely a conjecture.

1.4. Testing for correlation between serially correlated series.

The problem of testing for multiple correlation between the variate y_t and the variates x_t in (1.3.1.) is, of course, closely connected with the problem of the efficiency of estimation procedures for the β_j.

If the covariance matrix Γ of the ε_t is known then, as mentioned in # 1.3. a fully efficient estimation procedure can be developed and corresponding to this procedure there will be exact tests for multiple correlation, in the case where the ε_t are normally distributed. In fact if the matrix Λ is orthogonal (though not necessarily orthonormal) and

1.4.1. $$\Lambda \Gamma \Lambda' = I$$

the classical procedure applied to the vector Λy and the vectors in ΛX will yield the classical tests.

In the cases with which we will be concerned the covariance matrix $\underset{\sim}{\Gamma}$ will be of the form

(1.4.2.) $$\underset{\sim}{\Gamma} = \sigma^2 \sum_0^{n-1} \rho_j \underset{\sim}{U}^j$$

where σ^2 is the variance of the ε_t and

(1.4.3.) $$\underset{\sim}{U} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Then the covariance matrix of the straightforward least squares estimates of $\underset{\sim}{\beta}$ is

(1.4.4.) $$(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{\Gamma}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$$

(1.4.5.) $$= n^{-1}\sigma^2 \left\{ \underset{\sim}{M}_0^{-1} + \sum_{k=1}^{n-1} \rho_k \underset{\sim}{M}_0^{-1} \underset{\sim}{M}_k \underset{\sim}{M}_0^{-1} \right\}$$

where $\underset{\sim}{M}_0 = [c_{ij}(0)]$ and $\underset{\sim}{M}_k = [c_{ij}(k) + c_{ji}(k)]$

(1.4.6.)

and $$c_{ij}(k) = \frac{1}{n} \sum_{m=1}^{n-k} x_{i,m} x_{j,m+k}$$

If the ρ_k $(k \neq 0)$ are all zero (1.4.5.) reduces to the covariance matrix in the classic case. In general it will be different and the test of multiple correlation based on the least squares procedure will be invalid.

In the case where p is unity the variance of from (1.4.5.) reduces to

(1.4.7.) $$\frac{\sigma^2}{n s_x^2}\left\{1 + 2\sum_{k=1}^{n-1} \rho_k r_k'\right\}$$

where s_x^2 is the observed variance of the x_{1t} and r_k' is the observed k^{th} serial correlation of this variate. This suggests an approximate procedure (due to Bartlett (1935)). This is to test the correlation as having come from a sample of $n/\left\{1+2\sum_1^{n-1} \rho_k r_k'\right\}$ observations. On the null hypothesis the ρ_k can be estimated from the serial correlations of the y_t so that the number of degrees of freedom will be estimated by $n/\left\{1+2\sum_1^{n-1} r_k r_k'\right\}$ where r_k is the observed k^{th} serial correlation of the y_t. If both y_t and x_{1t} are Markovian then (1.4.7.) converges in probability, as n increases, to

(1.4.8.) $$\frac{\sigma^2}{n\sigma_x^2}\ \frac{1+\rho\rho'}{1-\rho\rho'}$$

where ρ and ρ' are the parameters in the two Markoff processes, so that the number of degrees of freedom can be estimated as (Bartlett (1935))

(1.4.9.) $$n\ \frac{1-r_1 r_1'}{1+r_1 r_1'}$$

In the case of a test for multiple correlation the corresponding procedure would be to replace the ρ_k with the r_k in (1.4.5.) and to use the inverse of this matrix as the matrix of the quadratic form (in the $\hat{\beta}_j$) in the numerator of $1-R^2$ (where R is the multiple correlation coefficient). For partial correlation the ρ_k would have to be estimated from the residuals as, on the null hypothesis, the serial correlations of the y_t and ε_t are, then, no longer the same. A test for multiple correlation of this form does not appear to have been used in practice.

The disadvantage of these procedures, apart from the inexactness of the tests involved, is that they are inefficient, since the least squares estimates of the β_j are inefficient. When it is reasonable to specify that the residuals are from a (fairly simple) stochastic process of a known type an alternative would be to estimate the ρ_k from the residuals from a least squares regression, use these to obtain an estimate of the parameters specifying the process, and use the resulting estimate of $\underset{\sim}{\Gamma}$ to transform the variates to a from where the residuals were (approximately) independent. The tests for multiple and partial correlation could then be carried out in terms of these transformed variates (Cochrane and Orcutt (1949)). Such a test would not, of course, be exact and could be very misleading, particularly if the nature of the process generating the errors is incorrectly specified (Watson (1951)).

It is clear from (1.4.8.) that if ρ and ρ' are small the variance of the estimate of β_1 will be little affected by the departure from independence. For example, for $\rho = \rho' = 0.2$ (1.4.8.) becomes $\left\{\frac{\sigma^2}{n\sigma_x^2} 1.08\right\}$ Any transformation that will remove most of the serial correlation will, therefore, recover most of the lost information and enable approximate tests to be used. In many cases in economics the serial correlations are high and they are nearly all positive. Insofar as the specification of the ϵ_t process (and the x_{jt} processes) as simple Markoff processes is appropriate the use of the first difference of the observations should then be a reasonable procedure (Cochrane and Orcutt (1949)).

None of the expedients mentioned above is very satisfactory for, apart from the fact that they are inexact, they all suffer from other disadvantages, being either inefficient or involving the use of only very rough approximations to the variance on the null hypothesis.

Quenouille (1949b) suggested the use of partial correlation coefficients for the case where the correlation between two series (x_t and y_t) is being considered. This procedure is applicable in any case where the process generating one of the two series can be represented by an autoregression. His procedure is as follows:

(1) Determine the order of the autoregression needed to represent the two series by the use of partial serial correlations. (The adequacy of the resulting autoregressiive representation could be checked by the test mentioned at the end of # 1.2.).

(2) Calculate partial correlations $r(x_t\, y_t \,|\, x_{t-1}, x_{t-2}, \cdots x_{t-h}; y_{t-1}, y_{t-2} \cdots y_{t-k})$ where h and k are the orders of the two autoregressive schemes and $r(x\, y \,|\, z)$ is the partial correlation between x and y with the effect of z removed.

(3) Treat the resulting statistic as a correlation from n - h - k observations.

If only one process could be adequately represented autoregressively the effects of lagged values of this variate only would be removed. Since only one of the two processes needs to be independent in order to validate the usual test of significance this is all that needs to be done in any case, from the point of view of the validity of the test. The power of these tests will be considered in Chapter 4.

Quenouille considered the case of a pair of, independent, simple Markoff processes* with serial

* The processes used were not in fact simple Markoff processes but approximations to such schemes obtained for example, by putting $x_n = \pm 1$; $P(x_n = 1 \,|\, x_{n-1} = 1) = P(x_n = -1 \,|\, x_{n-1} = -1) = 0.8$; $P(x_n = 1 \,|\, x_{n-1} = -1) = P(x_n = -1 \,|\, x_{n-1} = 1) = 0.2$ which approximates to a simple Markoff scheme with $\rho = 0.6$.

correlation 0.4 and 0.6 and by a sampling experiment showed that the mean and variance, on the null hypothesis, of his statistics were near to zero and $(n-2)^{-1}$ or $(n-3)^{-1}$ according as the effects of lagged values of only one or of both of the series were removed. For example, for 32 pairs of series of 20 observations the average correlation, $r(x_t y_t | x_{t-1}, y_{t-1})$, was -0.011 and the variance of the correlations was 0.069, (17^{-1} = 0.059). The results for the case where the effects of lagged values of both variates were removed were better than those for the case where the effects of lagged values of only one variate were removed.

The generalisation to the case of multiple correlation is clear though the validity of the approximation to the distribution on the null hypothesis has not been checked by a sampling experiment.

Finally, in this section mention may be made of a test suggested (amongst others) by Orcutt and James (1948) for econometric data. This test treats the correlation between two economic series of yearly data, of any length, as being equivalent to a correlation between two series of five or six observations, each series of observations being random in time. The basis of the argument for the test was a sampling experiment based on observations generated by a process of the form

$$(1.4.10) \quad (x_t - x_{t-1}) = 0.3(x_{t-1} - x_{t-2}) + \epsilon_t$$

Such a process is not stationary. The argument for treating all economic series of yearly observations as of this form was presented in a previous paper by Orcutt (1948), but was not convincing. It would seem better to employ no test at all, and to rely on subjective judgement, than to use a test founded on such an insecure basis as is this one.

Chapter 2.

The Asymptotic Efficiency of Tests.

2. 1. Introduction

It is often extremely difficult to determine the distribution of a test statistic on the alternative hypothesis, for usually the null hypothesis corresponds to the value zero for some parameter, of which the test statistic is an estimator, so that on the alternative hypothesis the frequency function of the observations is complicated by the presence of an additional parameter.

For this reason calculation of the power function for a test is often not easy. In this chapter an asymptotic criterion of the efficiency of a test, due essentially to Pitman (1948), will be described and justified. Because the criterion is asymptotic it is not wholly satisfactory but in the cases in which it will be applied in this thesis no alternative criterion is readily available. To calculate the power function would usually be very difficult.

2.2. Pitman's Theorem

Consider a test based on a statistic t_n, a function of n observations $x_1, x_2, \cdots, x_n$ with mean $\mu_n(\theta)$ and variance $\sigma_n^2(\theta)$. We will be concerned with the power of the test, based on the statistic t_n, of the hypothesis $\theta = \theta_0$ against the hypothesis

(2.2.1.) $$H_1 : \quad \theta_n = \theta_0 + k n^{-\frac{1}{2}} \qquad k > 0$$

We shall assume

(2.2.2.)

1. $\{\partial/\partial\theta \; \mu_n(\theta)\}$ exists and is greater than zero in some neighbourhood of $\theta = \theta_0$

2. $$\lim_{n\to\infty} \left\{ \frac{n^{-1/2}\, \partial/\partial\theta \; \mu_n(\theta)}{\sigma_n(\theta)} \right\}_{\theta=\theta_0} = c > 0$$

3. $$\lim_{n\to\infty} \frac{\mu_n(\theta_n)}{\mu_n(\theta_0)} = 1 \;; \quad \lim_{n\to\infty} \frac{\sigma_n(\theta_n)}{\sigma_n(\theta_0)} = 1$$

4. If $F_n(x|\theta)$ the distribution function of $\dfrac{t_n - \mu_n(\theta)}{\sigma_n(\theta)}$ then $F_n(x|\theta_0)$ and $F_n(x|\theta_0 + kn^{-1/2})$ tend to the standard normal distribution as n tends to infinity.

If

$$\Phi(\lambda_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\lambda_\alpha}^{\infty} e^{-\frac{1}{2}x^2}\, dx = \alpha$$

then the level of significance of the test based on the critical region

(2.2.3.) $$t_n \geqslant t_n(\alpha) , \qquad \frac{t_n(\alpha) - \mu_n(\theta_0)}{\sigma_n(\theta_0)} = \lambda_\alpha$$

tends to α as n increases.

The power of the test against the hypothesis $\theta = \theta_n$ is $P_n(\lambda_n) \rightarrow \Phi(\lambda_n)$

(2.2.4.) $$\lambda_n = \frac{\{\sigma_n(\theta_0)\lambda_\alpha + \mu_n(\theta_0) - \mu_n(\theta_n)\}}{\sigma_n(\theta_n)}$$

But $\mu_n(\theta_n) = \mu_n(\theta_0 + k n^{-1/2}) = \mu_n(\theta_0) + k n^{-1/2}\{\frac{\partial}{\partial\theta}\mu_n(\theta)\}_{\theta=\theta'}$

where $\theta_0 < \theta' < \theta_n$,

so that

$$\lim_{n\to\infty} \lambda_n = \lim_{n\to\infty} \frac{\sigma_n(\theta_0)\lambda_\alpha - k n^{-1/2}\{\frac{\partial}{\partial\theta}\mu_n(\theta)\}_{\theta=\theta'}}{\sigma_n(\theta_n)} = \lambda_\alpha - kc.$$

Then

(2.2.5.) $$\lim_{n\to\infty} P_n(\lambda_n) = \Phi(\lambda_\alpha - kc)$$

This proof is substantially the same as that given by Pitman (1948). Condition 4 is due to Noether (1955), (Pitman's corresponding condition required uniform convergence in θ), who also generalised the proof to the case where $\mu_n^{(j)}(\theta)$ is zero for $j = 0, \cdots, m$.

In the cases in which the criterion (to be defined below) is used in this thesis the validity of Pitman's theorem will be assumed. The statistics will all be regression coefficients or correlation coefficients and it seems that the limiting oscillations of the sequence of (cumulative) distribution functions of the corresponding statistics $\frac{t_n - \mu_n}{\sigma_n}$ will vanish at $\lambda_\alpha - kc$ in the sense that the limit of the oscillations of this sequence of distribution functions, in some neighbourhood of $\lambda_\alpha - k c$, will tend to zero with the neighbourhood. This will ensure the validity of Pitman's thoerem.

If the sequence of alternative hypotheses is given by

$$\theta_n = \theta_0 + k n^{-m} .$$

then (2.2.4.) becomes

$$\frac{\sigma_n(\theta_0)\lambda_\alpha - k n^{-m}\{\partial/\partial\theta\, \mu_n(\theta)\}_{\theta=\theta'}}{\sigma_n(\theta_n)}$$

and from (2.2.2.), condition 2, this converges to λ_α or diverges to $(-\infty)$ according as m is greater than or less than $\frac{1}{2}$. The power of the test, therefore, converges to α or unity so that it is only the case m = $\frac{1}{2}$ which is of interest.

Consider two test statistics t_{1n} and t_{2n} satisfying the condition of Pitman's theorem and having means $\mu_{1n}(\theta)$ and $\mu_{2n}(\theta)$ and variances $\sigma_{1n}^2(\theta)$ and $\sigma_{2n}^2(\theta)$ and with

(2.2.6.) $$\lim_{n\to\infty}\left\{\frac{n^{-1/2}\,\partial/\partial\theta\,\mu_{in}(\theta)}{\sigma_{in}(\theta)}\right\}_{\theta=\theta_0} = c_i \qquad (i=1,2)$$

Now if we have two sequences of alternatives,

(2.2.7.) $$H_{1i}: \quad \theta_0 + k_i n^{-1/2} \qquad (i=1,2)$$

then from (2.2.5.), (2.2.6.) and (2.2.7.) the two tests have, asymptotically, the same power against the same alternatives if

$$k_1 c_1 = k_2 c_2 \quad ; \quad k_1 n_1^{-1/2} = k_2 n_2^{-1/2}$$

$$\text{i.e.} \quad \frac{n_1}{n_2} = \left(\frac{c_2}{c_1}\right)^2 = \lim_{n\to\infty} \left\{ \frac{[\partial/\partial\theta\, \mu_{2n}(\theta)]^2/\sigma_{2n}^2(\theta)}{[\partial/\partial\theta\, \mu_{1n}(\theta)]^2/\sigma_{1n}^2(\theta)} \right\}_{\theta=\theta_0}$$

The ratio $\left(\frac{c_2}{c_1}\right)^2$ will be called the asymptotic relative efficiency of the test based on t_2 relative to that based on t_1 against the hypothesis specified by the parameter θ and will be indicated by

(2.2.8.) $\quad E(t_2, t_1 \mid \theta = \theta_0)$

The quantities $\lim_{n\to\infty} \left\{[\partial/\partial\theta\, \mu_{in}(\theta)]^2/\sigma_{in}^2(\theta)\right\}_{\theta=\theta_0}$ are called the efficacies of the statistics t_i, $(i=1,2)$.

The quantity (2.2.8.) evidently gives the asymptotic ratio of the sample sizes required for the two tests to give the same power against sequences of alternatives which are asymptotically identical and which tend to the null hypothesis at the same rate as $n^{-1/2}$ tends to zero. Asymptotically, for statistics with variance n^{-1}, these are the only alternatives of interest.

This criterion is, of course, due to Pitman (1948).

In order to make use of the criterion it is only necessary to know the variance on the null hypothesis and

(2.2.9.) $\quad \lim_{n\to\infty} \left\{\frac{\partial}{\partial\theta} \mathcal{E}(t_{in})\right\}_{\theta=\theta_0}$

Often it is simpler to evaluate $\left\{\frac{\partial}{\partial\theta}\left[\lim_{n\to\infty} \mathcal{E}(t_{in})\right]\right\}_{\theta=\theta_0}$ but then it will be necessary to show that the two limits are the same. This will certainly be so if the convergence of $\left\{\frac{\partial}{\partial\theta}\mathcal{E}(t_{in})\right\}$ is uniform in θ, at least in some neighbourhood of $\theta = \theta_0$.

As mentioned before the statistics with which we will be concerned in this thesis will all be reducible to correlation coefficients, $r(\underset{\sim}{x},\underset{\sim}{y})$, computed from vectors of observations $\underset{\sim}{x}, \underset{\sim}{y}$, having a joint normal distribution with covariance matrix $\Gamma(\theta)$. The limit (2.2.9.) can then always be evaluated directly (see for example Chapter 5). Nevertheless in what follows the limit will often be evaluated as

(2.2.10.) $$\left[\frac{\partial}{\partial\theta}\left\{\lim_{n\to\infty}\mathcal{E}(t_{in})\right\}\right]_{\theta=\theta_0}$$

to save the presentation of more extensive algebra.

2.3. An extension of Pitman's theorem

Let t_{1n} and t_{2n} be two statistics with means $\mu_{1n}(\theta)$ and $\mu_{2n}(\theta)$, variances $\sigma^2_{1n}(\theta)$ and $\sigma^2_{2n}(\theta)$ and correlation $\rho_n(\theta)$. We shall presume that t_{1n} and t_{2n} have distributions satisfying the conditions 1 to 3 of # 2.2 (with $c = c_1, c_2$) and in addition

3'. $$\lim_{n\to\infty}\frac{\rho_n(\theta_n)}{\rho_n(\theta_0)} = 1 \; ; \quad \theta_n = \theta_0 + kn^{-1/2}.$$

while 4 will be replaced by

4'. If $F_n(x,y|\theta)$ is the joint distribution function of $$\frac{t_{1n}-\mu_{1n}(\theta)}{\sigma_{1n}(\theta)}\;,\;\left\{\frac{t_{2n}-\mu_{2n}(\theta)}{\sigma_{2n}(\theta)} - \rho_n(\theta)\frac{t_{1n}-\mu_{1n}(\theta)}{\sigma_{1n}(\theta)}\right\}\left\{1-\rho_n^2(\theta)\right\}^{-1/2}$$

then the sequences $F_n(x,y|\theta_0)$ and $F_n(x,y,\theta_n)$ both converge to the bivariate normal distribution with zero means, unit variances and zero correlation.

If 4' holds then the distribution of

$$(2.3.1.)\quad R_n(\theta)=\left\{\left[\frac{t_{1n}-\mu_{1n}(\theta)}{\sigma_{1n}(\theta)}\right]^2+\left[\frac{t_{2n}-\mu_{2n}(\theta)}{\sigma_{2n}(\theta)}\right]^2-2\rho_n(\theta)\left[\frac{t_{1n}-\mu_{1n}(\theta)}{\sigma_{1n}(\theta)}\cdot\frac{t_{2n}-\mu_{2n}(\theta)}{\sigma_{2n}(\theta)}\right]\right\}(1-\rho_n^2(\theta))^{-1}$$

converges for $\theta=\theta_0$ and $\theta=\theta_n$ to the distribution of χ^2 with two degrees of freedom.

We may test the hypothesis $\theta=\theta_0$ by means of the critical region

$$R_n(\theta_0)\geqslant\chi^2_\alpha>0$$

where

$$\frac{1}{2}\int_{\chi^2_\alpha}^{\infty}e^{-\frac{1}{2}\chi^2}d\chi^2=\alpha$$

Then the power of this test against the alternative $\theta_n=\theta_0+kn^{-1/2}$ will be:

$$\int_{\chi^2_\alpha}^{\infty}G_n(R_n(\theta_0)|\theta_n)\,dR_n(\theta_0)$$

where $G_n(R_n(\theta_0)|\theta_n)$ is the frequency function of $R_n(\theta_0)$ when $\theta=\theta_n$. But from 3, 3', 4' and (2.3.1.) this is easily seen to converge to the distribution of non central χ^2 where the distance, d, from the origin is given by:

$$d^2=\lim_{n\to\infty}\left\{\left[\frac{\mu_{1n}(\theta_n)-\mu_{1n}(\theta_0)}{\sigma_{1n}(\theta_0)}\right]^2+\left[\frac{\mu_{2n}(\theta_n)-\mu_{2n}(\theta_0)}{\sigma_{2n}(\theta_0)}\right]^2-2\rho_n(\theta_0)\left[\frac{\mu_{1n}(\theta_n)-\mu_{1n}(\theta_0)}{\sigma_{1n}(\theta_0)}\cdot\frac{\mu_{2n}(\theta_n)-\mu_{2n}(\theta_0)}{\sigma_{2n}(\theta_0)}\right]\right\}(1-\rho_n^2(\theta_0))^{-1}$$

and from 2 this limit is:

$$(2.3.2.)\qquad k^2\,\frac{c_1^2+c_2^2-2\rho c_1c_2}{1-\rho^2}$$

where $\rho=\lim_{n\to\infty}\rho_n(\theta_0)$

Suppose now that we have two pairs of statistics $t_{1,1,n}$, $t_{1,2,n}$ and $t_{2,1,n}$, $t_{2,2,n}$ with means $\mu_{ijn}(\theta)$ variances $\sigma^2_{ijn}(\theta)$ and correlations (between the same pair) $\rho_{in}(\theta)$, $(i, j = 1, 2)$ and consider the sequences of alternatives:

$$H_i: \quad \theta = \theta_0 + k_i n^{-1/2}$$

Then if we put:

$$(2.3.3.) \qquad c_{ij} = \lim_{n\to\infty} \left\{ \frac{n^{-1/2}\, \partial/\partial\theta\, \mu_{ijn}(\theta)}{\sigma_{ijn}(\theta)} \right\}^2_{\theta=\theta_0}$$

and

$$(2.3.4.) \quad R_i(\theta_0) = \left\{ \left[\frac{t_{i1n}-\mu_{i1n}(\theta_0)}{\sigma_{i1n}(\theta_0)}\right]^2 + \left[\frac{t_{i2n}-\mu_{i2n}(\theta_0)}{\sigma_{i2n}(\theta_0)}\right]^2 - 2\rho_{in}(\theta_0)\left[\frac{t_{i1n}-\mu_{i1n}(\theta_0)}{\sigma_{i1n}(\theta_0)} \cdot \frac{t_{i2n}-\mu_{i2n}(\theta_0)}{\sigma_{i2n}(\theta_0)}\right]\right\}(1-\rho^2_{in}(\theta_0))$$

the tests of $\theta = \theta_0$ based on the R_i will have equal power against the same sequences of alternatives if:

$$k_1 n_1^{-1/2} = k_2 n_2^{-1/2}$$

and

$$k_1^2 \frac{c_{11}^2 + c_{12}^2 - 2\rho_1 c_{11} c_{12}}{1-\rho_1^2} = k_2^2 \frac{c_{21}^2 + c_{22}^2 - 2\rho_2 c_{21} c_{22}}{1-\rho_2^2}$$

where

$$\rho_i = \lim_{n\to\infty} \rho_{in}(\theta_0)\ ; \quad i = 1, 2.$$

Then

$$(2.3.5.) \qquad \frac{n_1}{n_2} = \frac{k_2^2}{k_1^2} = \frac{(c_{11}^2 + c_{12}^2 - 2\rho_1 c_{11} c_{12})(1-\rho_1^2)^{-1}}{(c_{21}^2 + c_{22}^2 - 2\rho_2 c_{21} c_{22})(1-\rho_2^2)^{-1}}$$

and this ratio can be computed from a knowledge of limit of the covariance matrix of the t_{i1} and t_{i2} at $\theta = \theta_0$ and the quantities $\left\{\lim_{n\to\infty} \partial/\partial\theta\, \mu_{ijn}(\theta)\right\}_{\theta=\theta_0}$ only.

As before the ratio $\frac{n_1}{n_2}$ may be called the asymptotic relative efficiency of the two tests and may be denoted by:

$E\{R_2, R_1 | \theta = \theta_0\}$

A number of comments may be made here.

(1) It is evident that this result may be directly generalised to any finite number of statistics t_{ij}, $(i=1,2; j=1,\cdots,p)$ and that the resulting asymptotic relative efficiency will be the ratio of two quadratic forms in the "efficacies" of the statistics in each set, the matrices of the quadratic forms being the inverses of the correlation matrices of the statistics in each set.

(2) This result has an immediate application to the comparison of a pair of multiple correlations for when the t_{ij} are regression coefficients the quantities R_i differ from the quadratic form in the regression coefficients in the numerator of the test statistic (used for the test of significance of a multiple correlation) only in that the sample variances used to weight the quadratic form have been replaced by the true variances on the null hypothesis. Asymptotically (for the sequences of alternatives considered) this replacement will have no effect.

(3) One cannot, however, use the ratio of the quadratic form (Q) in the efficacies of a set of t_j $(j=1,\cdots,p)$ to the efficacy of a statistic t (which satisfies the conditions of the original theorem of # 2.2) as a criterion of the relative power of the two tests. For

this relative power depends then not only on the quantities in the ratio but also on the form of the two limiting distributions. In fact for the set t_j the asymptotic power will be obtained from the distribution of non central χ^2 for p degrees of freedom while that for the statistic t will be obtained from the normal distribution. The two powers will be, asymptotically, $\bar{\Phi}(\lambda_\alpha - kc)$ and $G_p(k^2Q, \alpha)$ where $\bar{\Phi}(\lambda_\alpha - kc)$ has already been defined and $G_p(k^2Q, \alpha)$ is the power function of the χ^2 test when the square of the distance of the alternative from the origin is k^2Q, α is the level of significance and p is the number of degrees of freedom. The two power functions are no longer of the same analytical form and cannot, therefore, be made the same function by a proper choice of the limiting ratio of the sample sizes. For fixed k, p and α one can of course determine the two powers as functions of c and Q (making use of Tang's (1938) tables of the power function of the F test) but as k varies the relation between the two power functions will vary also.

(4) The same situation (as in (3) above) arises when tests based on two multiple correlations from different numbers of regressors are being compared.

Chapter 3.

An Exact Test for Serial Correlation.

3.1. Simple Markoff Processes

The fundamental property of a simple Markoff process is embodied in the relation

(3.1.1.) $P\{x_t \leq k \mid x_{t_1}, x_{t_2}, \cdots, x_{t_m}\} = P\{x_t \leq k \mid x_{t_1}\}$

where $t > t_1 > t_2 \cdots > t_m$.

It follows easily that if n is odd and the variates $x_{2t+1}, (t = 0, \cdots, \frac{n-1}{2})$ are held fixed the joint conditional density of the x_{2t} factors into the several conditional densities of each x_{2t} for fixed x_{2t-1}, x_{2t+1}.

If the process is

(3.1.2.) $x_t - \mu = \rho(x_{t-1} - \mu) + \varepsilon_t$

where $|\rho| < 1$ and ε_t is $N(0, \sigma^2)$ then the conditional frequency function of x_{2t} is

(3.1.3.) $f(x_{2t} \mid x_{2t-1}, x_{2t+1}) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp - \frac{1}{2\sigma_0^2}\left\{x_{2t} - a - bx_t'\right\}^2$

where

$$a = \frac{\mu(1-\rho)^2}{1+\rho^2}$$

$$b = \frac{\rho}{1+\rho^2}$$

$$\sigma_0^2 = \frac{\sigma^2(1-\rho^2)}{1+\rho^2}$$

$$x_t' = (x_{2t-1} + x_{2t+1})$$

This was pointed out by Ogawara (1951) who then observed that, from standard normal regression theory, the statistic

$$(3.1.4.)\quad \hat{b}_1 = \frac{\sum(x_t' - \bar{x}')(x_{2t} - \bar{x}_2)}{\sum(x_t' - \bar{x}')^2}$$

could be used to test a prior hypothesis concerning ρ or to obtain a confidence interval for ρ.

In (3.1.4.) the summations are from 1 to $\left[\frac{n-1}{2}\right]$ where $[x]$ means the integral part of x (it is no longer assumed that n is odd), and

$$(3.1.5.)\quad \bar{x}' = \left[\frac{n-1}{2}\right]^{-1}\sum x_t' \quad , \quad \bar{x}_2 = \left[\frac{n-1}{2}\right]^{-1}\sum x_{2t}$$

The test statistic for an hypothetical value ρ_0 corresponding to $b = b_0$, would of course be

$$F_{1,\left[\frac{n-1}{2}\right]-2} = \frac{(\hat{b}_1 - b_0)^2 \sum(x_t' - \bar{x}')^2}{\sum(x_{2t} - \hat{a} - \hat{b}_1 x_t')^2}\left\{\left[\frac{n-1}{2}\right] - 2\right\}$$

Here α is estimated, in the usual way, by

$$(3.1.6.)\quad \hat{\alpha} = \bar{x}_2 - \hat{b}_1 \bar{x}'$$

The variance of $\hat{b}_1$ in the conditional distribution of the x_{2t} is

$$(3.1.7.)\quad \frac{\sigma^2(1-\rho^2)}{1+\rho^2}\,\frac{1}{\sum(x_t' - \bar{x}')^2}$$

and as this converges in probability as n increases to

$$(3.1.8.)\quad \frac{\sigma^2(1-\rho^2)}{1+\rho^2}\,\frac{1}{n\sigma^2(1+\rho^2)} = \frac{1-\rho^2}{n(1+\rho^2)^2}$$

(since $\sum(x_t'-\bar{x}')^2$ is almost everywhere greater than zero) the conditional variance of $\hat{b}_1$ converges to (3.1.8.) also.

We then have

(3.1.9.) $$\lim_{n \to \infty} \left\{ \frac{[\partial/\partial\rho \, \mathcal{E}(\hat{b}_1)]^2}{\mathrm{Var}(\hat{b}_1)} \right\} = n \frac{1-\rho^2}{(1+\rho^2)^2}$$

The first serial correlation coefficient of the x_t, r_1, (see (1.2.9.)) is well known to have the asymptotic variance $\frac{1-\rho^2}{n}$ and this will clearly be asymptotically fully efficient. (Wald (1948), Whittle (1953)). The asymptotic efficiency of $\hat{b}_1$, relative to r_1 is

(3.1.10.) $$E(\hat{b}_1, r_1 / \rho = 0) = \left(\frac{1-\rho^2}{1+\rho^2}\right)^2$$

This function is shown for some values of ρ in table 1 below.

Table 1.

$E(\hat{b}_1, r_1 / \rho = 0)$

$\lvert\rho\rvert$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$E(\hat{b}_1, r_1 / \rho=0)$	1.00	0.96	0.85	0.70	0.52	0.36	0.22	0.12	0.05	0.01

The statistic $\hat{b}_1$ therefore provides an exact test of the hypothesis $\rho = 0$, which is asymptotically fully efficient against the hypothesis of a simple Markoff process.

It also provides an exact test of an hypothesis

(3.1.11.) $H_0: \rho = \rho_0 \neq 0$; x_t Markovian, but the asymptotic efficiency of this test is less than unity.

In this derivation it has been presumed that

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}E(r_1)\right\} = \frac{\partial}{\partial\rho}\left\{\lim_{n\to\infty}E(r_1)\right\} = 1$$

but this can easily be verified, for example, by a direct evaluation of the first limit.

If the null hypothesis is that given by (3.1.11.) another procedure would be to use the quantities

(3.1.12.) $$y_t = x_t - \rho_0 x_{t-1}$$

On the null hypothesis the variance of $\hat{b}_y$, where this statistic is given by (3.1.4.) with the x_t replaced by the corresponding y_t, will be n^{-1}. It is also easy to show that now

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}E(\hat{b}_y)\right\} = 1$$

Hence the asymptotic efficiency of the test based on this statistic is,

(3.1.13.) $$E(\hat{b}_y, r_1 | \rho) = 1 - \rho^2$$

which is shown in Table 2.

Table 2.

$E(\hat{b}_y, r_1 | \rho)$

$\lvert\rho\rvert$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$E(\hat{b}_y, r_1 \lvert \rho)$	1	0.99	0.96	0.91	0.84	0.75	0.64	0.51	0.36	0.19

The statistic $\hat{b}_y$ therefore provides a more efficient test of the hypothesis (3.1.11.) than $\hat{b}_1$. It will also provide a confidence interval for ρ. However, it does not provide an unbiassed estimator of a known function of ρ as does $\hat{b}_1$. As mentioned in # 1.2 (page 18) the presence of bias prevents the averaging of a number of estimates, of the same value of ρ , from otherwise differing samples.* It is, therefore, of interest to see if another unbiassed estimator of $\frac{\rho}{1+\rho^2}$ can be obtained, with a higher efficiency than $\hat{b}_1$.

A second estimator is available in the statistic

(3.1.14.) $$\hat{b}_2 = \frac{\Sigma x_{2t+1}(x_t'' - \bar{x}'')}{\Sigma(x_t'' - \bar{x}'')^2}$$

where $x_t'' = (x_{2t} + x_{2t+2})$ and the summations are from 1 to $\left[\frac{n-3}{2}\right]$.

The statistic

(3.1.15.) $$\hat{b} = \tfrac{1}{2}(\hat{b}_1 + \hat{b}_2)$$

will be an unbiassed estimator of $\frac{\rho}{1+\rho^2}$ and its variance will be $\frac{1}{4}\,\mathrm{Var}(\hat{b}_1) + \frac{1}{4}\,\mathrm{Var}(\hat{b}_2) + \frac{1}{2}\,\mathrm{Cov}(\hat{b}_1, \hat{b}_2)$

which, as n increases, tends to

$$\frac{1-\rho^2}{2n(1+\rho^2)^2} + \frac{1}{2}\,\mathrm{Cov}(\hat{b}_1, \hat{b}_2).$$

To the order n^{-1} the sample means may be neglected

* An example (due to E.J. Williams) has arisen in connection with the serial correlation of measurements along a log. The averaging of estimates from different logs, the only method of substantially increasing the number of observations, would not result in a consistent estimator if the estimates were obtained from the first serial correlation coefficient.

and the covariance of $\hat{b}_1$ and $\hat{b}_2$ becomes (by an obvious extension of the Theorem in Cramer (1946) page 366)

$$\frac{1}{(1+\rho^2)^2} \operatorname{Cov}(a_1, a_2) + \frac{\rho^2}{(1+\rho^2)^4} \operatorname{Cov}(c_1, c_2) - \frac{\rho}{(1+\rho^2)^3}\left\{\operatorname{Cov}(a_1, c_2) + \operatorname{Cov}(c_1, a_2)\right\}$$

where

$$a_1 = \frac{2}{\sigma^2 n} \sum x_{2t}(x_{2t-1} + x_{2t+1})$$

$$a_2 = \frac{2}{\sigma^2 n} \sum x_{2t+1}(x_{2t} + x_{2t+2})$$

$$c_1 = \frac{2}{\sigma^2 n} \sum (x_{2t-1} + x_{2t+1})^2$$

$$c_2 = \frac{2}{\sigma^2 n} \sum (x_{2t} + x_{2t+2})^2$$

By straightforward algebra we obtain, to the order n^{-1}

$$(3.1.16.) \quad \operatorname{Cov}(\hat{b}_1, \hat{b}_2) = \frac{2}{n}\left\{\frac{1}{(1+\rho^2)^2}\left[\frac{2+10\rho^2+6\rho^4-2\rho^6}{1-\rho^4}\right] + \frac{\rho^2}{(1+\rho^2)^4}\left[\frac{40\rho^2+40\rho^4-8\rho^6-8\rho^8}{1-\rho^4}\right] - \frac{2\rho}{(1+\rho^2)^3}\left[\frac{12\rho+20\rho^3+4\rho^5-4\rho^7}{1-\rho^4}\right]\right\}$$

The asymptotic variance of $\hat{b}$ therefore is

$$\operatorname{Var} \hat{b} \left[\frac{1-\rho^2}{(1+\rho^2)^2}\right]$$

and the variance of the estimator of ρ obtained by putting

$$\hat{\rho} = (1 - \sqrt{1-4\hat{b}^2})/2\hat{b}$$

is $(1-\rho^2)$.

This function was shown in Table 2.

The statistic $\hat{b}_1$ provides a test of an hypothetical value of ρ which has some attractive features, but these are not sufficient to make its use often worthwhile in practise.

1) The test of the hypothesis, $\rho = 0$ is exact, is based on a distribution of which tables are widely available and is asymptotically fully efficient. Other tests, however, are also available which are exact (see (1.2,7.), (1.2.8.) and (1.2.10.). Moreover these last statistics clearly give tests which are not only asymptotically fully efficient but are near to the most powerful test against one sided alternatives (this is usually what is required) for all values of ρ on that side of the origin. Though the significance points of these statistics are not so readily available, for larger samples accurate approximations can be obtained from the tables of the significance points of the correlation coefficient.

2) The statistic $\hat{b}_y$ will give a test against an hypothesised value of ρ , not equal to zero, which will be exact and which will be reasonably efficient if the value of ρ, to be tested, is not too large. A more efficient test will be obtained from the statistic z, where $r_1' = \sin z$. (See # 1.2, page 19). However, the distribution of this test statistic is not known exactly and, because of the bias in r_1 when the mean is not known, and the effects of departure from circularity in the

universe, the treatment of z as normal with mean $\sin^{-1}\rho$ and variance n^{-1} may be a poor approximation unless n is quite large. An exact test of equal efficiency to that based on $\hat{b}_y$ will be obtained from any of (1.2.7.), (1.2.8.) or (1.2.10.) with the x_t replaced by the corresponding y_t and good approximations to the significance points can be obtained from readily available tables.

In any case the test of a value of ρ, not zero, will not often be required.

3) Exact confidence intervals for ρ can be obtained from $\hat{b}_1$, $\hat{b}_y$, or the statistic mentioned at the end of the second last paragraph. For n fairly large the statistic z will, presumably, give a reasonable approximation and, moreover, the confidence interval may be much narrower because of the low efficiency of $\hat{b}_1$, $\hat{b}_y$, &c., when ρ is large. On the other hand when ρ is large the confidence interval based on z will also become very inaccurate (in the sense that the confidence coefficient may be very far from the assumed value).

4) Finally $\hat{b}$ (see (3.1.15.) is an unbiassed estimator of $\frac{\rho}{1+\rho^2}$ and, will therefore enable estimates based on otherwise inhomogeneous material to be combined. Again, however, this situation will not often arise.

3.2. Multiple Markoff Processes

Following Doob (1953) a process generated by a relation of the form (1.2.2.) may be called a (stationary) multiple Markoff process. This nomenclature is suggested by the relation

(3.2.1.) $P\{x_t \le k \mid x_{t_1}, x_{t_2}, \cdots, x_{t_m}\} = P\{x_t \le k \mid x_{t_1}, x_{t_2}, \cdots x_{t_h}\}$

where $t > t_1 > t_2 \cdots > t_h > \cdots > t_m$ and x_t is generated by (1.2.2.)

As in 3.1. it follows that for fixed values of $x_{k(h+1)-j}, x_{k(h+1)+j}$, $(k = 1, \cdots, n; j = 1, \cdots h)$ the variates $x_{k(h+1)}$ $(k = 1, 2, \cdots, n)$ are independent. The conditional frequency function of $x_{k(h+1)}$ is then

(3.2.2.) $$f(x_{k(h+1)} \mid x_{k(h+1)-j}, x_{k(h+1)+j}\,;\ j = 1, \cdots h)$$
$$= (2\pi\sigma_0^2)^{-1} \exp - \frac{1}{2\sigma_0^2}\left\{x_{k(h+1)} - \sum_0^h b_j x'_{jk}\right\}$$

where

(3.2.3.) $$\left\{\begin{array}{l} x_{0k} \equiv 1 \quad ; \quad b_0 = \mu\left(1 - \sum_1^h b_j\right) \\ x'_{jk} = (x_{k(h+1)-j} + x_{k(h+1)+j})\,;\ (j = 1, \cdots, h) \\ \sigma_0^2 = \dfrac{1 + a_1\rho_1 + a_2\rho_2 + \cdots + a_h\rho_h}{1 + a_1^2 + a_2^2 + \cdots + a_h^2}\,\sigma^2 \\ \begin{bmatrix} b_h \\ b_{h-1} \\ \vdots \\ b_1 \\ b_1 \\ \vdots \\ b_h \end{bmatrix} = \begin{bmatrix} 1 & \cdots & \rho_{h-1} & \rho_{h+1} & \cdots & \rho_{2h} \\ \rho_1 & \cdots & \rho_{h-2} & \rho_h & \cdots & \rho_{2h-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{h-1} & \cdots & 1 & \rho_2 & & \rho_{h+1} \\ \rho_{h+1} & \cdots & \rho_2 & 1 & \cdots & \rho_{h-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{2h} & \cdots & \rho_{h+1} & \rho_{h-1} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_h \\ \rho_{h-1} \\ \vdots \\ \rho_1 \\ \rho_1 \\ \vdots \\ \rho_h \end{bmatrix} \end{array}\right.$$

Here ρ_j is the j^{th} serial correlation of the x_t.

This result is again due to Ogawara (1951).
The third and fourth relations in (3.2.3.) follow easily from the following well known result (Rao (1952) page 54).

If $x_1, \cdots, x_p$ follow a p variate normal distribution with the covariance matrix

$$\mathcal{E}\begin{Bmatrix} x_1 \\ \vdots \\ x_k \\ \hline x_{k+1} \\ \vdots \\ x_p \end{Bmatrix}(x_1 \cdots x_k | x_{k+1} \cdots x_p) = \mathcal{E}\begin{Bmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{Bmatrix}(\underset{\sim}{x}_1' \; \underset{\sim}{x}_2') = \begin{bmatrix} \underset{\sim}{A} & \underset{\sim}{B} \\ \underset{\sim}{B}' & \underset{\sim}{C} \end{bmatrix}$$

Then for fixed $x_1, \cdots, x_k$ the conditional distribution of $x_{k+1}, \cdots, x_p$ is normal with covariance matrix

(3.2.4.) $$\underset{\sim}{C} - \underset{\sim}{B}'\underset{\sim}{A}^{-1}\underset{\sim}{B}$$

and expected value

(3.2.5.) $$\underset{\sim}{B}'\underset{\sim}{A}^{-1}\underset{\sim}{x}_1$$

The least squares estimate of b_h in the frequency function (3.2.2.) will provide a test of the hypothesis that the x_t process is of order (h - 1) against the hypothesis that it is of order h, since it is easily seen that b_h is zero when a_h is zero.

The variance of the estimator, $\hat{b}_h$, of b_h is, when a_h is zero and n observations are available,

$$Var(\hat{b}_h)_{a_h=0} = \left[\frac{n-h}{h+1}\right]^{-1} \frac{\sum_0^{h-1} a_j \rho_j}{\sum_0^{h-1} a_j^2} \cdot \sigma^2 \cdot C_{hh}$$

Here $a_0 = 1$ and C_{hh} is the element in the last row and column of the covariance matrix of the x'_{jk} $(j=1,\cdots,h)$

This matrix converges in probability to

$$2\sigma^2 \begin{bmatrix} 1+\rho_1 & \rho_1+\rho_3 & \cdots & \rho_{h-1}+\rho_{h+1} \\ & 1+\rho_4 & \cdots & \rho_{h-3}+\rho_{h+3} \\ & & & \\ & & & 1+\rho_{2h} \end{bmatrix}$$

Using the relations

(3.2.6.) $$\sum_0^{h-1} a_j \rho_{|k-j|} = 0 \qquad k>0$$

we easily obtain

(3.2.7.) $$\left\{\lim_{n\to\infty} \operatorname{Var} b_h\right\}_{a_h=0} = \left\{\frac{2n}{h+1}\sum_0^{h-1} a_j^2\right\}^{-1}$$

From the fourth relation in (3.2.3.) it is easily seen that

(3.2.8.) $$\mathcal{E}(\hat{b}_h) = b_h = \frac{\left\{\sum_0^{h-1} a_j \rho_j\right\}^h \Delta_1}{\Delta_{h+1,h+1}}$$

Here $\Delta_{h+1,h+1}$ is the determinantal value of the matrix whose inverse appears on the r.h.s. of the third relation in (3.2.3.) while

$$\Delta_1 = \begin{vmatrix} \rho_h & \rho_{h-1} & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{h-2} \\ \rho_2 & \rho_1 & \cdots & \rho_{h-3} \\ \cdot & \cdot & \cdots & \cdot \\ \rho_h & \rho_{h-1} & \cdots & \rho_1 \end{vmatrix}$$

From (3.2.6.) it is seen that Δ_1 vanishes when

a_h is zero so that to evaluate $\left(\frac{\partial b_h}{\partial a_h}\right)_{a_h=0}$ we have only to evaluate $\left(\frac{\partial \Delta_1}{\partial a_h}\right)_{a_h=0}$

But the relation

$$\begin{Bmatrix} a_1 \\ \vdots \\ a_h \end{Bmatrix} = -\begin{bmatrix} 1 & \cdots & \rho_{h-1} \\ \rho_{h-1} & \cdots & 1 \end{bmatrix} \begin{Bmatrix} \rho_1 \\ \vdots \\ \rho_h \end{Bmatrix}$$

follows from (3.2.6.) and differentiating both sides with respect to a_h and putting a_h equal to zero we obtain

$$\left(\frac{\partial \Delta_1}{\partial a_h}\right)_{a_h=0} = -\begin{vmatrix} 1 & \cdots & \rho_{h-1} \\ \rho_{h-1} & \cdots & 1 \end{vmatrix}$$

so that (3.2.8.) becomes, when a_h is zero

$$(3.2.9.) \quad \frac{-\left\{\sum_0^{h-1} a_j \rho_j\right\}^h \begin{vmatrix} 1 & \cdots & \rho_{h-1} \\ \rho_{h-1} & \cdots & 1 \end{vmatrix}}{\Delta_{h+1,\,h+1}} = -\sum_0^{h-1} a_j \rho_j \frac{\Delta}{\Delta_{h+1,\,h+1}}$$

where

$$\Delta = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{2h} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{2h-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \rho_{2h} & \rho_{2h-1} & \rho_{2h-2} & \cdots & 1 \end{vmatrix}$$

But $\Delta/\Delta_{h+1,\,h+1}$ is easily seen to be $\left(\sum_0^{h-1} a_j \rho_j\right)\left(\sum_0^{h-1} a_j^2\right)^{-1}$ so that (3.2.9.) becomes

$$(3.2.10.) \quad \left\{\frac{\partial b_h}{\partial a_h}\right\}_{a_h=0} = -\left\{\sum_0^{h-1} a_j^2\right\}^{-1}$$

Then we have from (3.2.7.) and (3.2.10.):

$$\lim_{n\to\infty}\left\{\frac{\left[\frac{\partial E(\hat{b}_k)}{\partial a_k}\right]^2}{V(\hat{b}_k)}\right\} = \frac{2n}{k+1}\left\{\sum_{0}^{k-1} a_j^2\right\}^{-1}$$

A most efficient test will clearly be obtained from the partial serial correlation coefficient $r_{1k\cdot 2,\cdots,k-1}$ for which the limit of the corresponding ratio will be n. We may, therefore, consider:

(3.2.11.) $$E\left(\hat{b}_k, r_{1k\cdot 2,\cdots,k-1} \mid a_k = 0\right) = \frac{2}{k+1}\left\{\sum_{0}^{k-1} a_j^2\right\}^{-1}$$

The test based on the statistic $\hat{b}_k$ is, therefore, always very inefficient for $k > 1$.

The other tests of interest which result from (3.2.2.) are:

(1) The test of independence of the x_t based on the multiple correlation between $x_{k(k+1)}$ and the x'_{jk} $(j = 1, \cdots, k)$. As shown in # 2.3. the criterion of the asymptotic efficiency of a test can be applied to a multiple correlation. In the present case it is clear, however, that the test based on this statistic will be very inefficient compared with the test based on the multiple correlation between x_t and the set $(x_{t-1}, x_{t-2}, \cdots x_{t-k})$. The distribution of this last statistic is of course known only very approximately. It

would be possible, under some circumstances, for the multiple correlation which follows from (3.2.2.) to give a more powerful test than that based on the first serial correlation coefficient, of course, but these circumstances are, perhaps, very unlikely to occur and even less likely to be recognised a priori.

(2) The test based on the partial multiple correlation between the $x_{k(k+1)}$ and the x'_{jk} $(j=p,\dots,k)$ with the effects of the x'_{jk} $(j=1,\dots,p-1)$ removed. The same comments as those made under (1) above may be made again here.

3.3. A critique of the method

The method which has been used in the two previous sections is essentially that of dividing the observations into two sets so that when one set is held fixed the other becomes a set of independent variates and the test statistic can be taken as a correlation between the two sets.

It is somewhat misleading to regard the loss of information as coming from the fixing of variates. For one thing test statistics can be taken whose distribution is exact when the variates are not fixed. An example is the correlation between x_{2t} and $x_{2t-1}+x_{2t+1}$ in # 3.1. which, on the null hypothesis, is distributed as an ordinary correlation from $\left[\frac{n-1}{2}\right]$ observations whether the x_{2t+1} are fixed or not.

A second objection to this point of view arises from the fact that by fixing fewer variates the tests may become less efficient. An example here is the fixing of x_{4t+1} and the use of the partial correlation between x_{4t} and x_{4t-2}, with the effects of x_{4t-1} removed, as a test statistic against the hypothesis that the process is of the form (1.2.2.) with h = 1 (the alternative being that h = 2). Here only one observation in 4 will be fixed but the test will be even less efficient than that presented in 3.2. since the efficiency will clearly be no greater than $\frac{1}{4}$ compared with $\frac{2}{3}(1+a_1^2)^{-1}$ (which is never less than $\frac{1}{3}$) from the test in 3.2. (see (3.2.11.)).

The loss of accuracy clearly arises from the impossibility of using some cross products, in the case where $h > 1$. For example in the second order case x_{3t} must be in one set and x_{3t-2}, x_{3t+2} in the other, but no matter which set x_{3t-1} and x_{3t+1} are put the cross product between them cannot be used if an exact test is to be simply obtained.

In some cases the loss of information can be identified with the introduction of additional parameters, as will be seen in Chapter 6.

Chapter 4.

Testing for Serial Correlation in the Residuals from a Regression.

4.1. An exact test against a simple Markoff alternative.

The method used in the previous chapter can be extended to give a test for serial correlation in the residuals from a regression of the form (1.3.1.). The condition (1) under (1.3.1.) will be retained, but (2) will be replaced by

(2)′ the residuals are generated by a process of the form

(4.1.1.) $$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \eta_t$$

where $|\rho_1| < 1$ and η_t comes from a process of independent random variates with zero mean and variance $\sigma_1^2(1-\rho_1^2)$*

Now for fixed $\varepsilon_{2t+1}, (t=0,\cdots,[\frac{n-1}{2}])$ and fixed x_t the conditional distribution of the y_{2t} becomes

(4.1.2.) $$(2\pi\sigma_0^2)^{-[\frac{n-1}{2}]} \exp -\frac{1}{2\sigma_0^2}\sum_1^{[\frac{n-1}{2}]}\left\{y_{2t} - \gamma_0 - \sum_{j=1}^{2p+1}\gamma_j z_{jt}\right\}^2 \prod_1^{[\frac{n-1}{2}]} dy_{2t}$$

where

(4.1.3.) $$\begin{cases} \sigma_0^2 = \dfrac{\sigma_1^2(1-\rho_1^2)}{1+\rho_1^2} & ; \quad \gamma_0 = \beta_0(1-2\gamma_{p+1}) \\ \gamma_j = \beta_j & ; \quad z_{jt} = x_{j,2t} \qquad (j=1,\cdots,p) \\ \gamma_{p+1} = \dfrac{\rho_1}{1+\rho_1^2} & ; \quad z_{p+1,t} = (y_{2t-1}+y_{2t+1}) \\ \gamma_j = -\gamma_{p+1}\beta_j & ; \quad z_{jt} = (x_{j,2t-1}+x_{j,2t+1}) (j=p+2,\cdots,2p+1) \end{cases}$$

In the case where p = 1 the relation between y_{2t}

---

* The notation ρ_1 and σ_1^2 rather than ρ and σ^2 is used to conform with a satisfactory notation for later chapters.

and the X_{jt} can be expressed diagrammatically as follows;

$$\begin{tikzcd}
y_t & \overset{2t-1}{\circ} \arrow[r, "\gamma_2"] \arrow[d, "-\gamma_1=-\beta_1"'] & \overset{2t}{\circ} & \overset{2t+1}{\circ} \arrow[l, "\gamma_2"'] \arrow[d, "-\gamma_1=-\beta_1"] \\
x_t & \underset{2t-1}{\circ} & \underset{2t}{\circ} \arrow[u, "\gamma_1=\beta_1"'] & \underset{2t+1}{\circ}
\end{tikzcd}$$

The fixing of the variates y_{2t-1} and y_{2t+1} directly affects the probability distribution of y_{2t} (with an influence measured by γ_2) and the fixing of x_{2t} influences y_{2t} directly also (with an effect measured by γ_1 which equals β_1). The fixing of x_{2t-1} and x_{2t+1} also influences the effects of y_{2t-1} and of y_{2t+1} on y_{2t} (this influence is measured by the product of $-\gamma_1$, and γ_2). The diagram is merely suggestive of course.

If the information involved in the knowledge that $\gamma_{j+p+1} = -\gamma_j\gamma_{p+1}, (j=1,\dots,p)$ is discarded, and the coefficients are estimated by the straightforward least squares procedure, these estimates will have the usual properties of least squares estimates when the residuals are normally distributed. They will not be maximum likelihood estimates but the computation of maximum likelihood estimates will involve the solution of non linear likelihood equations and exact tests will not be readily available.

Representing least squares estimates of the γ_j by $\hat{\gamma}_j$ the hypothesis, $\rho_1 = 0$, can be tested by the use of the statistic

$$(4.1.4.)\quad F_{1,\left[\frac{n-1}{2}\right]-2p-2} = \left\{\left[\tfrac{n-1}{2}\right]-2p-2\right\} \frac{\hat{\gamma}_{p+1}^2}{\sigma^{*2}} \frac{|L|}{|L_{p+1,p+1}|}$$

which has the F distribution with 1 and $[\frac{n-1}{2}]-2p-2$ degrees of freedom (Cramer (1946) page 553). Here σ^{*2} is the variance of the residuals from the least squares regression of y_{2t} on the z_{jt} while $\underset{\sim}{L}$ is the covariance matrix of the z_{jt} and $|\underset{\sim}{L}_{p+1,p+1}|$ is the cofactor of the element in the indicated row and column. The test is, of course, equivalent to the test of the partial correlation of y_{2t} and $(y_{2t-1}+y_{2t+1})$ with the effects of the $z_{jt}, (j \neq p+1)$ removed.

Tests of significance and confidence limits for the parameters β_j can similarly be obtained from statistics such as

$$(4.1.5.) \quad F_{1,[\frac{n-1}{2}]-2p-2} = \left\{[\tfrac{n-1}{2}]-2p-2\right\}\left\{(\hat{\gamma}_j-\beta_j)^2 \frac{|\underset{\sim}{L}|}{|\underset{\sim}{L}_{jj}|}\right\}\sigma^{*-2}$$

while the multiple correlation between y_t and the x_{jt} may be tested by the statistic

$$(4.1.6.) \quad F_{p,[\frac{n-1}{2}]-2p-2} = \left\{\frac{[\frac{n-1}{2}]-2p-2}{p}\right\}\left\{\sum_{i,j=1}^{p} \hat{\gamma}_i \hat{\gamma}_j \frac{|\underset{\sim}{\Lambda}_{ij}|}{|\underset{\sim}{\Lambda}|}\right\}\sigma^{*-2}$$

Here $\underset{\sim}{\Lambda}$ is the matrix formed from the first p rows and columns of $\underset{\sim}{L}^{-1}$ so that $\frac{\sigma_0^2}{[\frac{n-1}{2}]}\underset{\sim}{\Lambda}^{-1}$ is the covariance matrix of the $\hat{\gamma}_j, (j=1,\cdots,p)$

In order to examine the limiting variances and covariances of the estimates of the $\hat{\gamma}_j$ it is necessary to obtain the stochastic limit of the matrix $\underset{\sim}{L}$. If the vector $\{z_{jt}\}$ is partitioned as

$$\{z_{jt}\}' = \{\underset{\sim}{V}_{1t}', z_{p+1,t}, \underset{\sim}{V}_{2t}'\}$$

where $\underset{\sim}{V}_{1t}' = \{z_{1t}, z_{2t}, \cdots, z_{pt}\}$, $\underset{\sim}{V}_{2t}' = \{z_{p+2,t}, \cdots\cdots, z_{2p+1,t}\}$

then the covariance matrix $\underset{\sim}{L}$ is

$$(4.1.7.)\quad \underset{\sim}{L} = \left[\frac{n-1}{2}\right]^{-1}\sum_{1}^{\left[\frac{n-1}{2}\right]}\begin{bmatrix} (\underset{\sim}{V}_{1t}-\bar{\underset{\sim}{V}}_1)\underset{\sim}{V}_{1t}' & (\underset{\sim}{V}_{1t}-\bar{\underset{\sim}{V}}_1)z_{p+1,t} & (\underset{\sim}{V}_{1t}-\bar{\underset{\sim}{V}}_1)\underset{\sim}{V}_{2t}' \\ (z_{p+1,t}-\bar{z}_{p+1})\underset{\sim}{V}_{1t}' & (z_{p+1,t}-\bar{z}_{p+1})z_{p+1,t} & (z_{p+1,t}-\bar{z}_{p+1})\underset{\sim}{V}_{2t}' \\ (\underset{\sim}{V}_{2t}-\bar{\underset{\sim}{V}}_2)\underset{\sim}{V}_{1t}' & (\underset{\sim}{V}_{2t}-\bar{\underset{\sim}{V}}_2)z_{p+1,t} & (\underset{\sim}{V}_{2t}-\bar{\underset{\sim}{V}}_2)\underset{\sim}{V}_{2t}' \end{bmatrix}$$

Here $\bar{\underset{\sim}{V}}_i$ is the vector of arithmetic means of the variates in $\underset{\sim}{V}_{it}$ $(i = 1, 2)$.

Since, from (4.1.3.) and (1.3.1.),

$$z_{p+1,t} = y_{2t-1} + y_{2t+1} = 2\beta_0 + \sum_{j=1}^{p}\beta_j(x_{j,2t-1} + x_{j,2t+1}) + (\varepsilon_{2t-1} + \varepsilon_{2t+1})$$
$$= 2\beta_0 + \sum_{j=1}^{p}\beta_j z_{p+j+1,t} + (\varepsilon_{2t-1} + \varepsilon_{2t+1})$$

we have

$$z_{p+1,t} = 2\beta_0 + \underset{\sim}{\beta}_1'\underset{\sim}{V}_{2t} + (\varepsilon_{2t-1} + \varepsilon_{2t+1})$$

where $\underset{\sim}{\beta}_1'$ is the vector $\{\beta_1 \cdots\cdots \beta_p\}$ *

Under fairly weak restrictions on the nature of the process generating the x_{jt}, the matrix (4.1.7.) will have a limit in probability of the form

$$(4.1.8.)\quad \underset{\sim}{V} = \begin{bmatrix} \underset{\sim}{V}_0 & \underset{\sim}{V}_1\underset{\sim}{\beta}_1 & \underset{\sim}{V}_1 \\ \underset{\sim}{\beta}_1'\underset{\sim}{V}_1' & \begin{matrix}2\sigma_1^2(1+\rho^2) \\ +\underset{\sim}{\beta}_1'\underset{\sim}{V}_2\underset{\sim}{\beta}_1\end{matrix} & \underset{\sim}{\beta}_1'\underset{\sim}{V}_2 \\ \underset{\sim}{V}_1' & \underset{\sim}{V}_2'\underset{\sim}{\beta}_1 & \underset{\sim}{V}_2 \end{bmatrix}$$

Here $\underset{\sim}{V}_0$, $\underset{\sim}{V}_1$ (and its transpose) and $\underset{\sim}{V}_2$ are the stochastic limits of the matrices in the corners of $\underset{\sim}{L}$. Since the determinantal value of $\underset{\sim}{L}$ will be almost everywhere different from zero (if the joint distribution of the

---

* β_0 is no longer included in $\underset{\sim}{\beta}_1$ as it was in $\underset{\sim}{\beta}$ in (1.3.2.).

is not singular) the matrix $\underset{\sim}{L}^{-1}$ will converge in probability to the inverse of the matrix just written.

By elementary transformation on (4.1.8.) it is easily seen that

$$\frac{|\underset{\sim}{L}|}{|L_{p+1,p+1}|} \rightarrow 2\sigma_1^2(1+\rho_1^2)$$

The variance of $\hat{\gamma}_{p+1}$ therefore converges to

$$\frac{2}{n}\,\frac{\sigma_1^2(1-\rho_1^2)}{1+\rho_1^2}\cdot\frac{1}{2\sigma_1^2(1+\rho_1^2)} = \frac{1-\rho_1^2}{n(1+\rho_1^2)^2}$$

Since $\hat{\gamma}_{p+1}$ is an unbiassed estimator of $\rho_1/(1+\rho_1^2)$ we have

(4.1.9.) $$\lim_{n\to\infty}\left\{\frac{[\partial/\partial\rho_1\,\mathcal{E}(\hat{\gamma}_{p+1})]^2}{\mathrm{Var}(\hat{\gamma}_{p+1})}\right\} = n\,\frac{1-\rho_1^2}{(1+\rho_1^2)^2}$$

which is the same as (3.1.9.)

The test statistic $d = \frac{n-1}{n}\frac{\delta^2}{s^2}$ used by Durbin and Watson is asymptotically equal to $2(1-r_1)$ (see 1.2.9.). The variance of r_1 will depend on the matrix $\underset{\sim}{X}$ (See (1.3.2.)) but from the general expressions for the variance, (see Durbin and Watson (1951) page 164), it can be seen that var (r_1) will converge to n^{-1}. The asymptotic relative efficiency of $\hat{\gamma}_{p+1}$ and d as tests of the hypothesis $\rho_1 = 0$ therefore is

(4.1.10.) $$E(\hat{\gamma}_{p+1}, d \mid \rho_1 = 0) = 1.$$

Again here we have put $\lim_{n\to\infty}\{\partial/\partial\rho_1\,\mathcal{E}(r_1)\}_{\rho_1=0} = \frac{\partial}{\partial\rho_1}\{\lim_{n\to\infty}\mathcal{E}(r_1)\}_{\rho_1=0}$ but again the first quantity can be evaluated directly, though the algebra becomes very cumbersome.

The likelihood equation of the observations is

$$\log L = k - \tfrac{1}{2}\log|\underset{\sim}{\Gamma_1}| - \tfrac{1}{2}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})'\underset{\sim}{\Gamma_1}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})$$

where

$$\underset{\sim}{\Gamma_1} = \begin{bmatrix} 1 & \rho_1 & \rho_1^2 & \cdots & \rho_1^{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_1^{n-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \rho_1^{n-1} & \rho_1^{n-2} & \rho_1^{n-3} & \cdots & 1 \end{bmatrix}$$

and $\underset{\sim}{X}\underset{\sim}{\beta}$ was defined after (1.3.2.).

It is easy to see that the quantities $\mathcal{E}\left\{\frac{\partial^2 \log L}{\partial\rho_1\partial\beta_i}\right\}_{\rho_1=0}$ and $\mathcal{E}\left\{\frac{\partial^2 \log L}{\partial\rho_1\partial\sigma_1^2}\right\}_{\rho_1=0}$ are zero and that

(4.1.11.) $$\mathcal{E}\left\{\frac{\partial^2 \log L}{\partial\rho_1^2}\right\}_{\rho_1=0} = \mathcal{E}\Big\{(n-1)\frac{1+\rho_1^2}{1-\rho_1^2} + \tfrac{1}{2}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})'\underset{\sim}{\Gamma_1}^{-1}\left(\tfrac{\partial^2}{\partial\rho_1^2}\underset{\sim}{\Gamma_1}\right)\underset{\sim}{\Gamma_1}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})$$
$$-(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})'\underset{\sim}{\Gamma_1}^{-1}\left(\tfrac{\partial}{\partial\rho_1}\underset{\sim}{\Gamma_1}\right)\underset{\sim}{\Gamma_1}^{-1}\left(\tfrac{\partial}{\partial\rho_1}\underset{\sim}{\Gamma_1}\right)\underset{\sim}{\Gamma_1}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})\Big\}_{\rho_1=0}$$

Since the elements in the diagonal of $\left(\partial^2/\partial\rho_1^2\,\underset{\sim}{\Gamma_1}\right)_{\rho_1=0}$ all vanish this is

$$-(n-1) - \frac{1}{(\sigma^2)^3}\mathcal{E}\left\{(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})'\left(\tfrac{\partial}{\partial\rho_1}\underset{\sim}{\Gamma_1}\right)^2(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})\right\}_{\rho_1=0}$$

But the elements in the leading diagonal of $\left(\frac{\partial}{\partial\rho_1}\underset{\sim}{\Gamma_1}\right)^2_{\rho_1=0}$ are $(\sigma_1^4, 2\sigma_1^4, 2\sigma_1^4, \cdots, 2\sigma_1^4, \sigma_1^4)$ so that (4.1.11.) becomes $(n-1)-(2n-2) = -(n-1)$

Thus the only non zero element in the row containing $-\mathcal{E}\left\{\frac{\partial^2 \log L}{\partial\rho_1^2}\right\}_{\rho_1=0}$ in the information matrix is that corresponding to $-\mathcal{E}\left\{\frac{\partial^2 \log L}{\partial\rho_1^2}\right\}_{\rho_1=0} = n-1$ so that the element of the inverse, which gives the asymptotic variance of the maximum likelihood estimator (Wald (1948), Whittle (1953)) is $(n-1)^{-1}$. It follows that as test statistics of the hypothesis $\rho_1 = 0$ both d and $\hat{\gamma}_{p+1}$ are asymptotically fully efficient.

The statistic $\hat{\gamma}_{p+1}$ appears to give the best test at present available against serial correlation in the residuals from a regression because:

(1) The test is exact, provided only that the conditions (1) and (2) are fulfilled. The test statistic is, fundamentally, a correlation so that it is likely that the distribution of the ϵ_t can diverge fairly far from normal without the validity of this test being much affected. The statistic $\hat{\gamma}_{p+1}$ is the only statistic at present available giving an exact test.

(2) The test requires the use of tables which are very generally available.

(3) The statistic $\hat{\gamma}_{p+1}$ appears at first sight to suffer from the disadvantage that additional computations will have to be carried out if the straightforward least squares procedure is to be applied (to estimate the β_j). However, it will be shown in Chapter 5 that, for ρ_1 high, the estimates of the β_j obtained from the $\hat{\gamma}_j$ may be more efficient than those obtained from least squares procedures. In any case both are inefficient. Only in those situations where the test of serial correlation

based on $\hat{\gamma}_{p+1}$ accepts the null hypothesis will more computations be necessary than would be the case with the statistic d (if it also gave a non significant result). Finally, in a proportion of those cases where the result is not significant d will fall between the bounds tabulated by Durbin and Watson so that additional computations will be necessary to obtain the mean and variance of the test statistic and, thence, the approximation to the significance point from the appropriate Beta distribution.

(4) On the other hand the criterion of efficiency which has been used is asymptotic and is only suggestive for medium sized or small samples. For really small samples the effects from the estimation of p additional parameters will be felt, $(\gamma_{p+2}, \ldots, \gamma_{2p+1})$. The integer p will, presumably, be small for small samples (though this may not be always so in econometrics).

4.2. Examples.

The following examples illustrate the use of this statistic. The results are given in more detail than is necessary for the purposes of this chapter since they will also be used in Chapter 5.

Example (1). A series of 30 terms was chosen from

Kendall (1946, 1949) series 1 and another from his series 9. The commencing terms were 341 and 235. The two series were, therefore, generated by the relations:

$$x_t - 1.1\, x_{t-1} + 0.5\, x_{t-2} = \eta_{1t}$$
$$\epsilon_t - 0.7\, \epsilon_{t-1} = \eta_{2t}$$

where in the first case the disturbance is rectangular and in the second case normal.

The series $y_t = 1.4\, x_t + \epsilon_t$ was then formed. The data required for the test based on $\hat{\gamma}_{p+1}$ (p = 1, in this case) is as follows:

$$(4.2.1.)\begin{bmatrix}\hat{\gamma}_1\\ \hat{\gamma}_2\\ \hat{\gamma}_3\end{bmatrix} = \begin{bmatrix} 2.030 & -0.001 & -1.008 \\ & 0.041 & -0.112 \\ & & 0.940 \end{bmatrix} \begin{bmatrix} 6.945 \\ 38.642 \\ 11.527 \end{bmatrix} = \begin{bmatrix} 2.423 \\ 0.305 \\ -0.480 \end{bmatrix}$$

$\sigma^{*2} = 0.683$ $\qquad \left[\frac{n-1}{2}\right] - 2p - 2 = 10$

The matrix on the right hand side is the inverse of the matrix of cross products, $\sum_1^{14} (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j)$

The test statistic is $t_{10} = \dfrac{\sqrt{10}\; 0.305}{\{0.683 \times 14 \times 0.041\}^{1/2}} = 1.541$

Using a one sided 5% test this is not significant, the significance point being 1.812.

From the straightforward least squares analysis the regression coefficient is 2.791. The ratio d proves to be $d = \dfrac{33.504}{23.275} = 1.44$. From Table 4 in Durbin and Watson (1951) the upper and lower bounds for a 5% (one sided) test are 1.35 and 1.49. The additional computations required for the carrying out of a test based on the Beta distribution have not been done. In practice the approximate significance point falls very near to half way between the bounds and the result in this case is fairly certain to be not significant. (A low value of d indicates positive serial correlation).

This example agrees fairly well with the suggestion from the asymptotic relative efficiency of the two tests. Both estimates of the regression coefficient are very inaccurate, but this is partly due to the low variance of x_t compared with that of y_t.

In this example the "end effects" from dropping one observation when n is even and from the estimation of additional parameters is evident for the degrees of freedom available for the estimation of the test statistic, instead of being in the ratio 1:2, as they will be asymptotically, are in the ratio 1:2.7. In example 3 below (where 69 observations are available and there are two regressors) the ratio is 1:2.2.

On the other hand $\hat{\gamma}_{p+1}$ is always an unbiassed estimator of $\frac{\rho_1}{1+\rho_1^2}$ whereas, for small samples and $\rho_1 > 0$, there is probably a substantial upward bias in d which will reduce the power of the test based on this statistic.

<u>Example (2)</u>. Two sets of 70 consecutive observations were chosen from series 7 in Kendall (1949), the commencing terms being 191 and 316. These series are sufficiently far apart to justify their treatment as independent. Both series are then Gaussian simple Markoff processes with serial correlations 0.9. Representing the series by x_t and q_t a third series can be constructed as

$$y_t = 0.8x_t + q_t$$

The details of the test based on $\hat{\gamma}_2$ are as follows:

$$(4.2.2.)\begin{bmatrix}\hat{\gamma}_1\\ \hat{\gamma}_2\\ \hat{\gamma}_3\end{bmatrix} = \begin{bmatrix}0.0639 & -0.0017 & -0.0320\\ & 0.0023 & -0.0005\\ & & 0.0204\end{bmatrix}\begin{bmatrix}32.929\\ 201.613\\ 23.795\end{bmatrix} = \begin{bmatrix}0.994\\ 0.422\\ -0.472\end{bmatrix}$$

$$\sigma^{*2} = 0.491 \qquad \left[\frac{n-1}{2}\right] - 2p - 2 = 30$$

The test statistic is

$$t_{30} = \frac{\sqrt{30}\ \ 0.422}{\{0.491 \times 34 \times 0.0023\}^{\frac{1}{2}}} = 11.74$$

which is clearly very highly significant.

The regression coefficient from the least squares procedure is 0.417, which is much further from the true value than $\hat{\gamma}_1$.

The ratio d is 0.694.* The lower bound to the 1% point of a one sided test is 1.43 so that this also gives a very highly significant result.

---

* This value is slightly inaccurate as the estimate was based on the computations leading to (4.1.13.) and differs from the correct value due to the use of different mean corrections.

<u>Example 3.</u> Durbin and Watson (1951) used economic data due to A.R. Prest (1949) to demonstrate their method. The original data covers the period 1870–1938 and shows the logarithm of the consumption of spirits per head in the U.K. (y_t), the logarithm of real income per head (x_{1t}) and the logarithm of the relative price of spirits (x_{2t}). The result of the least squares regression is:

$$(y_t - \bar{y}) = -0.120(x_{1t} - \bar{x}_1) - 1.228(x_{2t} - \bar{x}_2)$$

The statistic d equals 0.249 and is highly significant. The $\hat{\gamma}_j$ are as follows:

(4.2.3.)

$$\begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \\ \hat{\gamma}_5 \end{bmatrix} = \begin{bmatrix} 175.459 & 0.855 & -2.063 & -177.455 & -4.872 \\ & 69.870 & -2.958 & 2.275 & -75.048 \\ & & 10.616 & 3.505 & 16.107 \\ & & & 187.762 & 0.794 \\ & & & & 98.166 \end{bmatrix} \begin{bmatrix} -0.636 \\ -1.818 \\ 2.401 \\ -0.621 \\ -1.804 \end{bmatrix} = \begin{bmatrix} 0.795 \\ -0.693 \\ 0.950 \\ -0.792 \\ 0.630 \end{bmatrix}$$

$\sigma_*^2 = 0.000232$ $\qquad \left[\frac{n-1}{2}\right] - 2p - 2 = 28$

$$t_{28} = \sqrt{28}\,\frac{0.950}{\{0.000232 \times 34 \times 10.616\}^{1/2}} = 17.38$$

which is highly significant.

On economic grounds the regression coefficients $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are more acceptable than the two least squares estimates as it is unlikely that the demand for spirits would fall as income rises.

In this case the lack of independence in the residuals may not be due so much to serial correlation as to a trend, due to the changing habits over the period (the decline of the "gin shops"), the upward trend in income having absorbed some of this effect in the least squares regression. For the alternative method the effect would be absorbed in $\hat{\gamma}_3$ of course.

4.3. <u>An alternative test.</u>

In # 4.1. a test of the hypothesis $\rho_1 = 0$ was obtained from $\hat{\gamma}_{p+1}$. The statistics $\hat{\gamma}_{p+j}\ (j>1)$ will also give tests of this hypothesis though the power of these tests will clearly depend upon the β_j and the covariance matrix of the Z_{jt}. In fact, since is asymptotically fully efficient, it follows from Fisher's well known theorem (Fisher (1924) page 446) that the asymptotic efficiency of $\hat{\gamma}_{p+j}\,(j>1)$ is ρ^2 where ρ is the correlation between $\hat{\gamma}_{p+1}$ and $\hat{\gamma}_{p+j}$ on the null hypothesis. In the case in which p = 1, for example, ρ can be shown to be given by

$$(4.3.1.)\quad \rho = -\left\{1 + \frac{\sigma_1^2}{\sigma_2^2}\,\frac{1}{[\beta_1^2(1+\rho_{2,2}) - 2\rho_{2,1}^2\beta_1^2]}\right\}^{-\frac{1}{2}}$$

where σ_2^2 is the variance of x_{1t} and $\rho_{2,1}$ and $\rho_{2,2}$ are the first and second serial correlations of this variate. The efficiency approaches unity as β_1 and σ_2^2 increase relative to σ_1^2 , as is to be expected.

It is clear that a linear combination of $\hat{\gamma}_{p+1}$ and the $\hat{\gamma}_{p+j}$ will always yield a less efficient test statistic than $\hat{\gamma}_{p+1}$, the best linear combination being that which gives all of the weight to $\hat{\gamma}_{p+1}$. The partial multiple correlation (between y_{3t} and the z_{jt} $(j \geq p+1)$ with the effects of the z_{jt} $(j < p+1)$ removed) is not, however, a linear combination of the $\hat{\gamma}_{p+j}$ $(j \geq p+1)$. Consider, however, the two statistics:

$$t_1 = \hat{\gamma}_2$$
$$(4.3.2.) \qquad t_2 = \left(\hat{\gamma}_3 - \rho \frac{s_3}{s_2} \hat{\gamma}_2 \right)$$

which arise from the case p = 1. Here s_2^2 and s_3^2 are the (true) variances of $\hat{\gamma}_2$ and $\hat{\gamma}_3$ respectively on the null hypothesis. Then t_1 and t_2 are independent and under the conditions of Pitman's theorem (# 2.2.) these will be asymptotically jointly and independently normally distributed. The test statistic gained by weighting the two statistics with their "efficacies" (the quantities $\lim\limits_{n \to \infty} \left\{ [\partial E(t_i)/\partial \rho_1]^2 / \operatorname{Var} t_i \right\}_{\rho_1 = 0}$) will give a critical region in the sample space of $\tau_1 = \frac{t_1}{\{\operatorname{Var} t_1\}^{1/2}_{\rho_1 = 0}}$, $\tau_2 = \frac{t_2}{\{\operatorname{Var} t_2\}^{1/2}_{\rho_1 = 0}}$ which will be the region "above" a straight line (for a one sided test). But asymptotically this straight line will be orthogonal to the line upon which the alternatives corresponding to $\rho_{(n)} = kn^{-1/2}$ will lie while for those alternative hypotheses (which will vary with k) the variances of τ_1 and τ_2 will be unity, and the correlation will still be zero (under the conditions of # 2.2.). It follows that, asymptotically, this linear combination gives the best test of these sequences of hypotheses (which are the only sequences of interest). From what has been said before the linear combination will, of course, reduce merely to $\hat{\gamma}_2$ so that no combination of $\hat{\gamma}_2$ and $\hat{\gamma}_3$, linear or otherwise, can give a better asymptotic test than $\hat{\gamma}_2$ gives and, in particular, the partial multiple correlation cannot.

When two arbitrary independent test statistics t_1 and t_2 are available the knowledge that the best linear combination will give asymptotically the best test from the two statistics is of no practical value, of course, since the best compounding coefficients will be unknown. If they are chosen so that the angle between the line bounding the critical region and the line defining the alternatives is small the asymptotic power of the test will become near to the level of significance. The virtue in the use of $\left\{\frac{t_1^2}{V\!ar\, t_1} + \frac{t_2^2}{V\!ar\, t_2}\right\}$ is that it gives, asymptotically, a region whose power is independent of the direction of the line prescribing these alternatives and is yet fairly powerful.

Chapter 5.

Exact Tests for Correlation between Time Series.

5.1. The efficiency of estimates of regression coefficients when the residual is Markovian.

From (1.4.5.) the covariance matrix of the regression coefficients, when a straightforward least squares procedure is used is:

(5.1.1.) $$n^{-1}\sigma_1^2 \underset{\sim}{M}_0^{-1}\left\{\sum_0^{n-1}\rho_1^k \underset{\sim}{M}_k\right\}\underset{\sim}{M}_0^{-1}$$

Here it will be more convenient to make mean corrections and to omit the coefficient β_0, so that $\underset{\sim}{M}_0$, for example, is the sample covariance matrix of the x_{jt}.

The least squares estimates will be indicated by $\hat{\beta}_j$.

The covariance matrix of the $\hat{\gamma}_j$ is, from (4.1.3.) and (4.1.8.) (asymptotically):

(5.1.2.) $$\frac{2\sigma_1^2(1-\rho_1^2)}{n(1+\rho_1^2)}\underset{\sim}{V}^{-1}$$

It is fairly clear that the maximum likelihood estimates will be asymptotically equivalent to the estimates which will be obtained if ρ_1 is known and is used to transform y_t and the x_{jt} so that the residuals from the regression become the η_t (see (4.1.1.)). The limiting form of the covariance matrix of the maximum likelihood estimates of the β_j will, therefore, be:

(5.1.3.) $$n^{-1}\sigma_1^2(1-\rho_1^2)\left[\underset{\sim}{M}_0(1+\rho_1^2)-\rho_1\underset{\sim}{M}_1\right]^{-1}$$

These estimates will be represented by the notation β_j^*.

Because of the non linearity of the likelihood equations, the lack of any small sample distribution theory for these estimates and the lack of knowledge concerning bias it is worthwhile considering the properties of $\hat{\beta}_j$ and $\hat{\gamma}_j$ even though they will, in general, be inefficient.

A fourth estimation procedure, which was discussed in # 1.4 (pages 30 and 31) is based on the first differences of the observations. The covariance matrix of the resulting estimates $\bar{\beta}_j$ is asymptotically:

(5.1.4.)

$$n^{-1}\sigma_1^2[2\underset{\sim}{M}_0-\underset{\sim}{M}_1]^{-1}\{2(1-\rho_1)[2\underset{\sim}{M}_0-\underset{\sim}{M}_1]+\sum_1^{n-1}(2\rho_1^k-\rho_1^{k-1}-\rho_1^{k+1})[2\underset{\sim}{M}_k-\underset{\sim}{M}_{k-1}-\underset{\sim}{M}_{k+1}]\}[2\underset{\sim}{M}_0-\underset{\sim}{M}_1$$

The estimates $\hat{\beta}_j$, $\hat{\gamma}_j$ and $\bar{\beta}_j$ are all unbiased, while the β_j^* are asymptotically unbiased.

For the moment we shall restrict ourselves to the $\hat{\gamma}_j$ for $0 < j \leq p$.

If all of the x_{jt} are independent in time the covariance matrices become:

(5.1.1.)' $\quad n^{-1}\sigma_1^2\ \underset{\sim}{M}_0^{-1}$

(5.1.2.)' $\quad \dfrac{2\sigma_1^2(1-\rho_1^2)}{n(1+\rho_1^2)}\underset{\sim}{V}_0^{-1}$

(5.1.3.)' $\quad \dfrac{\sigma_1^2(1-\rho_1^2)}{n(1+\rho_1^2)}\underset{\sim}{M}_0^{-1}$

(5.1.4.)' $\quad \dfrac{\sigma_1^2(1-\rho_1)(3-\rho_1)}{2n}\underset{\sim}{M}_0^{-1}$

$\underset{\sim}{M}_0^{-1}$ and $\underset{\sim}{V}_0^{-1}$ will clearly have the same limit in probability so that the asymptotic efficiencies of the various estimators in this case are:

$$E\{\hat{\beta}_j, \beta_j^* | \beta_j\} = \frac{1-\rho_1^2}{1+\rho_1^2}$$

(5.1.5.) $$E\{\hat{\gamma}_j, \beta_j^* | \beta_j\} = \frac{1}{2}$$

$$E\{\bar{\beta}_j, \beta_j^* | \beta_j\} = \frac{2(1+\rho_1)}{(1+\rho_1^2)(3-\rho_1)}$$

For $|\rho_1| > 3^{-\frac{1}{2}}$, $\hat{\beta}_j$ is less efficient than $\hat{\gamma}_j$. However, $\bar{\beta}_j$ is more efficient than $\hat{\gamma}_j$ for $\rho_1 > -0.18$ and is more efficient than $\hat{\beta}_j$ when $\rho_1 > 2-\sqrt{3} \approx 0.27$. $\hat{\beta}_j$ is fully efficient only when ρ_1 is zero, while $\bar{\beta}_j$ becomes fully efficient as ρ_1 tends to one.

If all of the x_{jt} are generated by independent stationary processes with variances $\sigma_{x_j}^2$ and k^{th} serial correlations ρ_{jk} the covariance matrices are asymptotically diagonal with non zero elements:

(5.1.1.)'' $$\frac{\sigma_1^2}{n}\frac{1}{\sigma_{x_j}^2}\left\{1+2\sum_1^{\infty}\rho_1^k\rho_{jk}\right\}$$

(5.1.2.)'' $$\frac{2\sigma_1^2}{n}\frac{1}{\sigma_{x_j}^2}\left\{\frac{1-\rho_1^2}{1+\rho_1^2}\cdot\frac{1+\rho_{j2}}{1+\rho_{j2}-2\rho_{j1}^2}\right\}$$

(5.1.3.)'' $$\frac{\sigma_1^2}{n}\frac{1}{\sigma_{x_j}^2}\left\{\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\rho_{j1}}\right\}$$

(5.1.4.)'' $$\frac{\sigma_1^2}{n}\frac{1}{\sigma_{x_j}^2}\frac{1-\rho_1}{1-\rho_{j1}}\left\{1-\frac{1-\rho_1}{1-\rho_{j1}}\sum_1^{\infty}\rho_1^{k-1}\left[\rho_{jk}-\frac{1}{2}(\rho_{j,k-1}+\rho_{j,k+1})\right]\right\}$$

Again it is clear that for ρ_1 sufficiently high,

$\hat{\gamma}_j$ will be more efficient than $\hat{\beta}_j$ and that $\bar{\beta}_j$ will be more efficient than either.

Consider the case where x_{jt} is generated by a process of the form (1.2.2.) with h = 2. Then it is well known, from the spectral theory of discrete processes, that:

$$\sigma_{x_j}^2 \sum_{-\infty}^{\infty} \rho_{jk} z^k = \frac{z^2}{(1+a_1 z + a_2 z^2)(z^2 + a_1 z + a_2)}$$

the Laurent series converging on an annulus including the unit circle.

It easily follows (Quenouille (1947a)) that:

$$1 + 2\sum_{1}^{\infty} \rho_{jk}\rho_1^k = \frac{1 - a_1\left(\frac{1-a_2}{1+a_2}\right)\rho_1 - a_2\rho_1^2}{1 + a_1\rho_1 + a_2\rho_1^2}$$

The asymptotic efficiencies of the $\hat{\beta}_j$, $\hat{\gamma}_j$ $(j \leq p)$ and $\bar{\beta}_j$ could now be written down from (5.1.1,2,3,4.)''.

In example 1 of # 4.2 we had $a_1 = -1.1$, $a_2 = 0.5$ and $\rho_1 = 0.7$.

The asymptotic efficiencies of $\hat{\beta}_1$ and $\hat{\gamma}_1$ then were:

$$E\{\hat{\beta}_1, \beta_1^* \mid \beta_1\} = 0.52$$

(5.1.6.)

$$E\{\hat{\gamma}_1, \beta_1^* \mid \beta_1\} = 0.28$$

This partly explains the very inaccurate estimates of the regression coefficient by the two estimators.

Finally, when x_{jt} (still independent of the other regressors) is generated by a simple Markoff process with serial correlation ρ_{j1} the variances become:

(5.1.1.)''' $\dfrac{\sigma_1^2}{n}\dfrac{1}{\sigma_{x_j}^2}\dfrac{1+\rho_1\rho_{j1}}{1-\rho_1\rho_{j1}}$

(5.1.2.)''' $\dfrac{2\sigma_1^2}{n}\dfrac{1}{\sigma_{x_j}^2}\dfrac{(1-\rho_1^2)(1+\rho_{j1}^2)}{(1+\rho_1^2)(1-\rho_{j1}^2)}$

(5.1.3.)''' $\dfrac{\sigma_1^2}{n}\dfrac{1}{\sigma_{x_j}^2}\dfrac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\rho_{j1}}$

(5.1.4.)''' $\dfrac{\sigma_1^2}{n}\dfrac{1}{\sigma_{x_j}^2}\dfrac{1-\rho_1}{1-\rho_{j1}}\left(\dfrac{3-\rho_1\rho_{j1}-\rho_1-\rho_{j1}}{2(1-\rho_1\rho_{j1})}\right)$

The asymptotic efficiencies then become:

$$E\{\hat{\beta}_j, \beta_j^*|\beta_j\} = \frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\rho_{j1}}\cdot\frac{1-\rho_1\rho_{j1}}{1+\rho_1\rho_{j1}}$$

(5.1.8.) $$E\{\hat{\gamma}_j, \beta_j^*|\beta_j\} = \frac{1}{2}\frac{(1+\rho_1^2)(1-\rho_{j1}^2)}{(1+\rho_{j1}^2)(1+\rho_1^2-2\rho_1\rho_{j1})}$$

$$E\{\bar{\beta}_j, \beta_j^*|\beta_j\} = \frac{2(1+\rho_1)(1-\rho_{j1})(1-\rho_1\rho_{j1})}{(1+\rho_1^2-2\rho_1\rho_{j1})(3-\rho_1\rho_{j1}-\rho_1-\rho_{j1})}$$

In example 2 of # 4.2 ρ_1 and $\rho_{1,1}$ both equalled 0.9 so that these ratios were 0.105, 0.5 and 0.974. In fact the estimate based on $\hat{\beta}_1$ was much further from the true value than that based on $\hat{\gamma}_1$. Evidently $\bar{\beta}_1$ would here be very near to the maximum likelihood estimate.

It is fairly evident that if an accurate point estimate is required better statistics than the $\hat{\gamma}_j$ can always be found. Even if the computations required for a maximum likelihood solution are too lengthy much of the information can usually be recovered by the use of a transformation. This is, of course, well known. The virtue in the $\hat{\gamma}_j$ lies in the fact that their distribution is known so that confidence intervals can be obtained and tests of significance carried out. In particular, the correlation between y_t and the x_{jt} can be tested exactly. This subject will be considered in the next section.

5.2. An exact test for correlation between two series when the residual is Markovian and independent of the regressor series.

We shall first consider the regression model:

$$(5.2.1.) \qquad y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

where

$$(5.2.2.) \begin{cases} (1)\ \varepsilon_t = \rho_1 \varepsilon_{t-1} + \eta_t \quad \text{and } \eta_t \text{ is } N\{0, \sigma_1(1-\rho_1^2)^{1/2}\} \\ (2)\ \varepsilon_t \text{ is independent of } x_s \text{ for all t and s.} \\ (3)\ x_t \text{ is generated by a process of the form} \\ \qquad x_t = \rho_2 x_{t-1} + \zeta_t \quad \text{where } \zeta_t \text{ is } N\{0, \sigma_2(1-\rho_2^2)^{1/2}\} \end{cases}$$

As was shown in # 4.1 an exact test for correlation between y_t and x_t can be based on the statistic $\hat{\gamma}_1$ (see (4.1.5.)). However, this test will not be fully efficient as can be seen from (5.1.7.). In fact the efficiency can never rise above 0.61 and can be very low for $|\rho_1|$ low and $|\rho_2|$ high.

One can, of course, find tests of the hypothesis that there is no correlation between y_t and x_t which are asymptotically fully efficient. Such a test could be based on the maximum likelihood solutions for the parameters involved. In fact, of course, one needs merely to use any consistent estimator of ρ_1, to transform the observations so that the residual process is approximately independent in time, in order to get an asymptotically fully efficient test. Apart from the fact that the asymptotic efficiency of a test statistic is an asymptotic criterion and, therefore, only of suggestive value for small samples it has an added disadvantage when applied to test statistics whose distribution is not known exactly in that for small samples it does not take into account any bias in the statistic (as an estimator) or any deviation of its true variance (on the null hypothesis) from its limiting form. If such a deviation in the variance is present the significance point used may be very far from the point appropriate to the level of significance required.

One procedure which has been fairly commonly used in the past has been to calculate the ordinary correlation coefficient between the series y_t and x_t and use (1.4.9.) to adjust the degrees of freedom so that a distribution (that of the correlation coefficient) with approximately the correct variance is being used. This procedure has two main disadvantages:

(1) The estimators of ρ_1 and ρ_2 obtained from the first serial correlations of the two series will be biased downwards (see 1.2.26.), if ρ_1 and ρ_2 are positive, so that the number of degrees of freedom will be biassed upwards and an insufficiently stringent test will be used. One can, of course, make a subjective adjustment for bias (see for example Moran (1952)), but this is clearly an unsatisfactory procedure.

(2) As can be seen from (5.1.7.) the test can be very inefficient for ρ_1 high. In fact if ρ_1 is high the test based on $\hat{\gamma}_1$ can be more efficient, as will be seen from Table 3 below.

Table 3.

$E\{\hat{\gamma}_1, \hat{\beta}_1 | \beta_1\}$

ρ_2 \ ρ_1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
0	2.28	1.06	0.69	0.54	0.50	0.54	0.69	1.06	2.28
0.2	1.52	0.77	0.54	0.46	0.46	0.54	0.75	1.25	2.90
0.4	0.85	0.47	0.36	0.33	0.36	0.42	0.69	1.26	3.20
0.6	0.38	0.24	0.20	0.20	0.24	0.32	0.63	1.06	3.05
0.8	0.11	0.08	0.08	0.09	0.11	0.16	0.29	0.66	2.28

The remaining statistic mentioned in # 5.1. is the estimator of the regression coefficient obtained from the first differences of the observations. In order to use this statistic it would be necessary to make some estimate of the variance of the estimator, using a formula such as (5.1.4.)''', and estimates of ρ_1 and ρ_2. The difficulty mentioned under (1) above will still apply, however, though the effect may often be now less serious since (5.1.4.)''' is less sensitive to small variations in ρ_1 and ρ_2 (when these are high and of the same sign) than is (5.1.1.)'''.

It is of some interest to examine the efficiency of the statistics suggested in Quenouille (1949b) (see # 1.4., pages 31 and 32).

For the present case Quenouille's most appropriate statistic would appear to be $r(x_t, y_t \mid x_{t-1})$ since only x_t will be Markovian if β_1 is not zero. Because of the results of later sections and Quenouille's findings with respect to the bias and variance of the statistics on the null hypothesis (see page 33, first paragraph) it is also interesting to examine the efficiency of $r(x_t, y_t \mid x_{t-1}, y_{t-1})$ Quenouille at no time specified the alternative hypotheses held in view but, as will be seen below, the alternative for which Quenouille's statistics are really appropriate is not (5.2.1.) but rather a situation where the correlation between two Markoff processes arises from a correlation between the random "shocks" of which they are linearly composed. Nevertheless, it may often be difficult to specify accurately beforehand the range of alternatives so that it is necessary to consider the efficiency of Quenouille's tests in the present case also. Moreover, the results of the sampling experiments, carried out by Quenouille for the case of no correlation between the series, are fully appropriate and, as mentioned in # 1.4., appear to indicate that the significance point obtained by treating the partial correlations as if they were obtained from separate series independent in time, will be near to the true significance point for the level of significance used. If the power of Quenouille's tests is fairly high, therefore, there is much to recommend them.

Asymptotically the variances of Quenouille's statistics on the null hypothesis will be n^{-1}. We shall evaluate $\lim_{n\to\infty}\left\{\frac{\partial}{\partial\beta_1}\mathcal{E}[r(x_t, y_t \mid x_{t-1})]\right\}$ directly though it is simpler to reverse the order of the two limit operations.

We have

$$\mathcal{E}\{r(x_t, y_t \mid x_{t-1})\} = k\int\cdots\int_{-\infty}^{\infty} r(x_t, y_t \mid x_{t-1}) \exp -\tfrac{1}{2}\{(\underset{\sim}{y}-\underset{\sim}{\beta_0}-\beta_1\underset{\sim}{x})'\underset{\sim}{\Gamma_1}^{-1}(\underset{\sim}{y}-\underset{\sim}{\beta_0}-\beta_1\underset{\sim}{x})+\underset{\sim}{x}'\underset{\sim}{\Gamma_2}^{-1}\underset{\sim}{x}\}\prod_1^n dx_t\,dy_t$$

where $\underset{\sim}{\Gamma_1}$ is the covariance matrix of the ξ_t, $\underset{\sim}{\Gamma_2}$ is the covariance matrix of the x_t, $\underset{\sim}{\beta_0}$ is the vector $\{\beta_0, \beta_0, \cdots \beta_0\}$ and $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are the vectors of the y_t and x_t.

Differentiating under the integral sign (as we may) and putting $\beta_1 = 0$ we obtain:

$$\left\{\frac{\partial}{\partial\beta_1}\mathcal{E}[r(x_t, y_t \mid x_{t-1})]\right\}_{\beta_1=0} = \mathrm{Cov}\{r(x_t, y_t \mid x_{t-1}),\ \underset{\sim}{y}'\Gamma_1^{-1}\underset{\sim}{x}\}_{\beta_1=0}$$

$$= \mathrm{Cov}\left\{r(x_t, y_t \mid x_{t-1}), \frac{1}{\sigma_1^2(1-\rho_1^2)}\left[(1+\rho_1^2)\sum_1^n x_t y_t - \rho_1\sum_1^n x_t y_{t-1} - \rho_1\sum_1^n x_{t-1} y_t\right]\right\}_{\beta_1=0}$$

ignoring certain end effects.

The limit of this may be obtained using an obvious extension of the theorem quoted in Cramer (1946), page 353, so that we obtain finally:

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\beta_1}\mathcal{E}(r(x_t,y_t|x_{t-1}))\right\}_{\beta_1=0} = \frac{\sigma_2}{\sigma_1}(1-\rho_2^2)^{1/2}$$

Similarly we obtain:

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\beta_1}\mathcal{E}(r(x_t,y_t|x_{t-1},y_{t-1}))\right\}_{\beta_1=0} = \frac{\sigma_2}{\sigma_1}\left\{\frac{1-\rho_2^2}{1-\rho_1^2}\right\}^{1/2}$$

Since the two statistics have the same limiting variance on the null hypothesis the second is always asymptotically the more efficient and from (5.1.3.)'''

$$E\{r(x_t,y_t|x_{t-1},y_{t-1}),\beta_1^*|\beta_1=0\} = \frac{1-\rho_2^2}{1+\rho_1^2-2\rho_1\rho_2} = \frac{1}{1+\frac{(\rho_1-\rho_2)^2}{1-\rho_2^2}}$$

Evidently $r(x_t,y_t|x_{t-1},y_{t-1})$ is asymptotically fully efficient if, and only if $\rho_1=\rho_2$. This is due to the fact that $\rho_1=\rho_2$ implies that y_t is also Markovian and vice versa (when $\beta_1 \neq 0$).

We also have from (5.1.2.)'''

$$E\{r(x_t,y_t|x_{t-1},y_{t-1}),\hat{\gamma}_1|\beta_1=0\} = 2\,\frac{1+\rho_2^2}{1+\rho_1^2} > 1 \quad ;|\rho_1|,|\rho_2|<1$$

and from (5.1.1.)'''

$$E\{r(x_t,y_t|x_{t-1},y_{t-1}),\hat{\beta}_1|\beta_1=0\} = \frac{(1+\rho_1\rho_2)(1-\rho_2^2)}{(1-\rho_1\rho_2)(1-\rho_1^2)}$$

This ratio is shown in Table 4 below for some values of ρ_1 and ρ_2.

Table 4.

$E\{r(x_t, y_t | x_{t-1} y_{t-1}), \hat{\beta}_1 | \beta_1 = 0\}$

ρ_2 \ ρ_1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
0	2.78	1.56	1.19	1.04	1	1.04	1.19	1.56	2.78
0.2	1.93	1.18	0.97	0.92	0.96	1.08	1.34	1.91	3.68
0.4	1.20	0.80	0.72	0.75	0.84	1.03	1.38	2.14	4.53
0.6	0.62	0.47	0.47	0.52	0.64	0.85	1.24	2.13	5.06
0.8	0.22	0.20	0.22	0.27	0.36	0.52	0.83	1.60	4.56

The statistic $r(x_t, y_t | x_{t-1}, y_{t-1})$ has considerable appeal as a test statistic of the hypothesis $\beta_1 = 0$. The partial correlation coefficient corresponding to $\hat{\gamma}_1$ has one advantage of course since it gives an exact test. However, unless $|\rho_1|$ is large relative to $|\rho_2|$ the test may be much less efficient than $r(x_t, y_t | x_{t-1}, y_{t-1})$

Another test statistic, whose distribution is exactly known is $\hat{\gamma}_3$. Since $\hat{\gamma}_1$ is not fully efficient it appears that some increase in efficieincy might be obtained from a combination of $\hat{\gamma}_1$ and $\hat{\gamma}_3$, for example the corresponding partial multiple correlation (of y_{2t} with x_{2t} and $(x_{2t-1} + x_{2t+1})$ with the effects of $(y_{2t-1} + y_{2t+1})$ removed).

The efficacy of $\hat{\gamma}_3$ can be easily evaluated as:

$$(5.2.3.)\quad \lim_{n\to\infty} \left\{ \frac{[\partial/\partial\beta_1 \, E(\hat{\gamma}_3)]^2}{Var(\hat{\gamma}_3)} \right\}_{\beta_1=0} = \frac{n\rho_1^2(1-\rho_2^2)}{(1-\rho_1^4)} \frac{\sigma_3^2}{\sigma_1^2}$$

and we have:

$$E\{\hat{\gamma}_1, \hat{\gamma}_3 \mid \beta_1 = 0\} = \frac{1+\rho_1^2}{2\rho_1^2(1+\rho_2^2)} > 1 \quad ; |\rho_1|, |\rho_2| < 1.$$

The correlation between the two statistics when β_1 is zero is:

$$(5.2.4.) \qquad \rho = -\frac{\sqrt{2}\,\rho_2}{(1+\rho_2^2)^{1/2}}$$

The efficacy of the best linear combination of the two statistics (which gives, asymptotically, the best test; see # 4.3.) will be the quadratic form in the two efficacies whose matrix is the inverse of the correlation matrix of the two statistics. We, therefore, have as the efficacy of $\hat{\gamma}$ (say) (where $\hat{\gamma}$ is the best linear combination of $\hat{\gamma}_1$ and $\hat{\gamma}_3$) from (5.2.1.)''', (5.2.3.) and (5.2.4.):

$$\lim_{n\to\infty} \left\{\frac{[\partial/\partial\beta_1 E(\hat{\gamma})]^2}{\operatorname{Var}(\hat{\gamma})}\right\}_{\beta_1=0} = \frac{n\sigma_2^2}{\sigma_1^2}\,\frac{1+\rho_2^2}{1-\rho_2^2}\left\{\frac{\rho_1^2(1-\rho_2^2)}{1-\rho_1^4} + \frac{(1+\rho_1^2)(1-\rho_2^2)}{2(1-\rho_1^2)(1+\rho_2^2)} - \frac{2\sqrt{2}\,\rho_2}{(1+\rho_2^2)^{1/2}}\,\frac{\rho_1(1-\rho_2^2)}{\sqrt{2}\,(1-\rho_1^2)(1+\rho_2^2)^{1/2}}\right\}$$

remembering that the square root of the efficacy of $\hat{\gamma}_3$ must be treated as negative if ρ_1 is positive. This quantity is:

$$(5.2.5.) \qquad \frac{n\sigma_2^2}{\sigma_1^2}\left\{\frac{\rho_1^2(1+\rho_2^2)}{1-\rho_1^4} + \frac{1+\rho_1^2-4\rho_1\rho_2}{2(1-\rho_1^2)}\right\}$$

We, therefore have:

$$(5.2.6.) \qquad E\{\hat{\gamma}, \hat{\gamma}_1 \mid \beta_1 = 0\} = \frac{1+\rho_2^2}{1-\rho_2^2}\left\{1 + \frac{2\rho_1^2(1+\rho_2^2)}{(1+\rho_1^2)^2} - \frac{4\rho_1\rho_2}{1+\rho_1^2}\right\} \geq 1$$

(as it clearly must be).

This is shown for some values of ρ_1 and ρ_2 in Table 5 below:

Table 5.

$E\{\gamma_3, \hat{\gamma}_1 \mid \beta_1 = 0\}$

ρ_2 \ ρ_1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
0	1.48	1.39	1.24	1.07	1	1.07	1.24	1.39	1.48
0.2	2.04	1.94	1.65	1.33	1.08	1	1.05	1.11	1.20
0.4	3.22	2.98	2.52	1.92	1.38	1.07	1	1.03	1.07
0.6	5.99	5.50	4.57	3.32	2.13	1.36	1.05	1	1.01
0.8	15.22	13.90	11.36	7.91	4.56	2.30	1.31	1.03	1

This table is of no direct practical value, of course, since the best compounding coefficients are never known in advance. In practice the (partial) multiple correlation corresponding to $\hat{\gamma}_1$, and $\hat{\gamma}_3$ will be used (for the reasons mentioned at the end of # 4.3.) and (5.2.5.) is still important in the determination of the asymptotic power of the resulting test (as was shown in # 2.3.). However, the comparison of the powers of this test and of the test depending on $\hat{\gamma}_1$, depends on the particular sequence considered. The double lines in Table 5 enclose a region of values of ρ_1 and ρ_2 which, from experience, seem very important. Here the best linear combination gives a test which is not very much more powerful than that based on $\hat{\gamma}_1$ alone and it seems, therefore, that throughout most or all of this region the power of the test based on the multiple correlation coefficient will be lower than that based on $\hat{\gamma}_1$.

The asymptotic powers can, of course, be compared but a large number of parameters are involved; $\alpha, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2$ and k. We shall consider only the case where $\rho_1 = 0.4, \rho_2 = 0.8$ and $\sigma_1^2 = \sigma_2^2$ so that (5.2.5.) becomes n(0.198) while the efficacy of $\hat{\gamma}_1$ is n(0.152) from (5.1.2.). We shall also restrict ourselves to the 5% level of significance and will consider a two sided test.

The asymptotic power function for $\hat{\gamma}_1$ against the alternative $\beta_1 = kn^{-1/2}$ is now:

(5.2.6.) $\Phi\{1.96 - k(0.152)^{1/2}\} + \Phi\{1.96 + k(0.152)^{1/2}\} = P(\hat{\gamma}_1)$

while for R, the partial multiple correlation, the asymptotic power function is:

(5.2.7.) $G_2(k^2(0.198), .05) = P(R)$

where $G_p(d^2, \alpha)$ is the power function of the χ^2 test for p degrees of freedom at the level α and for an alternative at a distance d from the origin. This function appears to be poorly tabulated. The tables used were those prepared by Tang (1938) and the values of k have been chosen to conform with this tabulation.

Table 6.

k	3.893	5.840	7.787	9.734	11.680
$P(\hat{\gamma}_1)$	0.329	0.623	0.858	0.966	0.995
$P(R)$	0.322	0.638	0.883	0.979	0.998

It appears that even in this case the asymptotic power of the test based on the partial multiple correlation is, at best, not much above that of the test based on $\hat{\gamma}_1$. Of course, if ρ_1 and ρ_2 were chosen so that $E(\hat{\gamma}, \hat{\gamma}_1 / \beta_1 = 0)$ were very high this would not be so but such a set of values of ρ_1 and ρ_2 is, perhaps, not met often in practice.

The situation, therefore, appears to be as follows:

(1) An asymptotically fully efficient test can, of course, always be gained from the maximum likelihood estimators or from the use of any consistent statistic (as an estimator of ρ_1) to transform x_t and y_t so that the residuals are approximately independent. No exact distribution theory is available and little is known of small sample biases in means and variances.

(2) Quenouille's statistic $r(x_t, y_t / x_{t-1}, y_{t-1})$ is fairly efficient for most values of ρ_1 and ρ_2, particularly if

they are not too far from equal. Its distribution is not known exactly but from Quenouille's sampling experiments it appears that its treatment as an ordinary partial correlation will be a reasonable procedure.

(3) The ordinary correlation coefficient is well known to be inefficient and its use is also hindered by the bias in the estimators of ρ_1 and ρ_2 (which affects the accuracy of the determination of the degrees of freedom to be used).

(4) The test based on the partial correlation between y_{2t} and x_{2t} with the effects of $(x_{2t-1} + x_{2t+1})$ and $(y_{2t-1} + y_{2t+1})$ removed is exact, and compares reasonably favourably with the ordinary correlation, with respect to efficiency, over a fair range of the values of ρ_1 and ρ_2 met in practice. It is always less efficient than $r(x_t, y_t \mid x_{t-1}, y_{t-1})$ so that the exactness of the distribution is its only advantage here.

(5) The test based on the (partial) multiple correlation of y_{2t} with x_{2t} and $(x_{2t-1}+x_{2t+1})$ with the effects of $(y_{2t-1} + y_{2t+1})$ removed can be asymptotically more powerful than the test considered in (4) above, but it can also be less powerful. When $|\rho_1 - \rho_2|$ is not very high little can be gained by using the multiple correlation and a good deal can be lost.

When x_t is not Markovian the comparison of the powers of the various tests becomes more complicated. Even when x_t comes from a second order autoregressive process the formulae resulting from (5.1.1,2,3,4.)''' and (5.1.6.) are quite complicated algebraic expressions. It appears, however, that no essentially new situation emerges and the asymptotic relative efficiencies of the various tests have not, therefore, been computed.

Example 1. The data is that of Example 2 of # 4.2. Some of the statistics considered in this section have been applied to this data and the results are shown in Table 7 below.

Table 7.

Statistic	Sample Value	Population Value	Test Statistic
$\hat{\gamma}_1$	0.994	0.8	$t_{30} = 5.27$ **
$\hat{\beta}_1$	0.417	0.8	$r_{20} = 0.316$
$r(x_t, y_t \mid x_{t-1})$	0.316	0.329	$r_{67} = 0.361$**
$r(x_t, y_t \mid x_{t-1}, y_{t-1})$	0.684	0.625	$r_{66} = 0.684$**

** Highly significant.

Here t_{30} indicates Student's t with 30 degrees of freedom while r_m is a correlation with m degrees of freedom.

The degrees of freedom appropriate to r_{20} were calculated from:

$$n' = 70\ \frac{1 - r_{1x}\, r_{1y}}{1 + r_{1x}\, r_{1y}} = 70\ \frac{1 - (0.755)(0.739)}{1 + (0.755)(0.739)} \approx 20$$

In fact, of course, the two serial correlations were 0.9.

Example 2. Here the data is that of Example 1 of # 4.2. The statistics $\hat{\beta}_1$, $\hat{\gamma}_1$, $r(x_t, y_t \mid x_{t-1}, y_{t-1})$ and the partial multiple correlation $R(y_{2t};x_{2t},x_{2t-1}+x_{2t+1}|y_{2t-1}+y_{2t+1})$ are considered below.

Table 8.

Statistic	Sample Value	Test Statistic	
$\hat{\beta}_1$	2.791	$r_{13} = 0.803$	**
$r(x_t, y_t \mid x_{t-1}, y_{t-1})$	0.636	$r_{26} = 0.636$	**
$R(y_{2t}; x_{2t}, x_{2t-1}+x_{2t+1} \mid y_{2t-1}+y_{2t+1})$	0.361	$F_{2,10} = 2.82$	
$\hat{\delta}_1$	2.423	$t_{10} = 1.82$	

5.3. An exact test for correlation between two simple Markoff processes.

Let

$$y_t - \mu_1 = \rho_1 (y_{t-1} - \mu_1) + \varepsilon_t \qquad |\rho_1| < 1$$

$$x_t - \mu_2 = \rho_2 (x_{t-1} - \mu_2) + \eta_t \qquad |\rho_2| < 1 \tag{5.3.1.}$$

be two simple Markoff processes with normally distributed disturbances having zero means and variances $\sigma_1^2(1-\rho_1^2)$ and $\sigma_2^2(1-\rho_2^2)$.

The correlation between the two processes can be prescribed in terms of the correlation properties of the residuals. Since these residuals come from processes of independent random variates this will be achieved by saying for what lags the residuals are correlated and what the correlations are.

We shall consider the case where the only non zero correlation is between ε_t and η_{t-m} and the series have been so lagged relative to each other that m is zero.

If the correlation between ξ_t and η_t is ρ then the correlation between x_t and y_s is:

$$(5.3.2.)\qquad \begin{aligned} t \geqslant s \quad & \frac{\rho\{(1-\rho_1^2)(1-\rho_2^2)\}^{1/2}}{1-\rho_1\rho_2}\rho_2^{t-s} && = \alpha\rho_2^{t-s} \text{ (say)} \\ t \leqslant s \quad & && = \alpha\rho_1^{s-t} \end{aligned}$$

The two series y_t and x_s will be jointly normally distributed for every t so that:

$$(y_t-\mu_1) = \alpha\frac{\sigma_1}{\sigma_2}(x_t-\mu_2) + \zeta_t$$

where ζ_t is independent of x_t and has mean zero and variance: $\sigma_1^2(1-\alpha^2)$.

Then a simple calculation shows that:

$$\mathcal{E}(\zeta_t\zeta_{t-s}) = \sigma_1^2(1-\alpha^2)\rho_1^s$$

It follows, since ζ_t is normally distributed, that it is generated by a simple Gaussian Markoff process with parameter ρ_1 (Doob (1953), page 233).

The disturbance θ_t in the process:

$$\zeta_t = \rho_1\zeta_{t-1} + \theta_t$$

is

$$\theta_t = \alpha\frac{\sigma_1}{\sigma_2}(x_{t-1}-\mu_2)(\rho_1-\rho_2) + \xi_t - \alpha\frac{\sigma_1}{\sigma_2}\eta_t$$

If $\rho_1 \neq \rho_2$ and $\rho \neq 0$ then θ_t depends on x_{t-1}, and is not wholly independent of the process x_t. The θ_t are, of course, independent.

When $\rho_1 = \rho_2$ the present case is equivalent to that considered in the # 5.2. When $\rho_1 \neq \rho_2$ the exact test of correlation there considered is still an exact test but the conclusions as to the power of the test and of the others there considered do not apply for when $\rho_1 \neq \rho_2$ condition

(2) of (5.2.2.) no longer holds.

The conditional distribution of the $y_{2t}, (t=1, \cdots, [1/2(n-1)])$ for fixed $x_t, (t=1, \cdots, n)$ and $y_{2t+1}, (t=0, \cdots, [1/2(n-1)])$ is the product of the conditional distributions of each y_{2t} for fixed $y_{2t-1}, y_{2t+1}, x_{2t-1}, x_{2t}$ and x_{2t+1}. This follows from the fact that for fixed y_{2t+1} and x_{2t+1} the y_{2t} and x_{2t} are serially independent (Ogawara (1951)). The correlation matrix of the variates $y_{2t-1}, y_{2t+1}, x_{2t-1}, x_{2t}, x_{2t+1}, y_{2t}$ (in that order) is:

$$\underset{\sim}{\Lambda} = \begin{bmatrix} \underset{\sim}{A} & \underset{\sim}{B} \\ \underset{\sim}{B}' & 1 \end{bmatrix}$$

where

$$\underset{\sim}{A} = \begin{bmatrix} 1 & \rho_1^2 & \alpha & \alpha\rho_2 & \alpha\rho_2^2 \\ \rho_1^2 & 1 & \alpha\rho_1^2 & \alpha\rho_1 & \alpha \\ \alpha & \alpha\rho_1^2 & 1 & \rho_2 & \rho_2^2 \\ \alpha\rho_2 & \alpha\rho_1 & \rho_2 & 1 & \rho_2 \\ \alpha\rho_2^2 & \alpha & \rho_2^2 & \rho_2 & 1 \end{bmatrix}$$

$$\underset{\sim}{B}' = (\rho_1 \quad \rho_1 \quad \alpha\rho_1 \quad \alpha \quad \alpha\rho_2)$$

Then, using the theorem mentioned in # 3.2 (page 54), we find that y_{2t} is conditionally distributed with mean $\sigma_1 \underset{\sim}{A}^{-1} \underset{\sim}{B} \underset{\sim}{V}$ and variance $\sigma_1^2(1 - \underset{\sim}{B}' \underset{\sim}{A}^{-1} \underset{\sim}{B})$

Here

$$\underset{\sim}{V}' = \left(\frac{y_{2t-1}}{\sigma_1} \quad \frac{y_{2t+1}}{\sigma_1} \quad \frac{x_{2t-1}}{\sigma_2} \quad \frac{x_{2t}}{\sigma_2} \quad \frac{x_{2t+1}}{\sigma_2} \right)$$

The vector $(\underset{\sim}{A}^{-1} \underset{\sim}{B})'$ is easily seen to be:

$$\frac{1}{1+\rho_1^2} \left\{ \rho_1 \quad \rho_1 \quad \frac{-\rho_2\alpha(1-\rho_1\rho_2)}{1-\rho_2^2} \quad \frac{\alpha(1-\rho_1^2\rho_2^2)}{1-\rho_2^2} \quad \frac{-\rho_1\alpha(1-\rho_1\rho_2)}{1-\rho_2^2} \right\}$$

while

$$(5.3.3.) \qquad \sigma_1^2\left(1 - \underset{\sim}{B}' \underset{\sim}{A}^{-1} \underset{\sim}{B}\right) = \sigma_1^2(1-\rho^2)\left(\frac{1-\rho_1^2}{1+\rho_1^2}\right)$$

For $\rho_1 = \rho_2$ these formulae are equivalent to those given in # 4.1, when it is remembered that there the variance of the residual ε_t was σ_1^2, while here the variance of the corresponding residual ζ_t is $\sigma_1^2(1-\rho^2)$.

When $\rho_1 \neq \rho_2$ two things are noticeable.

(1) The regression coefficient of x_{2t-1} is not equal to that of x_{2t+1} so that some of the symmetry present when $\rho_1 = \rho_2$ is now lost. This loss of symmetry is obviously due to the fact that:

$$\mathcal{E}\{(y_t-\mu_1)(x_{t-s}-\mu_2)\} \neq \mathcal{E}\{(y_t-\mu_1)(x_{t+s}-\mu_2)\}\ ;\ \rho_1 \neq \rho_2\ ;\ \rho \neq 0.$$

and correspondingly:

$$\mathcal{E}\{(y_t-\mu_1)(x_{t-s}-\mu_2) \mid y_{t-1}, y_{t+1}\} \begin{array}{l} = 0\ ,\ s>0 \\ \neq 0\ ,\ s \leq 0 \end{array}$$

(2) The coefficient of x_{2t} is no longer equal to $\alpha \dfrac{\sigma_1}{\sigma_2}$ but instead to

$$\frac{\alpha\,\sigma_1(1-\rho_1^2\rho_2^2)}{\sigma_2(1+\rho_1^2)(1-\rho_2^2)}$$

Since the coefficients of y_{2t-1} and y_{2t+1} are equal we may consider the four variates:

$$(5.3.4.)\qquad \begin{array}{lll} Z_{1t} = x_{2t} & \qquad & \gamma_1 = \dfrac{\alpha(1-\rho_1^2\rho_2^2)}{(1+\rho_1^2)(1-\rho_2^2)}\dfrac{\sigma_1}{\sigma_2} \\ Z_{2t} = y_{2t-1}+y_{2t+1} & & \gamma_2 = \dfrac{\rho_1}{1+\rho_1^2} \\ Z_{3t} = x_{2t-1} & & \gamma_3 = \dfrac{-\rho_2\alpha(1-\rho_1\rho_2)}{(1+\rho_1^2)(1-\rho_2^2)}\dfrac{\sigma_1}{\sigma_2} \\ Z_{4t} = x_{2t+1} & & \gamma_4 = \dfrac{-\rho_1\alpha(1-\rho_1\rho_2)}{(1+\rho_1^2)(1-\rho_2^2)}\dfrac{\sigma_1}{\sigma_2} \end{array}$$

whose coefficients (as regressors relative to y_{2t}) have been written down opposite them.

The covariance matrix of the corresponding least squares estimators is, asymptotically,

$$\frac{2\sigma_1^2(1-\rho^2)(1-\rho_1^2)}{n\,(1+\rho_1^2)}\begin{bmatrix} \sigma_2^2 & \alpha\sigma_1\sigma_2(\rho_1+\rho_2) & \rho_2\sigma_2^2 & \rho_2\sigma_2^2 \\ & 2\sigma_1^2(1+\rho_1^2) & \sigma_1\sigma_2(1+\rho_1^2) & \alpha\sigma\rho_2(1+\rho_2^2) \\ & & \sigma_2^2 & \rho_2\sigma_2^2 \\ & & & \sigma_2^2 \end{bmatrix}$$

and for $\rho=0$ this is:

$$(5.3.5.)\quad \frac{2\sigma_1^2(1-\rho_1^2)}{n\,(1+\rho_1^2)}\begin{bmatrix} \frac{1+\rho_2^2}{\sigma_2^2(1-\rho_2^2)} & 0 & \frac{-\rho_2}{\sigma_2^2(1-\rho_2^2)} & \frac{-\rho_2}{\sigma_2^2(1-\rho_2^2)} \\ & \frac{1}{2\sigma_1^2(1+\rho_1^2)} & 0 & 0 \\ & & \frac{1}{\sigma_2^2(1-\rho_2^2)} & 0 \\ & & & \frac{1}{\sigma_2^2(1-\rho_2^2)} \end{bmatrix}$$

We have, therefore:

$$(5.3.6.)\quad \lim_{n\to\infty}\left\{\frac{\left[\frac{\partial E(\hat{\gamma}_1)}{\partial\rho}\right]^2}{\operatorname{Var}(\hat{\gamma}_1)}\right\}_{\rho=0} = \frac{n\,(1+\rho_1\rho_2)^2}{2(1+\rho_1^2)(1+\rho_2^2)}$$

The partial correlation coefficient between y_{2t} and Z_{1t} (with the effects of z_{2t}, z_{3t} and z_{4t} removed) again gives an exact test of the hypothesis that the two series are independent. Again, however, the coefficients $\hat{\gamma}_3$ and $\hat{\gamma}_4$ also give tests of this hypothesis and we may calculate the efficacy of the best linear combination of these three statistics as in # 5.2. Once more indicating this combination by $\hat{\gamma}$ we have:

$$\lim_{n\to\infty}\left\{\frac{\left[\frac{\partial E(\hat{\gamma})}{\partial\rho}\right]^2}{\operatorname{Var}(\hat{\gamma})}\right\}_{\rho=0} = \frac{n}{2}$$

where this quantity is obtained from (5.3.4.) and (5.3.5.) by taking the quadratic form in the efficacies of the three statistic whose matrix is the inverse of the correlation matrix of the statistics. We have, therefore:

$$(5.3.7.)\quad E\left\{\hat{\gamma}_3\hat{\gamma}_1 \mid \rho=0\right\} = \frac{(1+\rho_1^2)(1+\rho_2^2)}{(1+\rho_1\rho_2)^2}$$

This reduces to unity for $\rho_1 = \rho_2$ as we already know it must (see Table 5). It is shown for some values of ρ_1 and ρ_2 in Table 9.

Table 9.

$E\left\{\hat{\gamma}_3\hat{\gamma}_1 \mid \rho=0\right\}$

ρ_2 \ ρ_1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
0	1.64	1.36	1.16	1.04	1	1.04	1.16	1.36	1.64
0.2	2.42	1.83	1.43	1.17	1.04	1	1.03	1.13	1.27
0.4	4.11	2.73	1.91	1.43	1.16	1.03	1	1.03	1.09
0.6	8.25	4.52	2.73	1.83	1.36	1.13	1.03	1	1.02
0.8	20.75	8.25	4.11	2.42	1.64	1.27	1.09	1.02	1

For the range of values of most practical interest (enclosed by double lines) it is evident from Table 9 (in combination with Table 6) that $\hat{\gamma}_1$ will provide a better test than the test based on the (partial) multiple correlation corresponding to $\hat{\gamma}_1$, $\hat{\gamma}_3$ and $\hat{\gamma}_4$.

In order to obtain the efficacies of Quenouille's statistics we need the limits of the rates of increase of their expected values on the null hypothesis. As in # 5.2 we have:

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}\mathcal{E}\left(r(x_t, y_t \mid x_{t-1})\right)\right\}_{\rho=0}$$

$$= \lim_{n\to\infty}\left\{\mathrm{Cov}\left(r(x_t, y_t \mid x_{t-1}),\ \underset{\sim}{y}'\,\Gamma_1^{-1}\Gamma_3\,\Gamma_2^{-1}\,\underset{\sim}{x}\right)\right\}_{\rho=0}$$

where $\underset{\sim}{\Gamma}_1$ is the covariance matrix of the y_t, $\underset{\sim}{\Gamma}_2$ is the covariance matrix of the x_t and

$$\underset{\sim}{\Gamma}_3 = \frac{1}{P}\left[\mathcal{E}\{(y_i-\mu_1)(x_j-\mu_2)\}\right]$$

The limit of the covariance may be evaluated in the same manner as before to give:

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}\mathcal{E}\left(r(x_t, y_t \mid x_{t-1})\right)\right\}_{\rho=0} = (1-\rho_1^2)^{1/2}$$

Similarly

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}\mathcal{E}\left(r(x_t, y_t \mid y_{t-1})\right)\right\}_{\rho=0} = (1-\rho_2^2)^{1/2}$$

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}\mathcal{E}\left(r(x_t, y_t \mid x_{t-1}, y_{t-1})\right)\right\}_{\rho=0} = 1.$$

Thus the last statistic is the only one which we need to consider and its efficacy is n.

The other variate considered in # 5.2 was the ordinary correlation coefficient (r) whose efficacy is, as before, $\frac{n(1-\rho_1\rho_2)}{1+\rho_1\rho_2}\ \frac{(1-\rho_1^2)(1-\rho_2^2)}{(1-\rho_1\rho_2)^2} = \frac{n(1-\rho_1^2)(1-\rho_2^2)}{1-\rho_1^2\rho_2^2}$

The maximum likelihood estimator will be asymptotically equivalent to the statistic which would be obtained if ρ_1 and ρ_2 were known and were used to transform to the residuals ξ_t and η_t (see (5.3.1.)). We, therefore, have:

$$\lim_{n\to\infty}\left\{\frac{\left[\frac{\partial\,\mathcal{E}(r^*)}{\partial\rho}\right]^2}{\mathrm{Var}(r^*)}\right\}_{\rho=0} = n$$

This can easily be checked by a direct evaluation of the information matrix when ρ is zero. We, therefore, have the asymptotic efficiencies of the three main statistic as:

$$(5.3.8.)\quad \begin{aligned} E\{\hat{\gamma}_1, r^* \mid \rho=0\} &= \frac{(1+\rho_1\rho_2)^2}{2(1+\rho_1^2)(1+\rho_2^2)} \leq 1/2 \\ E\{r, r^* \mid \rho=0\} &= \frac{(1-\rho_1^2)(1-\rho_2^2)}{1-\rho_1^2\rho_2^2} \leq 1 \\ E\{r(x_t, y_t \mid x_{t-1}, y_{t-1}), r^* \mid \rho=0\} &= 1. \end{aligned}$$

Here Quenouille's statistic is unquestionably superior to r. It is always at least twice as efficient as $\hat{\gamma}_1$, whose distribution is, however, known exactly. Since Quenouille's statistic, from the sampling experiments, gives a reasonably good approximate test it appears, however, that it should always be preferred to $\hat{\gamma}_1$.

Example: In Moran (1952) the correlations between the records of the numbers of game birds shot over a period of 73 years were considered. In this example the series for Caper and Ptarmigan are related.

Using the test given in Quenouille (1947a) (see # 1.2, page 19) we find that the quantities:

$$\chi^2_{11} = \frac{73}{(1-r_1^2)^2}\sum_{j=2}^{12} R_j^2$$

are for Ptarmigan 13.3 and for Caper 7.1. Neither is significant at any reasonable level. Here r_1 is the first serial correlation while the R_j were defined in (1.2.28.). It, therefore, seems that the two series can be treated as Markovian.

The results of some of the tests discussed above are shown in Table 10.

Table 10.

Statistic	Sample Value	Test Statistic
r	0.417	$r_{34} = 0.417$ *
$\hat{\gamma}_1$	0.444	$t_{31} = 2.484$ *
$r(x_t, y_t \mid x_{t-1}, y_{t-1})$	0.387	$r_{69} = 0.387$ **

* Significant at 5% point.
** Highly significant.

The two observed first serial correlations were 0.592 and 0.417 and these were adjusted to 0.65 and 0.55 by Moran to allow for bias. Using these values it appears from (5.3.8.) that $\hat{\gamma}_1$ and r should give tests of (roughly) the same asymptotic efficiency while $r(x_t, y_t \mid x_{t-1}, y_{t-1})$ should be over twice as efficient as either of these two. The results here appear to agree with this theory.

5.4. <u>The case where the regressor process is a second order autoregressive process.</u>

The tests for correlation between two series which have been presented in this chapter are exact provided that the residual from the regression of one variate on the other is a Markoff process. The efficiency of the tests have been examined in detail in the following cases:-

(1) The regressor process is a Markoff process wholly independent of the residual. In particular this covers

the case where the regressand is also a Markoff process with the same parameter as the regressor and the residual.

(2) The regressand and regressor processes are both Markovian. When the two parameters are equal this case is equivalent to the particular case mentioned under (1) above.

If neither of the two processes being correlated can reasonably be said to be Markovian then, strictly, the tests presented here are not applicable, for the residual can then be Markovian only if there is in truth a relation between the series. (One might apply the test as an approximation to an exact test of course).

A case which may arise in practice is that in which one process appears Markovian and the other an auto-regressive process of higher order. It is then possible to maintain the null hypothesis in a form which validates the test. The more interesting alternative hypothesis then appears to be one in which the correlation between the series arises from a correlation between the shocks of which they are linearly composed. In this section the efficiency of the test given in the previous section will be considered when one process is an autoregressive process of the second order. Now, of course, x_{t-2} and x_{t+2} will appear in the conditional mean value of

In fact no one is likely to carry out such a computationally burdensome regression as would be necessary to include their effects so that we shall study the same estimators as in # 5.3.

Let $(y_t - \mu_1) = \rho_1(y_{t-1} - \mu_1) + \varepsilon_t$
and $(x_t - \mu_2) + a_1(x_{t-1} - \mu_2) + a_2(x_{t-2} - \mu_2) = \eta_t$
be the two processes where:

(1) $|\rho_1| < 1$

(2) All roots of $z^2 + a_1 z + a_2 = 0$ are less than unity in absolute value.

(3) ε_t and η_t are jointly normally distributed with variances $\sigma_1^2(1-\rho_1^2)$ and $\sigma_2^2\left\{\frac{1-a_2}{1+a_2}\left[(1+a_2)^2 - a_1^2\right]\right\}$ and correlation ρ.

Then the correlation between y_t and x_t is:

$$\rho\left\{(1-\rho_1^2)\,\frac{1-a_2}{1+a_2}\left[(1+a_2)^2 - a_1^2\right]\right\}^{1/2}\sum_{j=0}^{\infty}\rho_1^j\, b_j$$

where b_j is the coefficient of η_{t-j} in the representation of x_t as a moving average of infinite extent.

This expression is (Quenouille, 1947a):

$$\rho\left\{(1-\rho_1^2)\,\frac{1-a_2}{1+a_2}\left[(1+a_2)^2 - a_1^2\right]\right\}^{1/2}(1 + a_1\rho_1 + a_2\rho_1^2)^{-1}$$

The mean in the conditional distribution of y_{2t} for fixed y_{2t-1}, y_{2t+1} and x_t will now involve x_{2t-2} and x_{2t+2} as well as the variates included in the regression given in the previous section. It is of interest to examine the effect on the efficiency of the tests presented in the last section of the divergence of the parameter a_2

from zero.

By straightforward methods it can be shown that:

$$E\{\hat{\gamma}_1, r^* \mid \rho = 0\} = \frac{(1 - a_1\rho_1 - a_2^2)^2}{2(1+\rho_1^2)(1+a_1^2-a_2^2)}$$

$$E\{r, r^* \mid \rho = 0\} = \frac{(1-\rho_1^2)\{(1+a_2)^2 - a_1^2\}}{\left\{\frac{1+a_2}{1-a_2}(1-a_2\rho_1^2) - a_1\rho_1\right\}(1+a_1\rho_1+a_2\rho_1^2)}$$

$$E\{r(x_t, y_t \mid x_{t-1}, y_{t-1}), r^* \mid \rho = 0\} = 1 - a_2^2 .$$

It is clear now that the statistic $r(x_t, y_t \mid x_{t-1}, x_{t-2}, y_{t-1})$ will give an asymptotically fully efficient test. However, it is unlikely that in practice the effect of more lagged values than the first will be removed. Nevertheless, Quenouille's statistic $r(x_t, y_t \mid x_{t-1}, y_{t-1})$ will still give a fairly efficient test provided $|a_2|$ is not too high. In Table 11 the ratio $E\{\hat{\gamma}_1, r(x_t, y_t \mid x_{t-1}, y_{t-1}) \mid \rho = 0\}$ is shown for certain values of ρ_1, ρ_2 and a_2. Here $\rho_2 = \frac{-a_1}{1+a_2}$

<u>Table 11.</u>

$$E\{\hat{\gamma}_1, r(x_t, y_t \mid x_{t-1}, y_{t-1}) \mid \rho = 0\} = \frac{\left\{1 + \frac{\rho_1\rho_2}{1-a_2}\right\}^2}{2(1+\rho_1^2)(1+\rho_2^2\frac{1+a_2}{1-a_2})}$$

a_2 \ ρ_1, ρ_2	$\rho_1 = \rho_2 = 0.4$	$\rho_1 = \rho_2 = 0.6$	$\rho_1 = \rho_2 = 0.8$	$\rho_1 = 0.6$, $\rho_2 = 0.8$	$\rho_1 = 0.8$, $\rho_2 = 0.6$
-0.6	0.50	0.51	0.52	0.54	0.47
-0.4	0.50	0.50	0.51	0.51	0.47
-0.2	0.50	0.50	0.50	0.51	0.48
0	0.50	0.50	0.50	0.49	0.49
0.2	0.50	0.50	0.50	0.48	0.51
0.4	0.50	0.51	0.52	0.48	0.54
0.6	0.52	0.54	0.58	0.50	0.60

For a_2 sufficiently near to unity the ratio just tabulated will be greater than unity but in this case, of course, both statistics give very inefficient tests.

Example: The data for Grouse and Blackgame from Moran (1952) may be used for illustration (see the example of # 5.3.). The Blackgame series appear to be Markovian but the first partial serial correlation for the Grouse series is 0.34 (from 73 observations) which is highly significant. The first serial correlation for Grouse is 0.64 and for Blackgame 0.43. Using $a_2 = 0.3$, $\rho_1 = 0.7$ and $\rho_2 = 0.5$ we obtain:

$$E\{\hat{\gamma}_1, r \mid \rho = 0\} = 1.02 \qquad E\{\hat{\gamma}_1, r(x_t, y_t \mid x_{t-1}, y_{t-1}) \mid \rho = 0\} = 0.47$$

The results for the various statistics are shown in Table 12.

Table 12.

Statistic	Sample Value	Test Statistic
r	0.347	r_S = 0.347 *
$\hat{\gamma}_1$	0.014	c_{31} = 2.56 *
$r(x_t, y_t \mid x_{t-1}, y_{t-1})$	0.327	r_Q = 0.327 **

* Significant at 5% point. ** Highly significant.

These results again appear to agree with the theory.

5.5. Exact tests for multiple correlation.

As shown in # 2.3. the asymptotic relative effieiency of two multiple correlations, as tests of an hypothesis $\theta = \theta_0$ (where all of the regression coefficients depend upon θ) can be satisfactorily defined as the ratio of two quadratic forms, when the sequence of alternatives considered is of the type $\theta_n = \theta_0 + kn^{-\frac{1}{2}}$. These quadratic forms (which may be termed the efficacies of the multiple correlations) are quadratic forms in the efficacies of the sample regression coefficients, from which each multiple correlation has been computed, the matrices of the forms being the inverses of the correlation matrices of the regression coefficients (at $\theta = \theta_0$).

In general, the regression coefficients will not be directly specified as functions of a single parameter but for the purposes of an asymptotic criterion they can be, formally, so specified the j^{th} coefficient β_j being put equal to $k_j\theta$.

If we write the covariance matrix of the β_j (at $\theta = \theta_0$) as $n^{-1}[m_{ij}]$ then the correlation matrix is $\left[\frac{m_{ij}}{\{m_{ii}m_{jj}\}^{1/2}}\right]$ and the efficacy of the corresponding multiple correlation is:

$$(5.5.1.)\quad \left\{\frac{k_i}{(m_{ii})^{1/2}}\right\}' \left[\frac{m_{ij}}{\{m_{ii}m_{jj}\}^{1/2}}\right]^{-1} \left\{\frac{k_i}{(m_{ii})^{1/2}}\right\} = \{k_i\}'[m_{ij}]^{-1}\{k_i\}$$

We shall consider the case where each regressor is Markovian and independent of the other regressors and the residual is Markovian also and independent of the regressors. Using the notation of # 5.1. the matrices for the various regression methods there considered were given by (5.1.1,2,3,4)'''.

We shall consider in particular the ratios of the efficacies of the multiple correlations corresponding to the ordinary least squares procedure and to the $\hat{Y}_j$ (as obtained from the procedure formulated in # 4.1.) for $j \leq p$ (where p is the number of regressors.)

Writing R for the ordinary multiple correlation and Γ for the partial multiple correlation computed from the $\hat{Y}_j$ $(j = 1, \cdots, p)$ we then have:

$$(5.5.2.) \quad E\{R, \Gamma \mid \theta = 0\} = 2\frac{\sum_j k_j^2 \dfrac{1-\rho_1\rho_{j1}}{1+\rho_1\rho_{j1}}}{\sum_j k_j^2 \dfrac{(1+\rho_1^2)(1-\rho_{j1}^2)}{(1-\rho_1^2)(1+\rho_{j1}^2)}}$$

It is again evident that for any fixed ρ_{j1} and k_j this will be less than unity for ρ_1 sufficiently near to unity.

The test based on Γ could be useful even though it is clearly inefficient (since (5.1.3.)''' is always less than (5.1.2.)''') for not much consideration has been so far given to tests of significance of multiple correlation when the residual from the regression is serially correlated. One can obtain an asymptotically fully efficient test by the use of a consistent estimator of ρ_1, but for small samples the biases present might be considerable. Again one could generalise Quenouille's statistics by taking the multiple partial correlation with the effects of lagged value of all variates involved removed, and one suspects that the small sample properties of this test (with respect to the variance on the null hypothesis) will be reasonably satisfactory. However, this has not yet been shown. Moreover, the $\hat{\gamma}_j$ at once give an unbiased estimate of the β_j together with an exact test (or a confidence interval). In any case the method which leads to Γ seems to have some advantages over the straightforward least squares procedure in econometrics (where ρ_1 is fairly certain to be high and positive). The least squares procedure has been recommended by Wold (1953).

A more thorough examination of the efficiency of Γ could be made and the inclusion of the remaining $\hat{\gamma}_j\ (j > p+1)$ could be considered but the formulae become forbiddingly complicated. The analysis could also be extended to the cases where all of the series are Markovian and the multiple correlation arises from a multiple correlation between the "shocks" of which the series are composed.

5.6. <u>Exact tests for correlation when the residual is an autoregressive process of order higher than the first.</u>

The whole analysis of # 5.1. to # 5.5. could be carried over to the case where the residual is generated by a process of the form (1.2.2.) with $h > 1$. However, it is evident from the results of # 3.2. that the exact tests will now become very inefficient and at the same time the computational procedures required in order to obtain exactness will become very great.

These facts, of course, underline a deficiency of the exact tests. They are only exact provided the residual is Markovian. This difficulty does not arise in the case of the test of significance of $\hat{\gamma}_{p+1}$ (test of the independence of the residual) since there the null hypothesis always validates this test (subject to normality of distribution, of course).

Chapter 6.

Exact Tests for Correlation in Vector Processes.

6.1. The canonical analysis of a vector Markoff process.

A vector of variates:

(6.1.1.) $$\underset{\sim}{u}_t' = (u_{1t}, u_{2t}, \ldots, u_{pt})$$

will be generated by a stationary vector Markoff process if it satisfies the relation:

(6.1.2.) $$\underset{\sim}{u}_t = \underset{\sim}{B}\, \underset{\sim}{u}_{t-1} + \underset{\sim}{\xi}_t$$

where

(a) $\underset{\sim}{B}$ is a square matrix of order p having all of its latent roots inside the unit circle.

(b) $\underset{\sim}{\xi}_t$ is a vector of p elements, ξ_{jt} which is generated by a process of independent random vectors. (Mann and Wald (1943)).

The $\underset{\sim}{\xi}_t$ process will be taken to be normal in what follows, with covariance matrix $\underset{\sim}{\Gamma}$, and there will be no loss of generality if the mean value of $\underset{\sim}{\xi}_t$ is taken as the null vector.

An overall test of the serial independence of the process could be obtained from a canonical analysis of the relation between $\underset{\sim}{u}_t$ and $\underset{\sim}{u}_{t-1}$ (Kendall (1946b) page 348). Apart from the neglect of information in cases where some of the elements of $\underset{\sim}{B}$ can be prescribed in advance, the tests based on a canonical analysis would have some appeal were the distributions of the test statistics known.

An alternative procedure which gives a test of the hypothesis that the $\underset{\sim}{u}_t$ process is independent can be obtained by an extension of the method given for scalar processes by Ogawara (1951) (see Chapter 3). This consists in carrying out a canonical analysis of the relation between the vectors $\underset{\sim}{u}_{2t}$ and the vectors:

$$\underset{\sim}{V}_{2t} = \begin{Bmatrix} \underset{\sim}{u}_{2t-1} \\ \underset{\sim}{u}_{2t+1} \end{Bmatrix}$$

The matrix of regression coefficients, of $\underset{\sim}{u}_{2t}$ on $\underset{\sim}{V}_{2t}$, will be (from (3.2.5.)):

$$(6.1.3.)\qquad \underset{\sim}{M} = (\underset{\sim}{B}\underset{\sim}{\Lambda} \quad \underset{\sim}{\Lambda}\underset{\sim}{B}') \begin{bmatrix} \underset{\sim}{\Lambda} & \underset{\sim}{\Lambda}(\underset{\sim}{B}')^2 \\ \underset{\sim}{B}^2\underset{\sim}{\Lambda} & \underset{\sim}{\Lambda} \end{bmatrix}^{-1}$$

where

$$\underset{\sim}{\Lambda} = \mathcal{E}(\underset{\sim}{u}_t \underset{\sim}{u}_t') = \sum_{j=0}^{\infty} \underset{\sim}{B}^j \underset{\sim}{\Gamma} (\underset{\sim}{B}')^j$$

The conditional covariance matrix of the $\underset{\sim}{u}_{2t}$ is (from (3.2.4.)):

$$(6.1.4.)\qquad \underset{\sim}{\Lambda} - [\underset{\sim}{B}\underset{\sim}{\Lambda} \quad \underset{\sim}{\Lambda}\underset{\sim}{B}'] \begin{bmatrix} \underset{\sim}{\Lambda} & \underset{\sim}{\Lambda}(\underset{\sim}{B}')^2 \\ \underset{\sim}{B}^2\underset{\sim}{\Lambda} & \underset{\sim}{\Lambda} \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{\Lambda}'\underset{\sim}{B}' \\ \underset{\sim}{B}\underset{\sim}{\Lambda} \end{bmatrix}$$

It is fairly easy to show that (6.1.3.) and (6.1.4.) reduce to:*

---

* The inverse exists because of (a) below (6.1.2.)

$$(6.1.5.)\quad \begin{cases} \left[\underset{\sim}{B} - \underset{\sim}{\Gamma} \underset{\sim}{B}' [\underset{\sim}{\Gamma} + \underset{\sim}{B}\underset{\sim}{\Gamma}\underset{\sim}{B}']^{-1} \underset{\sim}{B}^{2} \qquad \underset{\sim}{\Gamma}\underset{\sim}{B}'[\underset{\sim}{\Gamma} + \underset{\sim}{B}\underset{\sim}{\Gamma}\underset{\sim}{B}']^{-1} \right] \\ \qquad\qquad \underset{\sim}{\Gamma} - \underset{\sim}{\Gamma}\underset{\sim}{B}'[\underset{\sim}{\Gamma} + \underset{\sim}{B}\underset{\sim}{\Gamma}\underset{\sim}{B}']^{-1} \underset{\sim}{B}\underset{\sim}{\Gamma} \end{cases}$$

Since the validity of the tests, based on the canonical analysis, hold for arbitrary fixed vectors of regressors provided that the vectors of regressand are then normally and independently distributed the canonical correlation of $\underset{\sim}{u}_{2t}$ and $\underset{\sim}{V}_{2t}$ will give a test of the independence of the $\underset{\sim}{u}_{t}$ process. For fixed $\underset{\sim}{V}_{2t}$ the $\underset{\sim}{u}_{2t}$ are serially independent even if $\underset{\sim}{B}$ is not null so that the whole appartus of canonical analysis is available and not merely that part which tests the hypothesis that $\underset{\sim}{B}$ is null.

The canonical correlations ℓ_j between $\underset{\sim}{u}_{2t}$ and $\underset{\sim}{V}_{2t}$ are obtained as the roots of the determinantal equation:

$$(6.1.6.)\qquad \left| \underset{\sim}{C}_{21} \underset{\sim}{C}_{11}^{-1} \underset{\sim}{C}_{12} - \ell^{2} \underset{\sim}{C}_{22} \right| = 0$$

Here

$$(6.1.7.)\qquad \underset{\sim}{C}_{22} = \sum_{t=1}^{[\frac{n-1}{2}]} \underset{\sim}{u}_{2t} \underset{\sim}{u}_{2t}', \qquad \underset{\sim}{C}_{11} = \sum_{t=1}^{[\frac{n-1}{2}]} \underset{\sim}{V}_{2t} \underset{\sim}{V}_{2t}'$$

$$\underset{\sim}{C}_{21} = \sum_{t=1}^{[\frac{n-1}{2}]} \underset{\sim}{u}_{2t} \underset{\sim}{V}_{2t}'$$

where $[\frac{1}{2}(n-1)]$ is again used to indicate the integral part of $\frac{1}{2}(n - 1)$.

An overall test of significance of the ℓ_j is obtained from the quantity $\chi^2_{2p^2} = -\{[\frac{1}{2}(n-1)] - \frac{1}{2}(3p+1)\}\log_e \Lambda$ which is (approximately) distributed as χ^2 with $2p^2$ degrees of freedom (Bartlett (1938), Wald and Brookner (1941), Rao (1948)).

Here:

(6.1.8.) $$\Lambda = \frac{|C_{22} - C_{21} C_{11}^{-1} C_{12}|}{|C_{22}|} = \prod_1^k (1 - \ell_j^2)$$

When p equals 2 an exact test is available, for the quantity $F = \frac{1-\lambda^{1/2}}{\lambda^{1/2}} \frac{[\frac{1}{2}(n-1)] - 5}{4}$ is then distributed in the F distribution with 8 and $2([\frac{n-1}{2}] - 5)$ degrees of freedom. (Pearson and Wilks (1933)).

When p is greater than 2 no simple exact test is available but using the results of Wald and Brookner (1941), Rao (1948) obtained an asymptotic expansion for the distribution of χ^2_{2p} of the form:

(6.1.9.) $$P_{pq} + \frac{\gamma_2}{m^2}(P_{pq+4} - P_{pq}) - \frac{1}{m^4}\{\gamma_4(P_{pq+8} - P_{pq}) - \gamma_2^2(P_{pq+4} - P_{pq})\} + \dots$$

where, in the present case, $m = \{[\frac{1}{2}(n-1)] - \frac{1}{2}(3p+1)\}$, $q = 2p$,

$\gamma_2 = \frac{2p^2}{48}(5p^2 - 5)$, $\gamma_4 = \frac{\gamma_2^2}{2} + \frac{2p^2}{1920}\{91p^4 - 250p^2 + 150\}$

and P_{pq+s} is the distribution function of χ^2 for $pq+s$ degrees of freedom.

Thus the significance point may be obtained from the distribution of χ^2 alone. Bartlett's approximation corresponds to the use of the first term of the expansion.

Under some circumstances the largest root of the equation (6.1.4.) will provide a more powerful test of the serial independence of the process $\underset{\sim}{x}_t$ (for example if $\underset{\sim}{B}$ is of rank one). In certain circumstances (p = 2) an exact test is available for the largest root (Marriott (1952)). An approximate procedure in any case is to take

$$-\left\{\left[\frac{n-1}{2}\right] - \frac{1}{2}(3p+1)\right\}\log_e(1-\ell_1^2)$$

as χ^2_D where $D = 3p-1+\frac{1}{2}\{(p-1)(2p-1)\}^{2/3}$ (Marriott (1952)), where ℓ_1 is the largest root.

If this largest root is significant Bartlett (1938,1947) suggests that the remaining roots may be tested by using

$$\chi_1^2 = -\left\{[1/2(n-1)] - \frac{1}{2}(3p+1)\right\}\log_e\left\{\prod_2^p(1-\ell_i^2)\right\}$$

as χ^2 with (p - 1)(2p - 1) degrees of freedom. For large samples the largest root will correspond to the largest true canonical correlation, but if all of the canonical correlations in the population are zero the selection of the largest root will bias the value of χ_1^2 downwards so that the test will not be valid (Bartlett (1947)).

In the next section we shall examine the asymptotic efficiency of the procedure suggested here.

6.2. The asymptotic efficiency of the exact test.

It is well known that, under certain regularity conditions, if λ is a likelihood ratio then $-2\log_e\lambda$ is asymptotically distributed as χ^2 with s degrees of freedom, where s is the number of parameters whose values are specified by the null hypothesis which λ is designed to test. (Wilks (1938)). This follows, fundamentally, from the fact that, asymptotically,

$$-2\log_e\lambda = \sum_{i,j=1}^{s}(\hat{\theta}_i-\theta_i^0)(\hat{\theta}_j-\theta_j^0)c_{ij}$$

where θ_i^0 are the specified values, $\hat{\theta}_i$ the maximum likelihood estimators of the corresponding parameters (when all of the $r+s$ parameter values are estimated) and the c_{ij} are the elements of the inverse of the matrix in:

$$(6.2.1.) \qquad -\left[\mathcal{E}\left\{\frac{\partial^2\log L}{\partial\theta_i\partial\theta_j}\right\}\right]^{-1} \qquad (i,j=1,\ldots,r+s)$$

which corresponds to the θ_i^0 (i.e. the matrix in the bottom right hand corner if the suffixes $(r+1,\ldots,r+s)$ are allotted to the parameters specifying the null hypothesis.). Here L is the likelihood function.

We may consider the sequence of alternatives:

$$\theta_i = f_i(\theta)\ , \qquad \theta = kn^{-1/2} \quad (i=r+1,\ldots,r+s)$$

where now, without loss of generality, we have put the θ_i^0 equal to zero.

The asymptotic power of the test based on λ

will be, from # 2.3, $G_s(k^2 Q, \alpha)$ where Q is a quadratic form in the efficacies of the $\hat{\theta}_i$, $(i = r+1, \cdots, r+s)$ whose matrix is the inverse of the correlation matrix of the $\hat{\theta}_i$, on the null hypothesis.

The Λ criterion of Wilks, which appears in (6.1.8.) is of course a likelihood ratio. Equally if we took the corresponding criterion (λ_1 , say) obtained by treating the $\underset{\sim}{x}_t$ and the $\underset{\sim}{x}_{t-1}$ as separate sets of vectors and carrying out a canonical analysis, this would also lead to a test statistic asymptotically distributed as χ^2 (Of course the exact tests of Pearson and Wilks and Rao would no longer hold.) This follows from the fact that this criterion would still be, apart from certain end effects whose influences vanish in the limit, a likelihood ratio.

We may, therefore, use the results of # 2.3. to compare the power, asymptotically, of the test presented in # 6.1. and what appears to be the best overall test of the correlation of the vector process, i.e. the test based on λ_1 . The "overall" in the last sentence should be emphasised since no uniformly most powerful test will exist and under some circumstances a more powerful test than either of those discussed may be obtained by another (and perhaps a simpler) approach.

The null hypothesis now relates to the elements b_{ij} of $\underset{\sim}{B}$ and we shall put:

$$b_{ijn} = k_{ij}\,\theta_n \quad , \theta_n = k n^{-1/2}$$

The distance from the origin for the test based on λ_1 is now:

$$k^2 Q_1 = k^2 \lim_{n\to\infty}\left\{-n^{-1}\sum_{ijke=1}^{p} k_{ij} k_{ke}\, \mathcal{E}\left\{\frac{\partial^2 \log L}{\partial b_{ij}\partial b_{ke}}\right\}_{\underset{\sim}{B}=\underset{\sim}{0}}\right\}$$

since the $\hat{b}_{ij}$ will be asymptotically independent of the other unspecified parameters (those in $\underset{\sim}{\Gamma}$).

We have:

$$\mathcal{E}\left\{\frac{\partial^2 \log L}{\partial b_{ij}\partial b_{ke}}\right\}_{\underset{\sim}{B}=\underset{\sim}{0}} = \mathcal{E}\left\{-\tfrac{1}{2}\sum_t\left(\underset{\sim}{u}'_{t-1}\left[\frac{\partial}{\partial b_{ij}\partial b_{ke}}\,\underset{\sim}{B}'\underset{\sim}{\Gamma}^{-1}\underset{\sim}{B}\right]\underset{\sim}{u}_{t-1}\right)\right\}_{\underset{\sim}{B}=\underset{\sim}{0}}$$

$$= \mathcal{E}\left\{-\tfrac{1}{2}\sum_t 2\,\Gamma^{ik}\left(u_{j,t-1}\, u_{e,t-1}\right)\right\}_{\underset{\sim}{B}=\underset{\sim}{0}}$$

where Γ^{ik} is the element in the i^{th} row and the k^{th} column of $\underset{\sim}{\Gamma}^{-1}$.

$$= n\,\Gamma^{ik}\gamma_{je} \qquad (\underset{\sim}{\Gamma} = [\gamma_{ij}])$$

Hence:

$$k^2 Q_1 = k^2 \sum_{ijke=1}^{p}\left(k_{ij} k_{ke}\,\Gamma^{ik}\gamma_{je}\right)$$

In connection with the corresponding distance for the test presented in # 6.1. we require:

$$\frac{\partial}{\partial\theta}\left[\underset{\sim}{B} - \underset{\sim}{\Gamma}\underset{\sim}{B}'[\underset{\sim}{\Gamma}+\underset{\sim}{B}\underset{\sim}{\Gamma}\underset{\sim}{B}']^{-1}\underset{\sim}{B}^2 \qquad \underset{\sim}{\Gamma}\underset{\sim}{B}'[\underset{\sim}{\Gamma}+\underset{\sim}{B}\underset{\sim}{\Gamma}\underset{\sim}{B}']^{-1}\right]_{\underset{\sim}{B}=\underset{\sim}{0}}$$

$$= \left[[k_{ij}] \qquad \underset{\sim}{\Gamma}[k_{ij}]'\underset{\sim}{\Gamma}^{-1}\right]$$

The corresponding distance then is:

$$k^2 Q_2 = k^2\tfrac{1}{2}\sum_{ijke}\Gamma^{ik}\gamma_{je}\left\{k_{ij}k_{ke} + \left(\sum_{\mu\nu=1}^{p}\gamma_{i\nu}k_{\mu\nu}\Gamma^{\mu j}\right)\left(\sum_{\mu\nu=1}^{p}\gamma_{k\nu}k_{\mu\nu}\Gamma^{\mu e}\right)\right\}$$

If we write $\underset{\sim}{A} * \underset{\sim}{B}$ to indicate the sum of the elements obtained by multiplying each element of $\underset{\sim}{A}$ by the corresponding element of $\underset{\sim}{B}$, then:

$$\sum_{ijkl} \Gamma^{ik} \gamma_{jl} k_{ij} k_{kl} = [\underset{\sim}{\Gamma}^{-1} \underset{\sim}{K} \underset{\sim}{\Gamma}] * K \quad ; \quad K = [k_{ij}]$$

and

$$\sum_{ijkl} \Gamma^{ik} \gamma_{jl} \left(\sum_{\mu\nu=1}^{p} \gamma_{i\nu} k_{\mu\nu} \Gamma^{\mu j}\right)\left(\sum_{\mu\nu=1}^{p} \gamma_{k\nu} k_{\mu\nu} \Gamma^{\mu l}\right)$$

$$= [\underset{\sim}{\Gamma}^{-1} \underset{\sim}{\Gamma} \underset{\sim}{K}' \underset{\sim}{\Gamma}^{-1} \underset{\sim}{\Gamma}] * \underset{\sim}{\Gamma} \underset{\sim}{K}' \underset{\sim}{\Gamma}^{-1} = \underset{\sim}{K}' * \underset{\sim}{\Gamma} \underset{\sim}{K}' \underset{\sim}{\Gamma}^{-1} = \underset{\sim}{\Gamma}^{-1} \underset{\sim}{K} \underset{\sim}{\Gamma} * K$$

Hence $k^2 Q_2 = k^2 Q_1$

It follows, since $2p^2 > p^2$, that the test based on Λ_1 is always asymptotically more powerful than that given in # 6.1. When p = 1 the effect is clearly seen for now the latter test is that which results from ignoring the equality of the coefficients of x_{2t-1} and x_{2t+1} in (3.1.3.) so that a multiple correlation is used instead of a t test and the same kind of effect as that mentioned in # 4.3. reduces the power of the test. (A region "exterior" to a circle is used instead of the region "exterior" to a pair of lines orthogonal to the line on which the alternatives lie). In general (when a transformation to appropriate axes is made) a region exterior to a hypersphere in $2p^2$ dimensions is used instead of a region exterior to a hypercylinder (with base a p dimensional hypersphere) whose generators are orthogonal to the hyperplane in which the alternatives lie.

A condition which will ensure that the two matrices (into which the first matrix in (6.1.5.) is partitioned) should be equal is that $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$ should commute and $\underset{\sim}{B}$ should be symmetric. Since, in general, it will not be possible to prescribe this in advance the information lost through the doubling of the number of regression coefficients is not recoverable. The only circumstances under which this condition would be likely to be recognised in advance would be those which make both $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$ diagonal, but then it would be foolish to use a canonical analysis.

The loss of efficiency by the use of Ogawara's method can here by related to the necessity of introducing additional parameters, as was mentioned at the end of # 3.3.

6.3. The canonical analysis of a vector of residuals.

Consider now the multiple regression:

(6.3.1.) $$\underset{\sim}{y}_t = \underset{\sim}{D}\,\underset{\sim}{x}_t + \underset{\sim}{u}_t$$

where $\underset{\sim}{y}_t$ is a vector of p elements, $\underset{\sim}{x}_t$ is a vector of q elements and $\underset{\sim}{u}_t$ comes from the vector process (6.1.1.). Here $\underset{\sim}{D}$ is a $p\times q$ matrix of regression coefficients.

Then for fixed $\underset{\sim}{x}_t, (t=1,\cdots,n)$ and $\underset{\sim}{y}_{2t+1}, (t=0,\cdots,[\frac{n-1}{2}])$ the vector $\underset{\sim}{y}_{2t}$ has mean value:

(6.3.2.) $\underset{\sim}{D}\underset{\sim}{x}_{2t} + \underset{\sim}{M}\begin{Bmatrix} \underset{\sim}{y}_{2t-1} \\ \underset{\sim}{y}_{2t+1} \end{Bmatrix} - \underset{\sim}{M}\begin{bmatrix} \underset{\sim}{D} & \underset{\sim}{O} \\ \underset{\sim}{O} & \underset{\sim}{D} \end{bmatrix}\begin{Bmatrix} \underset{\sim}{x}_{2t-1} \\ \underset{\sim}{x}_{2t+1} \end{Bmatrix}$

where $\underset{\sim}{M}$ is given by the first relation in (6.1.5.) and $\underset{\sim}{O}$ is the null matrix $(p \times q)$.

Moreover, $\underset{\sim}{y}_{2t}$ is now serially independent with covariance matrix given by the second matrix in (6.1.5.).

One can now carry over the analysis of Chapter 4. The serial independence of the $\underset{\sim}{u}_t$ can be tested by a partial canonical analysis of the relation between $\underset{\sim}{y}_{2t}'$ and $(\underset{\sim}{y}_{2t-1}' \ \ \underset{\sim}{y}_{2t+1}')$ when the effects of the vector variates $\underset{\sim}{x}_{2t}'$ and $(\underset{\sim}{x}_{2t-1}' \ \ \underset{\sim}{x}_{2t+1}')$ are removed.

Let us write:

(6.3.3.) $$\begin{cases} \underset{\sim}{C}_{22} = \sum_t \underset{\sim}{y}_{2t}\underset{\sim}{y}_{2t}' & \underset{\sim}{C}_{2y} = \sum_t \underset{\sim}{y}_{2t}(\underset{\sim}{y}_{2t-1}' \ \ \underset{\sim}{y}_{2t+1}') \\ \underset{\sim}{C}_{yy} = \sum_t \begin{Bmatrix} \underset{\sim}{y}_{2t-1} \\ \underset{\sim}{y}_{2t+1} \end{Bmatrix}(\underset{\sim}{y}_{2t-1}' \ \ \underset{\sim}{y}_{2t+1}') & \underset{\sim}{C}_{xx} = \sum_t \begin{Bmatrix} \underset{\sim}{x}_{2t} \\ \underset{\sim}{x}_{2t-1} \\ \underset{\sim}{x}_{2t+1} \end{Bmatrix}(\underset{\sim}{x}_{2t}' \ \underset{\sim}{x}_{2t-1}' \ \underset{\sim}{x}_{2t+1}') \\ \underset{\sim}{C}_{2x} = \sum_t \underset{\sim}{y}_{2t}(\underset{\sim}{x}_{2t}' \ \underset{\sim}{x}_{2t-1}' \ \underset{\sim}{x}_{2t+1}') & \underset{\sim}{C}_{yx} = \sum_t \begin{Bmatrix} \underset{\sim}{y}_{2t-1} \\ \underset{\sim}{y}_{2t+1} \end{Bmatrix}(\underset{\sim}{x}_{2t}' \ \underset{\sim}{x}_{2t-1}' \ \underset{\sim}{x}_{2t+1}') \end{cases}$$

where all summations are from 1 to $[\tfrac{1}{2}n-1]$.

Then if:

(6.3.4.) $$\begin{cases} \underset{\sim}{C}_{22\cdot x} = \underset{\sim}{C}_{22} - \underset{\sim}{C}_{2x}\underset{\sim}{C}_{xx}^{-1}\underset{\sim}{C}_{x2} \\ \underset{\sim}{C}_{yy\cdot x} = \underset{\sim}{C}_{yy} - \underset{\sim}{C}_{yx}\underset{\sim}{C}_{xx}^{-1}\underset{\sim}{C}_{xy} \\ \underset{\sim}{C}_{2y\cdot x} = \underset{\sim}{C}_{2y} - \underset{\sim}{C}_{2x}\underset{\sim}{C}_{xx}^{-1}\underset{\sim}{C}_{xy} \end{cases}$$

the partial canonical correlations ℓ_j are found as the roots of the determinantal equation:

(6.3.5.) $$\left| \underset{\sim}{C}_{2y\cdot x}\underset{\sim}{C}_{yy\cdot x}^{-1}\underset{\sim}{C}_{2y\cdot x}' - \ell^2 \underset{\sim}{C}_{22\cdot x} \right| = 0$$

then Wilk's test criterion is:

$$\lambda_2 = \frac{|C_{22\cdot x} - C_{2y\cdot x} C_{yy\cdot x}^{-1} C_{2y\cdot x}'|}{|C_{22\cdot x}|}$$

and $\chi^2_{2p^2} = -\{[\frac{1}{2}(n-1)] - 3q - \frac{1}{2}(3p+1)\}\log_e \lambda_2$ is approximately distributed as χ^2 with $2p^2$ degrees of freedom, and more exact tests may be obtained by using (6.1.9.) with a suitable modification of m (or the appropriate modification of the F test given by Pearson and Wilks for the case p = 2).

It is clear that, asymptotically, the relation between the power of this test, as compared with that obtained by carrying out a partial canonical analysis of the relation between y_t and y_{t-1}, with the effects of x_t removed, will be the same as the relation between the power of the test given in # 6.1. (as compared with the test based on the canonical analysis of the relation between u_t and u_{t-1}).

In addition to this test just described, (6.3.2.) also leads to a test of the independence of y_t and x_t, this test being obtained by taking the partial canonical correlation of y_{2t}' and x_{2t}' with the effects of (y_{2t-1}' y_{2t+1}') and (x_{2t-1}' x_{2t+1}') removed, or alternatively the partial canonical correlation of y_{2t}' and (x_{2t}' x_{2t-1}' x_{2t+1}') with the effects of (y_{2t-1}' y_{2t+1}') removed.

With regard to this last test a similar situation to that met in # 5.2. will be met here. Some additional information relative to the independence of $\underset{\sim}{y}_t$ and $\underset{\sim}{x}_t$ will, in general, be available from the estimated coefficients of $\underset{\sim}{x}_{2t-1}$ and $\underset{\sim}{x}_{2t+1}$, but the use of this information (since $\underset{\sim}{M}$ will not be known a priori) will require that instead of writing the matrix of regression coefficients as:

$$[\underset{\sim}{I} \quad \underset{\sim}{M}] \begin{bmatrix} \underset{\sim}{D} & \underset{\sim}{O} & \underset{\sim}{O} \\ \underset{\sim}{O} & \underset{\sim}{D} & \underset{\sim}{O} \\ \underset{\sim}{O} & \underset{\sim}{O} & \underset{\sim}{D} \end{bmatrix}$$

this will have to be replaced by a general $p \times 3q$ matrix and, unless the additional information from $\underset{\sim}{x}_{2t-1}$ and $\underset{\sim}{x}_{2t+1}$ is large, the test will be less powerful than that based on the partial canonical correlation of $\underset{\sim}{y}_{2t}$ and $\underset{\sim}{x}_{2t}$ alone because of the inefficient weighting of the coefficients of $\underset{\sim}{x}_{2t}$ in the analysis.

The calculation of the asymptotic power of either of the two tests just discussed could now be carried out, once a prescription of the nature of the process generating the vector $\underset{\sim}{x}_t$ was made, following the method used in # 6.2. The analysis is fairly straightforward but will not be effected here. Its only purpose would be to compare the power of the test with that obtained from a canonical analysis of the relation between ($\underset{\sim}{y}_t - \underset{\sim}{B}\,\underset{\sim}{y}_{t-1}$) and

($\underset{\sim}{x}_t - \underset{\sim}{B}\, \underset{\sim}{x}_{t-1}$), which will be asymptotically equivalent to the maximum likelihood procedure. We know that the last mentioned procedure will be superior (though computationally difficult and with unknown small sample properties) and a numerical comparison of the two power functions will be made difficult by the necessity to interpolate in the tables of the power function for the test, even when the most suitable alternatives (from the point of view of the tables of this function) are chosen.

Two extensions of the subject matter of this section may be considered. The first of these corresponds to the case discussed in # 6.1. when the matrix $\underset{\sim}{B}$ is diagonal and it is necessary to test whether $\underset{\sim}{\Gamma}$ is diagonal also, or is the direct sum of a number of submatrices. This is the natural extension of the test discussed in # 5.3. The procedure of this section may clearly be modified to deal with this case. (See Wilks (1935)).

The second extension relates to a relation of the form:

$$\underset{\sim}{y}_t = \underset{\sim}{D}\, \underset{\sim}{x}_t + \sum_{j=0}^{s} \underset{\sim}{E}_j\, \underset{\sim}{y}_{t-j} + \underset{\sim}{u}_t$$

Such systems of relations are fundamental in econometrics and procedures for estimation based on maximum likelihood, or on some modified maximum likelihood procedure, have been considered in the case where $\underset{\sim}{u}_t$ is serially independent. (See Koopmans, 1950). When $\underset{\sim}{u}_t$ is of the form (6.1.1.) we may consider the regression of $\underset{\sim}{y}_{2t}$ on:

$$\underset{\sim}{D}\,\underset{\sim}{x}_{2t} + \sum_{j=0}^{s} \underset{\sim}{E}_j\, \underset{\sim}{y}_{2t-j} + \underset{\sim}{M}\begin{Bmatrix} \underset{\sim}{y}_{2t-1} \\ \underset{\sim}{y}_{2t+1} \end{Bmatrix} - \underset{\sim}{M}\begin{bmatrix} \underset{\sim}{D} & \underset{\sim}{O} \\ \underset{\sim}{O} & \underset{\sim}{D} \end{bmatrix}\begin{Bmatrix} \underset{\sim}{x}_{2t-1} \\ \underset{\sim}{x}_{2t+1} \end{Bmatrix} - \sum_{j=0}^{s} \underset{\sim}{M}\begin{bmatrix} \underset{\sim}{E}_j & \underset{\sim}{O} \\ \underset{\sim}{O} & \underset{\sim}{E}_j \end{bmatrix}\begin{Bmatrix} \underset{\sim}{y}_{2t-j-1} \\ \underset{\sim}{y}_{2t-j+1} \end{Bmatrix} \qquad (6.3.6.)$$

The same procedures as those applied in the case $\underset{\sim}{B} = \underset{\sim}{O}$ will give estimates of $\underset{\sim}{D}$ and the $\underset{\sim}{E}_j$ in the present case, and one expects these latter estimates to be more efficient under some circumstances than those obtained by treating $\underset{\sim}{B}$ as null when, in fact, it is not.*

No exact test procedures (for the significance of the coefficients in $\underset{\sim}{D}$ or the $\underset{\sim}{E}_j$) have been developed so far even for the case where $\underset{\sim}{B}$ is null. Any that are developed could, of course, be applied directly to (6.3.6.).

The computations for even modified maximum likelihood procedures (see Chapter IX of Koopmans (1950)) are computationally very burdensome and the computations for (6.3.6.) will be very great indeed.

6.4. Multiple Markoff processes.

A particular case of (6.1.1.) is obtained from the multiple Markoff process (1.2.2.) with h equalling p

* New identification problems arise, however.

(with ξ_t on the r.h.s. replaced for convenience by η_t)
when we put:

$$\underset{\sim}{u}_t' = (x_{tp}, x_{tp+1}, \cdots\cdots, x_{(t+1)p-1})$$

$$\underset{\sim}{B} = \left[\begin{matrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ a_p & a_{p-1} & a_{p-2} & & a_1 \end{matrix}\right] p$$

(6.4.1.)

$$\underset{\sim}{\xi}_t = (\eta_{tp}\ \eta_{tp+1} \cdots \eta_{(t+1)p-1}) \begin{bmatrix} 1 & -a_1 & \cdots & -a_{p-1} \\ 0 & 1 & \cdots & -a_{p-2} \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

In order that the process should be stationary it is necessary and sufficient that (1.2.2.)' (with h = p) should have all of its roots inside the unit circle.

An overall test of serial correlation is obtained from the canonical analysis of the relation between $\underset{\sim}{u}_{2t}'$ and $(\underset{\sim}{u}_{2t-1}'\ \underset{\sim}{u}_{2t+1}')$. It is easy to see that in the present case the first s and last s columns of $\underset{\sim}{M}$ will be null when (1.2.2.) is of order p - s instead of p. As a result a test of the hypothesis that the x_t process is of order p - s can be obtained by carrying out a partial canonical analysis of the relation between the vectors $\underset{\sim}{u}_{2t}'$ and $\underset{\sim}{V}_{2t}' = (x_{(t-1)p+s}, \cdots, x_{tp-1}, x_{(t+1)p}, \cdots, x_{(t+2)p-(s+1)})$ with the effects of $\underset{\sim}{w}_t'$ removed where

$$\underset{\sim}{w}_t' = (x_{(t-1)p}, \cdots\cdots, x_{(t-1)p+s}, x_{(t+2)p+s}, \cdots\cdots, x_{(t+2)p-1})$$

Let us consider the second order case which is simple and illuminating.

The matrix $\underset{\sim}{M}$ now becomes:

$$\begin{bmatrix} \rho_2 & \rho_1 & \rho_2 & \rho_3 \\ \rho_3 & \rho_2 & \rho_1 & \rho_2 \end{bmatrix} \begin{bmatrix} 1 & \rho_1 & \rho_4 & \rho_5 \\ \rho_1 & 1 & \rho_3 & \rho_4 \\ \rho_4 & \rho_3 & 1 & \rho_1 \\ \rho_5 & \rho_4 & \rho_1 & 1 \end{bmatrix}^{-1}$$

where

$$\rho_j = \mathcal{E}(x_t x_{t-j}) / \mathcal{E}(x_t^2)$$

It is easy to see that $\underset{\sim}{M}$ is of the form:

$$\begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ b_4 & b_3 & b_2 & b_1 \end{bmatrix}$$

But the correlation matrix of the vector $\underset{\sim}{u}_{2t}$ is:

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}$$

so that the discriminant functions corresponding to the two non zero canonical correlations can be immediately written down as ($x_{2t} + x_{2t+1}$) and ($x_{2t} - x_{2t+1}$) and do not need to be estimated. The two non zero canonical regressions then are the two multiple regressions:

$$(6.4.2.) \quad \begin{aligned} (x_{2t} + x_{2t+1}) &= (b_1 + b_4)(x_{2t-2} + x_{2t+3}) + (b_2 + b_3)(x_{2t-1} + x_{2t+2}) + \theta_{1t} \\ (x_{2t} - x_{2t+1}) &= (b_1 - b_4)(x_{2t-2} - x_{2t+3}) + (b_2 - b_3)(x_{2t-1} - x_{2t+2}) + \theta_{2t} \end{aligned}$$

Here θ_{1t} and θ_{2t} are independent serially and of each other and it is evident from the orthogonality relations between the variates in the two regressions that the tests based on the two correlations will be independent and can, therefore, be simply combined.

An analogous situation does not arise with processes of higher order than two for then the correlation matrix of the regressand vector depends upon more than one unknown parameter and its latent vectors cannot be prescribed in advance.

We can easily determine the efficacies of the estimates of ($b_1 + b_4$) and ($b_1 - b_4$) in (6.4.2.) (which give tests of the hypothesis that a_2 in (1.2.2.) is zero). These are:

$$\lim_{n\to\infty} \left\{ \frac{[\partial/\partial a_2 \, \mathcal{E}\,(\overline{b_1 + b_4})]^2}{V\!ar(\overline{b_1 + b_4})} \right\}_{a_2=0} = \frac{n}{4} \, \frac{1+a_1}{1+a_1^3}$$

$$\lim_{n\to\infty} \left\{ \frac{[\partial/\partial a_2 \, \mathcal{E}\,(\overline{b_1 - b_4})]^2}{V\!ar(\overline{b_1 - b_4})} \right\}_{a_2=0} = \frac{n}{4} \, \frac{1-a_1}{1-a_1^3}$$

Here $(\overline{b_1 + b_4})$, for example, is the sample regression coefficient of $(x_{2t-2} + x_{2t+3})$ in the first regression in (6.4.2.).

The efficacy of the best linear combination of $(\overline{b_1 + b_4})$ and $(\overline{b_1 - b_4})$ is easily shown to be:

$$\lim_{n\to\infty} \left\{ \frac{[\partial/\partial a_2 \, \mathcal{E}(\hat{c})]^2}{V\!ar(\hat{c})} \right\}_{a_2=0} = \frac{n}{2} \, \frac{1-a_1^4}{1-a_1^6}$$

where $\hat{c}$ is this best linear combination. We already know from (3.2.11.) that the efficacy of the estimate $\hat{b}_2$ obtained by Ogawara's original method is:

$$\frac{2n}{3} \, \frac{1}{1+a_1^2}$$

We have, therefore,

$$E\{\hat{c}, \hat{b}_2 \mid a_2 = 0\} = \frac{3}{4} \, \frac{(1+a_1^2)(1-a_1^4)}{1-a_1^6} < 1$$

(since $|a_1| < 1$ is implied by the stationarity condition).

The power of any exact test based on $(\overline{b_1 + b_4})$ and $(\overline{b_1 - b_4})$ will be even lower, of course, since the best compounding coefficients will not be known.

It would be possible to compare directly the multiple correlation procedure of # 3.2. with that resulting from the present section, but it is again clear what the results will be. The approach of the present section is no improvement on that of # 3.2. and the first order case, in relation to Ogawara's method, appears to stand in an unique position in that only there is it possible to obtain asymptotically fully efficient tests.

PART II

APPLIED RESEARCH

Chapter 7

Testing for Singularities in Rainfall Data

7.1. Meteoric dust as a cause of rainfall singularities

It would be natural to expect that the seasonal variation in daily mean rainfall for a certain station would show a fairly smooth pattern. In connection with work being carried out to determine methods of artificial rain stimulation Bowen (1953) noticed what appeared to be singularities in the pattern for the Sydney station; that is certain abrupt peaks in the mean seasonal variation. A comparison of the Sydney figures with figures from a number of other stations in the southern hemisphere* appeared to show that similar peaks were also present in the seasonal patterns of these other stations, at least for the month of January. Again a graph of the number of days upon which "heavy" falls of rain had occurred in the British Isles in January and early February, over a period from 1919 to 1949, showed peaks near to or at the dates of the Sydney peaks.

---

* These other stations were Durban, Perth, Alice Springs, Brisbane, Auckland and Christchurch.

The widespread nature of the phenomenon led Bowen to search for an extra-terrestrial cause and this he located in the passage of showers of meteors through the atmosphere. He conjectured that it would take approximately 30 days for the smaller meteoric dust (which would not be burnt up on entering the atmosphere) to fall, from the height at which it is initially concentrated by the abrupt fall in velocity as it enters the atmosphere, to a height where it would meet the larger, potential rain making, cloud structures. Bowen was then able to name certain principal meteor streams whose dates of maximum intensity preceded the dates of the Sydney rainfall singularities (for November, December, January and February) by about 30 days. He, therefore, suggested that these rainfall singularities were fundamentally due to the provision, by the meteoric dust, of the suitable rain forming nuclei without which certain types of cloud formations may not turn to rain.

The theory was criticised by Martyn (1954) who pointed out the danger of basing conclusions on rainfall data without satisfactory statistical tests, due to the nature of the distribution of daily rainfall. (This will be discussed in the next section). Martyn also showed that one of the main peaks in the Sydney rainfall

(January 13th for the period 1902 to 1949) could be entirely accounted for by one exceptional rainfall of 708 points on January 13th, 1911, a rainfall which from meteorological records could be directly attributed to a travelling cyclone of standard pattern. The theory was also criticised by Swinbank (1954) on physical grounds. Swinbank suggested that at the rain forming level there would not be a sufficient concentration of particles to upset the stability of rain clouds and that, in fact, the time of fall from the initial entering of the atmosphere by the particle to its descent to the upper troposphere would rather be from 50 days to one or two years than 30 days. He also pointed out that the rainfall peaks were mainly due to one or two days of heavy rain and that the years in which these occurred varied from station to station. (For example, the Sydney peak at January 13th for the period 1900 to 1949 was due almost entirely to a very heavy fall in 1911, whereas the Durban peak on this date over the same period was mainly due to heavy falls in 1931 and 1933. For Auckland and Christchurch the main cause was a heavy fall in 1940). Swinbank also inferred that the stations chosen were not representative.

The theory was tested in a statistical manner by Neumann (1954) and Oliver and Oliver (1955). Neumann considered a number of stations in the Palestine area and noted that the peaks for Tel Aviv for January and early February agreed fairly well with those for Sydney and thus could be claimed by Bowen to support his theory. Neumann then "shifted" each year's data (for Tel Aviv) in a random manner (for example starting the 1924 year at December, 26th, instead of January, 1.) and recomputed the mean rainfalls. Peaks of the same magnitude were observed in the rearranged data which suggested that the effect was a chance one and not due to any systematic extra terrestrial effect. Oliver and Oliver considered data for Kiambu in Kenya, which again showed peaks on dates near to the dates of Sydney peaks. They took as a measure of association between the peak dates and the dates of meteor showers the sum of the number of days between each rainfall peak and the nearest meteor shower date (date of maximum intensity). This measure was computed for every possible lag of one series relative to the other. A low value of the measure clearly indicates association. The lowest value was obtained for a lag of about 60 days, but in general the graph of the

365 values of the measure had no sufficiently marked trough to indicate, subjectively, a strong association.

Both of these statistical tests were essentially subjective. Because of the criticisms of the theory it became necessary to carry out a proper statistical test and this was done by the writer at Bowen's request.

The test could be approached in different ways. One would involve a consideration of the data for a number of stations. At the time the stations for which daily figures were available were those mentioned in the footnote on page 129 (in addition to Sydney). A completely satisfactory test using all of this data would not be easy to obtain for ~~three~~ reasons. Firstly the stations may not have been chosen in an objective manner and the importance of the element of subjectivity might be difficult to assess. Secondly some of the stations are sufficiently close together to be affected by the same meteorological phenomena, on occasions, (this would certainly be so for Brisbane and Sydney or Christchurch and Auckland) even if the phenomena were not extra terrestrial but rather fairly local. The nature of the association between the stations might be a very complicated one, which reflected, for example, only the common effects of very large meteorological disturbances. A third difficulty arose from the fact that for the stations mentioned in the footnote on page 129, statistics were immediately available in some cases for only two months of the year, while in all cases only the records of the last 50 years were immediately available. For Sydney 94 years of continuous record was immediately available.

A second method of procedure would be to correlate, in some way, the dates of peaks in one or more stations with the dates of meteor showers. Here again two difficulties occur. The knowledge available of the dates of meteor streams is not entirely satisfactory as until recently these were mainly streams observed visually and at night. There are now, of course, a number of streams which have been identified by radar methods. Moreover, the information concerning the intensities of the streams does not appear to be very good. Secondly, since one must clearly lag the series of meteor showers relative to the peak rainfall dates one is faced with the alternative of a very weak test based on a large number of trial lags or a test based on a lag of approximately 30 days. However, the choice of a lag of 30 days by Bowen may have been made after reference to the data so that the true level of significance used in the latter test would be unknown.

It was decided, therefore, to consider the Sydney series alone and to test whether there was any evidence in the Sydney data for the presence of singularities. Since the Sydney rainfall series originally suggested the theory it appears reasonable to check whether the suggestion was obtained from what might be a real effect or could have been suggested by a lack of familiarity with the extraordinarily non normal rainfall distribution.

7.2. The nature of the data and of the null hypothesis.

The distribution of daily rainfall at Sydney is J - shaped as is shown by Table 13 on page 136.

The period in late October and November, from which the distribution in Table 13 was obtained, was chosen because at this time of the year the seasonal effect is small and 22 days were used to attain 2000 observations (actually 2068). The numbers in the column headed f_o are the observed class frequencies. Those in the column headed f_e' are obtained from the distribution with frequency function:

$$(7.2.1.) \qquad \frac{(0 \cdot 013)^{0 \cdot 105}}{\Gamma(0 \cdot 105)} e^{-0 \cdot 013} x^{0 \cdot 105 - 1}$$

The χ^2 (with 15 degrees of freedom) is 7.8 so that the fit is good. The distribution (7.2.1.) was fitted by Das by maximum likelihood after truncation at 5 points and considering only the total number of days of rainfall in the region to the left of the point of truncation in the likelihood function.

Table 13

Daily Rainfall at Sydney, 17th October to 7th, November, 1859 to 1952.

Class Interval	f_o	f_e'	f_e''
0 - 5	1631	1638.5	(1614.0)
6 - 10	115	106.0	103.6
11 - 15	67	62.0	62.2
16 - 20	42	43.6	43.6
21 - 25	27	32.2	32.9
26 - 30	26	26.0	26.0
31 - 35	19	20.7	21.1
36 - 40	14	17.2	17.5
41 - 45	12	14.3	14.6
46 - 50	18	12.2	12.5
51 - 60	18	19.6	20.2
61 - 70	13	14.7	15.4
71 - 80	13	11.6	12.0
81 - 90	8	8.9	9.6
91 - 100	8	7.2	7.6
101 - 125	16	12.2	13.5
125 - 150	7	7.2	8.3
150 - 425	14	13.2	15.7

The numbers in the column headed f_e" were obtained by fitting by maximum likelihood using only the truncated distribution (neglecting the observations to the left of the point of truncation altogether). (See Das (1955)). Again the fit is good ($\chi^2 = 8.47$) and, moreover, the numbers in the class, 0 - 5 points, do not differ significantly from the observed frequency (an approximate test is given in Das' paper). This suggests that there is no singularity in the distribution at the origin, an hypothesis which might have been suggested by the very large number of zero frequencies and which Das' procedure was designed to test.

Finally from (7.2.1.) the distribution of the mean of 94 independent observations from this distribution can be obtained. (There are 94 years of rainfall data available.) This is:

$$(7.2.2.) \qquad \frac{(1{\cdot}128)^{9{\cdot}870}}{\Gamma(9{\cdot}870)}\, e^{-1{\cdot}128\bar{x}}\ \bar{x}^{(9{\cdot}870-1)}$$

Thus, except for the constant of scale, the mean daily rainfall (for October and November) over 94 years could be expected to be close to a χ^2 with 20 degrees of freedom and, therefore, by no means normal.

The variance of the distribution (7.2.1.) is

621.3 and the mean is 8.08. The largest observed rainfall was 425 points which is just over 16.5 standard deviation units from the mean. On the basis of normality such a deviation is of course incredible but, from the tables of the Incomplete Gamma function (Pearson (1922)) it appears that deviations as large as this or larger occur in the distribution (7.2.1.) with the probability 10^{-4}. In 2000 observations the event is not, therefore, exceptional. This highlights the extreme non-normality of the distribution and the danger of conclusions drawn, fundamentally from normal distribution theory, based on standard errors. It should be mentioned that there are no marked peaks in the daily mean rainfall over the period, from which the observations to which (7.2.1.) was fitted were taken. There do not appear to have been any meteor showers <u>observed</u> which, with a lag of 30 days, could have caused rainfall singularities in this period. (See Lovell and Clegg (1952) pages 78 and 99).

This non-normality is the first of a number of difficulties which affect a test of the significance of the rainfall singularities. A second is the lack of independence between rainfalls on days near each other in time. A final difficulty is the seasonal variation in the data which affects not only the daily means and variances but also the degree of dependence between days.

A study of the correlation for lags of up to 4 days suggests that:

$$\rho_{kj} \approx \rho_{1j}^{k} \qquad k = 1, 2, 3, 4.$$

Here ρ_{kj} is the correlation between rainfall on days j and j-k. For k > 4 the correlation appears to be, effectively, zero. For example, the first four lag correlations were computed from the observations over the 94 years between the dates October, 17th and November, 7th (2068 observations in all). Over this period the seasonal variation is small. The correlations were computed using the formulae:

$$r_k = \frac{22}{22-k} \; \frac{\sum_{i=1}^{94} \sum_{j=k+1}^{22} (x_{ij} x_{ij-k}) - 94(22-k)\bar{x}^2}{\sum_{i=1}^{94} \sum_{j=1}^{22} x_{ij}^2 - 94(22)\bar{x}^2} \qquad k = 1, 2, 3, 4$$

Here:

$$\bar{x} = \frac{1}{2068} \sum_{i=1}^{94} \sum_{j=1}^{22} x_{ij}$$

and x_{ij} is the rainfall on the jth day of the ith year. This particular formula was used mainly for computational convenience, the $\sum_{i=1}^{94} x_{ij} x_{ij-k}$ and $\sum_{i=1}^{94} x_{ij}$ having already been computed on the C.S.&.I.R.O. electronic computer so that the seasonal variation in the means, variances and lag correlations could be estimated. The statistics will provide consistent estimators of the corresponding

A formula such as $(22-k)^{-1}\sum_{j=k}^{22} r_{kj}$, where r_{kj} is the observed correlation between days j and j-k, would not be satisfactory since the bias will remain the same as the bias of the individual r_{kj}. The statistics r_k, together with r_1^k, are given below:

$r_1 = 0.219$; $r_2 = 0.057$; $r_3 = -0.007$; $r_4 = 0.048$

$r_1^2 = 0.048$; $r_1^3 = 0.011$; $r_1^4 = 0.002$

If these were computed from normally distributed observations their standard deviations would be about .02. The variance in the correlation coefficient depends upon the moments of the fourth order (Cramer (1946) page 359) and in the present case will be under-estimated by the formula, $\frac{(1-\rho^2)^2}{n}$, which applies approximately for a normal distribution.

The high value of r_4 (compared with r_1^4) is entirely due to one cross product, that between the rainfall of 181 points on October 25th, 1882, and 423 points on October 29th, 1882, which is probably fortuitous and a result of the extraordinarily skew parent distribution.*

It is interesting to note that the serial correlations r_k in the rainfall data are here estimated by ordinary correlations (space averages), the data for the several years being treated as realisations of a stochastic process. This is a rather extraordinary situation

* The intermediate days were rain free.

for usually only one realisation is available from which the serial correlation is estimated as a phase average.

The relation $\rho_{kj} = \rho_{1j}^{k}$, which may hold reasonably well for k small, at first sight suggests that the process can be represented by a model of the form:

(7.2.3.) $(x_j - m) = \rho(x_{j-1} - m) + \varepsilon_j \; ; \; |\rho| < 1$

Here m is the mean of the process, ρ the correlation between x_j and x_{j-1} and ε_j is an independent random process with zero mean and variance $\sigma^2(1-\rho^2)$.

It is at once evident that this process cannot be used to describe rainfall data for not only is there a seasonal variation in ρ, m and σ^2 but the variate x_j is necessarily positive so that ε_j would have to have a distribution admitting only positive values. This would, however, imply a smoothness in the series of x_j which would not be supported by the data. (A high x_j could not be followed by an x_{j+1} near to zero).

A modified process which suggests itself is:

(7.2.4.) $\left(\dfrac{x_j - m_j}{\sigma_j}\right) = \rho_{1j}\left(\dfrac{x_{j-1} - m_{j-1}}{\sigma_{j-1}}\right) + \varepsilon_j$

Here m_j is the mean and σ_j^2 the variance of x_j, ρ_{1j} is the correlation between x_j and x_{j-1} and the ε_j are independent random variates with zero means and variances $(1-\rho_{1j}^2)$.

It can be seen that the correlation between x_j

and x_{j-k} will be:

$$\rho_{kj} = \rho_{1j}\,\rho_{1j-1} \cdots\cdots \rho_{1j-k+1}$$

If the lag k is not too large and the seasonal effect is small, over short periods, then:

$$\rho_{kj} \approx \rho_{1j}^{k}$$

Again this process will not be satisfactory for the daily rainfalls (since they are positive) but it may provide a good approximation to the process generating the daily mean rainfall (over 94 years) since the range of variation of the sample mean below the true mean will be much greater. Moreover if the m_j, σ_j^2 and ρ_{1j} are equated to the means, variances and first serial correlations of daily rainfall for day j, (as estimated from a graduation of the observed values) the process (7.2.4.) will have, approximately, the same means, variances and serial correlations (for all lags) as the actual series.

A better approximation to the process generating the daily means could be got by including a term $\dfrac{x_{j-2} - m_{j-2}}{\sigma_{j-2}}$. However, the nearness of the r_{2j} to r_{1j}^2 suggests that the coefficient of this term will be very small for this coefficient will be, approximately, $\dfrac{r_{2j} - r_{1j}^2}{1 - r_{1j}^2}$. The additional computations involved would not be justified by the very small effect from the inclusion of the extra term.

The model of the process generating the daily means which will be used is, therefore:

$$(7.2.5.) \quad \sqrt{94}\ \frac{\bar{x}_j - m_j}{\left\{\sigma_j^2(1-\rho_j^2)\right\}^{1/2}} = \rho_j \sqrt{94}\ \frac{\bar{x}_{j-1} - m_{j-1}}{\left\{\sigma_{j-1}^2\,(1-\rho_j^2)\right\}^{1/2}} + \xi_j$$

Here $\bar{x}_j$ is the mean rainfall observed on the jth day of the year while m_j, σ_j^2 and ρ_j are the true means, variances and first serial correlations of rainfall on that day. The ξ_j come from a process with zero mean and unit variance.

The specification of the null hypothesis will be completed if the nature of the seasonal variation in the means, variances and correlation coefficients is laid down. All that needs to be said here is that these seasonal variations should be described by reasonably smooth curves such as a low order polynomial in j or a trigonometric polynomial formed from the first few harmonics. The choice among these alternatives depends on the goodness of their fit to the data, with the proviso that the fit should not be carried anywhere near to the point where individual singularities, confined to a period of a relatively few days, would be eliminated.

This restriction on the order of the polynomials means that a very large number of degrees of freedom are available for the estimation of the relatively few constants involved so that the $\hat{m}_j$, $\hat{\sigma}_j^2$ and $\hat{\rho}_{1j}$ given by the estimated curves can be treated as the exact values to a satisfactory degree of approximation and when used to transform the $\bar{x}_j$ into the corresponding $\hat{\xi}_j$ these last can be regarded as uncorrelated random variates with zero mean and unit variance. Since the means, $\bar{x}_j$, should have a distribution <u>near</u> to normality the sum $\chi^2 = \Sigma\hat{\xi}_j^2$ should be approximately distributed as χ^2 with (365 - p) degrees of freedom, where p is the number of constants fitted in the process of estimating the seasonal variation of the m_j, σ_j^2 and ρ_{1j}.

This test is the generalisation of the classical test of the homogeneity of a set of means (from observations subject to different treatments) which is obtained by comparing the variance within treatments with that between the means themselves. In the present case the comparisoh is being made between the within days variance of rainfall and the variance as estimated from the daily means. In addition, however, the data have had to be transformed to independence by the use of the ρ_{1j} and the within days variances made homogenous while the effect of the seasonal variation in the means has been removed so that it will not void the test against the effect of the meteor showers.

This test will be subject to certain qualifications, however, which follows from the non-normality of the daily means, for (7.2.2.) is still by no means normal and, in particular, there is positive kurtosis. This suggests that a transformation should have been made to reduce the distribution to something nearer normality. A log transformation could not be used because of the zero frequencies but the taking of cube or fourth roots would probably be fairly effective. In fact this was not done mainly because of the computations involved. There are nearly 35000 observations and even though over half are zeros and two thirds are less than 5 a large amount of work would still have been needed.

As a result of the non-normality the test will, however, be biased.

(1) The high fourth moment will increase the variance of the χ^2.

(2) The true nature of the underlying process generating the $\bar{x}_j$ will be more complicated than is indicated by (7.2.5.) so that the residuals $\hat{\xi}_j$ will not, in fact, be independent and terms such as:

$$\mathcal{E}(\hat{\xi}_i^2 \hat{\xi}_j^2) - \mathcal{E}(\hat{\xi}_i^2)\mathcal{E}(\hat{\xi}_j^2)$$

may be greater than zero (since the fourth moment of the $\hat{\xi}_j$ will be high). This will also increase the variance of the χ^2 test statistic.

If, therefore, the observed χ^2 is just beyond the critical point it will not follow that the result is significant at the chosen level since the true critical point may be further to the right by a sufficient amount to render this conclusion invalid. On the other hand if the observed χ^2 is to the left of the critical point it can be said with a fair degree of certainty that the result would still be not significant if the true significance point, appropriate to the level of significance used, were known.

7.3. Removal of seasonal variation.

In order to get a closer agreement between calendar and siderial years the leap year was reinserted in 1900 so that March 1st, 1900, become February 29th, 1900, and so on.

Since a strictly periodic seasonal effect was being removed it seemed clear that a trigonometric regression should be used to graduate the series. At the same time no attempt was made to fit a curve including terms with too short a period as this would tend to remove the effect being sought.

For both means and variances the year was divided into 24 periods of equal length and the first four harmonics were fitted to the resulting 24 figures. The intensities I_p of these four harmonics, expressed as multiples of $\frac{4s^2}{24}$ are shown in Table 14. Here s^2 is the estimate of the variance of the 24 figures.

The intensity corresponding to the period $^{24}/_p$ is derived from the relations (Kendall (1946)):

$$I_p = G_p \bar{G}_p$$
$$G_p = \sqrt{2/24} \sum_{j=1}^{24} u_j e^{\frac{2\pi i p j}{24}}$$

where $\bar{G}_p$ is the complex conjugate of G_p and u_j is the mean (or variance) on the jth day.

Table 14

$$K_j = I_p \left\{ \frac{4s^2}{24} \right\}^{-1}$$

Period	Means	Variances
24	8.53	9.57
12	.09	.06
8	.17	.10
6	.13	.24

The periodogram was computed no further since it would not be desirable to remove effects with too short a period in any case. In order to carry out an (approximate) test of significance of the intensities in Table 14 it would be necessary to make some estimate of the non singular component of the spectrum of the process (see for example Bartlett (1954)). However, in the present case any reasonable test will clearly declare the first intensity significant and the remainder as not significant.

The first harmonic was, therefore, fitted to the means and variances using the individual daily averages. The resulting curves were:

(7.3.1.) means: $$\hat{m}_j = 12.87 + 4.377 \sin\left(\frac{2\pi j}{365.25} + 0.527\right)$$

(7.3.2.) variances: $$\sigma_j^2 = 1755 + 1016 \sin\left(\frac{2\pi j}{365.25} + 0.511\right)$$

Here j = 1 on March 1st.

A simple harmonic was also fitted to the between days correlations the resulting curve being:

(7.3.3.) correlations: $$\hat{\rho}_j = 0.338 + 0.103 \sin\left(\frac{2\pi j}{365.25} - 0.285\right)$$

Insofar as the seasonal trends, (7.3.1,2,3.), are not fully adequate for the true seasonals the effect mentioned at the end of # 7.2. will be strengthened, of course.

Graphs of the means, variances and correlations together with the curves (7.3.1,2,3.) are given in appendices A, B and C.

7.4. The test of homogeneity of means.

The residuals:

(7.4.1.) $$\hat{\varepsilon}_j = \left(\frac{94}{1-\hat{\rho}_j^2}\right)^{1/2} \left\{ \frac{\bar{x}_j - \hat{m}_j}{\hat{\sigma}_j} - \hat{\rho}_j \, \frac{\bar{x}_{j-1} - \hat{m}_{j-1}}{\hat{\sigma}_{j-1}} \right\}$$

were formed. The $\hat{\epsilon}_j$ had a small positive mean (0.018) and a small negative serial correlation (- 0.05) both of which were far from significant.

The distribution of the $\hat{\epsilon}_j$ approached normality as could be expected but there was some positive skewness. The distribution is also shown, in appendix D.

There are 366 $\hat{\epsilon}_j$, one resulting from the 24 rainfalls on February, 29th. (For this $\hat{\epsilon}_j$ the factor$(24)^{\frac{1}{2}}$ will replace $(94)^{\frac{1}{2}}$ in (7.4.1.)). Three constants have been fitted to each of the means, variances and serial correlations so that:

$$\chi^2_{357} = \sum_1^{366} \hat{\epsilon}_j^2$$

may be treated as distributed as χ^2 with 357 degrees of freedom. The value of χ^2_{357} is 394.1. Using Fisher's approximation to the distribution of χ^2_p, for large p, the quantity:

$$\{2(394.1)\}^{1/2} - \{2(357)-1\}^{1/2} = 1.37$$

was computed. Treated as a standard normal variate and using one tail of the normal distribution this is well inside the 5% point (1.65).

The null hypothesis, therefore, cannot be rejected at this level of significance and the result of the test, while not disproving Bowne's hypothesis, cannot be said to justify it.

7.5. Alternative tests.

The test which has just been given might not be very powerful against the alternative of a few very widely scattered singularities since the effect of these in increasing χ^2_{357} may be lost among the accompanying "noise".

At first sight an alternative would be to pick out the high peaks in the residuals and test them as the largest among 366. For example the highest peak is on July, 23rd (24th, 1901 to 1952). This residual is 4.037 times its (trend) standard deviation. This lies almost exactly upon the 1% point for the largest out of 366 normally and independently distributed observations. In computing the probability of obtaining a largest observation as extreme as or more extreme than the largest observed, only the extreme tail area of the distribution of the individual observations is used. Moreover any error here will be multiplied in the process of computing the first probability by the number of observations (approximately). It is certain, from the shape of the parent distribution that the tail area used (on the basis of normality) will be too small but it is difficult to say by how much. How badly misleading the assumption of normality can be seen by considering the distribution of χ^2 for 102 degrees of freedom (the largest number of degrees of freedom available from Pearson (1922)). This distribution, to the eye, would be indistinguishable from the normal distribution yet the 5% point for the greatest out of 366 observations from χ^2_{102} is 4.2 standard deviation units from the mean so that on this basis the July, 23rd deviation is not significant at the 5% point rather than significant at the 1% point. Since the distribution of the daily means is near to χ^2_{20} (see # 7.2.) it is clear that the distribution of the $\hat{\xi}$ is much further from normal than χ^2_{102}. The large effect on the computed probabilities of small variations in the nature of the parent distribution appears to make the testing of the significance of the large observations an impossible task.

It is of some interest to examine the days for which the residuals are high. The 1% point was used as a criterion and all of the high residuals individually "significant", as normal deviates, at that point are shown in Table 15 together with the total of the four highest rainfalls on that day over the 94 years.

The highest rainfall for the 94 years for these 16 days is, on average, 23% of the total rainfall. Most or all of the deviation from the trend seems to be due to a relatively very few days of exceptionally heavy rain and this phenomenon is to some extent at least due to the very skew distribution of daily rainfall.

Table 15

Date (prior to 1900)		Total rainfall on this date over 94 years	Total of 4 highest rainfalls over 94 years	$\frac{\text{Col 2}}{\text{Col 1}}$ (100)
		Pts.	Pts.	%
Jan	11	1792	1170	65
	12	2293	1296	57
Feb	29	816 (24 days)	766	94
Mar	20	2603	1307	50
Apl	6	3293	1943	59
May	6	2382	1159	49
	25	2758	1412	51
July	23	2826	1043	37
Aug	2	1866	1148	62
	23	1635	613	37
Sept	24	1512	836	55
	28	1754	1067	61
Oct	12	1452	1066	73
	29	1186	743	63
Nov	17	1315	540	41
	18	1897	871	46
Dec	1	1505	905	60

REFERENCES

Aitken, A.C.	(1934-35)	On least squares and linear combination of observations. Proceedings Royal Society of Edinburgh 55, 42.
Anderson, R.L.	(1942)	Distribution of the serial correlation coefficient. Ann. Math. Statist. 13, 1.
Anderson, R.L. & T.W.	(1950)	Distribution of the circular serial correlation coefficient for residuals from a Fourier series. Ann. Math. Statist. 21, 59.
Anderson, T.W.	(1948)	On the theory of testing for serial correlation. Skandanavisk Aktuarietidskrift 31, 88.
Bartlett, M.S.	(1935)	Some aspects of the time correlation problem in regard to tests of significance. J.R. Statist. Soc. 98, 536.
" "	(1938)	Further aspects of the theory of multiple regression. Proc. Camb. Phil. Soc. 34, 33.
" "	(1946)	On the theoretical specification and sampling properties of autocorrelated time series. J.R. Statist. Soc. Supp. 8, 27.
" "	(1947)	Multivariate analysis. J.R. Statist. Soc. Suppl. 9, 176.
" "	(1954)	Problemes de l'analyse spectrale des series temporelles stationaires. Pubns. de l'institut de statistique de l'université de Paris, 3, 119.

Bartlett, M.S. & Diananda, P.H.	(1950)	Extension of Quenouille's test for autoregressive schemes. J.R. Statist. Soc. B. 12, 108.
Bowen, E.G.	(1953)	The influence of meteoritic dust on rainfall. Australian Journal of Physics. 6, 490.
Cochrane, D. & Orcutt, G.H.	(1949)	Application of least squares regression to relationships containing autocorrelated error terms. J. Amer. Statist. Assn. 44, 32.
Cramer, H.	(1946)	Mathematical Methods of Statistics, Princeton University Press.
Das, S.C.	(1955)	Fitting of truncated type III curves to daily rainfall data. Australian Journal of Physics. 8, 298.
Dixon, W.J.	(1944)	Further contributions to the problem of serial correlation. Ann. Math. Statist. 15, 119.
Doob, J.L.	(1953)	Stochastic Processes. John Wiley.
Durbin, J. & Watson, G.S.	(1950)	Testing for serial correlation in least squares regression I Biometrika 37, 409.
" "	(1951)	Testing for serial correlation in least squares regression II Biometrika 38, 159.
Fisher, R.A.	(1924)	The conditions under which χ^2 measures the discrepancy between observation and hypothesis. J.R. Statist. Soc. 87, 442.

Grenander, U. (1954) On the estimation of regre ion coefficients in the ca of an autocorrelated disturbance. Ann. Math. Statist. 25, 252.

Hart, B.I. (1942) Significance levels for th ratio of the mean square successive difference to t variance. Ann. Math. Statist. 13, 445.

Jenkins, G.M. (1954a) An angular transformation the serial correlation coefficient. Biometrika 41, 261.

" " (1954b) Tests of hypothesis in the linear autoregressive mode Biometrika 41, 405.

Kendall, M.G. (1946a) Contributions to the Study of Oscillatory Time Series National Institute of Economic and Social Resear Occasional Paper No. 1X.

" " (1946b) The Advanced Theory of Statistics, Vol. II. C. Griffin.

" " (1949) Tables of autoregressive series. Biometrika 36, 2

Koopmans, T.C. (1942) Serial correlation and qua ratic forms in normal vari ables. Ann. Math. Statist 13, 14.

" " (1950) Statistical Inference in Dynamic Economic Models. Edited by T.C. Koopmans fo the Cowles Commission. John Wiley.

Leipnik, R.B. (1947) Distribution of the serial correlation coefficient in a circularly correlated universe. Ann. Math. Statist. 18, 80.

Lovell, B. & Clegg, J.A. (1952) Radio Astronomy. Chapman & Hall.

Madow, W.G. (1945) Note on the distribution of the serial correlation coefficient. Ann. Math. Statist. 16, 308.

Mann, H.B. & Wald, S. (1943) On the statistical treatment of linear stochastic difference equations. Econometrica, 12, 143.

Marriott, F.H.C. (1952) Tests of significance in canonical analysis. Biometrika 39, 58.

Marriott, F.H.C. & Pope. (1954) Bias in the estimation of autocorrelations. Biometrika 41, 390.

Martyn, D.F. (1954) Comments on a paper by E.G. Bowen. Australian Journal of Physics. 7, 358.

Moran, P.A.P. (1950) A test for serial independence of residuals. Biometrika 37, 178.

" " (1952) A statistical analysis of game bird records. Journal of Animal Ecology 21, 154.

Neumann, J. (1954) Fluctuations of long period accumulations of daily rainfall amounts. Australian Journal of Physics 7, 522.

Noether, G.E. (1955) On a theorem of Pitman. Ann. Math. Statist. 26, 64.

Ogawara, M. (1951) A note on the test of serial correlation coefficient. Ann. Math. Statist. 22, 115.

Oliver, M.B. & V.J. (1955) Rainfall and stardust. Bulletin of the American Meteorological Society 36, 147.

Orcutt, G.H. (1948) A study of the autoregressive nature of the time series used in Tinbergen's model of the economic system of the United States; 1919-1932. J.R. Statist. Soc. B 10, 1.

Orcutt, G.H. & James, S.F. (1948) Testing the significance of correlation between time series. Biometrika 35, 397.

Pearson, E.S. & Wilks, S.S. (1933) Methods of statistical analysis suitable for k samples of two variables. Biometrika 25, 353.

Pearson, Karl. (1922) Tables of the Incomplete Gamma Function. Cambridge University Press.

Pitman, E.J.G. (1948) Non Parametric Inference. Notes for lectures delivered at University of North Carolina.

Prest, A.R. (1949) Some experiments in demand analysis. Review of Economic Statistics 31, 33.

Quenouille, M.H. (1947a) Notes on the calculation of the autocorrelations of autoregressive schemes. Biometrika 34, 365.

Quenouille, M.H. (1947b) A large sample test for the goodness of fit of autoregressive schemes. J.R. Statist. Soc. 110, 123.

" " (1948) Some results in the testing of serial correlation coefficients. Biometrika 35, 261.

" " (1949a) The joint distribution of serial correlation coefficients. Ann. Math. Statist. 20, 561.

" " (1949b) Approximate tests of correlation in time series. J.R. Statist. Soc. B 11, 68.

Rao, C.R. (1948) Tests of significance in multivariate analysis. Biometrika 35, 58.

" " (1952) Advanced Statistical Methods in Biometric Research. John Wiley.

Rubin, H. (1945) On the distribution of the serial correlation coefficient. Ann. Math. Statist. 16, 211.

Swinbank, W.C. (1954) Comments on a paper by E.G. Bowen. Australian Journal of Physics 7, 354.

Tang, P.C. (1938) Tables to facilitate the calculation of the second kind of error (P_{11}) in testing the general linear hypothesis. (Analysis of variance test.) Statistical Research Memoirs II, 126.

von Neumann, J.	(1941)	Distribution of the ratio of the mean square successive difference to the variance. Ann. Math. Statist. 12, 367.
Wald, A.	(1948)	Asymptotic properties of the maximum likelihood estimate of the unknown parameter of a discrete stochastic process. Ann. Math. Statist. 19, 40.
Wald, A. & Brookner, R.J.	(1941)	On the distribution of Wilks' statistic for testing independence of several groups of variates. Ann. Math. Statist. 12, 137.
Walker, A.M.	(1950)	Note on the generalisation of the large sample goodness of fit test for auto-regressive schemes. J.R. Statist. Soc. B 12, 102.
Watson, G.S.	(1951)	Serial Correlation in Regression Analysis. Unpublished Thesis issued in mimeographed form by University of North Carolina.
" "	(-)	On the joint distribution of the circular serial correlation coefficients. To be published in Biometrika.
Whittle, P.	(1951)	Hypothesis Testing in Time Series Analysis. Almqvist & Wicksells, Uppsala.
" "	(1953)	Estimation and information in stationary time series. Arkiv för Matematik 2, 423.
Wilks, S.S.	(1935)	Econometrica 3, 309.

Wilks, S.S. (1938) The large sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Statist. 9, 60.

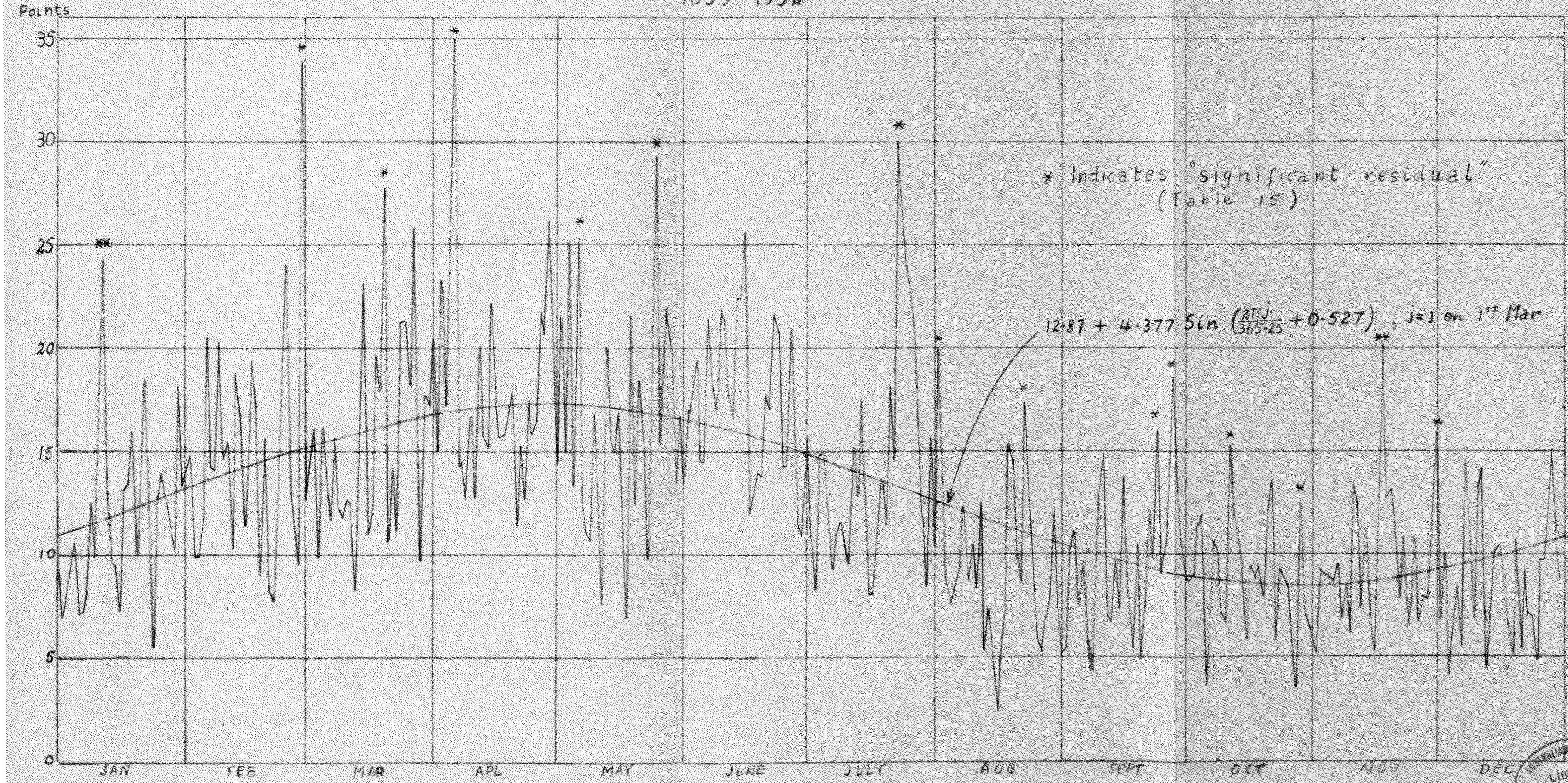
Wold, H. (1949) A large sample test for moving averages. J.R. Statist. Soc. B 11, 297.

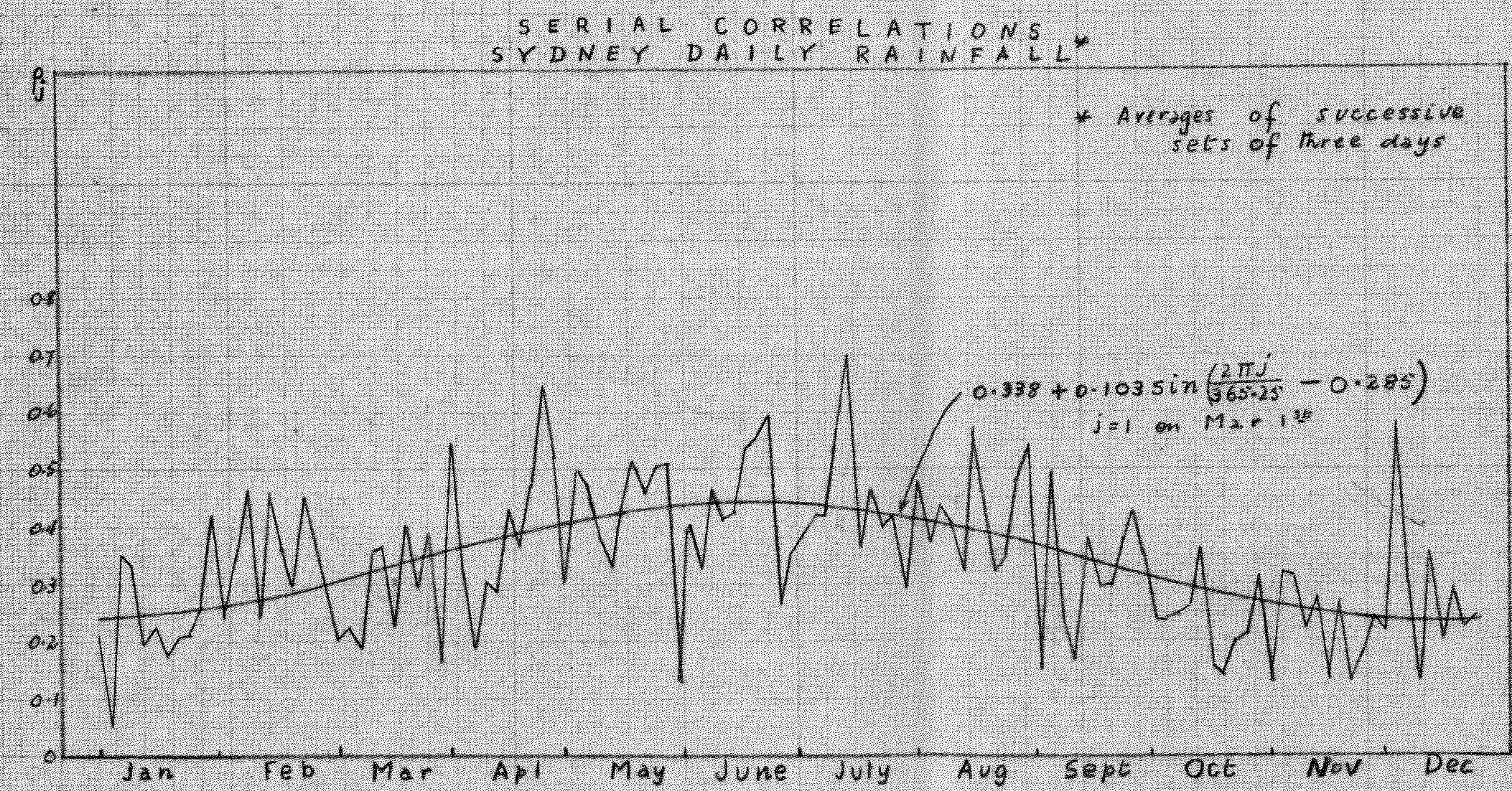
" " (1953) Demand Analysis. John Wiley.

Yule, G.U. (1927) On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. Phil. Trans. A. 226, 267.

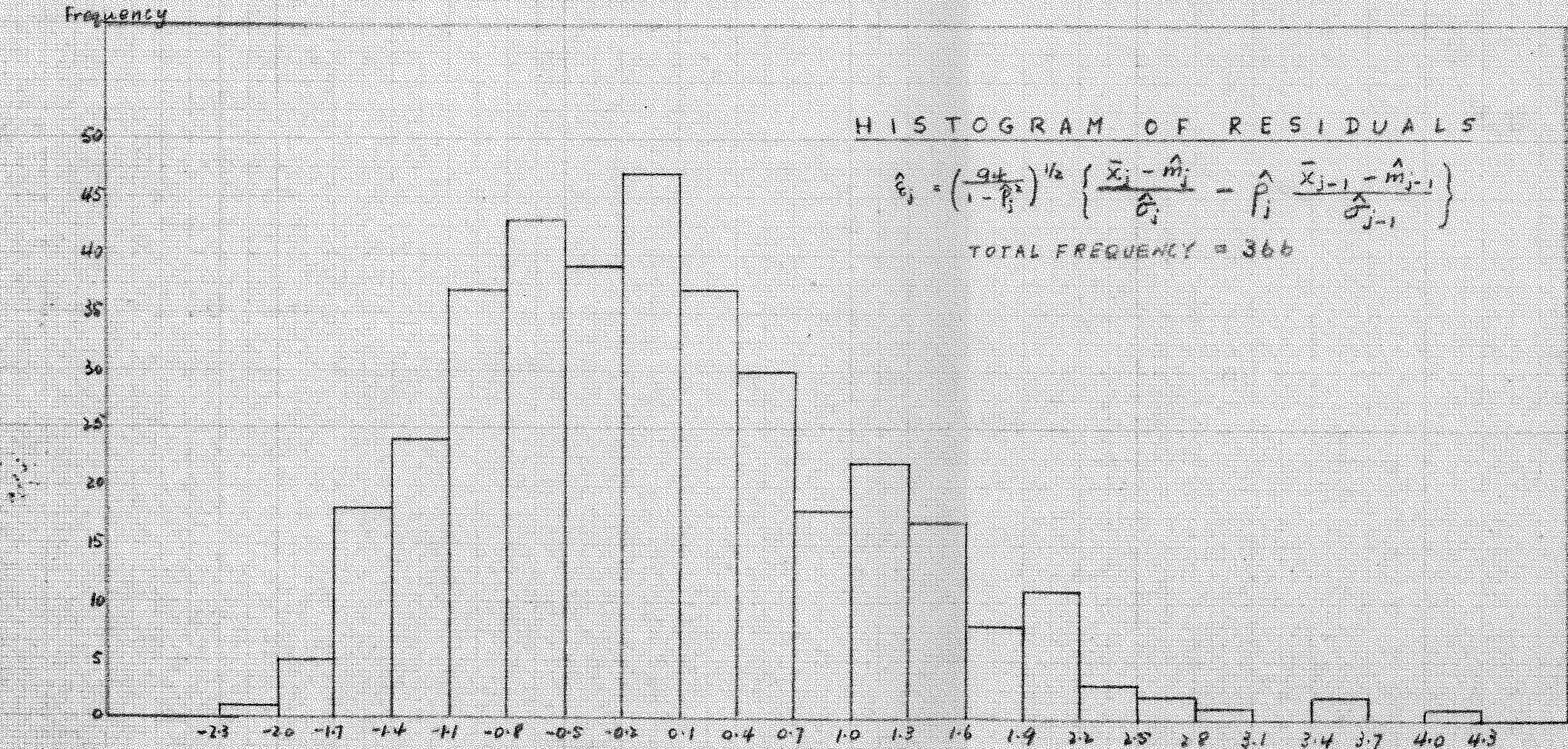
\- - - -

SYDNEY DAILY MEAN RAINFALL
1859-1952




SERIAL CORRELATIONS
SYDNEY DAILY RAINFALL*
* Averages of successive sets of three days
$0.338 + 0.103 \sin\left(\frac{2\pi j}{365.25} - 0.285\right)$
j=1 on Mar 1st
ρ_j
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Jan
Feb
Mar
Apl
May
June
July
Aug
Sept
Oct
Nov
Dec

APPENDIX D


HISTOGRAM OF RESIDUALS
$\hat{\varepsilon}_j = \left(\frac{.94}{1-\hat{\rho}_j^2}\right)^{1/2} \left\{ \frac{\bar{x}_j - \hat{m}_j}{\hat{\sigma}_j} - \hat{\rho}_j \frac{\bar{x}_{j-1} - \hat{m}_{j-1}}{\hat{\sigma}_{j-1}} \right\}$
TOTAL FREQUENCY = 366
Frequency
50
45
40
35
30
25
20
15
10
5
0
-2.3 -2.0 -1.7 -1.4 -1.1 -0.8 -0.5 -0.2 0.1 0.4 0.7 1.0 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3

VARIANCES OF SYDNEY DAILY RAINFALL*