



City Research Online

City St George's, University of London

Citation: Webb, G. I. (1973). Autocorrelations and general linear model.
(Unpublished Doctoral thesis, The City University)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/38442/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

AUTOCORRELATIONS AND THE GENERAL LINEAR MODEL

by

Gilbert Ian Webb

Thesis submitted for the degree of
Doctor of Philosophy in The City University

THE CITY UNIVERSITY LIBRARY,
81, JOHN STREET, LONDON, E.C.1.

Department of Mathematics,
The City University,
London.

March 1973

A B S T R A C T

Much previous work (e.g. Durbin and Watson 1950, 1951, 1971) concerning autocorrelated errors associated with the general linear model has been based upon the assumption that the autocorrelation structure follows the stationary first-order Markov process. Here, by consideration of experimental designs in which correlations between errors seem likely, a non-stationary process is postulated.

First, a modification of the likelihood ratio produces a statistic, designated g , for testing the hypothesis of first-order autocorrelation. This test is the most powerful invariant test against both non-stationary and stationary alternatives in the region of high autocorrelation. It is therefore more powerful than the Durbin-Watson test for these alternatives. Bounds are established for the g statistic under general conditions. In the particular case where an overall mean is included in the regression the bounds are identical to those of the Durbin-Watson statistic.

An account is given of computer simulations for the purpose of examining the properties of various estimation procedures suggested in the literature and some new ones. A method based on a simple Bayesian decision approach, involving the g statistic, is shown to be reasonably efficient and is comparatively easy to use in practice.

Qualitative and quantitative design problems are discussed. Williams' (1949) changeover designs, in which every treatment follows every other treatment, are shown to minimize the generalized variance of estimates of parameters for randomized block and latin square designs, and for positive and negative first-order alternatives. This result is as yet without analytic proof. The theory of a design procedure is given for the general quantitative factor experiment and is critically compared with the method proposed by Box and Guttman (1966).

Finally an alternative to the von Neumann (1941) ratio, based on the g statistic, is developed for testing for the presence of autocorrelation in an observed sequence with zero mean.

A C K N O W L E D G E M E N T S

My sincere thanks to Dr. I.I. Berenblut, who supervised this work, for his endless patience and wise direction. I would also like to thank my father for early encouragement to pursue my education, my wife, [REDACTED] for understanding during the course of the work which helped its progress, and [REDACTED] [REDACTED] for the high standard of typing.

I am grateful to the authorities of The City University and the Science Research Council for the facilities and finance respectively.

TABLE OF CONTENTS

Part	Section	Title	Page
1		Introduction	1
	1.1	The problem of autocorrelated errors	2
	1.2	Three case studies	5
	1.3	The mathematical model	10
2		Tests for autocorrelated errors	17
	2.1	Review of existing methods	18
	2.2	UMP tests based on least squares residuals	20
	2.3	Modification of the likelihood ratio to provide a new test	23
	2.4	Conditions under which the g test is most powerful	29
	2.5	Results from Invariance Theory	32
	2.6	Numerical examination of the power functions	37
	2.7	The numerical procedures	41
	2.8	Testing for negative autocorrelation	46
	2.9	Bounds tests using g	50
	2.10	Testing for second and higher order autocorrelations	61
3		Estimation procedures	66
	3.1	The inefficiency of least squares estimation	67
	3.2	Maximum likelihood estimation	77
	3.3	An estimator based on the g statistic	83

3.4	A Bayesian approach to the estimation of the parameters	86
3.5	A double dichotomy	90
3.6	Results of the computer simulations	99
4	Some design situations	103
4.1	A design criterion	104
4.2	Qualitative designs	106
4.3	The Box-Hunter design	111
4.4	Quantitative designs	117
4.5	Quantitative designs - two examples	122
5	Autocorrelated series	132
5.1	Two simple autocorrelation models	133
Appendix A		139
Appendix B		141
References		142

PART I

INTRODUCTION

1.1 The problem of autocorrelated errors

The results of many scientific experiments may be represented mathematically by an observed response being written in terms of an additive linear combination of factors. A model of this kind, known as the general linear model, is expressed in matrix notation as

$$\underset{\sim}{y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{u} , \quad (1)$$

where $\underset{\sim}{y}$ is an $n \times 1$ vector denoting the experimental observations or responses,

$\underset{\sim}{\beta}$ is the $k \times 1$ vector of unknown factor effects,

$\underset{\sim}{X}$ is the $n \times k$ design matrix, usually in this context a matrix of 1's and 0's indicating in any particular response the presence or absence of the contributing factors,

and $\underset{\sim}{u}$ is the $n \times 1$ vector of errors, caused by experimental inaccuracy and by any mis-specifications of the model.

We shall assume, without loss of generality, that $\underset{\sim}{X}$ is of full rank k . This model also has much application in economic time series and other regression situations where, of course, the elements of $\underset{\sim}{X}$ are not typically 1's and 0's.

In general the factors, or parameters, will include a subset of parameters representing treatment effects; the object of the experiment and the subsequent analysis is to make inferences about contributions of the various treatments to the observed responses.

When the assumption is made that the errors are independently distributed with zero mean and constant variance least squares estimators of the parameters, namely $(\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y}$, are available which are unbiased and have minimum variance. If the further assumption is made that the errors are independently normally distributed it is possible to make tests for the significance of the various subsets of parameters, and for components within the subsets, based on the F-distribution.

If the observations are being taken sequentially in time and from the same experimental unit (e.g. the same subject, field, body, apparatus or machine) the response at any particular time may be affected not only by the treatment applied at that time but by previous treatments. In this case it is expedient to incorporate residual effects of treatments into the model, retain the above assumptions about the errors and then make inferences in the usual way. However the assumption of independent errors in experiments and models of this type has been questioned by several writers. Finney (1956) suggested that an appropriate model in some bioassay experiments, where typically each subject is tested with more than one dose of the material under assay, might include residual treatment effects and autocorrelated errors. Box (1960) considers the problem of correlated errors in chemical processes where, 'due to material being held over in pipes, pumps and so forth, variation is introduced of the type where a low value at one time is followed by a high value at the next'. In another class of experiments, for example that considered by Box and Hunter (1958), page 156, treatments are applied at equally spaced intervals of time in the presence of an overall trend in the

observations. Any misrepresentations of this trend (e.g. by a wrong order polynomial in time) in the model will introduce the effect of autocorrelated errors. In fact Box (1960) distinguishes between situations where the errors are negatively autocorrelated due to carry-over and others where there is positive autocorrelation due to trend effects.

Cochrane and Orcutt (1949) considered the problem of autocorrelated errors in economic time series models and stimulated much subsequent work in this field. They demonstrated that in economic applications positive autocorrelation is not uncommon, and that least squares estimates of the parameters β may be very inefficient in the presence of high values of autocorrelation. In fact, in the discussion to a paper by Champernowne (1948), G.H. Orcutt on page 231 gives detailed reasons for suspecting high positive autocorrelation in the error terms of many economic models. The usual error structure adopted by authors dealing with these time series is the stationary first-order autocorrelation scheme.

The problem of the analysis of experiments represented by the linear model is examined when the errors are assumed to come from a first-order autocorrelated process. Because, typically, our applications consist of only short sequences of observations we do not impose the restriction of stationarity. In turn, with the non-stationary model primarily in mind, we consider testing for autocorrelated errors, estimation of the autocorrelation parameter and the other parameters, β , some particular design situations, and finally the problem of testing for autoregression in an observed sequence.

1.2 Three case studies

In this section we discuss three practical examples. These are drawn from the published literature and illustrate the type of situation where one might reasonably suspect the occurrence of autocorrelated errors. The autocorrelation may be due to either the physical nature of the experiment (Williams' data), or inadequacy of the model in representing the response (Box and Hunter's data), or perhaps to both (the Pear's data).

a. Williams' data is considered for several reasons, one of which is that it played an important role in motivating this research. It is an example of the carry-over situation - in fact Williams (1949) included residual effects of treatments as parameters in the model in an attempt to explain the carry-over effect on the response. However, this postulated model is unsatisfactory as we shall explain. The data is presented in Table 1. The usual notation for carry-over designs, e.g. Cochran and Cox (1957), Berenblut (1968), is adopted where the parameters are termed subject effects, period effects, treatment effects and residual effects of treatments.

The design matrix $\tilde{X}$ is given in fig. 1, and has been reduced to full rank using a method described in Appendix A. The six subjects are laboratory Lampén mills used for the evaluation of the effects of six different concentrations of the pulp additive (treatments) on the bursting strengths of the resulting sheets. The treatments were applied according to the Latin square in Table 2.

	Subject					
	1	2	3	4	5	6
1	56.7	58.5	55.7	57.3	53.7	58.1
2	53.8	60.2	60.7	57.7	57.1	55.7
3	54.4	61.3	56.7	55.2	59.2	58.9
4	54.4	54.4	59.9	58.1	58.9	56.6
5	58.9	59.1	56.6	60.2	58.9	59.6
6	54.5	59.8	59.6	60.2	59.6	57.5

The results in Williams' experiment

TABLE I

	Subject					
	1	2	3	4	5	6
1	C	E	A	B	F	D
2	F	C	D	A	E	B
3	B	D	E	F	A	C
4	E	F	B	D	C	A
5	D	A	F	C	B	E
6	A	B	C	E	D	F

The latin square used by Williams

TABLE 2

1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0
1	1	0	0	0	0	0	1	0	0	0	-1	-1	-1	-1	-1
1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0
1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1
1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0
1	1	0	0	0	0	-1	-1	-1	-1	-1	1	0	0	0	0
1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1
1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0
1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0
1	0	1	0	0	0	0	0	0	1	0	-1	-1	-1	-1	-1
1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0
1	0	1	0	0	0	-1	-1	-1	-1	-1	0	1	0	0	0
1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1
1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0
1	0	0	1	0	0	0	0	0	0	1	-1	-1	-1	-1	-1
1	0	0	1	0	0	-1	-1	-1	-1	-1	0	0	1	0	0
1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0
1	0	0	0	1	0	0	0	1	0	0	-1	-1	-1	-1	-1
1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
1	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0
1	0	0	0	1	0	-1	-1	-1	-1	-1	0	0	0	0	1
1	0	0	0	0	1	1	0	0	0	0	-1	-1	-1	-1	-1
1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1
1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0
1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0
1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0
1	0	0	0	0	1	-1	-1	-1	-1	-1	0	0	0	1	0
1	-1	-1	-1	-1	-1	1	0	0	0	0	0	0	0	1	0
1	-1	-1	-1	-1	-1	0	1	0	0	0	0	1	0	0	0
1	-1	-1	-1	-1	-1	0	0	1	0	0	0	0	1	0	0
1	-1	-1	-1	-1	-1	0	0	0	1	0	1	0	0	0	0
1	-1	-1	-1	-1	-1	0	0	0	0	1	0	0	0	0	1
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

The design matrix for the Williams' experiment

FIG. I

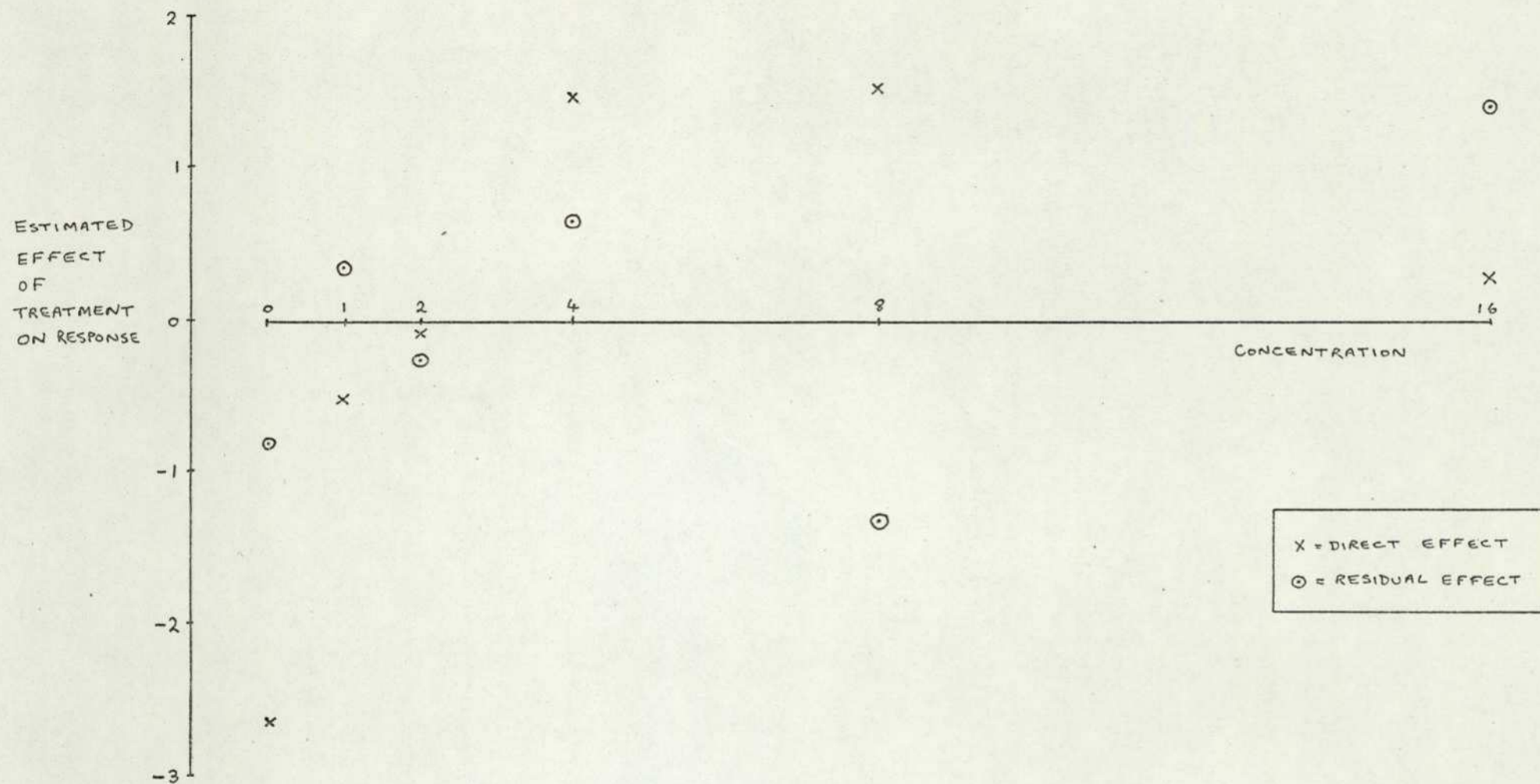


FIG. 2

PLOT OF TREATMENT ESTIMATES AGAINST CONCENTRATION
IN WILLIAMS' DATA

Using the F-tests both direct and residual effects are significant at the 1% level. When plotted against concentration of additive, fig. 2, the estimated direct effects lie approximately on a smooth curve, whereas the estimated residual effects do not. The erratic behaviour of the plotted residual effects, when they would be expected to vary continuously with concentration, casts doubts as to the physical validity of explaining the carry-over by means of residual effects of treatments. The data was considered by Berenblut (1968), pages 40-44, (in an analysis for non-additivity which produced a negative result) where details of the treatment concentrations and estimates may be found.

Following the remarks made by Box (1960) we conjecture that a model containing negatively autocorrelated errors might give a more adequate representation of the above data, possibly with or without the inclusion of residual effects of treatments.

b. Box and Hunter's data The data, given by Box and Hunter (1958) page 156, are the results of a pilot experiment conducted to ascertain the degree of a time dependent polynomial necessary to represent a trend in the response. The quantitative levels of the treatment factors in the main experiment are then chosen to be orthogonal to this polynomial. This case is presented to show the problem of designing in the presence of an overall trend in time where the effect of positive autocorrelation might occur if the actual trend is either a different degree polynomial or if it is not a polynomial and not adequately represented by a polynomial.

c. Pear's data This is an example taken from economic literature and will allow comparisons to be made between this work and that done by economic statisticians on the same subject of autocorrelated errors. The original data is presented by Hoos and Shear (1942), but has been considered subsequently by Hildreth and Lu (1960), Henshaw (1966) and Durbin and Watson (1971). The price of California Bartlett pears is regressed on an overall mean, a linear time trend, the supply of those pears, the supply of Washington and Oregon pears, and finally the non-agricultural income. All the parameters were significant. The presence of autocorrelated errors in this specified model might be suspected if for instance an important independent variable was missing - being unknown or immeasurable, or both. Alternatively the linear time trend may not adequately represent the actual time trend.

We now have some practical situations in mind, typical of various applications to which the developing theory can be critically applied.

1.3 The Mathematical Model

In the general linear model (1.1.1) we make the assumptions

1. $E(u) = 0$,
2. the elements of the design matrix X are fixed - e.g. there are no lagged dependent variables,

3. the dispersion or variance-covariance matrix of the observations is $D(\underline{y}) = E(\underline{u}\underline{u}') = \sigma^2 \underline{V}$.

Thus any correlation scheme is completely specified by the elements of the matrix $\underline{V}$. Making the further supposition that the errors are multivariate normal, the probability density function is then

$$|\underline{V}|^{-1/2} (2\pi\sigma^2)^{-n/2} \cdot \exp\left(-\frac{\underline{u}'\underline{V}^{-1}\underline{u}}{2\sigma^2}\right). \quad (1)$$

In the model errors occur at any time due to unknown and uncontrollable factors or mis-specifications of the true model. These errors are presumed to be small and related to the corresponding errors in previous experiments. The simplest possible relationship of this kind is the first-order autocorrelation scheme. At each experimental observation the total error is composed of an innovation of an independent error plus a proportion of the previous error. In typical experimental design situations where there is no history prior to the first observation then clearly in the first experimental observation the error consists entirely of the independent component. Therefore the first order autocorrelation process is defined as

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

$$\text{and} \quad u_1 = \epsilon_1,$$

where p is the number of periods, i.e. the number of observations taken on any one subject. If there are s subjects then the total number of observations $n = s \times p$. The model is defined in this way because correlations are only expected between errors on the same subject - for instance the Williams' data where $s = p = 6$ and

$n = 36$. Economic time series regressions fit into this model by taking $s = 1$. Experiments where unequal numbers of observations are taken on different subjects are not considered explicitly in this work, although by a suitable generalization of the notation the results could be extended to cover that case.

Thus the dispersion matrix is defined in the form

$$\underline{V} = \underline{V}_n(e) = \begin{pmatrix} \underline{V}_p(e) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \underline{V}_p(e) \end{pmatrix},$$

where all unspecified elements are zero, i.e. there are no correlations between errors on different subjects. In order to define the elements of $\underline{V}_p(e)$ one further assumption must be made concerning the autocorrelation scheme. The simplest alternatives to consider are either that the sequence of independent errors ϵ_i has constant variance σ^2 , or that the sequence u_i has constant variance σ^2 .

We shall adopt the former while observing that the latter assumption leads to the stationary first-order process as considered by Cochrane and Orcutt (1949), Durbin and Watson (1950, 1951, 1971), Von Neumann (1941), Henshaw (1966), R.M. Williams (1952) amongst others. Taking $\text{var}(\epsilon_i) = \sigma^2$, $i = 1, 2, \dots, p$, then

$$\underline{V}_p(e) = \begin{pmatrix} 1 & e & e^2 & \dots & e^{p-1} \\ e & 1+e^2 & e+e^3 & e^2+e^4 & \vdots \\ \vdots & e+e^3 & 1+e^2+e^4 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{p-1} & \vdots & \vdots & \vdots & 1+e^2+\dots+e^{2(p-1)} \end{pmatrix},$$

and

$$V_{\sim p}^{-1}(\epsilon) = \begin{pmatrix} 1+\epsilon^2 & -\epsilon & 0 & \dots & 0 \\ -\epsilon & 1+\epsilon^2 & -\epsilon & & \vdots \\ 0 & -\epsilon & 1+\epsilon^2 & -\epsilon & \vdots \\ \vdots & & & \ddots & \vdots \\ \vdots & & & & 1+\epsilon^2 & -\epsilon \\ 0 & \dots & \dots & \dots & -\epsilon & 1 \end{pmatrix}.$$

We shall call the autocorrelation scheme with the above assumption that $\text{var}(\epsilon_i) = \sigma^2$, all i , the non-stationary process, remarking that the variance of u_i remains finite as $i \rightarrow \infty$ provided $|\epsilon| < 1$.

It is easily verified that $V_{\sim p}^{-1}(\epsilon) = C_{\sim p}' C_{\sim p}$

$$\text{where } C_{\sim p} = C_{\sim p}(\epsilon) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -\epsilon & 1 & 0 & \vdots \\ 0 & -\epsilon & 1 & \vdots \\ \vdots & & & \ddots & \vdots \\ \vdots & & & & 1 & 0 \\ 0 & \dots & \dots & \dots & -\epsilon & 1 \end{pmatrix}. \quad (2)$$

Therefore $|V| = |V^{-1}|^{-1} = |C|^{-2} = 1$, for all p, n and ϵ .

In the stationary process the restriction that $|\epsilon| < 1$ is encountered, whereas in the non-stationary process ϵ can take any value. It is certainly possible to envisage situations where $\epsilon = 1$. For example the cumulative error structure considered by Cochrane and Orcutt (1949) and Mandel (1964), page 295, is in fact the non-stationary first-order process with $\epsilon = +1$.

In cases where ρ is known it is possible, using the matrix $C(\rho)$, to transform to the independent errors situation and then use least squares and the usual F-tests. Noting that $Cu = \tilde{\epsilon}$, where $\tilde{\epsilon} \sim N(0, \sigma^2 I_n)$, then from (1.1.1)

$$C\tilde{y} = C\tilde{X}\tilde{\beta} + \tilde{\epsilon} \quad (3)$$

Thus treating $C\tilde{y}$ as the new vector of observations and $C\tilde{X}$ as the design matrix the usual least squares situation obtains. Therefore the minimum variance estimate of $\tilde{\beta}$ is given by $\{(C\tilde{X})' (C\tilde{X})\}^{-1} (C\tilde{X})' (C\tilde{y})$, i.e.

$$\tilde{\beta} = (\tilde{X}' \tilde{V}^{-1} \tilde{X})^{-1} \tilde{X}' \tilde{V}^{-1} \tilde{y} \quad (4)$$

and the new residuals by

$$\tilde{u} = C\tilde{y} - C\tilde{X}\tilde{\beta} \quad (5)$$

When $\rho = 0$ the estimates of $\tilde{\beta}$ and the residuals become, in the usual notation, $\hat{\beta} = (X'X)^{-1} X'y$ and $\tilde{z} = y - X\hat{\beta}$.

It is of interest to note, in view of well known practice in economic time series analysis, e.g. Cochrane and Orcutt (1949), that for $\rho = 1$ premultiplying (1.1.1) by C to obtain (3) amounts to taking first differenced variables in both the observations and the regression vectors, except by this method the first rows remain unchanged.

In practice the value of ρ is not known except in special cases such as that mentioned above, or where it is assumed to be zero. When the magnitude of the autocorrelation is unknown three problems need to be considered:-

1. testing for the presence of autocorrelation,
2. estimation of the parameters in the model,
3. design procedures when the model has been extended to include autocorrelated errors.

The primary purpose of this work is to examine present methods and propose new solutions to these problems when the autocorrelation process under consideration is the non-stationary one.

The non-stationary first-order model has been used in two previous papers. Hildreth and Lu (1960) consider approximately the same likelihood, but from 'economic considerations' restrict the range of ρ to $(-1, 1)$; Zellner and Tiao (1965) work from a Bayesian viewpoint and adopt the same likelihood as Hildreth and Lu, but do not restrict the range of ρ . The model introduced here differs from that of the above authors in the treatment of the first observation. They consider the distribution of the first error to be dependent on a previous error, and then make various assumptions regarding this postulated dependence. This is reasonable in economic time series where there is usually a history of the series prior to the first observation in the model. In the experimental design applications where there is no treatment prior to the first on any particular subject it is more realistic to suppose that the error in the first response is not dependent on any previous postulated error.

Where possible in the subsequent theory, which will consider specifically the first-order process, we shall indicate extensions of the results to the second and higher order processes. This section

is concluded by generalization of the mathematical model to the r^{th} order autocorrelated error process, of which the first-order scheme defined above is a special case.

As before, assuming the variance of the independent component of errors to be constant σ^2 , the total error in the i^{th} observation is defined as

$$u_i = e_1 u_{i-1} + e_2 u_{i-2} + \dots + e_r u_{i-r} + \varepsilon_i, \quad (i = 1, 2, \dots, p),$$

regarding $u_0, u_{-1}, \dots, u_{1-r}$ as zero.

The associated dispersion matrix $\underline{V}$ and its inverse $\underline{V}^{-1}$ contain rather cumbersome elements, but the corresponding matrix $\underline{C}$ is uncomplicated, being

$$\underline{C}_p(e_1, e_2, \dots, e_r) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -e_1 & 1 & 0 & & \\ \vdots & -e_1 & 1 & & \\ -e_r & \vdots & & \ddots & \\ 0 & -e_r & & & 1 & 0 \\ \vdots & & & & & -e_1 & 1 \\ 0 & & & & & & -e_1 & 1 \end{pmatrix}.$$

The inverse of the dispersion matrix can be written

$$\underline{V}^{-1}(e_1, e_2, \dots, e_r) = \underline{C}'\underline{C}.$$

Quite generally $|\underline{V}| = |\underline{V}^{-1}| = 1$ for any values of $e_1, \dots, e_r, n, p$.

It may be easily verified that $\underline{C}\underline{u} = \underline{\varepsilon}$. Therefore, if the coefficients e_i are known the model can be transformed as in (3) to the independent error situation.

PART II

TESTS FOR AUTOCORRELATED ERRORS

2.1 Review of existing methods

The problem of testing for autocorrelation has received much attention from statisticians concerned with the analysis of economic time series. As remarked previously, in these applications the correlation has been generally assumed to take the form of a stationary first-order autoregressive process in the errors.

Von Neumann, Kent, Bellinson and Hart (1941) proposed a statistic (now generally known as the von Neumann ratio) suitable for testing for autocorrelation in an observed sequence. Durbin and Watson (1950) adapted this statistic for the purpose of testing for autocorrelated errors in the linear model by replacing the observed variables with the least squares residuals. Using results due to Anderson (1948) they were able to show that the test is uniformly most powerful (UMP) under certain conditions. Unfortunately the significance points of their test statistic depend on the design matrix. However they were able to derive upper and lower bounds to the critical values and published tables of these in (1951). They also outlined an approximate procedure to use in the event of the actual value of the statistic lying between the bounds and the test being inconclusive.

Koerts and Abrahamse (1968), using results due to Imhof (1961), gave a method for calculating the exact significance points, for any particular design matrix, of statistics of the type proposed by Anderson (1948) and adopted by Durbin and Watson (1950), i.e. ratios of quadratic forms in normal variables. More recently Durbin

and Watson (1971) indicated that a method due to Pan Je-Jian (1968) was preferable for evaluating tail end probabilities. In both procedures the amount of calculation involved is not trivial and requires the use of a computer.

Because the Durbin-Watson bounds test can give inconclusive results much work has been directed to the search for 'exact' tests - where there is no inconclusive region and the significance points can be tabulated once and for all. Theil and Nagar (1961) and Abrahamse and Koerts (1969) proposed von Neumann ratios of residuals based on best linear unbiased estimates. Recently Durbin (1970) suggested an exact test where the residuals are based on estimates of parameters obtained from a derived set of regressors, and Abrahamse and Louter (1971) proposed a statistic based upon a new class of estimators for the disturbance (error) vector. The calculation of the values of these statistics is much more complicated than that of the Durbin-Watson statistic, and requires the use of a computer.

These tests are all based on the stationary autocorrelated error process. We shall begin our treatment of the non-stationary process by applying Anderson's (1948) results.

2.2 UMP tests based on least squares residuals

T.W. Anderson (1948) showed that, for an $n \times k$ design matrix in which the columns are a linear combination of k latent vectors of the matrix $\underset{\sim}{\textcircled{M}}$, the statistic

$$r = \frac{\underset{\sim}{z}' \underset{\sim}{\textcircled{M}} \underset{\sim}{z}}{\underset{\sim}{z}' \underset{\sim}{z}}$$

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

$$K \exp \left[-\frac{1}{2\sigma^2} \left\{ (1+e^2) \underset{\sim}{u}' \underset{\sim}{u} - 2e \underset{\sim}{u}' \underset{\sim}{\textcircled{M}} \underset{\sim}{u} \right\} \right] \quad (1)$$

The density function associated with the stationary first order autocorrelation process is

$$\frac{(2\pi\sigma^2)^{-n/2}}{(1-e^2)^{(n-1)/2}} \exp \left[\frac{-1}{2\sigma^2} \cdot \left\{ (1+e^2) \sum_1^n u_i^2 - e^2(u_1^2 + u_n^2) - 2e \sum_2^n u_i u_{i-1} \right\} \right], \quad (2)$$

for which no UMP test exists, T.W. Anderson (1948). For simplicity of notation we are considering explicitly the case of one subject, but the theory holds generally for $s > 1$. By replacing (2) by a close approximation, namely

$$K_1 \exp \left[\frac{-1}{2\sigma^2} \left\{ (1+e^2) \sum_1^n u_i^2 - e(u_1^2 + u_n^2) - 2e \sum_2^n u_i u_{i-1} \right\} \right] \quad (3)$$

and applying Anderson's results, Durbin and Watson (1950) obtained their statistic (subscripted for clarity in this context)

$d_A = z'Az/z'z$, where

$$\tilde{A} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & 0 & & & \\ 0 & -1 & 2 & -1 & & & \\ \vdots & & & \ddots & \ddots & & \vdots \\ \vdots & & & & \ddots & \ddots & \vdots \\ \vdots & & & & & 2 & -1 \\ 0 & \cdots & \cdots & \cdots & \cdots & -1 & 1 \end{pmatrix}.$$

The test for zero autocorrelation provided by the d_A statistic is UMP against one sided alternatives defined by (3) if the k regression vectors (columns of $\tilde{X}$) are a linear combination of k latent vectors of $\tilde{A}$. The latent vector associated with the zero latent root of $\tilde{A}$ is the constant vector $\tilde{1}$, and this must be included in the model as a regression vector (Durbin and Watson (1971), page 10). If $s > 1$, i.e. there is more than one subject, then $\tilde{A}$ has s zero latent roots and the associated latent vectors are the subject means. In general these parameters would be included in the model from physical considerations, so the requirement that they are included is no real restriction.

It is worthy to note that (3) is a better approximation to (2) when ρ is positive - economic statisticians have been primarily concerned with testing for the one sided alternative of positive autocorrelation. In the case of testing for negative autocorrelation a closer approximation to (2) would be

$$K_2 \exp \left[\frac{-1}{2\sigma^2} \left\{ (1+\rho^2) \sum_i u_i^2 + \rho(u_1^2 + u_n^2) - 2\rho \sum_i u_i u_{i-1} \right\} \right]. \quad (4)$$

However the latent vectors associated with the zero latent roots then have the form $(1 -1 1 \dots -1 1)'$, which does not fit into the regression schemes of many physical situations.

Considering the hypothesis of non-stationary first-order autocorrelated errors, the density function becomes

$$(2\pi\sigma^2)^{-n/2} \cdot \exp \left[\frac{-1}{2\sigma^2} \left\{ (1+\rho^2) \sum_i u_i^2 - \rho u_n^2 - 2\rho \sum_i u_i u_{i-1} \right\} \right] \quad (5)$$

and again no UMP test exists. A close approximation to (5), for positive ρ , is

$$K_3 \exp \left[\frac{-1}{2\sigma^2} \left\{ (1+\rho^2) \sum_i u_i^2 - \rho u_n^2 - 2\rho \sum_i u_i u_{i-1} \right\} \right]. \quad (6)$$

Using Anderson's result, a test for the null hypothesis of $\rho = 0$ using the statistic $d_B = \underline{z}' \underline{B} \underline{z} / \underline{z}' \underline{z}$, where

$$\underline{B} = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & & & & \cdot \\ 0 & -1 & 2 & -1 & & & & \cdot \\ \cdot & & & \cdot & & & & \cdot \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & 2 & -1 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & -1 & 1 \end{pmatrix},$$

is UMP against one-sided alternatives given by (6) when the k regression vectors are a linear combination of any k latent vectors of $\underline{B}$. It is important to note that $\underline{B}$ and $\underline{V}^{-1}(1)$ are identical and that the latent roots are non-zero and positive. Bellman (1960), page 66, gave the latent roots of this matrix. In a form slightly more convenient for our use they are:

$$2 - 2 \cos \{ (2j-1)\pi / (2n+1) \}, \quad (j=1, 2, \dots, n). \quad (7)$$

For a test against negative autocorrelation a closer approximation to (5) than (6) would be

$$K_4 \exp \left[\frac{-1}{2\sigma^2} \left\{ (1+\epsilon^2) \sum_i^n u_i^2 + \epsilon u_n^2 - 2\epsilon \sum_i^n u_i u_{i-1} \right\} \right], \quad (8)$$

and the test given by the statistic $d_{-1} = \tilde{z}' \tilde{V}^{-1} (-1) \tilde{z} / \tilde{z}' \tilde{z}$ is UMP against one sided alternatives given by (8) when the k regression vectors are a linear combination of any k latent vectors of $\tilde{V}^{-1}(-1)$.

It is important to note these properties, but also to realize that in scientific experiments the latent vector conditions are unlikely to be satisfied. Furthermore, when the latent vector conditions do hold this is exactly the situation when a good test is least important, because here for any value of ϵ the least squares estimates $\hat{\tilde{\beta}}$ are best linear unbiased, being equal to $\tilde{\beta}$, e.g. see Durbin and Watson (1971), page 10.

2.3 Modification of the likelihood ratio to provide a new test

Taking the likelihood of the non-stationary process defined in (2.2.5) we have

$$L = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\tilde{y} - \tilde{X}\tilde{\beta})' \tilde{V}^{-1} (\tilde{y} - \tilde{X}\tilde{\beta}) \right\},$$

and the logarithm of the likelihood is

$$\ln L = -n/2 \ln(2\pi) - n/2 \ln \sigma^2 - (\tilde{y} - \tilde{X}\tilde{\beta})' \tilde{V}^{-1} (\tilde{y} - \tilde{X}\tilde{\beta}) / (2\sigma^2).$$

The maximum likelihood estimates of the parameters $\underline{\beta}$, $\underline{\epsilon}$ and σ^2 , denoted by $\underline{\tilde{\beta}}$, $\underline{\tilde{\epsilon}}$, $\underline{\tilde{\sigma}^2}$ respectively, are given in section 3.2. Using these it is easily verified that the likelihood ratio test for the hypothesis $\underline{\epsilon} = 0$ against alternatives of the form $\underline{\epsilon} = \underline{\epsilon}^* \neq 0$ is given by consideration of a statistic of the form

$$\frac{(\underline{y} - \underline{X}\underline{\tilde{\beta}})' \underline{V}^{-1}(\underline{\tilde{\epsilon}}) (\underline{y} - \underline{X}\underline{\tilde{\beta}})}{(\underline{y} - \underline{X}\underline{\hat{\beta}})' (\underline{y} - \underline{X}\underline{\hat{\beta}})} \quad (1)$$

Writing $\underline{V}^{-1}(\underline{\tilde{\epsilon}})$ as $\underline{V}^{-1}$, this ratio becomes

$$\frac{\underline{y}' (\underline{V}^{-1} - \underline{V}^{-1} \underline{X} (\underline{X}' \underline{V}^{-1} \underline{X})^{-1} \underline{X}' \underline{V}^{-1}) \underline{y}}{\underline{y}' \underline{M} \underline{y}}, \text{ where } \underline{M} = \underline{I} - \underline{X} (\underline{X}' \underline{X})^{-1} \underline{X}'.$$

The likelihood ratio statistic takes values between zero and one and the critical region is in the left hand tail. This is an extremely awkward expression to deal with as a test statistic, as for that purpose we should require to calculate its probability distribution. As a simplification we propose for use as a test statistic for positive autocorrelation, the following modification of (1) obtained by replacing $\underline{\tilde{\epsilon}}$ by $\underline{1}$ and hence $\underline{V}^{-1}(\underline{\tilde{\epsilon}})$ by $\underline{B}$:

$$g = \frac{(\underline{y} - \underline{X}\underline{\beta}^*)' \underline{B} (\underline{y} - \underline{X}\underline{\beta}^*)}{(\underline{y} - \underline{X}\underline{\hat{\beta}})' (\underline{y} - \underline{X}\underline{\hat{\beta}})} = \frac{\underline{y}' \underline{W} \underline{y}}{\underline{y}' \underline{M} \underline{y}}, \quad (2)$$

where $\underline{\beta}^* = (\underline{X}' \underline{B} \underline{X})^{-1} \underline{X}' \underline{B} \underline{y}$,

$$\underline{W} = \underline{B} - \underline{B} \underline{X} (\underline{X}' \underline{B} \underline{X})^{-1} \underline{X}' \underline{B}, \quad (3)$$

and, as before, $\underline{M} = \underline{I} - \underline{X} (\underline{X}' \underline{X})^{-1} \underline{X}'$. (4)

The test between the hypotheses $H_0 : \varrho = 0$ and $H_1 : \varrho = \varrho^* > 0$, with density function (2.2.5), is defined by:-

$$\omega = \{g \leq v\},$$

where ω is the critical region and v is determined so that $P(g \leq v/H_0) = \alpha$, the size of the test. It will be shown later that the range of g must lie between zero and four, a fact which is also true for the ranges of d_A and d_B .

If restrictions on the form of the design matrix are made, viz. the latent vector restraints as specified in the previous section for the tests given by d_A and d_B , the power properties of both the test statistics d_B and Durbin and Watson's d_A are inherent in the test given by the g statistic. This is proved by establishing equality of g to the statistics d_B and d_A when the design matrix takes the appropriate form:-

- (i) $g = d_B$ when the k regression vectors are a linear combination of k latent vectors of B . This follows from the well known fact (e.g. Watson (1967) page 1683) that in this case the minimum variance estimates coincide with the least squares estimates, i.e. $\tilde{\beta}^* = \hat{\tilde{\beta}}$ and

$$g = \frac{(\tilde{y} - \tilde{X}\hat{\tilde{\beta}})' \tilde{B}(\tilde{y} - \tilde{X}\hat{\tilde{\beta}})}{\tilde{y}' \tilde{M}_{\tilde{y}} \tilde{y}} = d_B.$$

- (ii) $g = d_A$ when the k regression vectors are a linear combination of k latent vectors of A , of which one must be the vector associated with the zero latent root, 1 , (i.e. a constant term must be fitted). In cases where the number of subjects is greater than one and autocorrelation is

assumed to occur only between errors on the same subject, the subject means are the latent vectors associated with the s zero latent roots. In the proof of equality of g and d_A which follows for $s = 1$, the extension of the result to the general case is straightforward.

Proof of (ii). Bearing in mind the definitions $g = \frac{y'Wy}{y'My}$ and $d_A = \frac{y'MAMy}{y'My}$, to prove equality of g and d_A it is sufficient to prove that $W = MAM$ under the given conditions.

Let the latent vectors of the symmetric matrix A be represented by the columns of the $n \times n$ matrix L . The latent vectors of a real symmetric matrix are orthogonal, e.g. Bellman (1960) page 36, so that $L'L = LL' = I_n$. Further, let the first column of L be the normalized unit vector 1 and the next $k - 1$ columns be those latent vectors which with 1 provide an orthogonal basis for X . Now

$$X = (1 \parallel X_0) = L_k G, \quad (5)$$

where G is a $k \times k$ non-singular matrix, and $L = (L_k \parallel L_{n-k})$. If Λ_{n-k} is the diagonal matrix of latent roots of A corresponding to the latent vectors L_{n-k} , then

$$L_{n-k}' A L_{n-k} = \Lambda_{n-k}, \quad (6)$$

and Λ_{n-k} is non-singular because it does not contain the zero latent root. The $n - 1$ positive latent roots of A are given by von Neumann (1941). The four following results can be easily

deduced:- remembering that $\tilde{B}^{-1} = \tilde{V}(1)$ then

$$\tilde{B}^{-1} \tilde{A} = \tilde{I} - \sqrt{n} \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}, \quad (7)$$

$$\tilde{W} = \tilde{B} - \tilde{B} \tilde{L}_k (\tilde{L}_k' \tilde{B} \tilde{L}_k)^{-1} \tilde{L}_k' \tilde{B} \quad (8)$$

$$\tilde{M} = \tilde{I} - \tilde{L}_k (\tilde{L}_k' \tilde{L}_k)^{-1} \tilde{L}_k' = \tilde{L}_{n-k} \tilde{L}_{n-k}', \quad (9)$$

because $\tilde{L} \tilde{L}' = \tilde{L}_k \tilde{L}_k' + \tilde{L}_{n-k} \tilde{L}_{n-k}' = \tilde{I},$

and $\tilde{M} \tilde{A} \tilde{M} = \tilde{L}_{n-k} \tilde{L}_{n-k}' \tilde{L}_{n-k} \tilde{L}_{n-k}' \quad (10)$

At this point we require the result of the following lemma, given in a slightly more general form than is strictly necessary here.

Lemma 1 $\tilde{V}^{-1} - \tilde{V}^{-1} \tilde{S} (\tilde{S}' \tilde{V}^{-1} \tilde{S})^{-1} \tilde{S}' \tilde{V}^{-1} = \tilde{T} (\tilde{T}' \tilde{V} \tilde{T})^{-1} \tilde{T}',$

where $\tilde{V}^{-1}$ (and hence $\tilde{V}$) is any $n \times n$ positive definite matrix and $\tilde{S}$ and $\tilde{T}$ are $n \times k$ and $n \times (n - k)$ matrices respectively such that if $\tilde{U} = \begin{pmatrix} \tilde{S} \\ \tilde{T} \end{pmatrix}$ then $\tilde{U}' \tilde{U} = \tilde{U} \tilde{U}' = \tilde{I}_n$. More explicitly, $\tilde{S}' \tilde{S} = \tilde{I}_k$, $\tilde{T}' \tilde{T} = \tilde{I}_{n-k}$ and $\tilde{S}' \tilde{T} = 0$.

Proof Because $\tilde{U}$ is an $n \times n$ orthonormal matrix any $n \times n$ matrix may be expressed as a linear combination of the columns of $\tilde{U}$. In particular,

$$\tilde{V}^{-1} - \tilde{V}^{-1} \tilde{S} (\tilde{S}' \tilde{V}^{-1} \tilde{S})^{-1} \tilde{S}' \tilde{V}^{-1} = \tilde{U} \tilde{R},$$

so that, where $\tilde{U} \tilde{R} = \begin{pmatrix} \tilde{S} \\ \tilde{T} \end{pmatrix} \begin{pmatrix} \tilde{R}_1 \\ \tilde{R}_2 \end{pmatrix},$

$$\tilde{V}^{-1} - \tilde{V}^{-1} \tilde{S} (\tilde{S}' \tilde{V}^{-1} \tilde{S})^{-1} \tilde{S}' \tilde{V}^{-1} = \tilde{S} \tilde{R}_1 + \tilde{T} \tilde{R}_2. \quad (11)$$

It remains to establish R_1 and R_2 . Multiplying both sides of (11) by S' gives $R_1 = 0$, and

$$V^{-1} - V^{-1} S (S' V^{-1} S)^{-1} S' V^{-1} = TR_2. \quad (12)$$

Multiplying both sides of (12) by $(T'VT)^{-1} T'V$ we find $R_2 = (T'VT)^{-1} T'$ and the desired result follows.

From (8) and application of Lemma 1,

$$W = L_{n-k} (L'_{n-k} B^{-1} L_{n-k})^{-1} L'_{n-k}. \quad (13)$$

Thus to prove that $W = MAM$, from (10) and (13) it is sufficient

$$\text{to prove that } (L'_{n-k} B^{-1} L_{n-k})^{-1} = \Lambda_{n-k}. \quad (14)$$

From (6) and because Λ_{n-k} is non-singular

$$L_{n-k} = A L_{n-k} \Lambda_{n-k}^{-1}.$$

Multiplying both sides by $L'_{n-k} B^{-1}$,

$$\begin{aligned} L'_{n-k} B^{-1} L_{n-k} &= L'_{n-k} B^{-1} A L_{n-k} \Lambda_{n-k}^{-1} \\ &= \left\{ I - \sqrt{n} L'_{n-k} \begin{pmatrix} 1 \\ 0 \end{pmatrix} L_{n-k} \right\} \Lambda_{n-k}^{-1}, \text{ from (7)} \\ &= \Lambda_{n-k}^{-1}, \end{aligned}$$

as $L'_{n-k} \mathbf{1} = 0$, $\mathbf{1}$ being a column of L_k . Thus

$(L'_{n-k} B^{-1} L_{n-k})^{-1} = \Lambda_{n-k}$, $W = MAM$, and equality of g and d_A is proved under the conditions stated in section 2.2 which give the Durbin-Watson test its optimum power properties.

Although intended as a test for autocorrelation in the non-stationary model these results (i) and (ii) suggest that the g statistic might have good power properties when used for testing

the null hypothesis of zero autocorrelation against both non-stationary and stationary alternatives. This point will be developed in section 2.5(ii) .

Analogous results obtain for the test against negative autocorrelation given by (1) with $\tilde{\rho}$ replaced by -1 . The properties of this test are summarized in section 2.9 after treatment of the g test for positive autocorrelation is completed.

2.4 Conditions under which the g test is most powerful

It is expedient now to ask if further properties can be derived for the proposed g test on the grounds that it is a simplification of the likelihood ratio test. For this purpose the hypothesis being tested is regarded as $H_0 : \rho = 0$, against $H_1 : \rho = 1$, i.e.

$$\left. \begin{aligned} H_0 : y &\sim N(X\beta, \sigma^2 I_n) \\ H_1 : y &\sim N(X\beta, \sigma^2 B^{-1}) \end{aligned} \right\} \quad (1)$$

Even in this form the test is not between simple hypotheses because the parameters β are unknown. It will be shown that there exist simple hypotheses (7) , obtained by transformation of (1) , for which the g test is most powerful (MP) . The general result is due to Kadiyala (1970) although, for later use, our method of proof is different.

The transformation is accomplished by choosing an $n \times (n - k)$ matrix $\tilde{Q}$ whose columns are orthonormal such that, where $\tilde{\xi} = \tilde{Q}'\tilde{y}$, then $E(\tilde{\xi}) = 0$. Following Durbin and Watson (1950), page 413, the design matrix $\tilde{X}$ may be written in the form $\tilde{S}\tilde{G}$ where $\tilde{S}$ is an $n \times k$ orthonormal matrix and $\tilde{G}$ is a $k \times k$ non-singular matrix. For example, $\tilde{S}$ could represent the Gram-Schmidt orthogonalization of $\tilde{X}$. Putting $\tilde{U} = (\tilde{S} \mid \tilde{T})$ as an $n \times n$ orthonormal matrix, $\tilde{U}'\tilde{U} = \tilde{U}\tilde{U}' = \tilde{I}_n$, and $\tilde{T}$ is any $n \times (n - k)$ orthonormal matrix such that $\tilde{S}'\tilde{T} = 0$. Therefore

$$\tilde{W} = \tilde{B} - \tilde{B}\tilde{S}(\tilde{S}'\tilde{B}\tilde{S})^{-1}\tilde{S}'\tilde{B} \quad (3)$$

$$\text{and} \quad \tilde{M} = \tilde{I} - \tilde{S}\tilde{S}' = \tilde{T}\tilde{T}' \quad (4)$$

From (3) and lemma 2.3.1

$$\tilde{W} = \tilde{T}(\tilde{T}'\tilde{B}^{-1}\tilde{T})^{-1}\tilde{T}' \quad (5)$$

The ensuing derivation of $\tilde{Q}$ differs from that in Berenblut and Webb (1973) in so far that it does not depend explicitly on a result due to Rao (1965b), page 358; the present argument seems preferable since it is more concise and direct.

In order to show that $\tilde{W}$ and $\tilde{M}$ can be diagonalized simultaneously using an orthonormal transformation it is necessary and sufficient, Bellman (1960), page 56, to show that $\tilde{W}$ and $\tilde{M}$ commute: from (4) and (5)

$$\tilde{W}\tilde{M} = \tilde{T}(\tilde{T}'\tilde{B}^{-1}\tilde{T})^{-1}\tilde{T}'\tilde{T}\tilde{T}' = \tilde{T}(\tilde{T}'\tilde{B}^{-1}\tilde{T})^{-1}\tilde{T}' = \tilde{M}\tilde{W}.$$

Now, where $\tilde{P}_1$ is the $(n - k) \times (n - k)$ orthonormal matrix of latent

vectors of $(\tilde{T}'\tilde{B}^{-1}\tilde{T})^{-1}$ it is easily seen that $\tilde{P}_1'\tilde{T}'\tilde{W}\tilde{T}\tilde{P}_1 = \tilde{J}$, the $(n-k) \times (n-k)$ diagonal matrix of latent roots of $(\tilde{T}'\tilde{B}^{-1}\tilde{T})^{-1}$, which are also the $n-k$ non zero latent roots of $\tilde{W}$, and $\tilde{P}_1'\tilde{T}'\tilde{M}\tilde{T}\tilde{P}_1 = \tilde{I}_{n-k}$. As $\tilde{P}_1'\tilde{P}_1 = \tilde{P}_1\tilde{P}_1' = \tilde{I}_{n-k}$, g may be written as

$$g = \frac{\tilde{y}'\tilde{T}\tilde{P}_1\tilde{P}_1'(\tilde{T}'\tilde{B}^{-1}\tilde{T})^{-1}\tilde{P}_1\tilde{P}_1'\tilde{T}'\tilde{y}}{\tilde{y}'\tilde{T}\tilde{P}_1\tilde{P}_1'\tilde{T}'\tilde{y}}.$$

Thus taking $\tilde{Q}$ as $\tilde{T}\tilde{P}_1/\sigma$, g becomes

$$g = \frac{\tilde{\xi}'\tilde{J}\tilde{\xi}}{\tilde{\xi}'\tilde{\xi}} = \frac{\sum_{i=1}^{n-k} d_i \xi_i^2}{\sum_{i=1}^{n-k} \xi_i^2}. \quad (6)$$

The distribution of $\tilde{\xi}$ is normal with mean given by

$E(\tilde{\xi}) = \tilde{P}_1'\tilde{T}'E(\tilde{y})/\sigma = \tilde{P}_1'\tilde{T}'S\tilde{G}\beta/\sigma = 0$, and dispersion matrix given by $D(\tilde{Q}'\tilde{y}) = \tilde{Q}'D(\tilde{y})\tilde{Q} = \tilde{P}_1'\tilde{T}'D(\tilde{y})\tilde{T}\tilde{P}_1$. Therefore under H_0 , $D(\tilde{\xi}) = \tilde{I}_{n-k}$ and under H_1 , $D(\tilde{\xi}) = \tilde{J}^{-1}$. Writing the hypotheses in terms of $\tilde{\xi}$ then

$$\left. \begin{aligned} H_0' : \tilde{\xi} &\sim N(0, \tilde{I}_{n-k}) \\ H_1' : \tilde{\xi} &\sim N(0, \tilde{J}^{-1}) \end{aligned} \right\} \quad (7)$$

Thus H_0 implies H_0' and H_1 implies H_1' , but the converses are not true as the transformation is irreversible, $\tilde{Q}$ being singular.

Kadiyala (1970) considered an approximation of the type (7) for

(1) and we make use of his results to state that the test given by the statistic g for the hypotheses defined in (7) is most powerful (by the Lehmann-Stein (1948) result) and unbiased. Moreover, no restrictions on the form of the design matrix have been made in this approximation. These results do not hold for the d_A and d_B tests, which are not particular cases of Kadiyala's statistic.

2.5 Results from Invariance Theory(i) Non-Stationary Model Durbin and Watson (1971)

considered the application of the theory of invariance to the problem of testing for autocorrelation, with particular reference to their d_A statistic. Because the autocorrelation structure is unchanged by transformations of the type $\tilde{y}^* = \gamma_0 \tilde{y} + \tilde{X} \tilde{\gamma}$, where $\tilde{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_k)'$, $-\infty < \gamma_i < \infty$, ($i = 1, 2, \dots, k$), and $0 < |\gamma_0| < \infty$, they applied the results of Lehmann (1959) using the approximation to the density of the stationary model given in (2.2.3). In this application of Lehmann's results we use the true densities. Consider the non-stationary process given by the density

$$(2\pi\sigma^2)^{-n/2} \cdot \exp\left(-\frac{1}{2\sigma^2} \tilde{u}' \tilde{V}^{-1} \tilde{u}\right) \quad (1)$$

and the test of the hypotheses defined by

$$\left. \begin{aligned} H_0 : \tilde{y} &\sim N(\tilde{X}\tilde{\beta}, \sigma^2 \tilde{I}_n) \\ \text{and } H_1 : \tilde{y} &\sim N(\tilde{X}\tilde{\beta}, \sigma^2 V(e^*)) \end{aligned} \right\} \quad (2)$$

i.e. $H_0 : \rho = 0$ against $H_1 : \rho = \rho^* > 0$. Proceeding along similar lines to Durbin and Watson (1970), pages 9 and 10 (but ^{noting} ~~correcting the mistake~~ that, under H_0 , $\tilde{B}$ is replaced by $\tilde{I}$ and $\hat{\tilde{u}}$ by $\tilde{z}$) it can be easily verified that the most powerful invariant test is given by the rejection region

$$g(\rho^*) = \frac{\hat{\tilde{u}}' \tilde{V}^{-1}(\rho^*) \hat{\tilde{u}}}{\tilde{z}' \tilde{z}} < c, \quad (3)$$

where $\hat{\underline{u}} = \left[\underline{I} - \underline{X} \{ \underline{X}' \underline{V}^{-1} (\epsilon^*) \underline{X} \}^{-1} \underline{X}' \underline{V}^{-1} (\epsilon^*) \right] \underline{y}$. Thus for the hypotheses (2.4.1), i.e. $H_1 : \epsilon^* = 1$, g provides the most powerful invariant test. This result is true of course independently of the form of the design matrix.

The UMP properties of the g test can be obtained very easily from the invariance theory approach. Durbin and Watson (1971) showed that their d_A test is an UMP invariant test under the conditions given in section 2.3.(ii) for equality of g and d_A , and so, under the same conditions, this property is also conferred upon the g test. The UMP property under the conditions given in section 2.3(i) is similarly derived, firstly noting that the approximation to the inverse of the dispersion matrix implied by (2.2.6),

$$\underline{\Omega}_{\sim}^{-1} = \begin{pmatrix} 1+\epsilon^2 & -\epsilon & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -\epsilon & 1+\epsilon^2 & -\epsilon & & & & & \vdots \\ 0 & \vdots & & & & & 1+\epsilon^2 & -\epsilon \\ \vdots & & & & & & -\epsilon & 1+\epsilon^2-\epsilon \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix},$$

can be written in the form $(1-\epsilon)^2 \underline{I}_{\sim} + \epsilon \underline{B}_{\sim}$. Because of the latent vector conditions $\hat{\underline{u}}_{\sim} = \underline{z}_{\sim}$, as before, and so (3) becomes

$$(1-\epsilon)^2 \frac{\underline{z}'_{\sim} \underline{z}_{\sim}}{\underline{z}'_{\sim} \underline{z}_{\sim}} + \epsilon \frac{\underline{z}'_{\sim} \underline{B}_{\sim} \underline{z}_{\sim}}{\underline{z}'_{\sim} \underline{z}_{\sim}} < c_1,$$

i.e. $d_B < c_2$.

Therefore, being equal to d_B under these conditions, the g statistic gives an uniformly most powerful invariant test when the columns of the design matrix $\tilde{X}$ are a linear combination of k latent vectors of $\tilde{B}$ and when the density being considered is

$$K \exp \left(- \frac{1}{2\sigma^2} \tilde{u}' \tilde{\Sigma}^{-1} \tilde{u} \right), \text{ a close approximation to (1) .}$$

If a test is required against an a priori value of ρ , say ρ^* , then $g(\rho^*)$ is the most powerful invariant statistic. However, if there is no obvious fixed value to choose for ρ^* , i.e. when testing against the composite alternative hypothesis of positive autocorrelation $H_1 : \rho = \rho^* > 0$, $g \{= g(1)\}$ has the advantageous power properties indicated above and in section 2.3 .

(ii) Stationary Model To pursue the remark made in section 1.3, if we followed the procedure of taking first differences and then used as a statistic the ratio of the residual sum of squares with first differenced and original variables, we would obtain

$$\frac{\tilde{y}'(\tilde{A} - \tilde{A}\tilde{X}\tilde{M}_1\tilde{X}'\tilde{A})\tilde{y}}{\tilde{y}'\tilde{M}_1\tilde{y}}, \text{ where } \tilde{M}_1 \text{ is a generalized}$$

inverse of $\tilde{X}'\tilde{A}\tilde{X}$. It is worthy of note that this ratio is the same as the g statistic with $\tilde{B}$ replaced by $\tilde{A}$. Under the condition that an overall mean is included in the regression, it is shown below that the statistic gives a most powerful invariant test for $\rho \rightarrow 1$ (the limit as ρ approaches 1 from below) in the case of the stationary process (2.2.2). This follows from Durbin and

Watson (1971), page 10, but remembering that their statement

$$\left(\frac{\hat{\underline{u}}' \hat{\underline{u}}}{\hat{\underline{u}}' \hat{\underline{B}} \hat{\underline{u}}} \right)^{\frac{1}{2}(n-k)} < c_1 \quad \text{should read}$$

$$\left(\frac{\hat{\underline{z}}' \hat{\underline{z}}}{\hat{\underline{u}}' \hat{\underline{B}} \hat{\underline{u}}} \right)^{\frac{1}{2}(n-k)} < c_1 .$$

Taking the exact density for the stationary process, (2.2.2), gives the rejection region as

$$(1 - e^2)^{-\frac{(n-1)/2}{2}} \cdot |X' U X|^{-\frac{1}{2}} \cdot \left(\frac{\hat{\underline{z}}' \hat{\underline{z}}}{\hat{\underline{u}}' \hat{\underline{U}}^{-1} \hat{\underline{u}}} \right)^{\frac{1}{2}(n-k)} > C ,$$

where $\sigma^2 \underline{U}$ is the dispersion matrix representing the stationary error process given explicitly in section 3.5. When $\underline{X}$ contains the unit vector $\underline{1}$, using the expansion of $|X' U X|$ in section 3.5 the terms in $(1 - e)$ cancel and the desired limit exists. Now, more remarkably, for regression involving an overall mean, this statistic will be shown to be identical to g . To prove equality it is necessary and sufficient to prove that $\underline{W} = \underline{W}^*$, where $\underline{W} = \underline{B} - \underline{B} \underline{X} (\underline{X}' \underline{B} \underline{X})^{-1} \underline{X}' \underline{B}$ and $\underline{W}^* = \underline{A} - \underline{A} \underline{X} \underline{M}_2 \underline{X}' \underline{A}$. As in section 2.4 we write $\underline{X}$ in the form $\underline{S} \underline{G}$, where here $\underline{X} = \begin{pmatrix} \underline{1} \\ \underline{X}_0 \end{pmatrix}$ and $\underline{S} = \begin{pmatrix} \underline{1} \\ \underline{S}_0 \end{pmatrix}$, i.e. the constant term is included as the first parameter in the regression scheme. Hence $\underline{W} = \underline{B} - \underline{B} \underline{S} (\underline{S}' \underline{B} \underline{S})^{-1} \underline{S}' \underline{B}$, $\underline{W}^* = \underline{A} - \underline{A} \underline{S} \underline{M}_2 \underline{S}' \underline{A}$ where $\underline{M}_2$ is a generalized inverse of $\underline{S}' \underline{A} \underline{S}$,

$$\text{and } \underline{S}' \underline{A} \underline{S} = \begin{pmatrix} \underline{1}' \\ \underline{S}_0' \end{pmatrix} \underline{A} \begin{pmatrix} \underline{1} \\ \underline{S}_0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \dots & \dots \\ 0 & \underline{S}_0' \underline{A} \underline{S}_0 \end{pmatrix} ,$$

Now, to show that $\underline{S}_0' \underline{A} \underline{S}_0$ is positive-definite and non-singular we have that $\underline{S}_0$ is orthogonal to $\underline{1}$, the latent vector of $\underline{A}$

associated with the zero latent root. Consequently $\tilde{S}_0$ may be written in general as $\tilde{L}_{n-1} \tilde{G}_1$, where $\tilde{G}_1$ is an $(n-1) \times (n-k-1)$ orthonormal matrix and $\tilde{L}_{n-1}$ is the matrix of the latent vectors of $\tilde{A}$ associated with the diagonal matrix of the $n-1$ positive latent roots $\tilde{\Lambda}_{n-1}$. Thus $\tilde{S}_0' \tilde{A} \tilde{S}_0 = \tilde{G}_1' \tilde{L}_{n-1}' \tilde{A} \tilde{L}_{n-1} \tilde{G}_1 = \tilde{G}_1' \tilde{\Lambda}_{n-1} \tilde{G}_1$, and is therefore positive-definite and hence non-singular. Using this result to write a generalized inverse of $\tilde{S}' \tilde{A} \tilde{S}$ as

$$\begin{pmatrix} 0 & \vdots & 0 \\ \hline 0 & \vdots & (\tilde{S}_0' \tilde{A} \tilde{S}_0)^{-1} \end{pmatrix},$$

we see that $\tilde{W}^* = \tilde{A} - \tilde{A} \tilde{S}_0 (\tilde{S}_0' \tilde{A} \tilde{S}_0)^{-1} \tilde{S}_0' \tilde{A}$. (4)

Here, lemma 2.3.1 cannot be used directly as $\tilde{A}$ is singular. However, $\tilde{W}^*$ may still be written in the form $\tilde{1} \tilde{R}_1 + \tilde{S}_0 \tilde{R}_2 + \tilde{T} \tilde{R}_3$, whence it is easily verified that $\tilde{R}_1 = 0$, $\tilde{R}_2 = 0$, and $\tilde{R}_3 = (\tilde{T}' \tilde{B}^{-1} \tilde{T})^{-1} \tilde{T}'$ by premultiplication of (4) by $(\tilde{T}' \tilde{B}^{-1} \tilde{T})^{-1} \tilde{T}' \tilde{B}^{-1}$ and use of (2.3.7). Finally $\tilde{W} = \tilde{T} (\tilde{T}' \tilde{B}^{-1} \tilde{T})^{-1} \tilde{T}'$, from (2.4.5) and so equality of $\tilde{W}^*$ and $\tilde{W}$ is established under the condition that an overall mean appears as a regression parameter.

Hence the g statistic provides a most powerful invariant test whether the process is stationary or non-stationary as $\rho \rightarrow 1$. In particular, because the d_A test is also an invariant test, it is concluded that the g test is more powerful than the Durbin-Watson test for high values of autocorrelation.

2.6 Numerical examination of the power functions

Durbin and Watson (1971) claim that their d_A test is a locally most powerful invariant test in the region of the null hypothesis, i.e. as $\rho \rightarrow 0$. There is a fault in the argument they put forward - on page 10, under $H_0 : \hat{u}_{\sim}$ should be replaced by $z_{\sim}$. However, the following investigations support their contention in an empirical way.

We consider the three practical examples given in section 1.2 and examine the power characteristics of the statistics d_B , d_A and g against both stationary and non-stationary alternatives. Particular interest is centred on whether the d_A test is in fact more powerful than the g test in the neighbourhood of the null hypothesis, and also on the values of ρ for which the g test becomes more powerful than the d_A test.

The numerical methods used to compute the significance points and powers of the three statistics are due to Koerts and Abrahamse (1968) using results due to Imhof (1960). The theory of the application of the method to the g statistic is outlined in the next section.

The powers of the three tests for various values of ρ in the alternative hypothesis are given in table 3 for the non-stationary case and in table 4 for the stationary case. The size of the tests is taken as .05 throughout.

Data	Statistic	Alternative value of ρ					
		0.1	0.3	0.5	0.7	0.9	1.0
Williams	d_B	.085	.199	.354	.506	.657	.753
	d_A	.084	.193	.342	.499	.677	.782
	g	.082	.190	.350	.530	.736	.850
Box and Hunter	d_B	.095	.251	.451	.613	.694	.704
	d_A	.095	.250	.453	.619	.701	.711
	g	.094	.248	.451	.619	.703	.713
Pears	d_B	.079	.171	.301	.442	.564	.590
	d_A	.078	.168	.296	.442	.575	.603
	g	.078	.169	.302	.455	.590	.618

Powers of the d_B, d_A and g tests for non-stationary alternatives

TABLE 3

Data	Statistic	Alternative value of ρ				
		0.1	0.3	0.5	0.7	0.9
Williams	d_B	.085	.202	.371	.549	.696
	d_A	.085	.202	.377	.567	.723
	g	.082	.200	.392	.611	.788
Box and Hunter	d_B	.095	.251	.452	.614	.695
	d_A	.095	.251	.455	.619	.701
	g	.094	.249	.453	.620	.703
Pears	d_B	.079	.172	.306	.452	.567
	d_A	.079	.171	.306	.458	.579
	g	.078	.172	.312	.470	.594

Powers of the d_B, d_A and g tests for stationary alternatives

TABLE 4

a. PEAR'S DATA A graph of the power functions for this data may be found in Berenblut and Webb (1973), fig. 1. The similarity between the power function characteristics for both non-stationary and stationary alternatives is noticeable. It is also clear that the g test is more powerful than the d_A test for values of ρ greater than about .3 in both cases. The d_A test is slightly more powerful for smaller values of ρ .

b. BOX AND HUNTER'S DATA Box and Hunter (1958) consider a cubic polynomial adequate to represent the trend and so the design matrix X is taken from the appropriate orthogonal polynomials in Fisher and Yates (1953), page 81. Thus $n = 21$ and $k = 4$. In this experiment there is a valid physical basis to postulate that the response at time $t = 0$ is 100%, so that inclusion of the overall mean in the model is debatable, Berenblut and Webb (1972). There is very little difference between the powers of d_A and g in this example. This is not unexpected when it is noted that the design matrix in the polynomial regression case spans a subspace close to that spanned by the k eigenvectors of A corresponding to the k smallest latent roots, Hannan (1960).

c. WILLIAMS' DATA This data is from a carry-over situation and so we would expect negative rather than positive autocorrelation. Because, in this section, we are interested in comparing the power properties of d_A and g {i.e. $g(1)$ }, which are tests specifically designed for positive autocorrelation, we defer treatment of the negative case until section 2.8 where the power properties of the test given by $g(-1)$ will be considered. Instead

the Williams design matrix is used as an example of a 6×6 Latin square in a test for positive autocorrelation. The parameters being fitted are subjects, periods, treatments and an overall mean. Thus $n = 36$, $k = 16$, and the design matrix is reduced to full rank as in section 1.2. In this example the differences in powers of the d_A and g tests are more marked, with the g test being decidedly more powerful for values of ρ of .5 and greater.

From these practical examples (and others considered in Berenblut and Webb (1973)) there is support for the Durbin and Watson claim that the d_A test is the locally most powerful invariant test in the region of the null hypothesis for the stationary alternative. However, whether this is a particularly desirable property of a test for autocorrelation is questionable, and we shall return to this point in section 3.1. The g test is seen to be most powerful as expected for high values of autocorrelation in both stationary and non-stationary alternatives.

2.7 The numerical procedures

From (2.4.6) $g = \sum_{i=1}^{n-k} \nu_i \xi_i^2 / \sum_{i=1}^{n-k} \xi_i^2$, where the

ν_i ($i = 1, 2, \dots, n-k$) are the non-zero latent roots of W and under $H_0: \xi \sim N(0, I_{n-k})$. The probability that g is less than some given value t is given by

$$P(g < t) = P(\sum_{i=1}^{n-k} \nu_i \xi_i^2 / \sum_{i=1}^{n-k} \xi_i^2 < t) = P\{\sum_{i=1}^{n-k} (\nu_i - t) \xi_i^2 < 0\}.$$

The significance point t_α for a given size of test α is found by solving iteratively for t the equation $P\{\sum (\nu_i - t)\xi_i^2 < 0\} = \alpha$. Using Imhof's (1961) result this reduces to finding a t for which

$$\frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{u \varrho(u)} du = \alpha,$$

where $\theta(u) = \frac{1}{2} \sum_1^{n-k} \tan^{-1}\{(\nu_i - t)u\}$

and $\varrho(u) = \left[\prod_1^{n-k} \{1 + (\nu_i - t)^2 u^2\} \right]^{1/4}$.

The integral may be evaluated by standard numerical techniques using the expressions given by Imhof for the truncation error and the limit of the quotient as $u \rightarrow 0$. The latent roots of $\tilde{W}$, ν_i , have only to be evaluated once. A suitable algorithm for this purpose is given, in Algol, by W. Barth et al (1967).

The power of the test of size α against $H_1: u \sim N(0, \sigma^2 V)$ is given by

$$\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \phi(u)}{u \mu(u)} du,$$

where $\phi(u) = \frac{1}{2} \sum_1^n \tan^{-1}(x_i u)$,

$$\mu(u) = \left\{ \prod_1^n (1 + x_i^2 u^2) \right\}^{1/4},$$

and x_i are the n latent roots of $\tilde{C}^{-1} \tilde{T}(\tilde{W} - t_\alpha \tilde{M}) \tilde{C}^{-1}$.

For the non-stationary alternative the inverse of $\tilde{C}$, defined in (1.3.2), is given by

$$\tilde{C}^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \rho & 1 & 0 & & \\ \rho^2 & \rho & 1 & & \\ \vdots & & & \ddots & \\ \vdots & & & & 1 & 0 \\ \rho^{n-1} & \dots & \dots & \dots & \rho & 1 \end{pmatrix}, \text{ and } \tilde{C}^{-1} \tilde{C}^{-1T} = \tilde{V}.$$

The appropriate matrix for the stationary alternative is obtained by multiplying the elements of the last $n - 1$ columns of this non-stationary $\tilde{C}^{-1}$ by $\sqrt{1 - \rho^2}$.

Thus, for any given design matrix $\tilde{X}$ and size of test α , the critical point and power of the g test for the hypotheses $H_0 : \rho = 0$ against non-stationary or stationary alternatives $H_1 : \rho = \rho^* \neq 0$ can be computed to any required accuracy. However in practice for particular examples it is possible to greatly reduce the amount of computation necessary for the significance test. The significance level of the test is found by taking for t the actual value of g from the data. The iterative procedure is not required and the integral need only be evaluated once.

This approach outlined above is based on that given by Koerts and Abrahamse (1968). The adaptation^{at} to the g statistic is given here because it was the method used to obtain our practical results. More recently Durbin and Watson (1971) give an alternative method based on results due to Pan Jie-jian (1968). They state that the method is preferable to the Imhof method particularly for small sample sizes and tail-end probabilities.

The calculation of the value of g from any particular set of data is, like that of d_A , a fairly straightforward procedure. The denominators are the same being the ordinary residual sum of squares. The numerator of g may be evaluated as the residual sum of squares of transformed dependent variables, $\tilde{C}y$, regressed on the correspondingly transformed design matrix $\tilde{C}X$, where $\tilde{C}$ is $\tilde{C}(1)$ defined in (1.3.2). This is the residual sum of squares of first differenced variables with y_i and $X_{i,j}$ ($j = 1, 2, \dots, k$) remaining unchanged.

In the particular case where X contains the column $\tilde{1}$ the calculation of the numerator can be simplified further by computing the residual sum of squares using the $n - 1$ first differenced variables and $k - 1$ parameters, y_i and $X_{i,j}$ ($j = 1, 2, \dots, k$) being discarded. This follows directly from the proof that $\tilde{W} = \tilde{W}^*$ in section 2.5 but, alternatively, it is more illustrative to consider the transformed design matrix

$$\tilde{C}X = \begin{pmatrix} n^{-\frac{1}{2}} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tilde{C}X_0, \quad \tilde{C}X_0 = \begin{pmatrix} X_0 \\ \vdots \\ X_0 \end{pmatrix},$$

without loss of generality putting the normalized unit vector $\tilde{1}$ as the first column of X , i.e. $X = (\tilde{1} \mid X_0)$. Reducing the columns of the transformed design matrix so that they are orthogonal to the first is accomplished by making zero the first row elements in the last $k - 1$ columns. This is equivalent to postmultiplication by a non-singular $k \times k$ matrix of the type

$$\begin{pmatrix} 1 & * & * & * & \cdot & \cdot & \cdot & * \\ 0 & 1 & 0 & & & & & 0 \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix}$$

The residual sum of squares is not affected by this post-multiplication and it is now evident that the least squares estimate of the first observation is in fact itself, i.e. $\hat{y}_1 = y_1$, leaving a zero residual. Therefore the first element of Cy , and the first row and column of CX may be ignored when evaluating the numerator of g in regression including an overall mean as the first parameter. In experimental design situations where there is more than one subject the numerator of g can be evaluated as the residual sum of squares of first differenced variables on each subject with the first observations and corresponding rows of X being ignored.

We are now able to use the g statistic as a test for autocorrelated errors in the practical examples. Testing for positively autocorrelated errors in the PEARS DATA the d_A test is significant at the 5% level, although the value of d_A lies between the tabulated bounds. The value of g is 1.085 and $P(g < 1.085)$ is 0.019, so that the g test is significant beyond the 5% level. In fact the 5% significance point is 1.27. It is worth noting that both an overall mean and linear time trend are included as parameters in this example: the inclusion of a linear time trend is not uncommon in economic models.

In the BOX AND HUNTER DATA with a cubic fit the g test is significant at the 0.1 % level. The value of g is 0.989 and $P(g < 0.989) = .0004$. As the object of the pilot experiment is to design the main experiment to be efficient in the estimation of the treatment effects the assumption of independent errors is fundamental to Box and Hunter's design procedures. The consequences of this rejection of the null hypothesis of independent errors will be examined later.

The WILLIAMS DATA is analysed in the next section where we consider a version of the g test appropriate for testing the hypothesis of negative autocorrelation.

2.8 Testing for negative autocorrelation

The results of section 2.5 suggest that as a test for negatively autocorrelated errors we use the statistic $g(-1)$, defined in (2.5.3), which is a most powerful invariant test when $\rho = -1$. In addition it has the property of providing a uniformly most powerful invariant test when the columns of the design matrix $\tilde{X}$ are a linear combination of k latent vectors of $V^{-1}(-1)$, and when the density being considered is (2.2.7), a close approximation to the density of the non-stationary process. Furthermore, it can be shown along similar lines to the method given in section 2.5(ii) for the positive case, that the statistic $g(-1)$ gives a most powerful invariant test in the region $\rho \rightarrow -1$

against the stationary alternative (2.2.2), provided that the vector $(1 - 1 \ 1 \dots -1 \ 1)'$ is included as a column of $\tilde{X}$. This provision corresponds to the requirement of fitting an overall mean in the positive autocorrelation case. However, whereas in many regressions it is usual to fit an overall mean, the inclusion of an arbitrary vector $(1 -1 \ 1 \dots -1 \ 1)'$ is not easily justifiable.

In regressions including an overall mean it seems reasonable to ask whether a test given by using the right hand tails of the statistics d_A or g would have good power properties against negative alternatives. In an attempt to answer this question using the Williams data design matrix we have computed the powers of the tests given by appropriate use of the three statistics $g(-1)$, $g(1)$ and d_A for values of ρ of -0.5 , -0.8 , -0.9 . These are given in tables 5 and 6 for the non-stationary and stationary alternatives respectively. The $g(-1)$ test is substantially more powerful than the tests given by using the right hand tails of the statistics d_A or $g(1)$. It is interesting to note by comparing tables 3 and 5 and tables 4 and 6 that, against the same absolute values of autocorrelation, the test for negative autocorrelation is more powerful than the test for positive autocorrelation.

Evaluation of $g(-1)$ for any particular set of observations is easily accomplished in a similar manner to $g(1)$. The denominator is again the ordinary residual sum of squares and the numerator is the residual sum of squares of transformed dependent variables, $\tilde{C}y$, regressed on the correspondingly transformed design matrix, $\tilde{C}X$, where $\tilde{C}$ is $C(-1)$ defined in (1.3.2).

Test	Alternative value of ρ		
	-.5	-.8	-.9
$d_A(1)$	.433	.753	.822
$g(1)$	.364	.590	.637
$d_A(-1)$	.439	.787	.861
$g(-1)$	.492	.907	.963

Powers of the d_A and g tests for alternatives
of non-stationary negative autocorrelation
with the Williams design

TABLE 5

Test	Alternative value of ρ		
	-.5	-.8	-.9
$d_A(1)$	.417	.781	.898
$g(1)$	.352	.601	.674
$d_A(-1)$	.464	.849	.920
$g(-1)$	.478	.923	.985

Powers of the d_A and g tests for alternatives
of stationary negative autocorrelation
with the Williams design

TABLE 6

These results are now applied using the $g(-1)$ statistic to test the WILLIAMS DATA for negatively autocorrelated errors. With an overall mean, subject, period and treatment effects included as parameters in the model the value of $g(-1)$ is 1.144 and $P(g(-1) < 1.144) = 0.383$. With residual effects also included in the model $g(-1) = 1.013$ and $P(g(-1) < 1.013) = 0.543$. Thus neither case is significant at even the 10% level.

Cochran and Cox (1957) present an example based on another Williams' design, namely

A	B	C	A	B	C	
B	C	A	C	A	B	, using the results
C	A	B	B	C	A	

of an experiment on the feeding of dairy cows. Here there are 14 estimable parameters, two of which are residual effects. When the residual effects are not included in the model $g(-1) = 0.308$ and $P\{g(-1) < .308 = .053\}$. The F test for the residual effects was also not quite significant at the 5% level. The $g(-1)$ value for the full model including residual effects parameters is 0.538 and $P\{g(-1) < .538 = .461\}$. The analysis of the dairy cow experiment indicates a connection between the models of negative autocorrelation and residual effects but the results of the paper mill analysis indicate the opposite, i.e. that the models of autocorrelation and residual effects are not interchangeable representations of carry-over situations. More work is necessary to clarify this issue.

2.9 Bounds tests using g

Durbin and Watson's d_A test has found much application because (i) the statistic is comparatively easy to calculate, (ii) the power properties possessed by d_A , and (iii) the authors were able to derive and tabulate upper and lower bounds to the significance points. For the practical worker without easy access to a computer program the 'bounds test' procedure provides a fairly quick and easy test for autocorrelation. In the case of the value of d_A falling in the inconclusive region between the tabulated bounds the aid of a computer has to be enlisted - even if using the beta approximation, according to Durbin and Watson (1971) page 4. Of course, if a computer is to be used at all it might as well be used for the g test on the grounds of power alone.

Because the 'bounds test' does find much use, particularly among econometricians, in this section bounds are derived on the significance points of the g test. Berenblut and Webb (1973) established bounds on the g statistic with and without a vector $\tilde{1}$ being included in the design matrix. We shall now generalize this approach. The class of design matrices being considered is the one where the first r of the k columns of $\tilde{X}$ are specified. That is, r of the regression vectors are known and without loss of generality these are placed first. Thus $\tilde{X} = (\tilde{X}_0 \vdots \tilde{X}_1)$ where $\tilde{X}_0$ is a known $n \times r$ matrix and $\tilde{X}_1$ is an unknown $n \times (k - r)$ matrix. In many economic applications $\tilde{2}$ of the k regression vectors would represent a constant term and a linear time trend. In some experimental design situations the known vectors might represent the overall mean, subject effects and period effects. On the other hand r would be

zero if a result was required which was completely independent of the form of the design matrix. These special cases where $r = 0, 1, 2$ will be examined in the light of the general result.

Again we write $\underset{\sim}{X} = \underset{\sim}{S} \underset{\sim}{G}$, where $\underset{\sim}{S}$ is orthonormal and $\underset{\sim}{G}$ non-singular for $\underset{\sim}{X}$ of full rank. Now $\underset{\sim}{S}$ may be partitioned $(\underset{\sim}{S}_0 \mid \underset{\sim}{S}_1)$, where $\underset{\sim}{S}_0$ has dimension $n \times r$. Define $\underset{\sim}{T}$ as any $n \times (n - k)$ matrix forming an orthonormal complement to $\underset{\sim}{S}$, i.e. $(\underset{\sim}{S} \mid \underset{\sim}{T})' (\underset{\sim}{S} \mid \underset{\sim}{T}) = (\underset{\sim}{S} \mid \underset{\sim}{T}) (\underset{\sim}{S} \mid \underset{\sim}{T})' = \underset{\sim}{I}_n$. Similarly define $\underset{\sim}{H}$ to be any $n \times (n - r)$ matrix orthonormal to $\underset{\sim}{S}_0$, i.e. $(\underset{\sim}{S}_0 \mid \underset{\sim}{H})' (\underset{\sim}{S}_0 \mid \underset{\sim}{H}) = (\underset{\sim}{S}_0 \mid \underset{\sim}{H}) (\underset{\sim}{S}_0 \mid \underset{\sim}{H})' = \underset{\sim}{I}_n$. Clearly $\underset{\sim}{T}$ may be written as

$$\underset{\sim}{T} = \underset{\sim}{H} \underset{\sim}{P}, \quad (1)$$

where $\underset{\sim}{P}$ is an $(n - r) \times (n - k)$ orthonormal matrix such that $\underset{\sim}{P}' \underset{\sim}{P} = \underset{\sim}{I}_{n-k}$. Within the given conditions choice of $\underset{\sim}{H}$ is arbitrary and does not affect the final results. From (2.4.6)

$$g = \sum_1^{n-k} \gamma_i \xi_i^2 / \sum_1^{n-k} \xi_i^2, \text{ where the } \gamma_i \text{ (} i = 1, 2, \dots, n - k \text{)}$$

are the latent roots of $(\underset{\sim}{T}' \underset{\sim}{B}^{-1} \underset{\sim}{T})^{-1}$. As $(\underset{\sim}{T}' \underset{\sim}{B}^{-1} \underset{\sim}{T})^{-1}$ is positive definite the latent roots of $(\underset{\sim}{T}' \underset{\sim}{B}^{-1} \underset{\sim}{T})$ are $1/\gamma_i$. For convenience the latent roots of $(\underset{\sim}{T}' \underset{\sim}{B}^{-1} \underset{\sim}{T})^{-1}$ are ordered from smallest to largest so that $0 < \gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_{n-k}$. From (1) it is evident that

$$\underset{\sim}{T}' \underset{\sim}{B}^{-1} \underset{\sim}{T} = \underset{\sim}{P}' \underset{\sim}{H}' \underset{\sim}{B}^{-1} \underset{\sim}{H} \underset{\sim}{P}. \quad (2)$$

Now $\tilde{H}$ is of full rank so that $\tilde{H}' \tilde{B}^{-1} \tilde{H}$ is positive definite.

Let the $n - r$ ordered latent roots of $\tilde{H}' \tilde{B}^{-1} \tilde{H}$ be $1/\mu_i$ where $0 < 1/\mu_{n-r} \leq 1/\mu_{n-r-1} \leq \dots \leq 1/\mu_1$. The relationship between the latent roots of $\tilde{T}' \tilde{B}^{-1} \tilde{T}$ and those of $\tilde{H}' \tilde{B}^{-1} \tilde{H}$ is given by

$$1/\mu_{n-r-i+1} \leq 1/\nu_i \leq 1/\mu_{n-k-i+1} \quad , \quad (i = 1, 2, \dots, n - k) \quad ,$$

and is obtained from a similar result in Kadiyala (1970) page 101, or from Hannan (1970) page 460. Taking reciprocals,

$$\mu_i \leq \nu_i \leq \mu_{i+k-r} \quad , \quad (i = 1, 2, \dots, n - k) \quad , \quad (3)$$

where the μ_i are the $n - r$ latent roots of $(\tilde{H}' \tilde{B}^{-1} \tilde{H})^{-1}$ ordered from smallest to largest. Thus it is clear that $g_L \leq g \leq g_U$,

$$\text{where } g_L = \frac{\sum_{i=1}^{n-k} \mu_i \xi_i^2}{\sum_{i=1}^{n-k} \xi_i^2} \quad \text{and} \quad g_U = \frac{\sum_{i=1}^{n-k} \mu_{i+k-r} \xi_i^2}{\sum_{i=1}^{n-k} \xi_i^2} \quad ,$$

and the significance points of g are bounded by the corresponding significance points of g_L and g_U .

Therefore to establish bounds to the significance points of the g statistic for any given $\tilde{X}_0$ and r it is only necessary to obtain $\tilde{H}$ as any $n \times (n - r)$ orthonormal matrix orthogonal to $\tilde{X}_0$. The latent roots of $(\tilde{H}' \tilde{B}^{-1} \tilde{H})^{-1}$ are evaluated either analytically or numerically and the bounds to the significance points can be computed using the techniques in section 2.7. Having derived the general result we proceed to investigate some particular cases.

(i) No conditions on $\tilde{X}$. Here, with $r = 0$, $\tilde{H}$ is chosen as the $n \times n$ identity matrix $\tilde{I}_n$. Then $(\tilde{H}' \tilde{B}^{-1} \tilde{H})^{-1} = \tilde{B}$ and the bounds on the g statistic are

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

where the θ_i are the latent roots of $\tilde{B}$ in ascending order of magnitude. From (2.2.7) they are given by

$\theta_i = 2 - 2 \cos \{(2i - 1)\pi / (2n + 1)\}$, ($i = 1, 2, \dots, n$). From this result the limits of the range of the g statistic are found. Clearly $\theta_1 \leq \gamma_i \leq g \leq \gamma_{n-k} \leq \theta_n$. Thus $0 < g < 4$, a result which was quoted in section 2.3.

The task of computing the significance points for the upper and lower bounds is now straightforward, using the methods of section 2.7. The bounds for the 5% level significance points are given in table 7, for values of $n = 12(1)30(5)50(10)100$ and $k = 1(1)6$, correct to the second decimal place.

(ii) For regressions including an overall mean. Let $\tilde{L}$ be the $n \times n$ orthonormal matrix representing the latent vectors of $\tilde{A}$. Further, let the first column of $\tilde{L}$ be $\tilde{1}$, the normalized unit vector corresponding to the zero latent root. Then $\tilde{L} = (\tilde{1} \vdots \tilde{L}_o)$ and

TABLE 7

n	k=1		k=2		k=3		k=4		k=5		k=6	
	g_L	g_U	g_L	g_U	g_L	g_U	g_L	g_U	g_L	g_U	g_L	g_U
12	0.91	1.23	0.76	1.45	0.62	1.70	0.48	1.98	0.35	2.28	0.24	2.60
13	0.96	1.25	0.81	1.45	0.68	1.67	0.54	1.92	0.42	2.20	0.31	2.48
14	0.99	1.27	0.86	1.45	0.73	1.65	0.60	1.88	0.48	2.13	0.37	2.39
15	1.03	1.29	0.90	1.45	0.78	1.64	0.65	1.85	0.54	2.07	0.43	2.31
16	1.06	1.30	0.94	1.45	0.82	1.63	0.70	1.82	0.59	2.02	0.48	2.24
17	1.09	1.32	0.98	1.46	0.86	1.62	0.75	1.79	0.64	1.98	0.53	2.18
18	1.12	1.33	1.01	1.46	0.90	1.61	0.79	1.77	0.69	1.95	0.58	2.13
19	1.14	1.35	1.04	1.47	0.94	1.61	0.83	1.76	0.73	1.92	0.63	2.09
20	1.16	1.36	1.07	1.47	0.97	1.60	0.87	1.74	0.77	1.90	0.67	2.06
21	1.19	1.37	1.09	1.48	1.00	1.60	0.90	1.73	0.81	1.88	0.71	2.03
22	1.20	1.38	1.12	1.49	1.02	1.60	0.93	1.73	0.84	1.86	0.75	2.00
23	1.22	1.39	1.14	1.49	1.05	1.60	0.96	1.72	0.87	1.84	0.78	1.98
24	1.24	1.40	1.16	1.50	1.07	1.60	0.99	1.71	0.90	1.83	0.82	1.96
25	1.26	1.41	1.18	1.50	1.10	1.60	1.01	1.71	0.93	1.82	0.85	1.94
26	1.27	1.42	1.20	1.51	1.12	1.60	1.04	1.70	0.96	1.81	0.88	1.92
27	1.29	1.43	1.21	1.51	1.14	1.60	1.06	1.70	0.98	1.80	0.91	1.91
28	1.30	1.44	1.23	1.52	1.16	1.60	1.08	1.70	1.01	1.79	0.93	1.90
29	1.31	1.45	1.24	1.52	1.17	1.61	1.10	1.69	1.03	1.79	0.96	1.89
30	1.33	1.46	1.26	1.53	1.19	1.61	1.12	1.69	1.05	1.78	0.98	1.88
35	1.38	1.49	1.32	1.55	1.26	1.62	1.20	1.69	1.14	1.76	1.08	1.84
40	1.42	1.52	1.37	1.57	1.32	1.63	1.27	1.69	1.21	1.75	1.16	1.82
45	1.46	1.54	1.41	1.59	1.37	1.64	1.32	1.69	1.27	1.75	1.22	1.80
50	1.49	1.57	1.45	1.61	1.40	1.65	1.36	1.70	1.32	1.75	1.28	1.80
60	1.53	1.60	1.50	1.64	1.47	1.67	1.43	1.71	1.40	1.75	1.36	1.79
70	1.57	1.63	1.54	1.66	1.51	1.69	1.48	1.72	1.45	1.75	1.42	1.79
80	1.60	1.65	1.58	1.68	1.55	1.70	1.52	1.73	1.50	1.76	1.47	1.79
90	1.62	1.67	1.60	1.69	1.58	1.72	1.56	1.74	1.53	1.76	1.51	1.79
100	1.65	1.68	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.77	1.54	1.79

Overall bounds to the 5% significance points of the g test

$$\underset{\sim}{L}' \underset{\sim}{A} \underset{\sim}{L} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & \lambda_1 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix} = \underset{\sim}{\Lambda} = \begin{pmatrix} 0 & \vdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \vdots & \underset{\sim}{\Lambda}_{n-1} \end{pmatrix},$$

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

$$\text{Conversely } \underset{\sim}{A} = \underset{\sim}{L} \underset{\sim}{A} \underset{\sim}{L}' = \underset{\sim}{L}_o \underset{\sim}{\Lambda}_{n-1} \underset{\sim}{L}_o' \quad (4)$$

Choosing $\underset{\sim}{H}$ to be $\underset{\sim}{L}_o$, in order to establish the bounds it is necessary to find the $n-1$ latent roots of $(\underset{\sim}{L}_o' \underset{\sim}{B}^{-1} \underset{\sim}{L}_o)^{-1}$.

From (4)

$$\underset{\sim}{B}^{-1} \underset{\sim}{A} = \underset{\sim}{B}^{-1} \underset{\sim}{L}_o \underset{\sim}{\Lambda}_{n-1} \underset{\sim}{L}_o'$$

so that

$$\underset{\sim}{L}_o' \underset{\sim}{B}^{-1} \underset{\sim}{A} \underset{\sim}{L}_o = \underset{\sim}{L}_o' \underset{\sim}{B}^{-1} \underset{\sim}{L}_o \underset{\sim}{\Lambda}_{n-1}$$

As $\underset{\sim}{\Lambda}_{n-1}$ is non-singular and using (2.3.8)

$$\underset{\sim}{L}_o' \underset{\sim}{B}^{-1} \underset{\sim}{L}_o = \underset{\sim}{L}_o' \underset{\sim}{B}^{-1} \underset{\sim}{A} \underset{\sim}{L}_o \underset{\sim}{\Lambda}_{n-1}^{-1} = \underset{\sim}{\Lambda}_{n-1}^{-1} \quad (5)$$

Therefore the $n-1$ latent roots of $(\underset{\sim}{L}_o' \underset{\sim}{B}^{-1} \underset{\sim}{L}_o)^{-1}$ are the λ_i , ($i = 1, 2, \dots, n-1$), and the bounds to the g statistic when an overall mean is fitted are

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

precisely the same bounds as for the d_A statistic. Thus the tables published in Durbin and Watson (1951) are directly applicable to bounds tests employing the g statistic.

(iii) Regressions including an overall mean and linear time trend.

As this class of design matrices is fairly common, particularly in economic models, it is considered a useful exercise to tabulate the relevant bounds to the 5% level significance points of the g statistic.

A particularly convenient choice of $\tilde{H}$ in this case is the orthonormal matrix of orthogonal polynomials, designated ξ_r ($r = 2, 3, \dots, n-1$) by Fisher and Yates (1953). Thus $\tilde{H}$ can be generated very easily in the computer using the usual recurrence relation. The $n - 2$ latent roots of $\tilde{H}' \tilde{B}^{-1} \tilde{H}$ are evaluated using any standard algorithm, e.g. W. Barth et al (1967), and the latent roots of $(\tilde{H}' \tilde{B}^{-1} \tilde{H})^{-1}$ are the reciprocals of these. Using the methods of section 2.7 the 5% significance points of the upper and lower bounding statistics have been tabulated to an accuracy of 2 decimal places in table 8. The corresponding bounds to the d_A statistic under these conditions are given by consideration of the latent roots of $\tilde{H}' A \tilde{H}$, and are not the same as the bounds of the g statistic. These bounds to the d_A statistic were computed for selected values of n and are given in table 9.

Most multiple regression computer package programs compute and print out the value of the d_A statistic. Modification of these programs to print out the value of the g statistic should be straightforward. Where an overall mean is included in the regression set the user would still consult the same tables in Durbin and Watson (1951). In the case where both the overall mean

n	k=3		k=4		k=5		k=6	
	g_L	g_U	g_L	g_U	g_L	g_U	g_L	g_U
12	1.15	1.56	0.99	1.85	0.82	2.16	0.66	2.50
13	1.17	1.54	1.02	1.80	0.87	2.08	0.71	2.38
14	1.20	1.54	1.05	1.77	0.91	2.02	0.77	2.29
15	1.22	1.53	1.08	1.74	0.94	1.97	0.81	2.21
16	1.24	1.53	1.11	1.72	0.98	1.93	0.85	2.15
17	1.26	1.53	1.13	1.70	1.01	1.89	0.89	2.10
18	1.27	1.53	1.16	1.69	1.04	1.86	0.93	2.05
19	1.29	1.53	1.18	1.68	1.07	1.84	0.96	2.02
20	1.30	1.53	1.20	1.67	1.10	1.82	0.99	1.99
21	1.32	1.53	1.22	1.66	1.12	1.81	1.02	1.96
22	1.33	1.54	1.24	1.66	1.14	1.79	1.04	1.94
23	1.35	1.54	1.26	1.66	1.16	1.78	1.07	1.92
24	1.36	1.54	1.27	1.65	1.18	1.77	1.09	1.90
25	1.37	1.55	1.29	1.65	1.20	1.76	1.12	1.88
26	1.38	1.55	1.30	1.65	1.22	1.76	1.14	1.87
27	1.39	1.55	1.31	1.65	1.24	1.75	1.16	1.86
28	1.40	1.56	1.33	1.65	1.25	1.74	1.17	1.85
29	1.41	1.56	1.34	1.65	1.27	1.74	1.19	1.84
30	1.42	1.56	1.35	1.65	1.28	1.74	1.21	1.83
31	1.43	1.57	1.36	1.65	1.29	1.73	1.22	1.82
32	1.44	1.57	1.37	1.65	1.31	1.73	1.24	1.82
33	1.44	1.57	1.38	1.65	1.32	1.73	1.25	1.81
34	1.45	1.58	1.39	1.65	1.33	1.73	1.27	1.81
35	1.46	1.58	1.40	1.65	1.34	1.72	1.28	1.80
36	1.47	1.59	1.41	1.65	1.35	1.72	1.29	1.80
37	1.47	1.59	1.42	1.65	1.36	1.72	1.30	1.79
38	1.48	1.59	1.43	1.65	1.37	1.72	1.31	1.79
39	1.49	1.60	1.43	1.66	1.38	1.72	1.32	1.79
40	1.49	1.60	1.44	1.66	1.39	1.72	1.33	1.78
45	1.52	1.61	1.47	1.67	1.43	1.72	1.38	1.78
50	1.54	1.63	1.50	1.67	1.46	1.72	1.42	1.77
55	1.56	1.64	1.53	1.68	1.49	1.72	1.45	1.77
60	1.58	1.65	1.55	1.69	1.51	1.73	1.48	1.77
65	1.60	1.66	1.57	1.70	1.53	1.73	1.50	1.77
70	1.61	1.67	1.58	1.70	1.55	1.73	1.52	1.77
75	1.62	1.68	1.60	1.71	1.57	1.74	1.54	1.77
80	1.64	1.69	1.61	1.72	1.59	1.74	1.56	1.77
90	1.66	1.70	1.63	1.73	1.61	1.75	1.59	1.78
100	1.67	1.71	1.65	1.74	1.63	1.76	1.61	1.78

Rounds to the 5 % significance points of the
g test when an overall mean and linear time
trend are included in the regression

TABLE 8

n	k=3		k=4		k=5		k=6	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.23	1.54	1.09	1.75	0.96	1.98	0.82	2.22
20	1.31	1.54	1.21	1.68	1.10	1.83	1.00	1.99
30	1.42	1.57	1.35	1.65	1.28	1.74	1.21	1.83
40	1.49	1.60	1.44	1.66	1.39	1.72	1.34	1.79
50	1.54	1.63	1.50	1.67	1.46	1.72	1.42	1.77

Rounds to the 5 % significance points of the
 d_n test when an overall mean and linear time
trend are included in the regression

TABLE 9

and linear time trend are fitted the narrower bounds given in table 8 are relevant - with the corresponding smaller probability of an inconclusive test. It is interesting to note from table 8 that the effect of including the linear time trend in $\tilde{X}_0$ is to increase the lower bound, with the upper bound being only slightly decreased. Table 10 gives the probabilities of a significant result (i.e. the effective powers) of the g bounds test with the PEARS DATA, comparing the use of the Durbin-Watson lower bound to the use of the table 8 lower bound.

For completeness we give some results for the bounds of the $g(-1)$ statistic. It can be easily verified that the latent roots of $\tilde{V}^{-1}(-1)$ are identical to those of $\tilde{B}$, although the latent vectors are different, see section 4.5. Therefore, for $r = 0$, the bounds of $g(-1)$ and g are identical. The bounds for $g(-1)$ and g for the cases $r = 1, 2$ above are not identical, but may be derived using the same methods after choosing the appropriate $\tilde{H}$ and calculating the latent roots of $(\tilde{H}' \tilde{V}(-1) \tilde{H})^{-1}$.

The results and methods of this section are given for the particular case where $s = 1$, i.e. the observations form a single uninterrupted sequence from the same subject. Of course, as in the case of the Williams data, when $s > 1$ the methods are still relevant, but, for particular values of n , the bounds will not be the same for different values of s .

	Value of the autocorrelation coefficient ρ					
	0.0	0.5	0.7	0.8	0.9	1.0
$P(g < .74)$	.0013	.0289	.0774	.1178	.1644	.1905
$P(g < .98)$	.0099	.1159	.2262	.2945	.3603	.3931

This table presents the probabilities, for given values of ρ , of obtaining a conclusive significant result with the g test on the Pears' data using the Durbin-Watson lower bound (.74) and the table 8 lower bound (.98).

TABLE 10

2.10 Testing for second and higher order autocorrelations

We have adopted the first-order autocorrelated error process in an attempt to represent the effects of carry-over and unspecified trends by use of a simple autocorrelation structure. Second and higher order autocorrelation processes might of course provide a closer representation of some of the actual error structures encountered in practice. It is of interest to see if these higher order processes are amenable to analysis.

The general r^{th} order autocorrelation scheme is defined in section 1.3. Correspondingly, we generalize the definition of the statistic g :

$$g(e_1^*, e_2^*, \dots, e_r^*) = \frac{\hat{\tilde{u}}' \tilde{V}^{-1} \hat{\tilde{u}}}{\tilde{z}' \tilde{z}} \quad , \quad (1)$$

where $\tilde{V}^{-1} = \tilde{V}^{-1}(e_1^*, e_2^*, \dots, e_r^*)$, from (1.3.4)

$$\tilde{\beta} = (\tilde{X}' \tilde{V}^{-1} \tilde{X})^{-1} \tilde{X}' \tilde{V}^{-1} \tilde{y} \quad ,$$

$$\hat{\tilde{u}} = \tilde{y} - \tilde{X} \tilde{\beta} \quad ,$$

and $\tilde{z}$ is the vector of least squares residuals.

However, we shall begin by considering the simple r^{th} order scheme where, for one subject ($s = 1$) ,

$$u_i = \rho_r u_{i-r} + \varepsilon_i \quad , \quad (i = 1, 2, \dots, n) \quad (2)$$

and, as before, take $\varepsilon \sim N(0, \sigma^2 I_n)$. A stationary process would have again been obtained by assuming $\text{var}(u_i) = \sigma^2$.

Writing the hypotheses to be tested as

$$H_0 : \underset{\sim}{u} \sim N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I}_n)$$

and

$$H_1 : \underset{\sim}{u} \sim N\{\underset{\sim}{0}, \sigma^2 \underset{\sim}{V}(0, 0, \dots, 0, \rho_r^*)\} \quad , \quad (3)$$

it is easily seen, using the methods of section 2.5, that the test given by $g(0, 0, \dots, 0, 1)$ is a most powerful invariant test in the region $\rho_r \rightarrow 1$, for the stationary as well as the non-stationary alternative. Again, when the regression vectors are a linear combination of k latent vectors of $\underset{\sim}{V}^{-1}(0, 0, \dots, 0, 1)$, the statistic gives an UMP invariant test. The reasons for the similar properties of the tests given by $g(1)$ and $g(0, \dots, 0, 1)$ become clear when the implications of reordering the observations are considered. If for instance in the Williams example, the observations are renumbered in order across the rows (rather than down subjects) the first order autocorrelation model on six subjects becomes a simple sixth order autocorrelation on one subject.

Wallis (1972) adopts the Durbin-Watson test to test for simple fourth order autocorrelation. The results we give for the d_A test in the case of the Williams data correspond, in fact, to the Wallis version of the Durbin-Watson statistic. The optimum power property of the Durbin-Watson-Wallis test is that, for a close approximation to the stationary density function, it is an UMP invariant test when the k regression vectors are a linear combination of k latent vectors of the appropriate version of the matrix $\underset{\sim}{A}$, under the condition that the latent vectors associated with the zero latent roots are included as regressors. Of course, this property is again

inherent in the $g(0, \dots, 0, 1)$ test, and the bounds for the g statistic under the above conditions are identical to those published by Wallis (1972) in Table 2.

If, in the hypotheses given by (3), the value of ρ_r^* was negative then a test given by the statistic $g(0, \dots, 0, -1)$ produces properties analogous to those of $g(-1)$. Box (1960) considers, however, that higher-order autocorrelation coefficients are more likely to be positive due to the overriding influence of variabilities in trend effects, even in the carry-over situation where the first-order coefficient is negative. Thus we are led to an interesting question of what statistic should be used to test for positive second-order autocorrelation when ρ_1 is taken to be negative. On the other hand in economic applications where the first-order coefficient has been accepted as positive there is a similar problem. At this stage it is interesting to note that in the previous examples considered of alternative hypotheses of simple r^{th} -order autocorrelation, including the first-order case, the optimum values of the parameters of the statistic $g(\rho_1^*, \rho_1^*, \dots, \rho_r^*)$ have been at the limits of the autocorrelation coefficients in the corresponding stationary process, viz. $\rho_1^* = \dots = \rho_{r-1}^* = 0$ and $\rho_r^* = 1$. The limits of the second-order stationary process are given by R.M. Williams (1952) in the form of the inequalities

$$-1 < \rho_1 < 1$$

and

$$-(1 - \rho_1) < \rho_2 < 1 - \rho_1.$$

Thus the point $(\rho_1^*, \rho_2^*) = (0, 1)$ is on the boundary, while the points given by $(1, 1)$ and $(-1, 1)$ lie outside the stationarity region.

The theory of invariance gives the result that the test given by a statistic of the form $g(e_1^*, e_2^*)$ is a most powerful invariant test for the hypotheses defined as

$$H_0 : \underset{\sim}{u} \sim N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I}_n) ,$$

$$H_1 : \underset{\sim}{u} \sim N\{\underset{\sim}{0}, \sigma^2 \underset{\sim}{V}(e_1^*, e_2^*)\} .$$

Therefore, if e_1^* and e_2^* are known, even approximately, a priori, these values should be used in the test statistic. In general the values of the coefficients are unknown and the problem of choice of the test statistic parameters is unsolved. In an attempt to throw some light on the matter we have computed powers of the tests given by the statistics $g(-1, 1)$, $g(0, 1)$ and $g(-.5, .5)$ against various non-stationary alternatives for the design matrix representing the Williams model. These are presented in table 11. The size of the tests was 0.05. After appropriate definition of the matrix $\underset{\sim}{V}(e_1, \dots, e_r)$, and hence $\underset{\sim}{W}$, the numerical methods presented in section 2.7 for the first-order model are applicable to the more general r^{th} order model. It is of interest to note that the test given by $g(0, 1)$ performs well, even for the alternatives $(-1, 1)$ and $(-.7, .3)$.

Statistic	Critical region	The values of the coefficients (c_1, c_2)				
		(-1,1)	(0,1)	(-.5,.5)	(-.7,.3)	(-.3,.7)
$g(-1,1)$	< 1.625	1.000	.797	.886	.914	.860
$g(0,1)$	< 1.100	1.000	.890	.912	.929	.900
$g(-.5,.5)$	< 0.889	1.000	.844	.935	.957	.910

The powers of the tests given by $g(-1,1)$, $g(0,1)$ and $g(-.5,.5)$
for certain second-order alternative hypotheses,
using the Williams' design.

TABLE II

PART III

ESTIMATION PROCEDURES

3.1 The inefficiency of least squares estimation

For a given value of the autocorrelation coefficient it is well known, e.g. Anderson (1971), that the best linear unbiased estimates of the parameters are given by

$$\tilde{\beta}(e) = (\tilde{X}' \tilde{V}^{-1} \tilde{X})^{-1} \tilde{X}' \tilde{V}^{-1} \tilde{y} \quad (1)$$

The covariance matrix of this estimator is $(\tilde{X}' \tilde{V}^{-1} \tilde{X})^{-1}$, and the generalized variance is defined by Rao (1965a) as $|\tilde{X}' \tilde{V}^{-1} \tilde{X}|^{-1}$. The least squares estimator $\hat{\beta} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' \tilde{y}$ is not the same as $\tilde{\beta}(e)$ unless $e = 0$ or in the very special circumstances that $\tilde{X}$ can be written as a linear combination of k latent vectors of $\tilde{V}$. Thus, in general, the least squares estimator is not the best linear unbiased estimator. The usual method of comparing estimators of β is by means of the criterion of efficiency, defined in terms of $\hat{\beta}$ and $\tilde{\beta}(e)$ as the ratio of their generalized variances:-

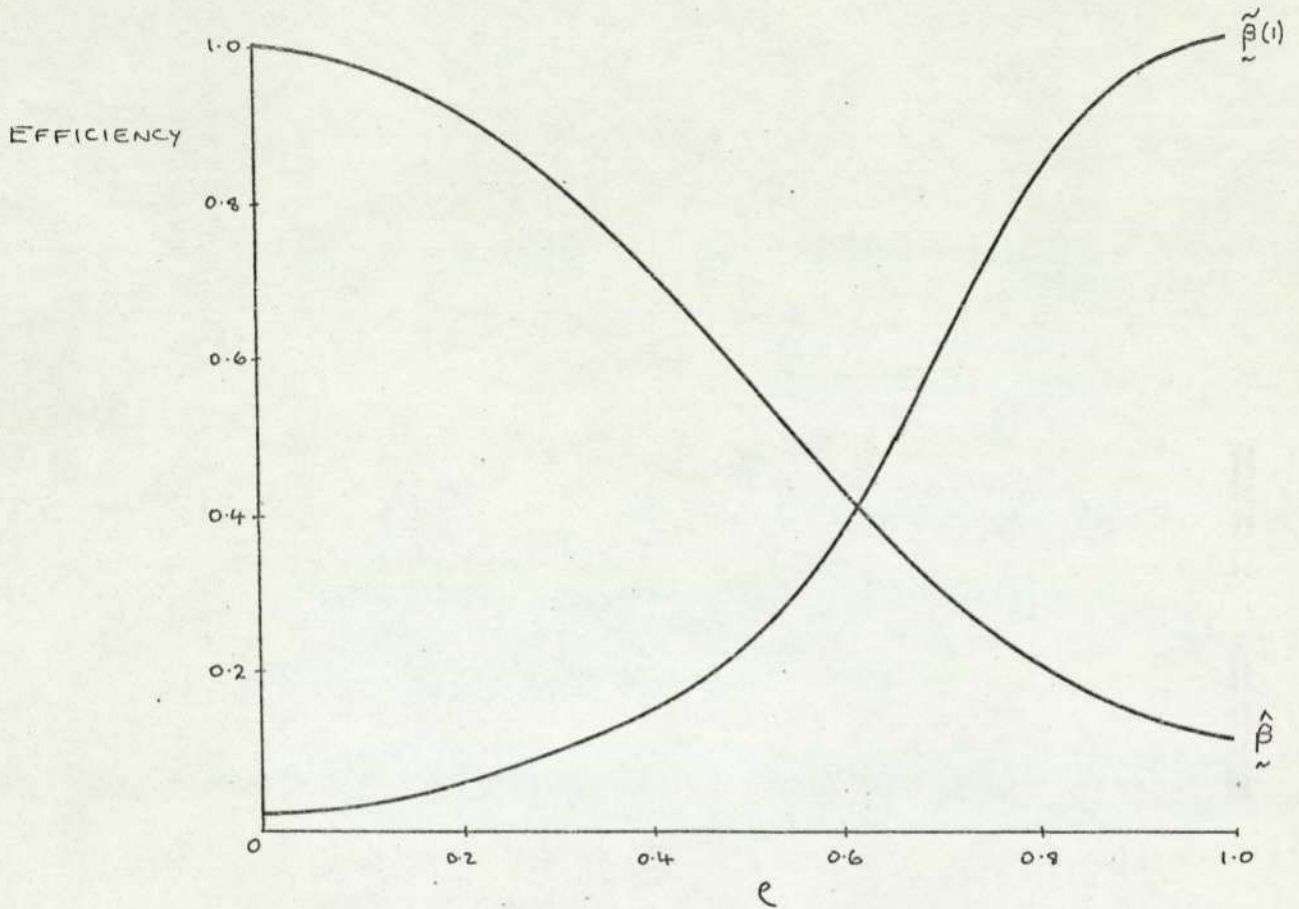
$$\text{Eff}(\hat{\beta}) = \frac{|\tilde{X}' \tilde{X}|^2}{|\tilde{X}' \tilde{V} \tilde{X}| \cdot |\tilde{X}' \tilde{V}^{-1} \tilde{X}|} \quad (2)$$

for example, given by Watson (1955). In the case where $\tilde{X} = \tilde{L}_k \tilde{G}$, when $\tilde{L}_k$ is an $n \times k$ matrix of k latent vectors of $\tilde{V}$ and $\tilde{G}$ is a $k \times k$ non-singular matrix, the estimator $\hat{\beta} = \tilde{\beta}$ and so the least squares estimator is fully efficient, i.e. $\text{Eff}(\hat{\beta}) = 1$.

This condition is necessary and sufficient for the least squares estimator to be fully efficient, Watson (1967). The latent vectors

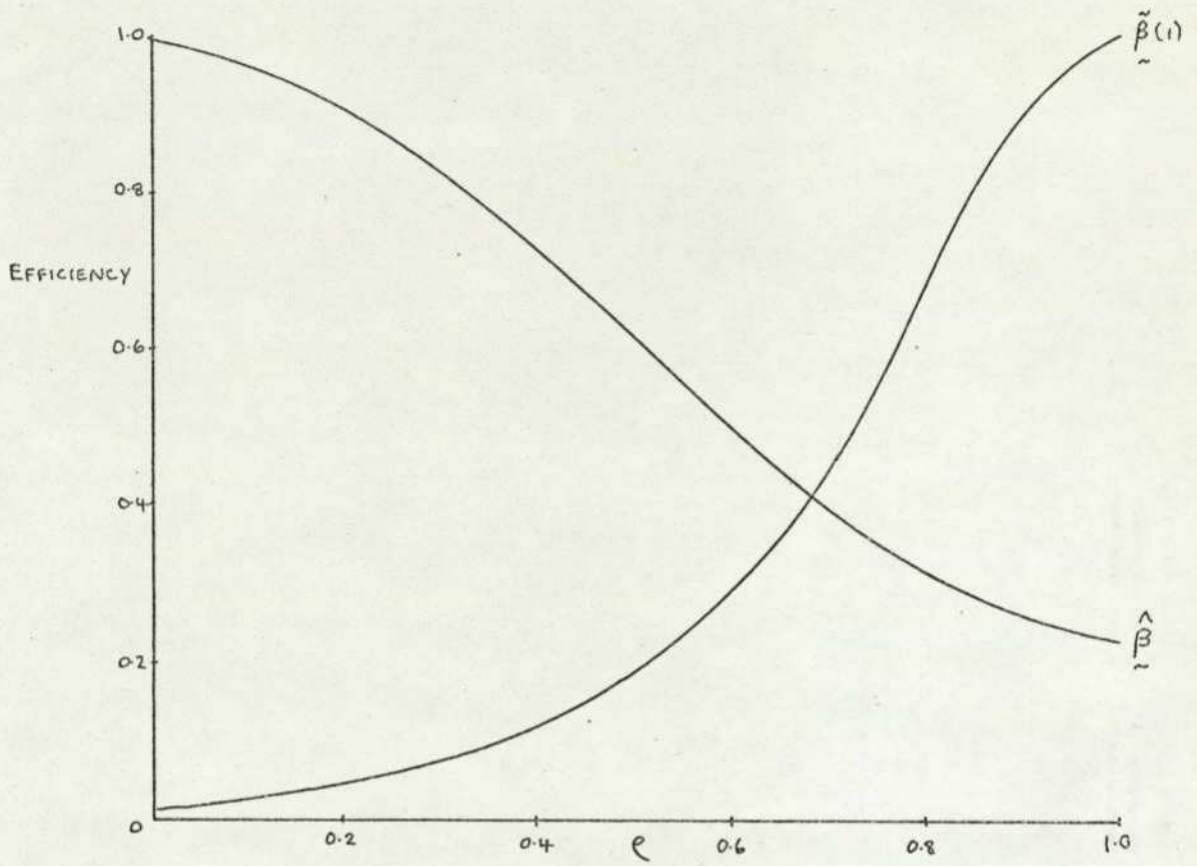
are seldom, in practice, the regression vectors or columns of the design matrix and so the question arises: how inefficient can the least squares estimates become? The problem of finding lower bounds to the efficiency of $\hat{\beta}$ is examined and reviewed by Anderson (1971) for a general dispersion matrix V . A valid, attainable lower bound has only been found for the case $k = 1$. For the three practical examples given in section 1.2 the efficiencies of the least squares estimates have been computed and are given in Figs. 3, 5, 7, where the autocorrelation is taken from the non-stationary process. In order to compare the results with the stationary process the efficiency of the least squares estimator for the PEARS DATA is illustrated in Fig. 4, and is seen to be very similar to the efficiency for the non-stationary alternative. It is evident from all the graphs that even for only moderately high values of ρ a serious loss in efficiency can result from using least squares estimates.

It is worth mentioning at this stage the case of polynomial regression. Hannan (1960) shows that least squares estimates are asymptotically efficient in polynomial regression, for the stationary process when $|\rho| < 1$. We have not been concerned in this thesis with asymptotic properties, but rather with the properties which are relevant to everyday sample sizes. Thus for the Box-Hunter polynomial regression, where $n = 21$, it is seen that least squares is certainly not fully efficient even for only moderately high values of ρ , in both the non-stationary process (fig. 5) and the stationary process (fig. 6).



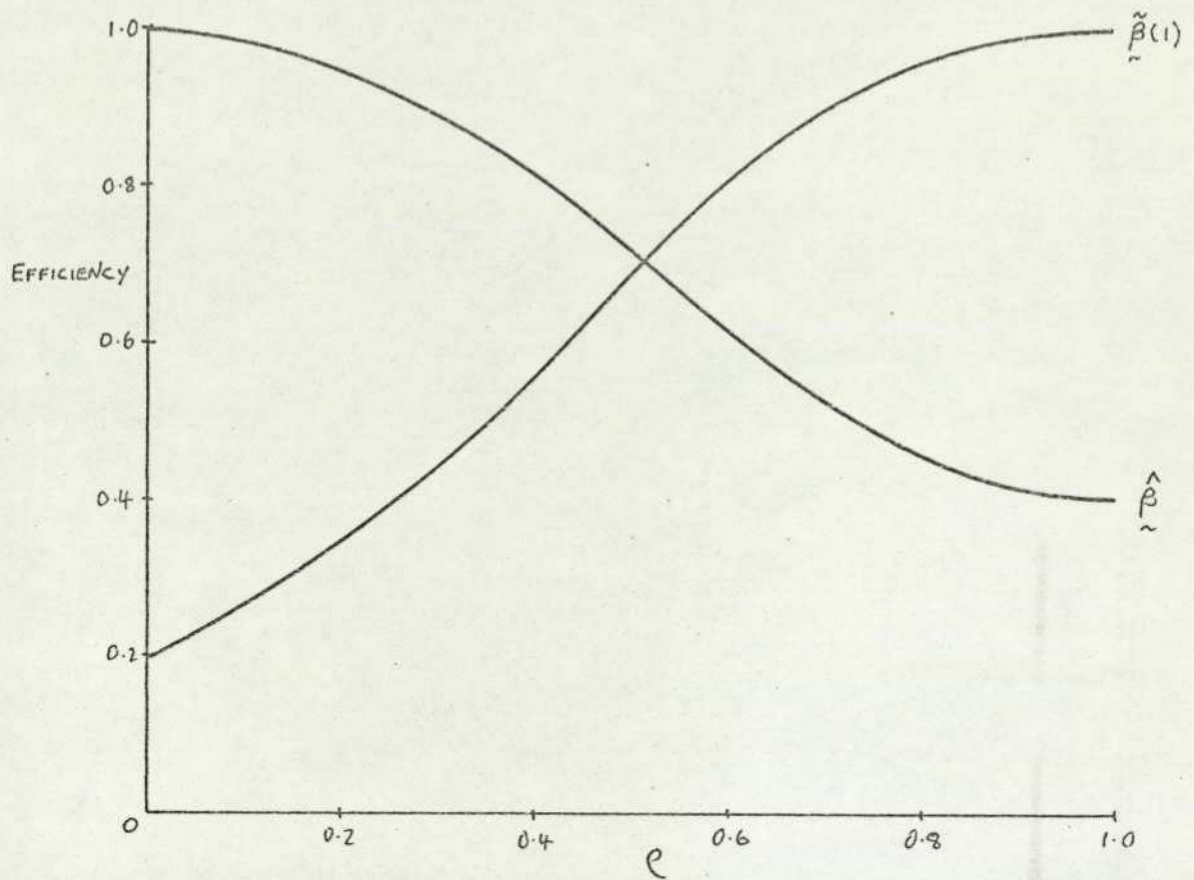
GRAPHS OF EFFICIENCY AGAINST ρ IN THE NON-STATIONARY PROCESS FOR THE PEARS DATA.

FIG.3



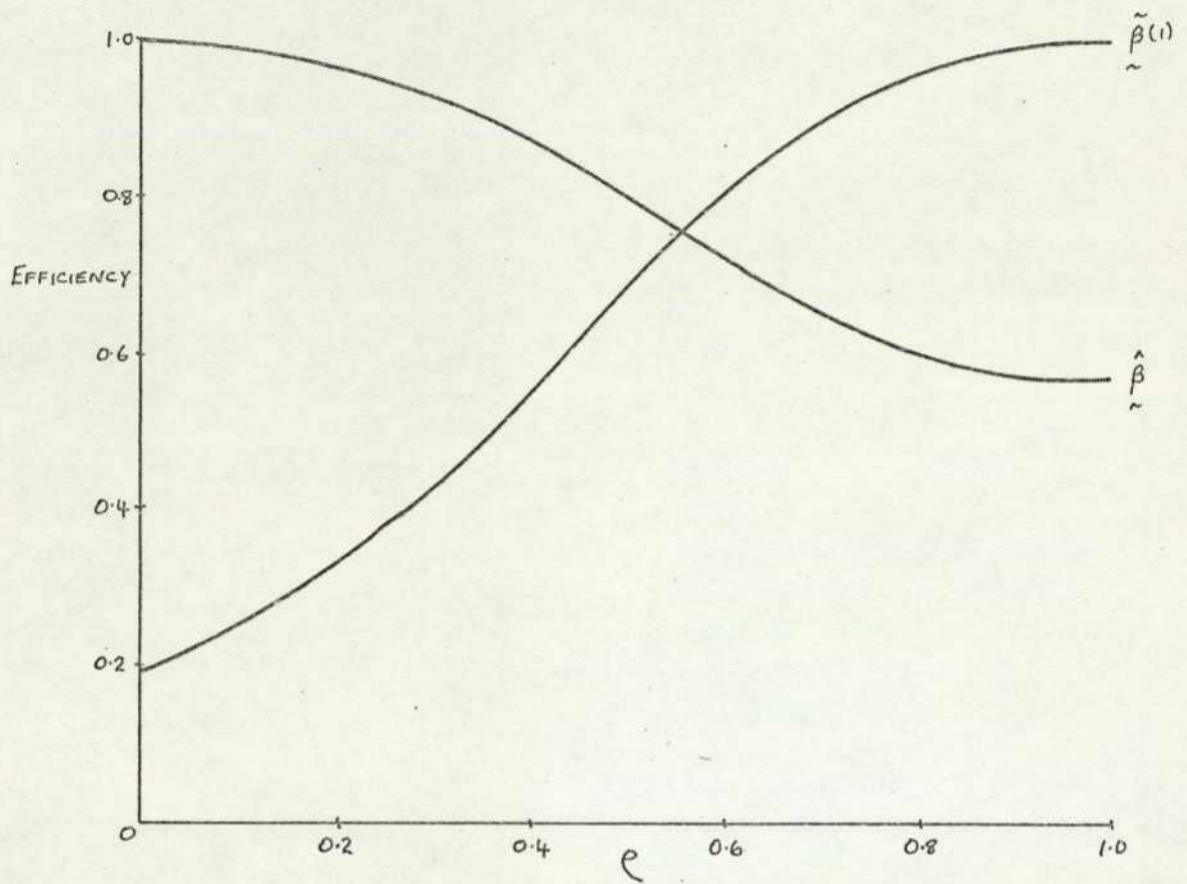
GRAPHS OF EFFICIENCY AGAINST ρ IN THE
STATIONARY PROCESS FOR THE PEAR'S DATA

FIG.4



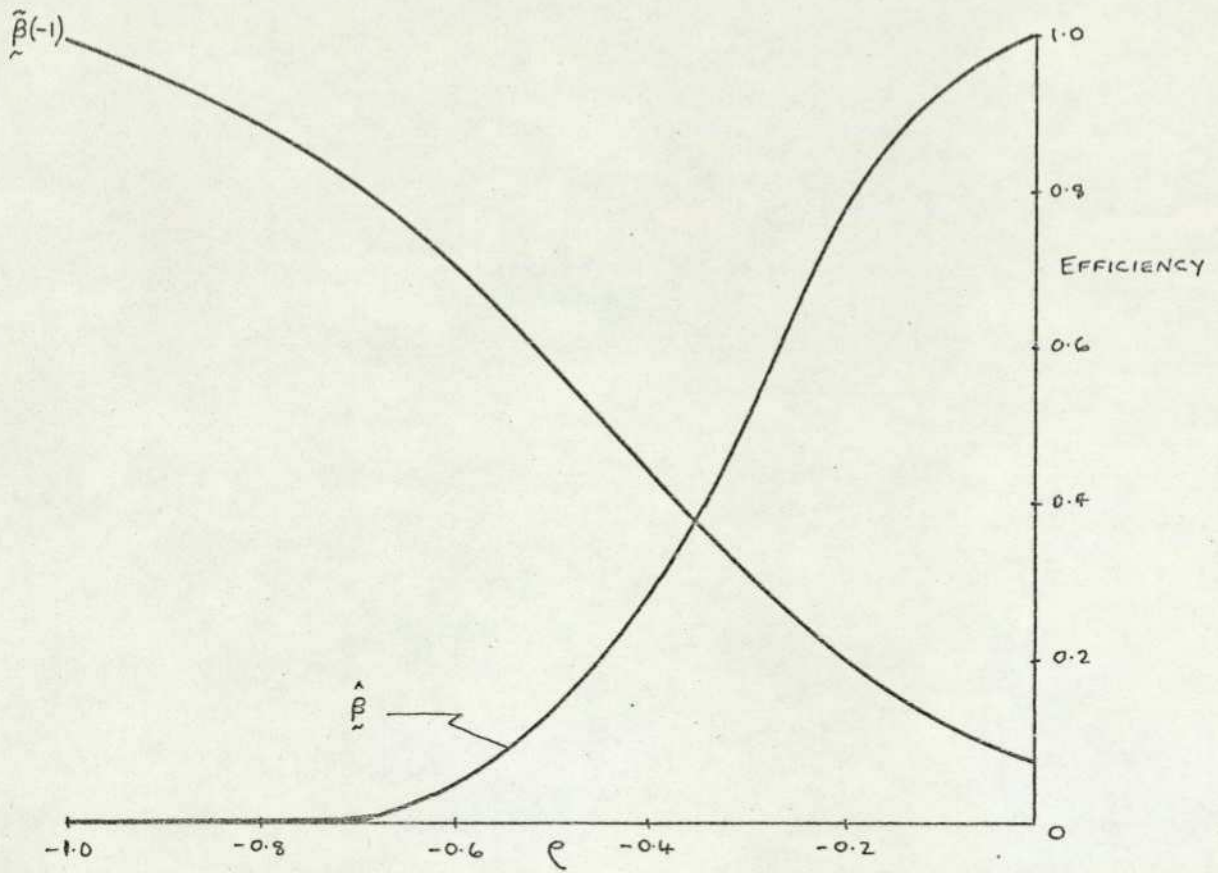
GRAPHS OF EFFICIENCY AGAINST ρ IN THE
NON-STATIONARY PROCESS FOR THE BOX-HUNTER DATA

FIG.5



GRAPHS OF EFFICIENCY AGAINST ρ IN THE
STATIONARY PROCESS FOR THE BOX-HUNTER DATA

FIG. 6



GRAPHS OF EFFICIENCY AGAINST ρ IN
THE NON-STATIONARY PROCESS FOR THE WILLIAMS' DATA

FIG.7

The danger of using least squares estimates in situations where there is a suspected autocorrelation structure in the errors is clear. Consequently the problem arises of what approach to use to ensure maximum efficiency for the estimates of the parameters β when ρ is unknown. An intuitive approach would be to estimate ρ by some means and then insert this value in equation (1) to obtain an estimate of β . The maximum likelihood procedure is basically this, although of course ρ and β are estimated simultaneously. Alternatively, a Bayesian method could be used - either to find ρ in the above procedure or to estimate β directly. A simpler method would be to base the estimate of ρ on the statistic g in the same way as Henshaw (1965) uses the d_A statistic.

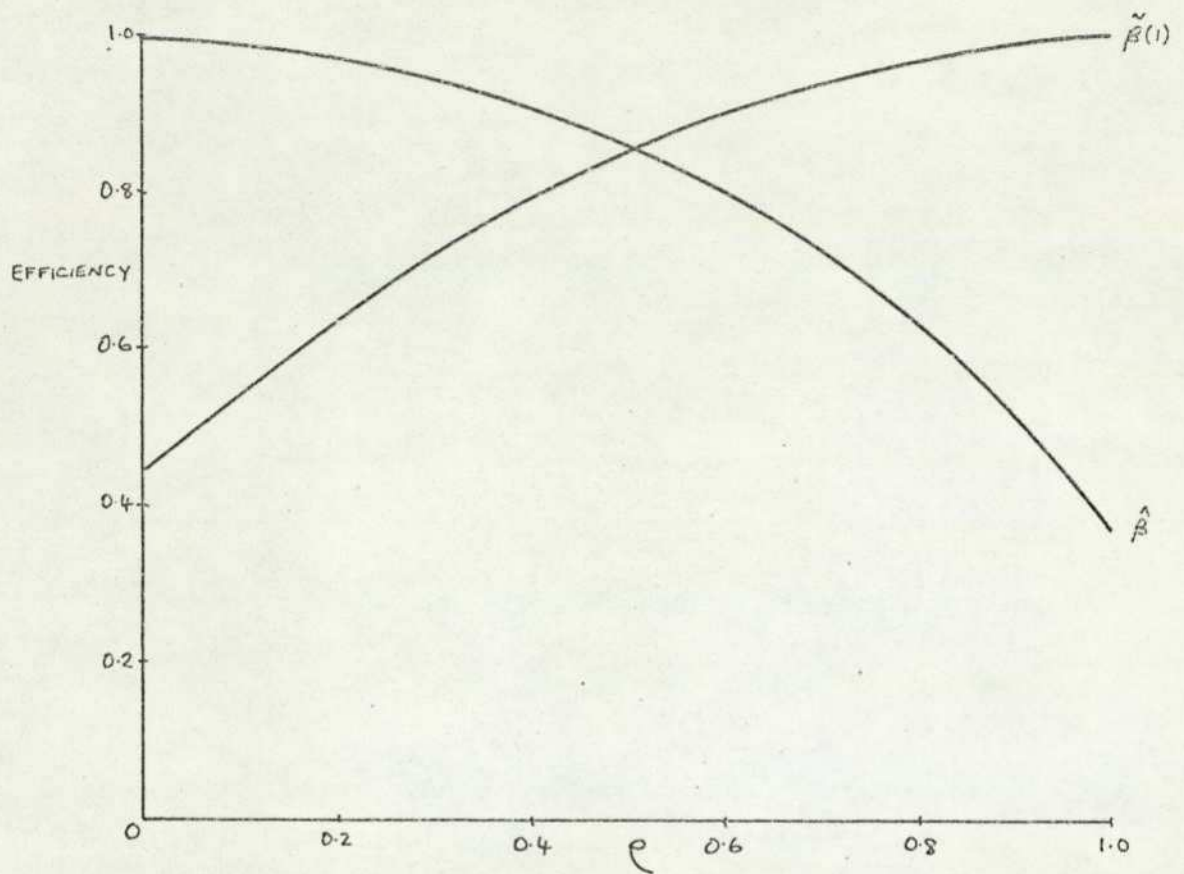
On the other hand, if the autocorrelation is in fact caused by some mis-specification of the model the problem should be tackled at this point. For instance, the effect of autocorrelation due to the existence of an unrecognized trend in the dependent variable may be eliminated by the inclusion in the model of the correct trend parameter. Three well known methods of dealing with the problem of autocorrelation include the use of trend terms, first differences and lagged dependent variables. However, Hildreth and Lu (1960) warn against the indiscriminate use of any particular one of these methods, saying that each situation must be treated on its merits. The inclusion of a quadratic trend in the design matrix of the PEARS DATA produces a g value which is not significant. But, even so, it is debatable which is the best procedure to adopt in situations of this type - adding the other trend parameter, using lagged variables, or even using the autocorrelation model.

However, the intention of this work is not to consider alternative forms of regression structure but to assume that the autocorrelation model adequately represents the data and to investigate various estimation procedures available under this assumption. The practical procedures we shall consider are not amenable to analytic enquiry and so in order to evaluate the properties of the various estimates we are led to use Monte-Carlo simulation techniques. Unfortunately the design matrices of the three examples given in section 1.2 were too big for simulation purposes from computer time considerations and so a much smaller design matrix was used. The model taken was

$$\underset{\sim}{y} = \underset{\sim}{X}\beta_1 + \underset{\sim}{u}$$

where $\underset{\sim}{X}' = (-1 -1 -1 1 1 1) \quad .$

This represents a simple experiment to evaluate the effect of a treatment applied at two levels where the response for the treatment applied at the mean level was known a priori. Here $n = 6$ and $k = 1$. The efficiency of the least squares estimate of β_1 is shown in Fig. 8. This design matrix was chosen in preference to the unit vector $\underset{\sim}{1}$ because the latter may not be a typical example as it is the latent vector associated with the zero latent root of $\underset{\sim}{A}$. Various methods of estimation are considered in the next four sections, and in section 3.6 the methods are compared by means of the Monte-Carlo simulations performed on the computer using the above design matrix.



GRAPHS OF EFFICIENCY AGAINST ρ IN
THE NON-STATIDNARY PROCESS FOR THE
SIMULATION MODEL WHERE $X' = (-1 -1 -1 \ 1 \ 1)$

FIG.8

This section is concluded by a closer examination of the efficiency curves presented in figs. 3-8. These curves confirm the intuitive thought that least squares estimation is highly efficient in the region of $\rho = 0$, but performs badly in the region $\rho \rightarrow 1$. Therefore, if the main concern is to guard against the possibility of using least squares when it is of low efficiency it is important to use a test of the null hypothesis which has high power for alternative hypotheses of high values of autocorrelation. Moreover, because least squares estimates are nearly fully efficient in the region of ρ around zero, a test for autocorrelated errors is not desirable if it has relatively highest probability of rejecting the null hypothesis in the region of $\rho = 0$. Therefore the property of being the locally most powerful invariant test in the region of the null hypothesis is not a recommendation for the d_A test.

3.2 Maximum likelihood estimation

The logarithm of the likelihood function of the multivariate normal density given by (2.2.5) is

$$\ln L = -n/2 \ln(2\pi) - n/2 \ln \sigma^2 - \frac{1}{2} \ln |V| - \frac{\tilde{u}' \tilde{V}^{-1} \tilde{u}}{2\sigma^2}, \quad (1)$$

where $\tilde{u} = \tilde{y} - \tilde{X} \tilde{\beta}$. Authors dealing with the stationary first-order autocorrelation scheme encounter a difficulty in that the determinant of their dispersion matrix is not independent of ρ .

In fact for the stationary case $|\underline{\underline{V}}| = (1 - \rho^2)^{n-s}$, where s is the number of subjects. Durbin and Watson (1971) page 15 indicate an approximate procedure for sufficiently large n . As mentioned in section 1.3, the non-stationary model has the very convenient property that $|\underline{\underline{V}}| = 1$ for all values of ρ , s and n , and so $\ln|\underline{\underline{V}}| = 0$.

The maximum likelihood estimates of the parameters $\underline{\underline{\beta}}$, σ^2 and ρ are given by simultaneous solution of the equations

$$\frac{\partial \ln L}{\partial \underline{\underline{\beta}}} = \frac{-\underline{\underline{X}}' \underline{\underline{V}}^{-1} (\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\beta}})}{\sigma^2} = \underline{\underline{0}}, \quad (2)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\beta}})' \underline{\underline{V}}^{-1} (\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\beta}})}{2\sigma^4} = 0, \quad (3)$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{-(\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\beta}})' \underline{\underline{D}}(\rho) (\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\beta}})}{2\sigma^2} = 0, \quad (4)$$

where

$$\underline{\underline{D}}(\rho) = \begin{pmatrix} 2\rho & -1 & 0 & \dots & 0 \\ -1 & 2\rho & -1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & 2\rho & -1 \\ 0 & \dots & \dots & -1 & 0 \end{pmatrix}.$$

Denoting the maximum likelihood estimates as $\underline{\underline{\tilde{\beta}}}$, $\underline{\underline{\tilde{\sigma}^2}}$ and $\underline{\underline{\tilde{\rho}}}$,

$$\text{then } \underline{\underline{\tilde{\beta}}} = (\underline{\underline{X}}' \underline{\underline{V}}^{-1} \underline{\underline{X}})^{-1} \underline{\underline{X}}' \underline{\underline{V}}^{-1} \underline{\underline{y}}, \quad (5)$$

$$\underline{\underline{\tilde{\sigma}^2}} = (\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\tilde{\beta}}})' \underline{\underline{V}}^{-1} (\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\tilde{\beta}}}) / n, \quad (6)$$

$$\text{and } \underline{\underline{\tilde{\rho}}} = \frac{(\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\tilde{\beta}}})' \underline{\underline{D}}(0) (\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\tilde{\beta}}})}{(\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\tilde{\beta}}})' \underline{\underline{E}} (\underline{\underline{y}} - \underline{\underline{X}} \underline{\underline{\tilde{\beta}}})}, \quad (7)$$

where $\tilde{V}^{-1} = V^{-1}(\tilde{\rho})$ and E is the diagonal matrix

$$\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & \ddots & \\ & & & 2 & \\ & & & & 0 \end{pmatrix}.$$

It is clear that the estimation of $\tilde{\beta}$ and $\tilde{\rho}$ is therefore independent of the estimation of σ^2 . To solve (5) and (7) $\tilde{\beta}$ is substituted in (7) and values of $\tilde{\rho}$ satisfying the resulting equation may be found using a computer. To decide whether any particular solution is a maximum or minimum the Hessian, $\tilde{H}$, can be evaluated. It may be verified that

$$\tilde{H} = \frac{-1}{2\sigma^2} \begin{pmatrix} 2\tilde{X}'\tilde{V}^{-1}\tilde{X} & -\tilde{X}'D(\tilde{\rho})(\tilde{y}-\tilde{X}\tilde{\beta}) & 0 \\ -(\tilde{y}-\tilde{X}\tilde{\beta})'D(\tilde{\rho})\tilde{X} & (\tilde{y}-\tilde{X}\tilde{\beta})'E(\tilde{y}-\tilde{X}\tilde{\beta}) & 0 \\ 0 & 0 & n/\sigma^2 \end{pmatrix},$$

and the condition for a maximum is that $|\tilde{H}| < 0$. However, because of the problem of possible multiple maxima it is illustrative to view the estimation of $\tilde{\beta}$ and $\tilde{\rho}$ from a slightly different angle. Write (2) and (4) as

$$\left. \frac{\partial \ln L(\tilde{\beta}, \tilde{\rho}, \sigma^2)}{\partial \tilde{\beta}} \right|_{\tilde{\rho}, \sigma^2} = 0$$

$$\text{and} \quad \left. \frac{\partial \ln L(\tilde{\beta}, \tilde{\rho}, \sigma^2)}{\partial \tilde{\rho}} \right|_{\tilde{\beta}, \sigma^2} = 0.$$

Instead of solving (2) and (4) simultaneously an equivalent approach is to take $\tilde{\beta}$ from (2) and then to solve for $\tilde{\rho}$ the equation

$$\left. \frac{\partial \ln L(\tilde{\beta}, \tilde{\rho}, \sigma^2)}{\partial \tilde{\rho}} \right|_{\sigma^2} = 0. \quad (8)$$

$$\text{Proof: } \left. \frac{\partial \ln L(\tilde{\beta}, e, \sigma^2)}{\partial e} \right]_{\sigma^2} = \left. \frac{\partial \ln L(\tilde{\beta}, e, \sigma^2)}{\partial \tilde{\beta}} \right]_{e, \sigma^2} \times \frac{d\tilde{\beta}}{de} + \left. \frac{\partial \ln L(\tilde{\beta}, e, \sigma^2)}{\partial e} \right]_{\tilde{\beta}, \sigma^2}$$

$$\text{Now } \left. \frac{\partial \ln L(\tilde{\beta}, e, \sigma^2)}{\partial \tilde{\beta}} \right]_{e, \sigma^2} = 0, \quad \text{from (2),}$$

$$\left. \frac{\partial \ln L(\tilde{\beta}, e, \sigma^2)}{\partial e} \right]_{\tilde{\beta}, \sigma^2} = 0, \quad \text{from (4),}$$

and as $\frac{d\tilde{\beta}}{de}$ is finite, the result is proved.

Substituting $\tilde{u} = y - X\tilde{\beta}$ in (1) it is seen that the maximum likelihood estimation of e may be regarded as minimization of the Generalized Residual Sum of Squares, where

$$\text{GRSS} = \tilde{u}' V^{-1} \tilde{u} = y' (V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}) y. \quad (9)$$

There is no guarantee of a unique minimum to this function, and hence of a unique maximum to the likelihood function. An example where there are two minima to the function GRSS is given in fig. 9. The maximum likelihood estimate is $\tilde{e} = -.45$, with another local maximum at $e = 1.02$. The data in this example was

$$\begin{aligned} \tilde{y} &= (-.70 \quad -.54 \quad .36 \quad -.98 \quad .44 \quad 2.65)' \text{ with design matrix} \\ \tilde{X} &= (-1 \quad -1 \quad -1 \quad 1 \quad 1 \quad 1)'. \end{aligned}$$

Hildreth and Lu (1960) use a similar method of maximum likelihood estimation, but their treatment of the first observation is different and they restrict the permitted range of $\tilde{e}$ to $(-1, 1)$ from 'economic considerations'. In general GRSS is the ratio of a polynomial of degree $2k + 2$ in e to a polynomial of

FIG. 9

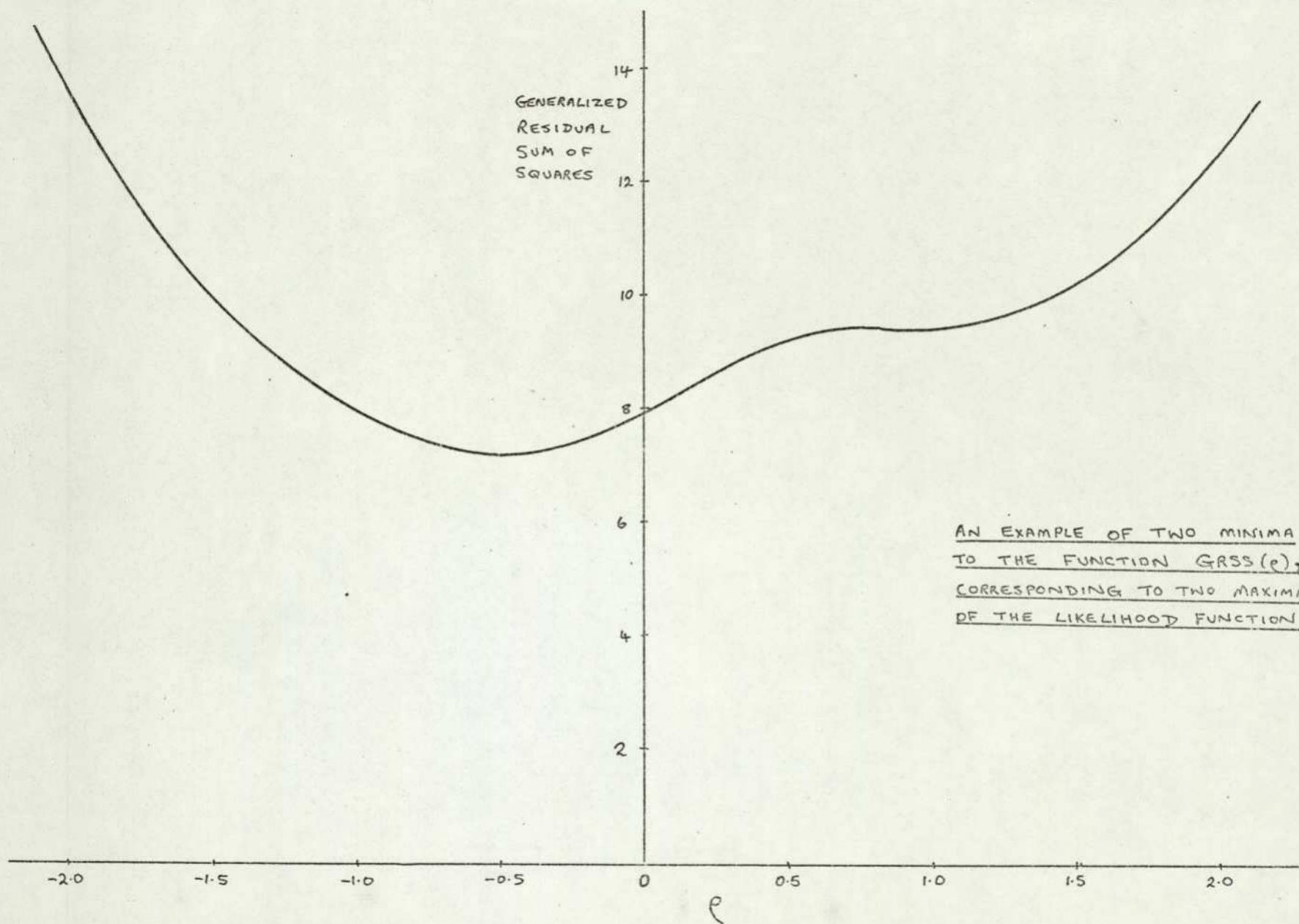
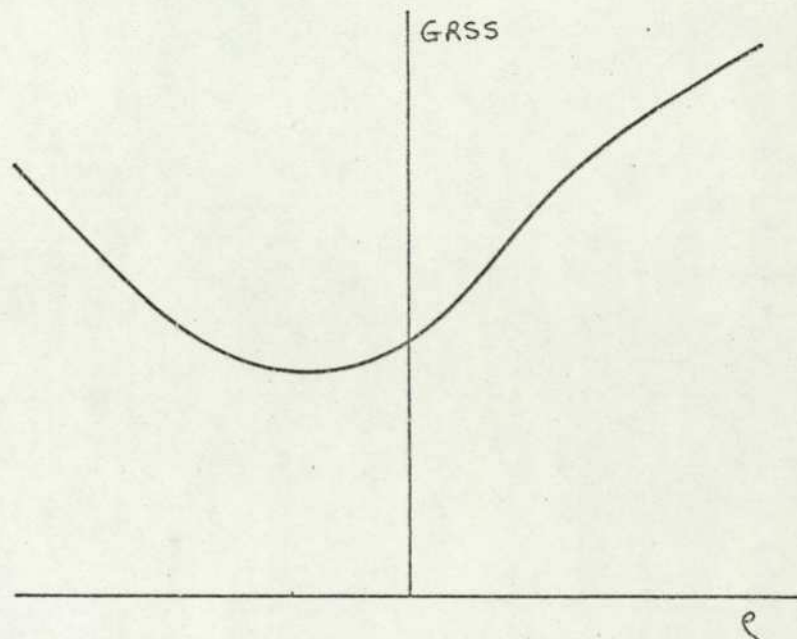
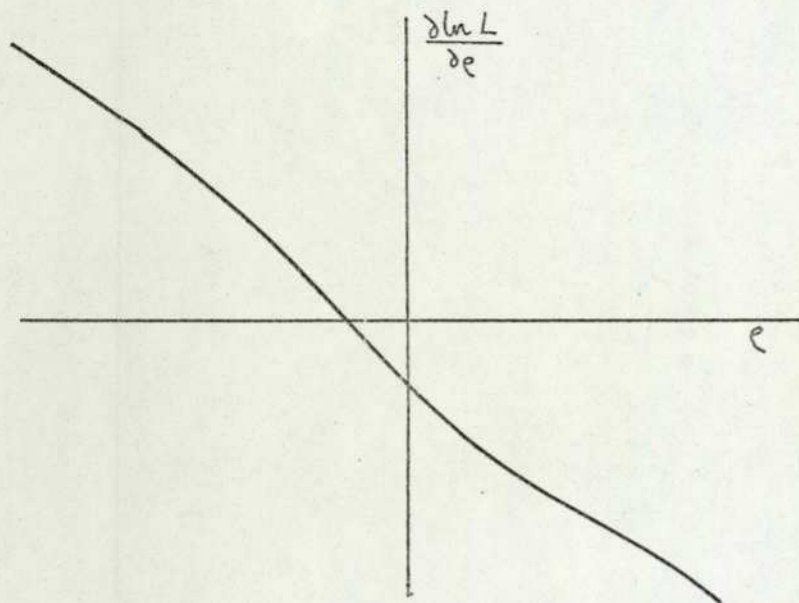


FIG.10



SKETCHES OF THE TWO FUNCTIONS USED
TO OBTAIN THE MAXIMUM LIKELIHOOD ESTIMATE OF ρ
FOR A TYPICAL DATA SET.

degree $2k$ in ϱ . In practice evaluation of the coefficients of ϱ in the expansion of GRSS becomes intractable for even the simplest design matrices such as $\tilde{X} = (-1 \ -1 \ -1 \ 1 \ 1 \ 1)'$. In order to gain some insight into the small sample properties of the maximum likelihood estimates the computer simulation with the above design matrix was performed. The results of this simulation are summarized in section 3.6. As there is no guarantee of a unique minimum the program evaluates the expression of the left hand side of (4) for various initial values of ϱ : in this case $-10, -1, -0.5, 0, 0.5, 1, 10$. This grid was chosen after close examination of 50 sets of simulation data. A typical sketch of the expression and its associated GRSS is given in fig. 10. As maxima of the likelihood function correspond to zeros of the function (4) with negative slope, only these zeros are evaluated accurately. If there is more than one such zero of (4) the value of ϱ associated with the smallest GRSS is the maximum likelihood estimate.

3.3 An estimator based on the g statistic

Theil and Nagar (1961) and Henshaw (1966) suggest the estimator $(\pi_1 + \pi_{n-k} - 2d_A)/(\pi_{n-k} - \pi_1)$ for ϱ , where π_1 is the smallest non-zero latent root of $\tilde{MAM}$ and π_{n-k} the largest. They also give expressions for the mean and variance of the estimate under the null hypothesis. This estimator has the advantage over the maximum likelihood estimator of being comparatively easy to evaluate.

An estimator based on a similar linear transformation of g to the range $(-1, 1)$ would be

$$\hat{e} = \frac{v_1 + v_{n-k} - 2g}{v_{n-k} - v_1},$$

where v_1 and v_{n-k} are the smallest and largest permissible values of g , defined as latent roots of $\tilde{W}$ in section 2.4. For the non-stationary autocorrelation scheme it is not possible of course to transform g to the actual range of e $(-\infty, \infty)$ with a linear transformation. The mean and variance of $\hat{e}$ are easily calculated from the mean and variance of g , and under the null hypothesis $H_0: e = 0$ this is straightforward using a similar method to that for the mean and variance of d_A , e.g. Durbin and Watson (1950). However as the region of interest is for high values of e , where good properties for the estimates are most important, we wish to obtain the mean and variance of g for high values of e .

Analytically no headway was made except in the case when e was 1. Then

$$g = \frac{\tilde{u}'\tilde{W}\tilde{u}}{\tilde{u}'\tilde{M}\tilde{u}} = \frac{\tilde{\epsilon}'\tilde{C}^{-1}\tilde{W}\tilde{C}^{-1}\tilde{\epsilon}}{\tilde{\epsilon}'\tilde{C}^{-1}\tilde{M}\tilde{C}^{-1}\tilde{\epsilon}},$$

where $\tilde{\epsilon} \sim N(0, \sigma^2 I)$ and $\tilde{C}^{-1} = \tilde{C}(1)^{-1}$. From (2.4.3) and (2.4.5) $\tilde{C}^{-1}\tilde{W}\tilde{C}^{-1} = \tilde{C}^{-1}\tilde{T}(\tilde{T}'\tilde{B}^{-1}\tilde{T})^{-1}\tilde{T}'\tilde{C}^{-1}$ and $\tilde{C}^{-1}\tilde{M}\tilde{C}^{-1} = \tilde{C}^{-1}\tilde{T}\tilde{T}'\tilde{C}^{-1}$. Now $\tilde{B} = \tilde{C}'\tilde{C}$ so that $\tilde{B}^{-1} = \tilde{C}^{-1}\tilde{C}^{-1}$. Thus $\tilde{C}^{-1}\tilde{W}\tilde{C}^{-1}$ and $\tilde{C}^{-1}\tilde{M}\tilde{C}^{-1}$ commute and therefore can be simultaneously reduced to diagonal form using the same orthonormal transformation, Bellman (1960) page 56.

Clearly the matrix $\tilde{C}^{-1} \tilde{W} \tilde{C}^{-1}$ is idempotent and has $n - k$ non-zero latent roots of unity. The non-zero latent roots of $\tilde{C}^{-1} \tilde{M} \tilde{C}^{-1}$ are the same as the latent roots of $\tilde{T}' \tilde{C}^{-1} \tilde{C}^{-1} \tilde{T} = \tilde{T}' \tilde{B}^{-1} \tilde{T}$. Thus when $\varrho = 0$ under H_0 :

$$g = \frac{\sum_{i=1}^{n-k} \nu_i \xi_i^2}{\sum_{i=1}^{n-k} \xi_i^2}, \quad (1)$$

and when $\varrho = 1$ under H_1 :

$$g = \frac{\sum_{i=1}^{n-k} \phi_i^2}{\sum_{i=1}^{n-k} \frac{1}{\nu_i} \phi_i^2}, \quad (2)$$

where the ν_i are the $n - k$ latent roots of $(\tilde{T}' \tilde{B}^{-1} \tilde{T})^{-1}$, and ξ, ϕ are both $N(0, I_{n-k})$. The moments of g in (1) under H_0 can be obtained as the ratio of the moments of numerator and denominator because g is distributed independently of its denominator, e.g. Hannan (1970) page 343. Unfortunately the distribution of g in (2) under H_1 is independent not of its denominator but of its numerator and the moments of the ratio cannot be written as the ratio of the moments of numerator and denominator.

In order to find the mean and variance of g and hence $\hat{\varrho}$ for any particular ϱ and design matrix it is necessary to resort to numerical techniques. Denoting the distribution function of g as $F(g)$, obtained from Imhof's (1961) result and section 2.7, the first two moments of g are given by

$$\mu_1' = \int_{y_1}^{y_{n-k}} g \frac{dF(g)}{dg} dg = y_{n-k} - \int_{y_1}^{y_{n-k}} F(g) dg ,$$

and

$$\mu_2' = \int_{y_1}^{y_{n-k}} g^2 \frac{dF(g)}{dg} dg = y_{n-k}^2 - 2 \int_{y_1}^{y_{n-k}} gF(g) dg .$$

The mean and variance of $\hat{e}$ for various actual values of e and for the design matrix $(-1 \ -1 \ -1 \ 1 \ 1 \ 1)'$ are given in table 12.

3.4 A Bayesian approach to the estimation of the parameters

In order to make inferences about e or β all that is required are the respective posterior distributions, $p(e|y)$ or $p(\beta|y)$. An estimate of β could either be the mean of $p(\beta|y)$ or alternatively could be taken from (3.1.1) after obtaining an estimate of e from $p(e|y)$. Zellner and Tiao (1964) also considered the non-stationary model from the Bayesian viewpoint although, being concerned with economic time series in the same way as Hildreth and Lu (1960), they assumed the presence of an observation y_0 taken prior in time to the first observation y_1 . We have not made this assumption as in experimental design applications it seems unrealistic.

As before, the likelihood of the non-stationary process is

$$p(\underline{y} | \underline{\beta}, \sigma^2, \underline{e}) = (2\pi\sigma^2)^{-n/2} \cdot \exp \left\{ \frac{-(\underline{y} - \underline{X}\underline{\beta})' \underline{V}^{-1} (\underline{y} - \underline{X}\underline{\beta})}{2\sigma^2} \right\}.$$

In Bayesian terminology, by extending the conversation so that

$$p(\underline{e} | \underline{y}) = \int_{\sigma^2} p(\underline{e} | \underline{y}, \sigma^2) \cdot p(\sigma^2 | \underline{y}) \cdot d\sigma^2$$

$$\text{and } p(\underline{e} | \underline{y}, \sigma^2) = \int_{\underline{\beta}} p(\underline{e} | \underline{y}, \sigma^2, \underline{\beta}) \cdot p(\underline{\beta} | \underline{y}, \sigma^2) \cdot d\underline{\beta},$$

and using Bayes theorem it can be easily verified that

$$p(\underline{e} | \underline{y}) = K_1 \int_{\sigma^2=0}^{\infty} \int_{\underline{\beta}=-\infty}^{\infty} p(\underline{y} | \underline{e}, \sigma^2, \underline{\beta}) p(\underline{e} | \sigma^2, \underline{\beta}) p(\underline{\beta} | \sigma^2) p(\sigma^2) d\underline{\beta} d\sigma^2.$$

Assuming vague prior knowledge of σ^2 , $\underline{\beta}$ and $\underline{e}$, and that knowledge of any would not influence our knowledge of the others, the prior distributions are taken as, Jeffreys (1948),

$$p(\sigma^2) d\sigma^2 \propto \frac{d\sigma^2}{\sigma^2},$$

$$p(\underline{\beta} | \sigma^2) d\underline{\beta} \propto d\underline{\beta},$$

$$\text{and } p(\underline{e} | \sigma^2, \underline{\beta}) d\underline{e} \propto d\underline{e}.$$

Noting that $(\underline{y} - \underline{X}\underline{\beta})' \underline{V}^{-1} (\underline{y} - \underline{X}\underline{\beta}) = (\underline{y} - \underline{X}\underline{\tilde{\beta}})' \underline{V}^{-1} (\underline{y} - \underline{X}\underline{\tilde{\beta}}) + (\underline{\tilde{\beta}} - \underline{\beta})' \underline{X}' \underline{V}^{-1} \underline{X} (\underline{\tilde{\beta}} - \underline{\beta})$,
where $\underline{\tilde{\beta}} = (\underline{X}' \underline{V}^{-1} \underline{X})^{-1} \underline{X}' \underline{V}^{-1} \underline{y}$, then

$$\int_{\underline{\beta}} p(\underline{y} | \underline{e}, \sigma^2, \underline{\beta}) d\underline{\beta} \propto \exp \left\{ \frac{-(\underline{y} - \underline{X}\underline{\tilde{\beta}})' \underline{V}^{-1} (\underline{y} - \underline{X}\underline{\tilde{\beta}})}{2\sigma^2} \right\} \cdot (\sigma^2)^{-n/2 + k/2 - 1} \cdot |\underline{X}' \underline{V}^{-1} \underline{X}|^{-1/2},$$

from use of the fact that $\underline{\tilde{\beta}} \sim N\{\underline{\beta}, \sigma^2 (\underline{X}' \underline{V}^{-1} \underline{X})^{-1}\}$. The last integration with respect to σ^2 is straightforward and yields the

posterior distribution as

$$p(\underline{e}|\underline{y}) = K_2 \cdot |\underline{X}'\underline{V}^{-1}\underline{X}|^{-1/2} \cdot \left\{ (\underline{y}-\underline{X}\underline{\beta})' \underline{V}^{-1} (\underline{y}-\underline{X}\underline{\beta}) \right\}^{-n/2 + k/2} \quad (1)$$

The mean of the posterior (1) is taken as an estimate of $\underline{e}$, and this estimate may be substituted in equation (3.1.1) to provide an estimate of $\underline{\beta}$. The performance of these estimators in the simulations is reported in section 3.6.

To make inferences about $\underline{\beta}$ directly the posterior $p(\underline{\beta}|\underline{y})$ should be used. This can be inferred from the expression for $p(\underline{e}|\underline{y})$ and is

$$p(\underline{\beta}|\underline{y}) = K_3 \int_{\underline{\sigma}^2} \int_{\underline{e}} p(\underline{y}|\underline{\beta}, \underline{\sigma}^2, \underline{e}) p(\underline{\beta}|\underline{\sigma}^2, \underline{e}) p(\underline{e}|\underline{\sigma}^2) p(\underline{\sigma}^2) d\underline{e} d\underline{\sigma}^2.$$

Again assuming vague prior knowledge of $\underline{\beta}$, $\underline{\sigma}^2$ and $\underline{e}$,

$$p(\underline{\beta}|\underline{\sigma}^2, \underline{e}) d\underline{\beta} \propto d\underline{\beta}$$

$$p(\underline{e}|\underline{\sigma}^2) d\underline{e} \propto d\underline{e}$$

$$\text{and} \quad p(\underline{\sigma}^2) d\underline{\sigma}^2 \propto \frac{d\underline{\sigma}^2}{\underline{\sigma}^2}.$$

$$\text{Thus } p(\underline{\beta}|\underline{y}) = K_4 \int_{\underline{\sigma}^2=0}^{\infty} (\underline{\sigma}^2)^{-n/2} \int_{\underline{e}=-\infty}^{\infty} \exp\left(\frac{-\underline{u}'\underline{V}^{-1}\underline{u}}{2\underline{\sigma}^2}\right) d\underline{e} d\underline{\sigma}^2,$$

$$\text{where } \underline{u} = \underline{y} - \underline{X}\underline{\beta}, \text{ and } \underline{u}'\underline{V}^{-1}\underline{u} = \left(\sum_1^{n-1} u_i^2\right) \underline{e}^2 - \left(2 \sum_1^{n-1} u_i u_{i+1}\right) \underline{e} + \sum_1^n u_i^2.$$

On completing the square in the exponential it may be verified that

$$\int_{-\infty}^{\infty} \exp\left(\frac{-\underline{u}'\underline{V}^{-1}\underline{u}}{2\underline{\sigma}^2}\right) d\underline{e} \propto \exp\left(\frac{-N}{2\underline{\sigma}^2}\right) \cdot \left(\frac{\underline{\sigma}^2}{\sum_1^{n-1} u_i^2}\right)^{1/2},$$

where
$$N = \sum_{i=1}^n u_i^2 - \frac{\left(\sum_{i=1}^{n-1} u_i u_{i+1} \right)^2}{\sum_{i=1}^{n-1} u_i^2} .$$
 After the

integration by σ^2 the posterior becomes

$$p(\beta | y) = K_4 \left(\frac{1}{\sum_{i=1}^{n-1} u_i^2} \right)^{\frac{1}{2}} \left\{ \frac{\sum_{i=1}^n u_i^2 \cdot \sum_{i=1}^{n-1} u_i^2 - \left(\sum_{i=1}^{n-1} u_i u_{i+1} \right)^2}{\sum_{i=1}^{n-1} u_i^2} \right\}^{-\frac{(n-1)}{2}} . \quad (2)$$

In general, to estimate β using the posterior mean entails two k -fold integrations. This is certainly the most difficult estimation procedure considered and is not likely to be used in practice. In order to observe its performance as an estimator it is included in the simulation, where for $X = (-1 \ -1 \ -1 \ 1 \ 1 \ 1)'$, $k = 1$.

This section is completed by examination of the computing procedures necessary to obtain the estimates of e and β . Explicitly the estimate of e from (1) is

$$e_* = \frac{\int_{-\infty}^{\infty} e |X' V^{-1} X|^{-\frac{1}{2}} \left\{ (y - X\beta)' V^{-1} (y - X\beta) \right\}^{-n/2 + k/2} de}{\int_{-\infty}^{\infty} |X' V^{-1} X|^{-\frac{1}{2}} \left\{ (y - X\beta)' V^{-1} (y - X\beta) \right\}^{-n/2 + k/2} de} .$$

To overcome the difficulty of the limits of integration the transformation $e = \tan t$ is made, so that $de = \sec^2 t \, dt$ and the limits of integration become $\pm \pi/2$. The term in the braces is the GRSS. From section 3.2 this is the ratio of a polynomial of degree $2k + 2$ to one of degree $2k$ in e . Therefore the integrand approaches zero as $e \rightarrow \infty$ for $n - k \geq 2$. For the model in the simulations the estimate of β_1 is similarly obtained from (2). The same substitution is used, i.e. $\beta_1 = \tan t$, to transform the range of integration onto $\pm \pi/2$, and again the integrand becomes zero at the limits.

3.5 A double dichotomy

The Bayesian methods of the previous section would certainly not be regarded as easy to operate. From the graphs of the efficiencies in section 3.1, i.e. figs. 3-8, we have seen that least squares estimates are fairly efficient except for high values of ρ . If ρ is taken to be unity the best linear unbiased estimator of β is, from (3.1.1), $\tilde{\beta}(1) = (\tilde{X}'\tilde{B}\tilde{X})^{-1}\tilde{X}'\tilde{B}y$. This is similar to the Cochrane and Orcutt (1949) estimator, i.e. the least squares estimator using first differenced variables, but here the first observation and row of $\tilde{X}$ are retained. The efficiency of this estimator is shown on the graphs, figs. 3-8, and is seen to have an efficiency approaching unity for $\rho \rightarrow 1$ in both non-stationary and stationary alternatives. For any particular value of ρ the efficiency of $\tilde{\beta}(1)$ is defined as the ratio of the generalized variance of $\tilde{\beta}(\rho)$ to the generalized variance of $\tilde{\beta}(1)$, i.e.

$$\text{Eff}\{\tilde{\beta}(1)\} = \frac{|\tilde{X}'\tilde{V}^{-1}\tilde{X}|}{|\tilde{X}'\tilde{B}\tilde{X}|^{-1} \cdot |\tilde{X}'\tilde{B}\tilde{V}\tilde{B}\tilde{X}| \cdot |\tilde{X}'\tilde{B}\tilde{X}|^{-1}} = \frac{|\tilde{X}'\tilde{B}\tilde{X}|^2}{|\tilde{X}'\tilde{V}^{-1}\tilde{X}| \cdot |\tilde{X}'\tilde{B}\tilde{V}\tilde{B}\tilde{X}|}.$$

Clearly, by definition, the estimator $\tilde{\beta}(1)$ is fully efficient in the non-stationary process when $\rho = 1$. However, the reason why the estimator has optimum efficiency for the stationary process when $\rho \rightarrow 1$ is not so clear and we give a proof of the result

$\lim_{\rho \rightarrow 1} \text{Eff}\{\tilde{\beta}(1)\} = 1$ at the end of this section. The alternative method used to deal with high autocorrelation in the stationary process of taking first differences and discarding the first observation and row of the design matrix, e.g. Cochrane and Orcutt (1949), Stone (1948), is unsatisfactory because it cannot estimate an overall mean.

The suggestion of this section is to simplify the problem and treat it as a double dichotomy. The action to be taken is either to accept that ρ is approximately zero and use least squares, or to accept that ρ is high, in the region of one and use least squares on first differenced variables, i.e. $\tilde{\beta}(1)$. The two alternative decision are $D = (d_1, d_2)$ where

d_1 is the decision to use $\hat{\beta}$,
and d_2 is the decision to use $\tilde{\beta}(1)$.

It is clear from the efficiency curves plotted in figs. 3-8 that correctly deciding between $\hat{\beta}$ and $\tilde{\beta}(1)$ would be a considerable improvement over the direct procedure of using least squares.

Nevertheless, it will be seen that there is a region between $\rho = 0$ and $\rho = 1$ where neither estimator $\hat{\beta}$ nor $\tilde{\beta}(1)$ are highly efficient in general. Against this criticism of the method must be weighed the large saving in computation by not having to estimate ρ and, as we shall see, by not having to perform a significance test.

The problem remains to make this decision between d_1 and d_2 . For simplicity we are considering, a posteriori, only the values $\rho = 0$ and $\rho = 1$, so that the parameter space becomes $\Theta = (\rho = 0, \rho = 1)$. The loss table for the double dichotomy may be written:

	$\rho = 0$	$\rho = 1$
d_1	0	l_{12}
d_2	l_{21}	0

Now the correct decision theory procedure, e.g. De Groot (1970), is to choose that decision which minimizes the expected loss.

For the two decisions the respective expected losses are

$$\begin{aligned} d_1 &: l_{12} \cdot p(\epsilon = 1|\underline{y}) \\ \text{and} \quad d_2 &: l_{21} \cdot p(\epsilon = 0|\underline{y}) \end{aligned}$$

Thus the decision d_1 to use least squares is taken if

$$l_{12} \cdot p(\epsilon = 1|\underline{y}) \leq l_{21} \cdot p(\epsilon = 0|\underline{y}) \quad , \quad (1)$$

and d_2 is taken otherwise. From the previous section, equation (3.4.1), it is clear that

$$\begin{aligned} p(\epsilon = 0|\underline{y}) &= K \cdot |\underline{X}'\underline{X}|^{-\frac{1}{2}} \cdot \left\{ (\underline{y} - \underline{X}\hat{\underline{\beta}})' (\underline{y} - \underline{X}\hat{\underline{\beta}}) \right\}^{-(n-k)/2} \cdot p(\epsilon = 0) \\ \text{and} \quad p(\epsilon = 1|\underline{y}) &= K \cdot |\underline{X}'\underline{B}\underline{X}|^{-\frac{1}{2}} \cdot \left\{ (\underline{y} - \underline{X}\tilde{\underline{\beta}})' \underline{B} (\underline{y} - \underline{X}\tilde{\underline{\beta}}) \right\}^{-(n-k)/2} \cdot p(\epsilon = 1), \end{aligned}$$

where our prior knowledge about $\underline{\beta}$, σ^2 is assumed vague and the prior distribution of ϵ is assumed to be independent of $\underline{\beta}$ and σ^2 . Substituting in (1), d_1 is the appropriate decision if

$$\begin{aligned} \left\{ \frac{(\underline{y} - \underline{X}\tilde{\underline{\beta}})' \underline{B} (\underline{y} - \underline{X}\tilde{\underline{\beta}})}{(\underline{y} - \underline{X}\hat{\underline{\beta}})' (\underline{y} - \underline{X}\hat{\underline{\beta}})} \right\}^{\frac{-(n-k)}{2}} &\leq \frac{l_{21} \cdot p(\epsilon = 0) \cdot |\underline{X}'\underline{X}|^{-\frac{1}{2}}}{l_{12} \cdot p(\epsilon = 1) \cdot |\underline{X}'\underline{B}\underline{X}|^{-\frac{1}{2}}} \\ \text{i.e. if } g \geq h \text{ where } h &= \left\{ \frac{l_{21} \cdot p(\epsilon = 0) \cdot |\underline{X}'\underline{X}|^{-\frac{1}{2}}}{l_{12} \cdot p(\epsilon = 1) \cdot |\underline{X}'\underline{B}\underline{X}|^{-\frac{1}{2}}} \right\}^{\frac{2}{n-k}} \quad (2) \end{aligned}$$

Thus the decision procedure rests on the g statistic. In fact there is a remarkable similarity in algebra between the Bayesian approach with vague priors and the invariance theory approach.

Here, however, interest is not now centred on finding significance points of the g statistic but only on evaluating h .

Assuming no prior knowledge to the contrary we put $p(\varrho = 0) = p(\varrho = 1)$ and the problem remains to find l_{12} and l_{21} . We wish to choose the losses such that they are both realistic and mathematically convenient. The Bayesian statistician has similar reasons for using conjugate priors. If least squares is not used when $\varrho = 0$ a suitable loss would seem to be one which was inversely proportional to the generalized standard deviation of the least squares estimate, i.e. $l_{21} = \frac{\text{constant}}{|\tilde{X}'X|^{-\frac{1}{2}}}$.

In other words, the more precise the least squares estimate the greater the loss arising from not using it. Similarly the loss l_{12} of not using $\tilde{\beta}(1)$ when $\varrho = 1$ is taken as inversely proportional to the generalized standard deviation of $\tilde{\beta}(1)$, i.e. $l_{12} = \frac{\text{constant}}{|\tilde{X}'BX|^{-\frac{1}{2}}}$.

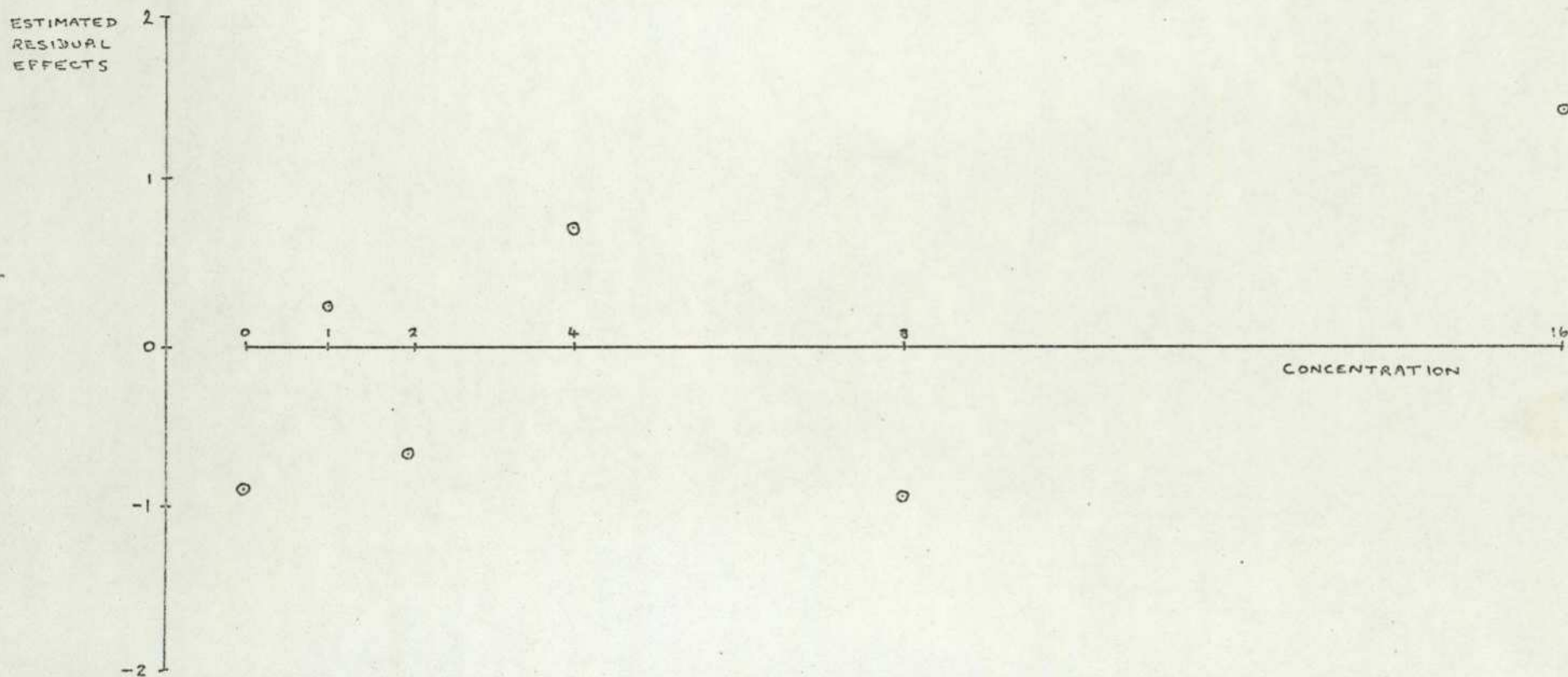
With these assumptions about the priors and losses it is seen that $h = 1$ and the decision to use first differenced or original variables is based upon whether or not $g < 1$. A further argument for using this value of h is that in any case it is the limiting value as $n - k$ becomes large, i.e. $\lim_{n-k \rightarrow \infty} h \rightarrow 1$, provided that none of the losses or prior probabilities is zero.

If $h = 1$ is regarded as a significance point to the g statistic it corresponds in general to a rather high level of significance under the null hypothesis $H_0 : \varrho = 0$. Perusal of the Durbin and Watson (1951) tables shows that for values of $n - k > 20$ the point $g = 1$ lies below even the lower bounds of the 5% significance points. Therefore there is a low probability of this procedure deciding in favour of taking first differences when in fact $\varrho = 0$ and the errors are independent.

Practically, this decision procedure amounts to comparing the residual sum of squares using the original variables to the residual sum of squares using first differenced variables. The estimates of the parameters to be used are those which provide the smallest residual sum of squares. This procedure is easy to operate and compared to a bounds test has the advantage of always being able to arrive at a decision.

To make a decision between $\rho = 0$ and $\rho = -1$, it can be shown in a similar fashion that the criterion is whether $g(-1) < 1$, and this would be a suitable choice when guarding against the possibility of using least squares estimates when they are highly inefficient in the presence of negative autocorrelation. The appropriate estimates to take as the alternative to least squares are $\tilde{\beta}(-1) = \{X'V^{-1}(-1)X\}^{-1} X'V^{-1}(-1)y$, which are the least squares estimates using the transformed variables $C(-1)X$ and $C(-1)y$, or the 'first summed' variables. From the results given in section 2.8 for the WILLIAMS DATA it appears that the probability of $g(-1)$ being less than one, unlike the positive case above, is not very small, reflecting (from fig. 7) the fact that the least squares estimates become very inefficient for even moderately negative values of autocorrelation. Because the value of $g(-1)$ in the Williams example is 1.013 it is tempting to examine the new estimates $\tilde{\beta}(-1)$ to see if the apparent randomness of the estimates of residual effects of treatments against treatment concentration is affected. The estimates $\tilde{\beta}(-1)$ are plotted in fig. 11, where it is seen that there is little change from the least squares estimates given in fig. 2.

FIG. II



PLOT OF THE ESTIMATES OF RESIDUAL EFFECTS
OF TREATMENTS AGAINST CONCENTRATION FOR
THE DATA FROM WILLIAMS' EXPERIMENT

A proof is now given that the estimator $\tilde{\beta}(1)$ has efficiency approaching unity for $\rho \rightarrow 1$ when the autocorrelation is taken to be of the stationary structure. Denoting the dispersion matrix representing the stationary autocorrelation scheme by $\sigma^2 \tilde{U}$, then

$$\tilde{U} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{pmatrix}$$

and

$$\tilde{U}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1+\rho^2 & -\rho \\ 0 & \cdots & \cdots & -\rho & 1 \end{pmatrix}.$$

For the stationary process the efficiency of $\tilde{\beta}(1)$ is defined therefore as

$$\text{Eff}\left\{\tilde{\beta}(1)\right\} = \frac{|(\tilde{X}' \tilde{U}^{-1} \tilde{X})^{-1}|}{|(\tilde{X}' \tilde{B} \tilde{X})^{-1}|^2 | \tilde{X}' \tilde{B} \tilde{U} \tilde{B} \tilde{X} |} \quad (3)$$

Difficulty arises at the limit when the matrix $\tilde{X}' \tilde{U}^{-1} \tilde{X}$ becomes singular for regressions including an overall mean. That the result is not generally true for regressions omitting an overall mean can easily be seen by putting $\tilde{X}' = (1 \ 0 \ 0 \ \dots \ 0)$, in which case

$$\text{Eff}\left\{\tilde{\beta}(1)\right\} = 4(1-\rho^2)/(5-4\rho).$$

It is of special interest to show that the result is true for the case when only an overall mean is fitted, i.e. $\tilde{X} = \underline{1}$. Here, for convenience, $\underline{1}$ is the unit vector of 1's, and is not normalized. Clearly the denominator of (3) is identically unity for any value of ρ . The numerator is

$$(1 - \rho^2) / \{(1 - \rho)(n - \overline{n-2}\rho)\} = (1 + \rho) / \{n - (n-2)\rho\}.$$

Therefore the efficiency of the estimator of the overall mean, given by $\tilde{\beta}(1) = y_1$, approaches unity as $\rho \rightarrow 1$ in the stationary autocorrelation process.

To prove that the result is true in general for regressions including an overall mean, $\tilde{X}$ is written in the form $(\underline{1} \mid \tilde{X}_0)$ where

$$\tilde{X}_0 = \begin{pmatrix} 0 & \dots & 0 \\ \dots & \dots & \dots \\ & & \tilde{X}_1 \end{pmatrix}, \text{ and } \tilde{X}_1 \text{ is of dimension}$$

$(n-1) \times (k-1)$. There is no loss of generality as the value of the efficiency (3) is unaltered by such reparameterizations.

From Rao (1965a),

$$\begin{vmatrix} \tilde{A} & \tilde{B} \\ \tilde{B}' & \tilde{C} \end{vmatrix} = |\tilde{A}| \cdot |\tilde{C} - \tilde{B}'\tilde{A}^{-1}\tilde{B}|. \quad (4)$$

Using (4) and lemma (2.3.1),

$$|\tilde{X}'\tilde{U}^{-1}\tilde{X}| = (1 - \rho^2)^{-k} \cdot |\underline{1}'\tilde{U}^{-1}\underline{1}| \cdot |\tilde{X}_0' \tilde{L}_0 (\tilde{L}_0' \tilde{U}^{-1} \tilde{L}_0)^{-1} \tilde{L}_0' \tilde{X}_0|,$$

where $\tilde{L}_0$ is defined in 2.9(ii) as the matrix of the $(n-1)$ latent vectors of $\tilde{A}$ other than $\underline{1}$.

Therefore

$$|X'U^{-1}X|^{-1} = \frac{(1-e^2)^k}{(1-e)\{n-(n-2)e\}} |X_o' L_o (L_o' U^{-1} L_o)^{-1} L_o' X_o|^{-1} \quad (5)$$

Now

$$\begin{aligned} |X' B U B X| &= \begin{vmatrix} 1' B U B 1 & \vdots & 1' B U B X_o \\ \vdots & \ddots & \vdots \\ X_o' B U B 1 & \vdots & X_o' B U B X_o \end{vmatrix} \\ &= |X_o' B (U - U B 1 1' B U) B X_o| \quad , \text{ from (4),} \end{aligned}$$

$$\text{and } 1' B U = (1 \quad e \quad e^2 \quad \dots \quad e^{n-1}) \quad .$$

Thus

$$U - U B 1 1' B U = \begin{pmatrix} 0 & \vdots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \vdots & (1-e^2) \cdot V_{n-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & 0 & \dots & 0 \end{pmatrix} = (1-e^2) W_1$$

$$\begin{aligned} \text{and } |X' B U B X| &= (1-e^2)^{k-1} \cdot |X_o' B W_1 B X_o| \\ &= (1-e^2)^{k-1} \cdot |X_1' B_{n-1} V_{n-1} B_{n-1} X_1| \quad . \quad (6) \end{aligned}$$

Now from (3), (5), (6)

$$\begin{aligned} \text{Eff}\{\tilde{\beta}(1)\} &= \frac{(1-e^2)^k \cdot |X_o' L_o (L_o' U^{-1} L_o)^{-1} L_o' X_o|^{-1} \cdot |X' B X|^{-2}}{(1-e)\{n-(n-2)e\} \cdot (1-e^2)^{k-1} \cdot |X_1' B_{n-1} V_{n-1} B_{n-1} X_1|} \\ &= \frac{(1+e) \cdot |X_o' L_o (L_o' U^{-1} L_o)^{-1} L_o' X_o|^{-1} \cdot |X_1' B_{n-1} X_1|^{-2}}{\{n-(n-2)e\} \cdot |X_1' B_{n-1} V_{n-1} B_{n-1} X_1|} \quad . \end{aligned}$$

Taking the limit,

$$\begin{aligned} \text{Eff}\{\tilde{\beta}(1)\} &= \frac{2 \cdot |X_o' A X_o|^{-1} \cdot |X_1' B_{n-1} X_1|^{-2}}{2 \cdot |X_1' B_{n-1} X_1|} = 1 \quad . \\ \lim_{e \rightarrow 1} & \end{aligned}$$

Therefore the estimator $\tilde{\beta}(1)$, adopted from consideration of the non-stationary process, is also efficient for high values of autocorrelation ($\rho \rightarrow 1$) in the stationary autocorrelated error structure when the regression includes an overall mean as a parameter.

The Cochrane-Orcutt (1949) method of using the $n - 1$ first differenced variables is criticised by Kadiyala (1966). Appendix B examines Kadiyala's comments in view of the results of this section.

3.6 Results of the computer simulations

This section is concerned with presenting and interpreting the results of the Monte-Carlo simulation undertaken as a study to compare the different estimators of ρ and β referred to in previous sections. The simulation model is given in section 3.1 and has a design matrix $\tilde{X} = (-1 \ -1 \ -1 \ 1 \ 1 \ 1)'$. Without loss of generality, the value of the parameter β_1 is set to zero so that the observations $\tilde{y}$ are conveniently put equal to the errors $\tilde{u}$. All the estimators of ρ are in some way or other based on the generalized residual sum of squares (3.2.9). Now GRSS remains unchanged by transformations of the type $\tilde{y}^* = \tilde{y} + \tilde{X}\gamma$, i.e. $\tilde{y}^{*'}\tilde{W}\tilde{y}^* = \tilde{y}'\tilde{W}\tilde{y}$, and therefore the estimates of ρ are not affected by changes in the settings of β . Further, it can be easily verified that for all the given estimators of β the result of a transformation such as $\tilde{y}^* = \tilde{y} + \tilde{X}\gamma$ is to change the estimate of β by an increment γ . Hence no generality is lost by taking $\beta = 0$. The autocorrelated error processes are generated from random normal deviates $\tilde{\epsilon}$ taken from the RAND (1955) tables. Three values of ρ were considered, 0, .8 and 1. For each value 1000 simulations were made. Thus

the values of the means, variances and mean square errors of the various estimates given in table 12 are each based on a sample size of 1000. Most of the simulations were carried out on the City University's ICL 1905 computer but the simulations for the Bayesian estimates, being very time consuming, were done on the much faster CDC 6600 of London University.

Least squares The means and variances of the least squares estimates of β_1 agree fairly closely with the theoretical values, which are given in brackets. As seen from fig. 8 the efficiency of least squares estimation is poor for high values of autocorrelation.

Maximum likelihood It is not surprising to find the maximum likelihood estimates with a higher variance than the other estimates for a model in which the number of observations is 6. The occurrence of dual maxima was rare, there being only 11 such cases in the 3000 simulations. One typical example was given in section 3.2.

Estimates based on g The means and variances of the estimator of ρ agree closely with the theoretical values given in brackets and calculated by the methods given in section 3.3. The variances of the estimates of β_1 are smaller than the corresponding variances of the maximum likelihood estimates. Moreover, far less computational effort is required than for the maximum likelihood estimates.

Bayesian estimates As seen from the table these two estimates of β_1 have the smallest variances for the higher values of autocorrelation. However, because of the amount of computation involved in obtaining the estimates these methods cannot seriously be considered as practical alternatives.

			Least Squares	Maximum Likelihood	Based on g statistic	Bayesian estimate of ρ	Bayesian estimate of β_1	$\tilde{\beta}_1(0)$ if $g < 1$ else $\hat{\beta}_1$
$\rho = 0$	ρ	mean	***	-.130	-.145 (-.159)	.022	***	***
		var	***	.283	.171 (.167)	.211	***	***
		mse	***	.300	.192	.212	***	***
	β_1	mean	.011 (0.0)	-.006	.017	-.025	-.021	.012
		var	.167 (.167)	.181	.186	.200	.197	.180
		mse	.167	.181	.186	.200	.197	.180
$\rho = .8$	ρ	mean	***	.577	.450 (.440)	.597	***	***
		var	***	.312	.192 (.196)	.210	***	***
		mse	***	.362	.315	.251	***	***
	β_1	mean	.009 (0.0)	.017	.015	.038	.037	.017
		var	.339 (.349)	.276	.257	.255	.246	.260
		mse	.339	.277	.258	.256	.248	.260
$\rho = 1$	ρ	mean	***	.766	.603 (.600)	.773	***	***
		var	***	.325	.181 (.170)	.220	***	***
		mse	***	.380	.338	.271	***	***
	β_1	mean	-.008 (0.0)	-.019	-.012	.031	.031	-.010
		var	.506 (.511)	.252	.241	.225	.221	.239
		mse	.506	.252	.241	.226	.222	.239

Results of the computer simulations to compare various
methods of estimation. The figures given in brackets are
theoretical values of the corresponding means and variances.

TABLE 12

The final column of table 11 gives the means, variances and mean squared errors obtained when using least squares estimates on either original or first differenced variables, depending whether or not $g \geq 1$. Of the alternatives to least squares estimation this method requires the least amount of computational effort. Although no general conclusions can be drawn from this one particular model, the practical results given in table 12 indicate that the estimation procedure suggested in section 3.5 deserves attention in future as the simplest alternative to least squares. Clearly the main danger to the efficiency of this procedure occurs when ϵ is in the mid-way region between 0 and 1, where both $\hat{\beta}$ and $\tilde{\beta}(1)$ may be inefficient. Alternatively, the method of estimation based on the statistic g is suggested as it requires much less computational effort than the maximum likelihood or Bayesian methods.

PART IV

SOME DESIGN SITUATIONS

4.1 A design criterion

The choice of a design criterion has been considered by many authors, e.g. Wald (1943), Tocher (1952), Chernoff (1953) and Kiefer (1959), but at present there is no universally accepted method. For example, Box and Guttman (1966) whose results we shall examine in section 4.4, select $\tilde{X}$ so as to minimize the trace of the variance-covariance matrix of the estimates (i.e. the sum of variances). This criterion does not select designs independently of non-orthonormal transformations (e.g. rescaling) of the parameters, a property which seems intuitively desirable on the grounds that reparameterization of the model does not affect the physical nature of the experiment and so should not affect the choice of design. A similar comment is made by M.J. Box (1972). A criterion which does possess this property is the maximization of the determinant $|\tilde{X}'\tilde{X}|$. In the case where the errors are independent $|\tilde{X}'\tilde{X}|^{-1}$ is the generalized variance, e.g. Rao (1965a), of the least squares estimates. The design method is to choose that matrix $\tilde{X}$ which minimizes this generalized variance. The criterion and its properties are discussed in Box and Draper (1971). When the errors are autocorrelated the generalized variance becomes $|\tilde{X}'V^{-1}\tilde{X}|^{-1}$. If the value of ρ is known a priori then we should choose that $\tilde{X}$ which minimizes $|\tilde{X}'V^{-1}\tilde{X}|^{-1}$. However, rarely is ρ known, but it may be suspected that $\rho \neq 0$ and a test for autocorrelation is made. If, after testing for positive (or negative)

autocorrelation the strategy suggested in section 3.5 is adopted, least squares on either original or first differenced variables is used. Thus designs are chosen according to the following procedure: from the set of designs with minimum variance when $\rho = 0$ select the subset which minimizes the generalized variance of the estimates when ρ is assumed to be 1 and first differenced variables are used, i.e. choose those $\tilde{X}$ which minimize $|\tilde{X}'B\tilde{X}|^{-1}$. Similarly, when guarding against negative autocorrelation the generalized variance to be minimized with respect to $\tilde{X}$ would be $|\tilde{X}'V^{-1}(-1)\tilde{X}|^{-1}$.

These formal generalized variances are applicable only if the design matrix $\tilde{X}$ is of full rank. In the case where the design matrix is $n \times p$ where $p > k$ it may be reduced to full rank k , as in Appendix A, by a suitable reordering of the columns and any transformation such as

$$\tilde{X}_p G = (\tilde{X} \begin{smallmatrix} \vdots \\ 0 \end{smallmatrix}), \text{ where } \tilde{X} \text{ is } n \times k \text{ of rank } k.$$

The model becomes

$$E(\tilde{y}) = \tilde{X}_p \tilde{\theta} = (\tilde{X} \begin{smallmatrix} \vdots \\ 0 \end{smallmatrix}) \begin{pmatrix} \tilde{\beta} \\ \tilde{\theta}^* \end{pmatrix} = \tilde{X} \tilde{\beta},$$

where the $\tilde{\beta}$ are any k estimable functions, determined by

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\theta}^* \end{pmatrix} = \tilde{G}^{-1} \tilde{\theta}.$$

Having selected which estimable functions are to be considered different designs may be compared on the basis of the generalized variance of these estimable functions. The initial choice of $\tilde{\beta}$ (the set of estimable functions) is immaterial because all other

sets of estimable functions can be derived from this first set by means of a non-singular transformation, and the $|\tilde{X}' \tilde{V}^{-1} \tilde{X}|$ criterion selects designs independently of such transformations. For example, suppose we have two designs (reduced to full rank) determined by $\tilde{X}_1$ and $\tilde{X}_2$ so that

$$E(\tilde{y}) = \tilde{X}_1 \beta \quad \text{and} \quad E(\tilde{y}) = \tilde{X}_2 \beta.$$

Let $\tilde{X}_1$ be a better design than $\tilde{X}_2$ for this set of parameters, i.e. $|\tilde{X}_1' \tilde{V}^{-1} \tilde{X}_1| > |\tilde{X}_2' \tilde{V}^{-1} \tilde{X}_2|$. Consider the reparameterization $\beta^* = \tilde{G}^{-1} \beta$, where $\tilde{G}^{-1}$ is non-singular, so that $E(\tilde{y}) = \tilde{X}_1 \tilde{G} \beta^*$ and $E(\tilde{y}) = \tilde{X}_2 \tilde{G} \beta^*$. The criteria become $|\tilde{G}' \tilde{X}_1' \tilde{V}^{-1} \tilde{X}_1 \tilde{G}|$ and $|\tilde{G}' \tilde{X}_2' \tilde{V}^{-1} \tilde{X}_2 \tilde{G}|$. Thus where $|\tilde{X}_1' \tilde{V}^{-1} \tilde{X}_1| > |\tilde{X}_2' \tilde{V}^{-1} \tilde{X}_2|$

we have $|\tilde{G}|^2 |\tilde{X}_1' \tilde{V}^{-1} \tilde{X}_1| > |\tilde{G}|^2 |\tilde{X}_2' \tilde{V}^{-1} \tilde{X}_2|$,

so that $|\tilde{G}' \tilde{X}_1' \tilde{V}^{-1} \tilde{X}_1 \tilde{G}| > |\tilde{G}' \tilde{X}_2' \tilde{V}^{-1} \tilde{X}_2 \tilde{G}|$,

and the choice of design is unaffected by the reparameterization.

4.2 Qualitative designs

The Williams' experiment is an example of an experimental design situation where one might postulate negative autocorrelation in the errors caused by carry-over effects. Autocorrelations are assumed to occur only between errors in observations on the same subject, as defined in section 1.3. The special Williams' designs balanced for residual effects are latin squares, but if the period

effects are ignored we have block designs with a special sequence for treatments. The significance points of the g test will of course vary according to the design. However, using the methods of section 2.9 it is possible to evaluate bounds to the significance points of the statistics g and $g(-1)$ for use with the randomized blocks and latin squares. In the notation of section 2.9 the matrix $\tilde{X}$ is taken to represent the overall mean and subject effects in the randomized blocks, and the overall mean, subject effects and period effects in the latin squares. Table 13 gives the bounds to the significance points of the $g(-1)$ and $g(1)$ statistics where the level of significance is 5%.

The problem of choosing qualitative designs using the criterion adopted in the previous section of minimizing the generalized variances $|\tilde{X}'\tilde{V}^{-1}(-1)\tilde{X}|^{-1}$ or $|\tilde{X}'\tilde{B}\tilde{X}|^{-1}$ over a particular set of $\tilde{X}$'s (e.g. randomized blocks or latin squares) has not been amenable to analytic solution. For any given class of designs it is possible to evaluate the determinants and thus find the optimum designs. Unfortunately this is only technically feasible for the smallest designs. A simplification can be made in the evaluation of $|\tilde{X}'\tilde{B}\tilde{X}|$ when an overall mean and subject effects are parameters in the model. Partitioning $\tilde{X}$ into $(\tilde{X}_0 \vdots \tilde{X}_1)$, where $\tilde{X}_0$ represents the overall mean and subject effects, given for the 4×4 design by

$$\tilde{X}'_0 = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix},$$

n	r	k	g		g(-1)	
			lower bound	upper bound	lower bound	upper bound
16	4	7	0.90	1.78	0.34	0.98
16	7	10	0.72	2.14	0.21	1.19
25	5	9	1.05	1.83	0.55	1.15
25	9	13	0.90	1.95	0.44	1.22
36	6	11	1.17	1.85	0.73	1.29
36	11	16	1.06	1.91	0.64	1.32
49	7	13	1.27	1.86	0.88	1.38
49	13	19	1.19	1.90	0.81	1.41
64	8	15	1.35	1.87	1.00	1.46
64	15	22	1.20	1.90	0.94	1.47

Bounds to the significance points of the g and g(-1)
statistics for randomized block and latin square designs

TABLE 13

n	k	$g(1)$	$g(-1)$
16	7	1.101	0.423
16	10	0.995	0.306

The 5 % significance points of the g and $g(-1)$ statistics for 4×4 randomized blocks ($k=7$) and latin squares ($k=10$) when the treatments are arranged according to a Williams' design

TABLE 14

then
$$|\tilde{X}'\tilde{B}\tilde{X}| = |\tilde{X}_1' \tilde{A}\tilde{X}_1|$$

Proof
$$(\tilde{X}'\tilde{B}\tilde{X}) = \begin{pmatrix} \tilde{X}_0' \tilde{B}\tilde{X}_0 & \tilde{X}_0' \tilde{B}\tilde{X}_1 \\ \tilde{X}_1' \tilde{B}\tilde{X}_0 & \tilde{X}_1' \tilde{B}\tilde{X}_1 \end{pmatrix}$$

For the 4×4 design

$$\tilde{X}_0' \tilde{B}\tilde{X}_0 = I_4 \text{ and}$$

$$\tilde{X}_0' \tilde{B}\tilde{X}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Using (3.5.4) ,

$$|\tilde{X}'\tilde{B}\tilde{X}| = |\tilde{X}_0' \tilde{B}\tilde{X}_0| \cdot |\tilde{X}_1' (\tilde{B} - \tilde{B}\tilde{X}_0\tilde{X}_0'\tilde{B})\tilde{X}_1|$$

i.e.
$$|\tilde{X}'\tilde{B}\tilde{X}| = |\tilde{X}_1' \tilde{A}\tilde{X}_1|$$

This result furnishes a saving in computer time for the evaluation of $|\tilde{X}'\tilde{B}\tilde{X}|$.

A computer evaluation was carried out for the class of 4×4 designs, for randomized blocks and latin squares, and for $|\tilde{X}'\tilde{B}\tilde{X}|^{-1}$ and $|\tilde{X}'\tilde{V}^{-1}(-1)\tilde{X}|^{-1}$. The result was that the same designs minimized the generalized variances for both randomized blocks and latin squares and for positive and negative alternatives. These designs were in fact the Williams (1949) designs: namely those designs for which every treatment follows every other treatment, i.e.

A B C D	A B C D	A B C D	A B C D	A B C D	A B C D
B D A C	B C D A	C A D B	C D B A	D A B C	D C A B
C A D B	D A B C	B D A C	D C A B	B C D A	C D B A
D C B A	C D A B	D C B A	B A D C	C D A B	B A D C

We are unable to prove the result analytically, but further computer work has shown that the same result appears to be true for the 6×6 designs. Finally, for all the 4×4 designs given above there are the same significance points. These are given in table 14 for a level of significance of 5% and for the 4×4 randomized blocks and latin squares where $k = 7$ and 10 respectively.

4.3 The Box-Hunter design

Box and Hunter (1958) considered the design problem of choosing an optimum sequence of settings for quantitative treatment factors on the basis of information from a pilot experiment (which we have called the Box and Hunter data) in which there is a variation of yield attributable to a time dependent parameter. Box and Hunter establish the degree of polynomial which represents the time trend using the usual F-tests and then choose the design vectors to be orthogonal to these trend polynomials. The F-tests depend for their validity on the assumption that the errors at any fitting are independent. However, misrepresentations of the trend by a polynomial of too low an order will cause the effects of positively autocorrelated errors in which case the validity of the use of the F-tests is questionable. Instead the use of a test for autocorrelation

is appropriate. Now if the degree of the polynomial is increased the positive autocorrelation induced by under-parameterization is reduced, and a suitable choice of degree of polynomial would be the lowest for which the g test proves non-significant. For this purpose the 5% significance points of the g test are given in table 15 for various values of n and k , the degree of the polynomial being $k - 1$.

A further point arises with this example regarding the expected value of the response at time $t = 0$. To illustrate this we must first define the mathematical model for this general type of experiment. Let $f, g, \dots$ be the levels of several treatment factors $F, G, \dots$ in a sequence of experiments. Particular levels of the factors are to be regarded as fixed, viz. $f_0, g_0, \dots$; under these conditions and at time $t = 0$ the expected response has a theoretical value μ_0 . At any time t the expectation of the response y_t is written in the form

$$E(y_t - \mu_0) = \mu + \sum_{i=1}^p \alpha_i t^i + \beta(f_t - f_0) + \gamma(g_t - g_0) + \dots;$$

the pilot experiment, in which the factors are at their standard values, gives information about the suitable value for μ and in the main experiment changes are made in the settings $f_t, g_t, \dots$, $t = 1, 2, \dots, n$, so that the parameters $\beta, \gamma, \dots$ can be estimated efficiently. In this formulation, μ is clearly the systematic bias in the measurements y_t .

Considering Box and Hunter's data there is good reason to believe from the nature of the experiment that the theoretical response at time $t = 0$ is 100%. Thus taking $\mu_0 = 100$, we

n	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
10	1.20	1.30	1.55	1.83	2.15	2.49
11	1.22	1.30	1.53	1.76	2.03	2.32
12	1.24	1.31	1.51	1.72	1.95	2.20
13	1.25	1.33	1.51	1.70	1.89	2.10
14	1.27	1.34	1.50	1.67	1.85	2.04
15	1.29	1.35	1.50	1.66	1.82	1.99
16	1.31	1.36	1.50	1.65	1.80	1.95
17	1.32	1.37	1.50	1.64	1.78	1.92
18	1.33	1.38	1.51	1.63	1.76	1.89
19	1.35	1.39	1.51	1.63	1.75	1.87
20	1.36	1.40	1.51	1.62	1.74	1.85
21	1.37	1.41	1.52	1.62	1.73	1.84
22	1.38	1.42	1.52	1.62	1.72	1.82
23	1.39	1.43	1.53	1.62	1.72	1.81
24	1.40	1.44	1.53	1.62	1.71	1.80
25	1.41	1.45	1.53	1.62	1.71	1.80
26	1.42	1.46	1.54	1.62	1.70	1.79
27	1.43	1.47	1.54	1.62	1.70	1.78
28	1.44	1.47	1.55	1.62	1.70	1.78
29	1.45	1.48	1.55	1.62	1.70	1.77
30	1.46	1.49	1.56	1.63	1.70	1.77
35	1.49	1.52	1.58	1.64	1.70	1.76
40	1.52	1.54	1.59	1.65	1.70	1.75
45	1.54	1.56	1.61	1.66	1.70	1.75
50	1.54	1.58	1.62	1.67	1.71	1.75

The 5% significance points of the g test when the columns of $\bar{X}$ represent a polynomial trend about an overall mean

TABLE 15

n	k=1	k=2	k=3	k=4	k=5	k=6
10	1.18	1.42	1.69	1.99	2.30	2.62
11	1.20	1.41	1.65	1.91	2.18	2.46
12	1.22	1.42	1.63	1.85	2.03	2.33
13	1.24	1.42	1.61	1.81	2.02	2.23
14	1.26	1.42	1.60	1.78	1.97	2.16
15	1.28	1.43	1.59	1.76	1.93	2.10
16	1.30	1.44	1.58	1.74	1.90	2.06
17	1.31	1.44	1.58	1.72	1.87	2.02
18	1.33	1.45	1.58	1.71	1.85	1.99
19	1.34	1.46	1.58	1.70	1.83	1.96
20	1.36	1.46	1.58	1.70	1.82	1.94
21	1.37	1.47	1.58	1.69	1.80	1.92
22	1.38	1.48	1.58	1.68	1.79	1.90
23	1.39	1.48	1.58	1.68	1.78	1.88
24	1.40	1.49	1.58	1.68	1.77	1.87
25	1.41	1.50	1.58	1.68	1.77	1.86
26	1.42	1.50	1.59	1.67	1.76	1.85
27	1.43	1.51	1.59	1.67	1.76	1.84
28	1.44	1.51	1.59	1.67	1.75	1.83
29	1.45	1.52	1.59	1.67	1.75	1.83
30	1.46	1.52	1.60	1.67	1.75	1.82
35	1.49	1.55	1.61	1.67	1.74	1.80
40	1.52	1.57	1.62	1.68	1.73	1.79
45	1.54	1.59	1.64	1.68	1.73	1.78
50	1.57	1.61	1.65	1.69	1.73	1.78

The 5 % significance points of the g test when the columns
of $\tilde{x}$ represent a polynomial trend through the origin

TABLE 16

would wish to test the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. Appropriate orthogonal polynomials, designated γ_i , for this purpose are given in Berenblutt and Webb (1972). The g test is used to establish the degree of polynomial consistent with the assumption of independent errors. For polynomial regressions without an overall mean, i.e. using the γ_i , the 5% significance points of the g test are given in table 16. In this example a fifth order polynomial is the lowest for which the g test is not significant at the 5% level, for regressions both with and without an overall mean - a different result to that obtained by Box and Hunter who accepted a cubic fit. In fact the F values given in fig. 12 show that sixth order polynomial fits are significant. Finally, performing the F test with $k = 5$ does not give a significant result, thus accepting $H_0 : \mu = 0$.

The use for the factor settings of any set of vectors orthogonal to the time trend polynomials will minimize the generalized variance, $|\tilde{X}'\tilde{X}|^{-1}$. Box and Hunter took as a basis for the design space the 17 orthogonal polynomials other than $\xi_0, \xi_1, \xi_2, \xi_3$, and used the method of angular randomization to select the design vectors. Here, having established that a fifth order polynomial through the origin gives a representation of the trend consistent with the hypothesis of independent errors, the design space is different. It is the set of 16 vectors $\gamma_6, \gamma_7, \dots, \gamma_{21}$. The usual process of angular randomization can be applied to obtain the desired design vectors. Alternatively, the design vectors can be taken with elements as random normal

k	Regression through origin		Regression about overall mean	
	$g(1)$	F'_s value associated with γ_k	$g(1)$	F'_{Ht} value associated with ξ_{k-1}
1	0.255***	-----	0.041***	-----
2	0.322***	193.5***	0.245***	3799***
3	0.723***	230.8***	0.330***	154.3***
4	0.745***	4.06	0.989***	237.9***
5	1.818	110.2***	1.399**	24.14***
6	3.103	16.24**	2.261	27.74***
7	3.103	-----	3.089	8.26*

* indicates significance at 5% level
 ** indicates significance at 1% level
 *** indicates significance at 0.1% level

Results of tests on the Box-Hunter data

FIG.12

deviates and made orthogonal to the trend polynomials by the Gram-Schmidt process.

4.4 Quantitative designs

The general quantitative factor design problem may be regarded as determining a matrix $\tilde{X}_1$ to represent the factor settings, where the full design matrix is $\tilde{X} = (\tilde{X}_0 \mid \tilde{X}_1)$ and the dispersion matrix is $\sigma^2 \tilde{V}$. Assuming $\tilde{X}_0$ is $n \times r$, the r parameters associated with $\tilde{X}_0$ could for example represent an overall mean and subject effects, or trend effects. In the last section the specific instance was considered where $\tilde{X}_0$ represented a polynomial time trend in terms of the first r orthogonal polynomials and where the errors were assumed to be independent.

For convenience the columns of $\tilde{X}_1$ are normalized, i.e. the sum of squares of elements in each column is unity. This is not a restriction because the design criterion used here, $|\tilde{X}' \tilde{V}^{-1} \tilde{X}|$, selects designs independently of changes of scale of the parameters, and consequently the normalized levels of any particular factor settings can be rescaled to cover the range of interest in the experiment. Box and Guttman (1966) use a similar formulation of the problem. However they impose the additional assumption that $\tilde{X}_1$ is orthonormal and orthogonal to $\tilde{X}_0$, while using as a design criterion the trace of $(\tilde{X}' \tilde{X})^{-1}$, which does not select designs independently of changes of scale of the parameters.

As indicated in section 4.1 we shall, as our design procedure, consider those designs which provide estimates of the parameters with minimum generalized variance when $\underline{V} = \underline{I}$ (independent errors) and select from these the subset with minimum generalized variance when $\underline{V} = \underline{B}$ for positive autocorrelation, and $\underline{V} = \underline{V}(-1)$ when the errors are taken to be negatively autocorrelated. Thus from those designs which minimize $|\underline{X}'\underline{X}|^{-1}$ we select those which minimize $|\underline{X}'\underline{V}^{-1}(\pm 1)\underline{X}|^{-1}$ with the constraint that the columns of $\underline{X}_1$ are normalized.

Define $\underline{H}$ as any $n \times (n - r)$ orthonormal matrix orthogonal to $\underline{X}_0$, i.e. $\underline{X}_0'\underline{H} = \underline{0}$ and $\underline{H}'\underline{H} = \underline{I}_{n-r}$. Together $\underline{X}_0$ and $\underline{H}$ provide a basis for the set of $n \times n$ matrices so that the unknown design vectors may be written

$$\underline{X}_1 = (\underline{X}_0 : \underline{H}) \begin{pmatrix} \underline{G}_0 \\ \dots \\ \underline{G}_1 \end{pmatrix} = \underline{X}_0 \underline{G}_0 + \underline{H} \underline{G}_1, \quad (1)$$

and

$$|\underline{X}'\underline{X}|^{-1} = \begin{vmatrix} \underline{X}_0'\underline{X}_0 & \dots & \underline{X}_0'\underline{X}_0\underline{G}_0 \\ \vdots & \ddots & \vdots \\ \underline{G}_0'\underline{X}_0'\underline{X}_0 & \dots & \underline{G}_0'\underline{X}_0'\underline{X}_0\underline{G}_0 + \underline{G}_1'\underline{G}_1 \end{vmatrix}^{-1}. \quad (2)$$

From (2) and (3.5.4) it is easily seen that

$$|\underline{X}'\underline{X}|^{-1} = |\underline{X}_0'\underline{X}_0|^{-1} \cdot |\underline{G}_1'\underline{G}_1|^{-1}.$$

Therefore to minimize $|\underline{X}'\underline{X}|^{-1}$ it is only necessary to minimize $|\underline{G}_1'\underline{G}_1|^{-1}$ because $\underline{X}_0$ is known and fixed.

$$\text{Now } \tilde{X}_1' \tilde{X}_1 = \begin{pmatrix} | & * & \dots & * \\ * & | & & \\ \vdots & & \ddots & \\ \vdots & & & | \\ * & \dots & \dots & | \end{pmatrix} = \tilde{G}_0' \tilde{X}_0' \tilde{X}_0 \tilde{G}_0 + \tilde{G}_1' \tilde{G}_1, \quad (3)$$

and the diagonal elements of $\tilde{G}_1' \tilde{G}_1$ must be less than or equal to unity. The minimum of $|\tilde{G}_1' \tilde{G}_1|^{-1}$ occurs when $\tilde{G}_1$ is orthonormal and the diagonal elements of $\tilde{G}_1' \tilde{G}_1 (= I_{k-r})$ are unity. Thus, from (3), the diagonal elements of $\tilde{G}_0' \tilde{X}_0' \tilde{X}_0 \tilde{G}_0$ are zero and, because $\tilde{X}_0$ is full rank r , this can only occur when $\tilde{G}_0 = 0$. That is, when the errors are taken to be independent and homoscedastic, the matrix $\tilde{X}_1$ which minimizes $|\tilde{X}' \tilde{X}|^{-1}$, the generalized variance of the least squares estimates, should be chosen from the set of orthonormal matrices which are orthogonal to $\tilde{X}_0$. This result is the same as one of the basic assumptions of Box and Guttman (1966). However, they continue by using the trace criterion and come to the conclusion that if the dispersion matrix $\tilde{V}$ was known completely the $k - r$ latent vectors of $\tilde{V}$ associated with the smallest latent roots, $\tilde{L}_1$ say, should be used as a basis for the design vectors $\tilde{X}_1$, i.e. $\tilde{X}_1 = \tilde{L}_1 \tilde{G}_1$. But their method is not possible in general since $\tilde{L}_1$ and $\tilde{X}_0$ are not always orthogonal.

Returning to our design procedure, it is required to minimize $|\tilde{X}' \tilde{V}^{-1} \tilde{X}|^{-1}$ when $\tilde{X}_1$ takes the form $\tilde{H} \tilde{G}_1$. This determinant can be written

$$|\tilde{X}' \tilde{V}^{-1} \tilde{X}|^{-1} = \begin{vmatrix} \tilde{X}_0' \tilde{V}^{-1} \tilde{X}_0 & \dots & \tilde{X}_0' \tilde{V}^{-1} \tilde{X}_1 \\ \vdots & \ddots & \vdots \\ \tilde{X}_1' \tilde{V}^{-1} \tilde{X}_0 & \dots & \tilde{X}_1' \tilde{V}^{-1} \tilde{X}_1 \end{vmatrix}.$$

Then using (3.5.4)

$$|\tilde{X}'\tilde{V}^{-1}\tilde{X}|^{-1} = |\tilde{X}_0'\tilde{V}^{-1}\tilde{X}_0|^{-1} \cdot |\tilde{X}_1'\tilde{V}^{-1}\tilde{X}_1 - \tilde{X}_1'\tilde{V}^{-1}\tilde{X}_0(\tilde{X}_0'\tilde{V}^{-1}\tilde{X}_0)^{-1}\tilde{X}_0'\tilde{V}^{-1}\tilde{X}_1|^{-1},$$

and lemma (2.3.1) gives

$$|\tilde{X}'\tilde{V}^{-1}\tilde{X}|^{-1} = |\tilde{X}_0'\tilde{V}^{-1}\tilde{X}_0|^{-1} \cdot |\tilde{X}_1'\tilde{H}(\tilde{H}'\tilde{V}\tilde{H})^{-1}\tilde{H}'\tilde{X}_1|^{-1}.$$

Remembering that $\tilde{H}'\tilde{X}_1 = \tilde{G}_1$ the answer is given by minimizing

$|\tilde{G}_1'(\tilde{H}'\tilde{V}\tilde{H})^{-1}\tilde{G}_1|^{-1}$ with respect to $\tilde{G}_1$. Clearly $\tilde{G}_1$ is chosen as the $k - r$ normalized latent vectors of $(\tilde{H}'\tilde{V}\tilde{H})^{-1}$ corresponding to the largest latent roots, satisfying the restriction that $\tilde{G}_1$ be orthonormal because the latent vectors of a symmetric matrix are orthogonal. This can be seen easily from the result (2.9.3).

The value of the determinant remains unchanged by orthonormal transformations of the columns of $\tilde{G}_1$, and so randomization is accomplished by the methods of Box and Hunter (1958) and Box and Guttman (1966) or by postmultiplication of $\tilde{G}_1$ by any $(k - r) \times (k - r)$ orthonormal matrix $\tilde{P}$.

Thus, in general, the design solution for the factor levels can be written

$$\tilde{X}_1 = \tilde{H}\tilde{G}_1\tilde{P}, \quad (4)$$

where $\tilde{H}$ is any $n \times (n - r)$ orthonormal matrix orthogonal to $\tilde{X}_0$, the columns of $\tilde{G}_1$ are the $k - r$ normalized latent vectors of $\tilde{H}'\tilde{V}\tilde{H}$ corresponding to the smallest latent roots, and $\tilde{P}$ is any $(k - r) \times (k - r)$ orthonormal randomization matrix. It is easily shown that $\tilde{X}_1$ is unaffected by choice of $\tilde{H}$ by considering two different orthonormal matrices $\tilde{H}$ and $\tilde{H}_1$ both orthogonal to $\tilde{X}_0$. The

relationship between $\underline{H}$ and $\underline{H}_1$ is in general $\underline{H} = \underline{H}_1 \underline{K}$ where $\underline{K}$ is an $(n - r) \times (n - r)$ orthonormal matrix, and

$(\underline{H}_1' \underline{V} \underline{H}_1)^{-1} = \underline{K}' (\underline{H}' \underline{V} \underline{H})^{-1} \underline{K}$. If the latent vectors of $(\underline{H}' \underline{V} \underline{H})^{-1}$ are denoted by $\underline{L} \{= (\underline{L}_0 : \underline{G}_1)\}$ the latent vectors of $(\underline{H}_1' \underline{V} \underline{H}_1)^{-1}$ are given by $\underline{K} \underline{L}$. Therefore $\underline{X}_1 = \underline{H}_1 \underline{K} \underline{G}_1 \underline{P} = \underline{H} \underline{G}_1 \underline{P}$ and the design vectors are unaltered by choice of $\underline{H}$.

Following Box (1960) we distinguish between carry-over situations where it is suspected that negative autocorrelated errors occur and trend situations where positive autocorrelation would be the appropriate assumption. In general the actual values of ρ are unknown, but from section 3.1 it is seen that greatest importance should be attached to guarding against high absolute values of autocorrelation. Therefore for positive autocorrelation situations we take $\underline{V} = \underline{V}(1) = \underline{B}^{-1}$, and for negative autocorrelation cases $\underline{V} = \underline{V}(-1)$. Thus, knowing $\underline{X}_0$ and $\underline{V}$ the quantitative factor levels are given by $\underline{X}_1$ from (4), obtained either analytically or numerically. This design procedure produces efficient estimates of the parameters when the errors are independent ($\rho = 0$) and when the autocorrelation is high ($\rho \rightarrow 1$) provided the correct decision about $\rho = 0$ or 1 has been made. This decision may be based on a significance test using the g or $g(-1)$ statistics, or simply on g or $g(-1)$ using the methods of section 3.5.

4.5 Quantitative designs - two examples

Using the methods of the previous section we examine two particular cases where analytic solution for the design vectors is possible. These are (i) where $k - s$ treatments factors are applied to s subjects with $\underline{X}_o$ representing the s subject effects, and (ii) the Box-Hunter design situation where $\underline{X}_o$, the slow trend with time in the response, may be represented by suitable latent vectors of $\underline{B}$ or $\underline{A}$.

The latent roots of $\underline{B}$ are given by (2.2.7) in ascending order as

$$\theta_j = 2 - 2 \cos \left\{ (2j-1)\pi / (2n+1) \right\} \quad , \quad (j = 1, 2, \dots, n) \quad .$$

It can be found, using the method given by Von Neumann (1941) for the latent vectors of the matrix $\underline{A}$, that the i^{th} element of the normalized latent vector corresponding to the j^{th} latent root is given by $\{2/(2n+1)\} \sin \{i \cdot (2j-1)\pi / (2n+1)\}$. The frequency of the j^{th} latent vector is clearly $\frac{2j-1}{2(2n+1)}$. To establish the latent roots and latent vectors of $\underline{V}^{-1}(-1)$ it should first be noted that $\underline{V}^{-1}(-1) = \underline{J}\underline{B}\underline{J}$, where $\underline{J}$ is the diagonal matrix

$$\begin{pmatrix} 1 & & & \\ & -1 & & \\ & & 1 & \\ & & & \ddots \\ & & & & (-1)^{n-1} \end{pmatrix} \quad ; \quad \text{clearly } \underline{J}^{-1} = \underline{J}' = \underline{J} \quad .$$

If $\underline{x}$ is a latent vector of $\underline{B}$, $\underline{B}\underline{x} = \lambda\underline{x}$ and $\underline{J}\underline{V}^{-1}(-1)\underline{J}\underline{x} = \lambda\underline{x}$, i.e. $\underline{V}^{-1}(-1)\underline{J}\underline{x} = \lambda\underline{J}\underline{x}$. Therefore the latent roots of $\underline{V}^{-1}(-1)$ are identical to those of $\underline{B}$, and the normalized latent vector of $\underline{V}^{-1}(-1)$ associated with the j^{th} latent root is defined by its i^{th} element as $2/\sqrt{(2n+1)} \cdot (-1)^{i-1} \cdot \sin\{i \cdot (2j-1) \cdot \pi / (2n+1)\}$, and has a frequency of $(n - j + 1)/(2n + 1)$.

In an experiment where $k - s$ treatment factors are applied to s subjects the matrix $\underline{X}_0$ would represent the s subject effects with $r = s$. This has the implicit effect of including an overall mean as a parameter. In fact when $s = 1$ and all the treatments are applied to one experimental unit $\underline{X}_0 = \underline{1}$. If it is thought for example that slow changes in subject means was inducing positive autocorrelation, then an analytic solution to the design problem is readily obtainable. Here, as in section 2.9(ii), $\underline{H}$ is chosen to be $\underline{L}_0$, remembering that where $s > 1$, $\underline{L}_0$ is the matrix of $n - s$ latent vectors of $\underline{A}_n$ associated with the non-zero latent roots, where

$$\underline{A}_n = \begin{pmatrix} \underline{A}_p & \underline{0} & \dots & \dots & \underline{0} \\ \vdots & \underline{A}_p & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \underline{0} & \dots & \dots & \dots & \underline{A}_p \end{pmatrix}.$$

From (2.9.5) $(\underline{L}_0' \underline{B}^{-1} \underline{L}_0)^{-1} = \underline{\Lambda}_{n-s}$; the latent vectors of this diagonal matrix are given by

$$\underline{K} = \begin{pmatrix} \underline{K}_1 \\ \vdots \\ \underline{K}_2 \\ \vdots \\ \vdots \\ \vdots \\ \underline{K}_{p-1} \end{pmatrix},$$

where $\tilde{K}_1, \tilde{K}_2, \dots, \tilde{K}_{p-1}$ are any $s \times s$ orthornormal matrices. The design vectors are then specified by taking $\tilde{G}_1$ as the columns of $\tilde{K}$ corresponding to any set of $k - s$ largest latent roots of $\tilde{\Lambda}_{n-s}$. This set may not be unique if $s > 1$ in which case each latent root has a multiplicity of s and choice of vectors from the $\tilde{K}_i$ offers extra scope for randomization. After choosing the actual randomization matrix $\tilde{P}$, the design vectors are given by $\tilde{X}_1 = \tilde{L}_0 \tilde{G}_1 \tilde{P}$. The latent vectors of $\tilde{A}_{\tilde{n}}$ are the latent vectors of $\tilde{A}_{\tilde{p}}$ with $n - p$ zeros appropriately positioned. The latent vectors of $\tilde{A}_{\tilde{p}}$ are given by Durbin and Watson (1971). The i^{th} element of the normalized j^{th} vector is

$$\sqrt{(2/n)} \cos \{(2i-1)j\pi/2n\}, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, p-1.$$

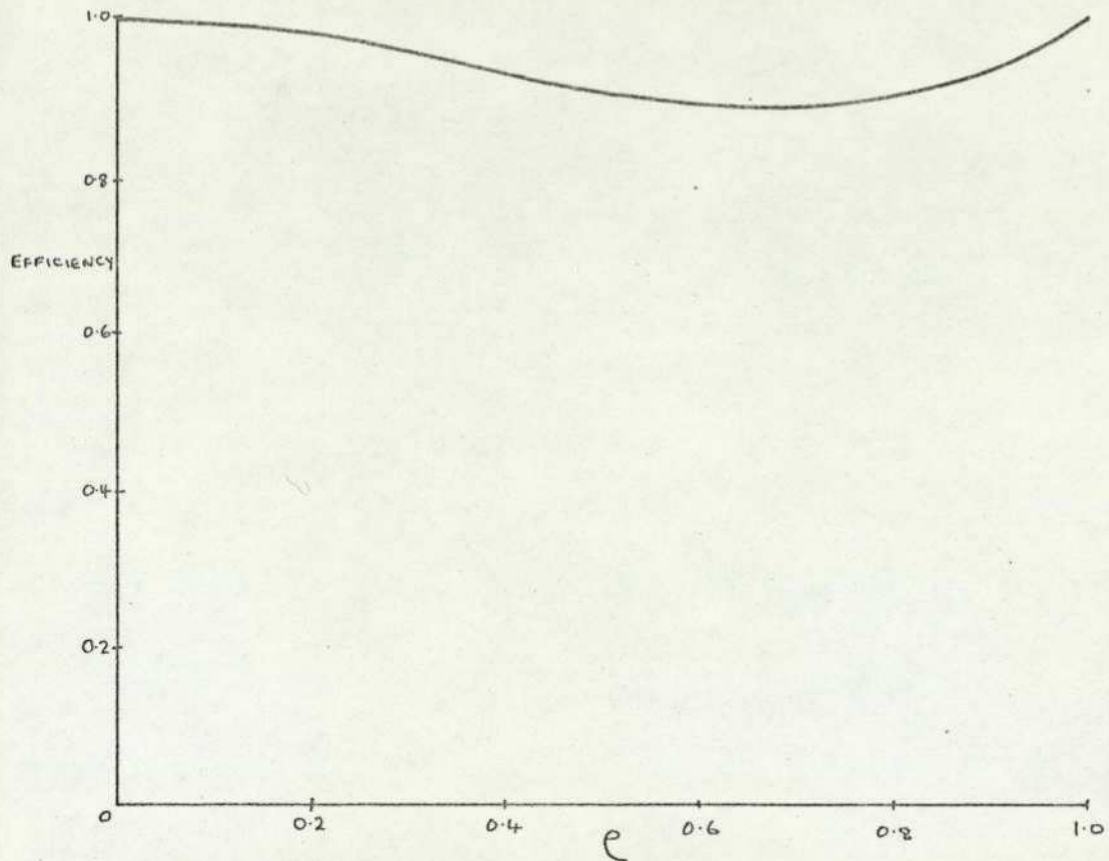
Thus the design problem of choosing the columns of $\tilde{X}_1$ as $k - s$ treatment factor vectors when $\tilde{X}_0$ represents the s subject means has been solved explicitly for the situation where there is suspected positive autocorrelation. Two interesting points arise from this solution. The g test for positive autocorrelation in this case is approximately UMP because the design matrix is now a linear combination of k latent vectors of $\tilde{A}$ (section 2.2). If the autocorrelation process is assumed to be stationary it is not necessary to make a decision on whether or not to use least squares because for these regression vectors the least squares estimator is efficient for autocorrelations approaching unity, i.e. $\rho \rightarrow 1$. We shall return to this point after considering the second particular case.

The Box-Hunter example is typical of design situations where there is a slow trend with time in the response. As the latent vectors associated with the smallest latent roots of $\tilde{A}$ and $\tilde{B}$ are also slowly changing with time (i.e. have low frequency) it is of interest to consider the possibility of using these vectors to represent the time trend. The elements of the latent vectors of $\tilde{A}$ are cosine terms and are suitable for regressions about an overall mean; the elements of the latent vectors of $\tilde{B}$ are sine terms and suitably represent regressions through the origin. Unless there are good physical reasons to believe the time trend is in fact polynomial, use of the latent vectors has certain advantages.

When the expected response is known at time $t = 0$ under the standard conditions this known value may be subtracted from all the observed responses after which it is desirable that the fitted regression vectors also pass through the origin. The real time trend will generally be more complicated than a linear combination of particular orthogonal polynomials or latent vectors. Thus the misrepresentation of the time trend will induce positive autocorrelation in the error structure. But as the true trend and the fitted trend both pass through the origin it is reasonable to assume the variance of the errors increases with departure from the origin. Consequently there seems to be an innate compatibility between the use of the latent vectors of $\tilde{B}$ to represent regressions through the origin and adoption of the non-stationary autocorrelation process, where the variances of the overall errors increase with time. On the other hand when the expected response does not pass through the

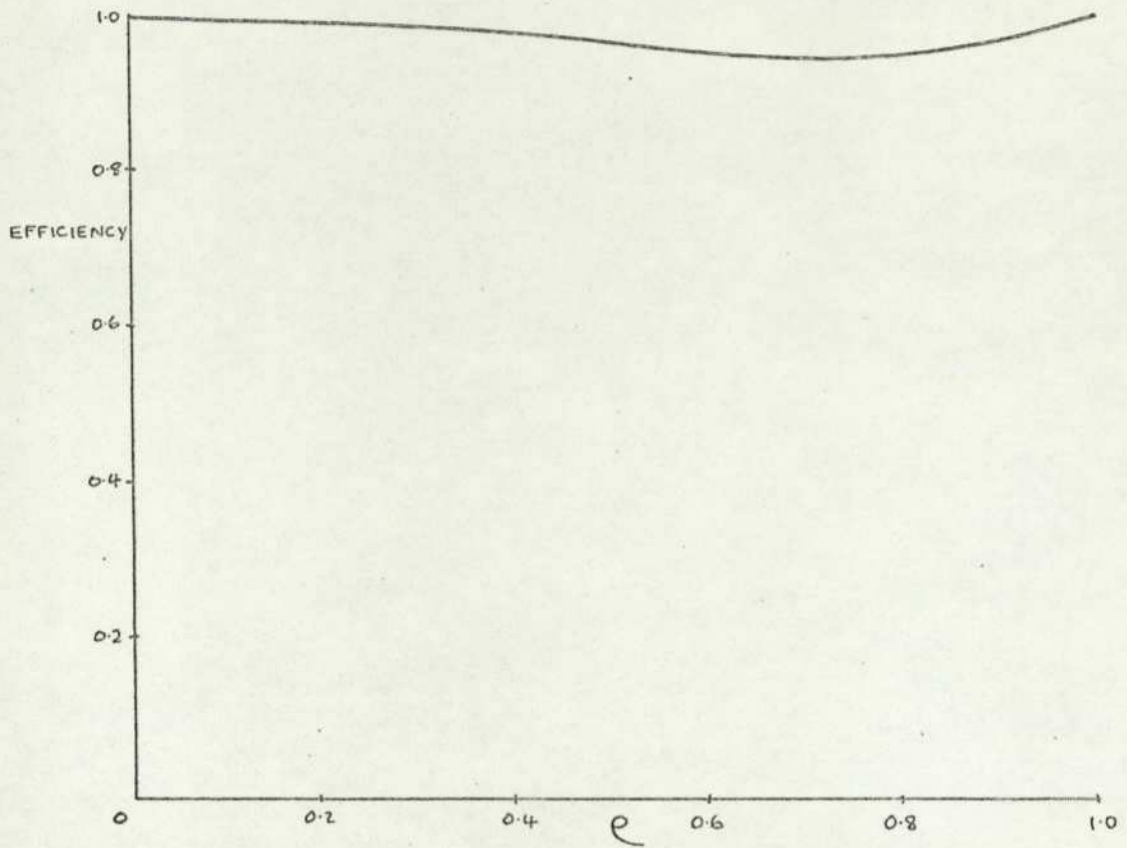
origin and an overall mean is included as a parameter in the regression there appears to be no reason to assume other than $\text{var}(u_i) = \sigma^2$ ($i = 1, 2, \dots, n$), i.e. the stationary autocorrelation process. For high values of autocorrelation, as $\rho \rightarrow 1$, $\mathbf{V}^{-1}(\rho) \rightarrow \mathbf{A}$, and the latent vector of $\mathbf{A}$ associated with its zero latent root is $\mathbf{1}$. Thus use of the latent vectors of $\mathbf{A}$ to represent the trend seems to be consistent with the adoption of the stationary process to represent the error structure.

In any case if the regression vectors representing the trend are taken to be the latent vectors of $\mathbf{A}$ associated with the smallest latent roots the resultant design vectors are the same whether the autocorrelation process is taken to be stationary or non-stationary. Let the n latent vectors of $\mathbf{A}$ be $\mathbf{L} = (\mathbf{L}_0 : \mathbf{L}_1)$, where the r latent vectors used to represent the trend are $\mathbf{L}_0$. Then, for the non-stationary process, the design solution given by (4.4.4) becomes $\mathbf{X}_1 = \mathbf{L}_1 \mathbf{G}_1 \mathbf{P}$, where $\mathbf{G}_1 = \begin{pmatrix} 0 \\ \vdots \\ \mathbf{I}_{k-r} \end{pmatrix}$ and $\mathbf{P}$ is any $(k-r) \times (k-r)$ orthonormal randomization matrix. This can be easily verified by use of the result (2.3.14). Thus the $k-r$ design vectors for treatment levels would be $\mathbf{L}_1^* \mathbf{P}$, where $\mathbf{L}_1^*$ is the $n \times (k-r)$ matrix of latent vectors of $\mathbf{A}$ associated with the $k-r$ largest latent roots. These design vectors are also optimum for the stationary process where the limit as $\rho \rightarrow 1$ of the generalized variance must be considered, and in this case the least squares estimates are fully efficient for $\rho \rightarrow 1$. When $n = 21$ and $r = k = 4$, as in the Box-Hunter example, the efficiency of least squares for various values of ρ is given in fig. 13 for the case when the regressors are latent vectors of $\mathbf{A}$. It is



GRAPH OF THE EFFICIENCY OF LEAST SQUARES
AGAINST ρ IN THE STATIONARY PROCESS
WHEN THE REGRESSION VECTORS ARE
LATENT VECTORS OF $\underline{A}$

FIG.13



GRAPH OF THE EFFICIENCY OF LEAST
SQUARES AGAINST ρ IN THE NON-STATIONARY
PROCESS WHEN THE REGRESSION VECTORS
ARE LATENT VECTORS OF $\underline{B}$

FIG.14

seen that nowhere is the least squares estimator less than 95% efficient, whereas for polynomial regression, fig. 6, the least squares estimator becomes seriously inefficient for high values of autocorrelation.

A similar result obtains for time trends through the origin when the non-stationary autocorrelation process is taken to represent the error structure. In this case the design solution (4.4.4) gives the $k - r$ design vectors as any orthonormal combination of the $k - r$ latent vectors of $\tilde{B}$ associated with the largest latent roots. Again, under these assumptions, the least squares estimator of $\tilde{\beta}$ is fully efficient when $\rho \rightarrow 1$, as well as, of course, in the region $\rho = 0$. Fig. 14 shows the efficiency of the least squares estimator against values of ρ in the non-stationary model when $n = 21$ and the regression vectors are the latent vectors of $\tilde{B}$ associated with the four smallest latent roots.

For the quantitative design problem where the time trend may be represented by latent vectors of $\tilde{A}$ or $\tilde{B}$ the solution for the design vectors is $\tilde{X}_i = \tilde{L}^* \tilde{P}$, where $\tilde{L}^*$ is the $n \times t$ matrix of latent vectors of $\tilde{A}$ or $\tilde{B}$ respectively associated with the largest latent roots and $\tilde{P}$ is any $t \times t$ orthonormal randomization matrix. If all that is required are the estimates of the treatment parameters these are given by $\tilde{P}' \tilde{L}^* \tilde{y}$ and there is no need to establish the value of r , the number of latent vectors to represent the trend.

When the experimental errors are negatively autocorrelated and there is a time trend in the response the design problem is not amenable to solution using the above methods. In this case the latent vectors associated with the smallest latent roots of the dispersion matrices are of high frequency and hence are not appropriate for representing slowly changing trends. However, even if negative autocorrelation was present due to some type of carry-over effect the process of fitting the trend could tend to induce positive autocorrelation. In the results of an experiment given by Box (1952), page 55, he postulates the presence of negative autocorrelation with a fifth order polynomial to represent the trend. However, the $g(-1)$ test does not produce a significant result with this or any degree of polynomial up to ninth order. In fact the g test for positive autocorrelation is significant at the five per cent level for a quadratic fit but not for a cubic fit. The usual F tests also indicate a cubic trend, the sums of squares associated with the fourth and fifth order polynomials not being significant. The g , $g(-1)$ and F values are given in fig. 15. There is no evidence from the $g(-1)$ tests to support Box's assumption of a polynomial trend with negatively autocorrelated errors of the structure considered in this work.

k	$g(1)$	$g(-1)$	F_k^1 value associated with ξ_{k-1}
1	0.634**	2.724	----
2	0.628***	2.715	0.10
3	1.460*	2.318	34.8***
4	2.285	1.631	7.22*
5	2.634	1.175	1.34
6	2.683	1.114	0.40

* indicates significance at 5% level
 ** indicates significance at 1% level
 *** indicates significance at 0.1% level

Results of tests on the Box(1052) example

FIG.15

PART V

AUTOCORRELATED SERIES

5.1 Two simple autocorrelation models

The methods developed earlier are now applied to give some particular results for two simple first-order autocorrelation models. The first is the linear model

$$\underline{y} = \underline{1}\mu + \underline{u} \quad , \quad (1)$$

where $y_i = \mu + u_i$

and $\underline{y} \sim N(\underline{1}\mu, \sigma^2 \underline{V})$.

In this notation the unit vector $\underline{1}$ is not normalized. The usual statistic used for testing for autocorrelated errors in this model is the von Neumann (1941) ratio of the mean square successive difference to the variance, denoted by δ^2/s^2 . We now prove the result that $\{(n-1)/n\} \delta^2/s^2 = d_A = g$ when $\underline{X} = \underline{1}$.
By definition

$$\left(\frac{n-1}{n}\right) \delta^2/s^2 = \frac{\sum_1^{n-1} (y_{i+1} - y_i)^2}{\sum_1^n (y_i - \bar{y})^2} \quad ,$$

$$d_A = \frac{(\underline{y} - \underline{1}\hat{\mu})' \underline{A} (\underline{y} - \underline{1}\hat{\mu})}{(\underline{y} - \underline{1}\hat{\mu})' (\underline{y} - \underline{1}\hat{\mu})}$$

and $g = \frac{(\underline{y} - \underline{1}\tilde{\mu})' \underline{B} (\underline{y} - \underline{1}\tilde{\mu})}{(\underline{y} - \underline{1}\hat{\mu})' (\underline{y} - \underline{1}\hat{\mu})} \quad .$

From 2.3(ii) $g = d_A$ under the condition $\bar{X} = 1$, and $d_A = \{(n-1)/n\} \delta^2 / s^2$ because $\sum_{i=1}^{n-1} (y_{i+1} - y_i)^2 = \bar{y}' A \bar{y}$ and $(\bar{y} - \bar{1}\bar{\mu})' A (\bar{y} - \bar{1}\bar{\mu}) = \bar{y}' A \bar{y}$, $\bar{1}' A$ being 0. Thus, from section 2.5, the von Neumann ratio has the properties of the g statistic, when used for testing against stationary and non-stationary one-sided alternatives, of being the most powerful invariant test in the region $\rho = 1$ and an UMP invariant test for alternatives represented by close approximations to the respective density functions. The significance points of this statistic have been tabulated by Owen (1962).

If a test for negative autocorrelation is required the most powerful invariant test against the non-stationary alternative of $\rho = -1$ is obtained by using the $g(-1)$ statistic, as in section 2.8. The result does not extend to the case of ^{the} stationary process because the latent vector associated with the zero latent root of the inverse of the dispersion matrix is not included as a design vector. The critical values of this $g(-1)$ test, for a significance level of 5%, are given in table 17 with values of $n = 10(1)30(5)50(10)100$.

An even simpler structure than (1) is obtained when the mean μ is zero and the observations form a non-stationary first-order autoregressive process, i.e. $\mu = 0$ and $y_i = u_i$. Thus

$$y_i = \rho y_{i-1} + \varepsilon_i, \quad (i = 2, 3, \dots, n) \quad (2)$$

$$y_1 = \varepsilon_1,$$

$$\text{and} \quad \text{var}(\varepsilon_i) = \sigma^2, \quad (i = 1, 2, \dots, n).$$

n	Critical value		n	Critical value
10	0.814		25	1.258
11	0.869		26	1.273
12	0.916		27	1.288
13	0.959		28	1.301
14	0.997		29	1.314
15	1.032		30	1.326
16	1.064		35	1.379
17	1.092		40	1.422
18	1.119		45	1.457
19	1.143		50	1.487
20	1.166		60	1.534
21	1.187		70	1.571
22	1.206		80	1.601
23	1.225		90	1.625
24	1.242		100	1.645

The 5 % significance points of the $g(-1)$ test

when an overall mean only is fitted, i.e. $\bar{X} = \bar{y}$

TABLE 17

By the Kadiyala (1970) - Lehmann and Stein (1948) result, the statistic

$$g(\rho^*) = \frac{\tilde{y}' \tilde{V}^{-1}(\rho^*) \tilde{y}}{\tilde{y}' \tilde{y}} \quad (3)$$

gives the most powerful test of the hypothesis $H_0 : \tilde{y} \sim N(0, \sigma^2 I)$ against the alternative $H_1 : \tilde{y} \sim N(0, \sigma^2 V(\rho^*))$. When $\rho^* = 1$ the alternative being tested is of course the simple random walk. For the close approximation to the non-stationary density given by (2.2.6) the $g(1) = g$ statistic above gives an UMP test against the one sided alternative of positive autocorrelation. Similar results obtain for tests against one-sided alternatives of negative autocorrelation with the statistic $g(-1)$. The significance points of the statistics g and $g(-1)$ here are identical because of equality of the latent roots of B and $\tilde{V}^{-1}(-1)$, which was proved in section 4.5. The 5% critical values are given in table 18. The Kadiyala (1970) - Lehmann and Stein (1948) result cannot be applied to the stationary process for the limit as $\rho \rightarrow 1$ because, at the limit, the inverse of the corresponding dispersion matrix (U^{-1} in section 3.5) and the likelihood ratio do not exist.

The likelihood function of the non-stationary process (2) is

$$L = (2\pi\sigma^2)^{-n/2} \cdot \exp \left(-\frac{\tilde{y}' \tilde{V}^{-1} \tilde{y}}{2\sigma^2} \right). \quad (4)$$

n	Critical value		n	Critical value
4	0.54		23	1.31
5	0.66		24	1.32
6	0.74		25	1.33
7	0.82		26	1.35
8	0.88		27	1.36
9	0.93		28	1.37
10	0.98		29	1.38
11	1.02		30	1.39
12	1.06		35	1.43
13	1.09		40	1.47
14	1.12		45	1.50
15	1.15		50	1.53
16	1.18		60	1.57
17	1.20		70	1.60
18	1.22		80	1.63
19	1.24		90	1.65
20	1.26		100	1.66
21	1.28		150	1.73
22	1.29		200	1.76

The 5 % significance points of the g and $g(-1)$
tests for autocorrelation in an observed
sequence with zero mean

TABLE 18

Solution of $\partial \ln L / \partial \rho = 0$ gives the maximum likelihood estimator of ρ as

$$\tilde{\rho} = \frac{\sum_{i=1}^{n-1} y_i y_{i+1}}{\sum_{i=1}^{n-1} y_i^2}.$$

This statistic differs from $1 - \frac{1}{2}g$ only by $\frac{1}{2}y_n^2$ in the numerator and y_n^2 in the denominator. However $\tilde{\rho}$ can take any value while $1 - \frac{1}{2}g$ lies strictly between -1 and 1 . If the problem is regarded as one of predicting y_{n+1} given $y_1, y_2, \dots, y_n$, the posterior distribution of y_{n+1} given the data y , the likelihood (4) and vague (Jeffreys, 1948) prior knowledge of ρ (on $-\infty, \infty$) and σ^2 , is

$$p(y_{n+1} | y) \propto \left(\sum_{i=1}^n y_i^2 \right)^{-\frac{1}{2}} \cdot \left(\sum_{i=1}^{n+1} y_i^2 - \frac{\sum_{i=1}^n y_i y_{i+1}}{\sum_{i=1}^n y_i^2} \right)^{-n/2}.$$

Taking the mode of this posterior distribution as an estimate of y_{n+1} gives the simple result $\tilde{\rho} y_n$.

Appendix A

In experimental design models the design matrices are typically not of full rank and difficulties arise in computer applications because the $\tilde{X}'\tilde{X}$ and $\tilde{X}'\tilde{V}^{-1}\tilde{X}$ matrices are singular. The usual solutions to this problem are either to use generalized inverses or to consider only estimable functions of the parameters. In this work we have used a simple method of reducing the design matrix to full rank which was suggested by Berenblut (1968). The usual restrictions on the estimates of the parameters are incorporated into the reduced design matrix.

Taking the Williams design, section 1.2 , as an illustrative example the parameter vector is

$$\tilde{\beta} = (\mu, s_1, s_2, s_3, s_4, s_5, p_1, \dots, p_5, t_1, \dots, t_5, s_6, p_6, t_6)^T,$$

where μ = overall mean ,
 s_i = subject i effect,
 p_i = period i effect ,
 t_i = treatment i effect .

The design matrix associated with these parameters is not of full rank. The parameters have been ordered without loss of generality, so that the first 16 vectors form a basis for the design matrix. The procedure may be summed up as follows: by consideration of the usual restrictions on the estimates of the parameters, viz.

$\sum \hat{s}_i = \sum \hat{p}_i = \sum \hat{t}_i = 0$, the original design matrix $\tilde{X}$ may be replaced by the alternative $\tilde{X}^*$, given in fig. 1, which is of full rank. The parameters estimated by use of $\tilde{X}^*$ in conjunction with the above restrictions are

$$(\mu \quad s_1 \quad \dots \quad s_5 \quad p_1 \quad \dots \quad p_5 \quad t_1 \quad \dots \quad t_5) \quad . \quad (1)$$

To verify these statements first consider the model

$$E(\tilde{y}) = \tilde{X}\tilde{\beta} = \tilde{X}\tilde{G}\tilde{G}^{-1}\tilde{\beta} = (\tilde{X}^* : 0)' \tilde{\beta}^* \quad ,$$

where $\tilde{\beta}^* = (\mu^* \quad s_1^* \quad \dots \quad s_5^* \quad p_1^* \quad \dots \quad p_5^* \quad t_1^* \quad \dots \quad t_5^*)^T$.

The matrix $\tilde{G}$ is chosen so that $\tilde{X}\tilde{G} = (\tilde{X}^* : 0)$. Evaluation of $\tilde{G}^{-1}$ produces $\tilde{\beta}^*$ with elements

$$\mu^* = \mu + \sum s_i/6 + \sum p_i/6 + \sum t_i/6$$

$$s_i^* = (5s_i - \sum_{j \neq i} s_j)/6 \quad , \quad i = 1, 2, \dots, 5 \quad ,$$

$$s_6^* = \sum s_i/6 \quad ,$$

and p and t of similar form.

In this reparameterization it is clear that s_6^* , p_6^* and t_6^* cannot be estimated. Imposing the restrictions $s_6^* = p_6^* = t_6^* = 0$ gives $\hat{s}_i^* = \hat{s}_i$, $\hat{p}_i^* = \hat{p}_i$ and $\hat{t}_i^* = \hat{t}_i$, $i = 1, 2, \dots, 5$.

The least squares or minimum variance estimates for (1) may now

be obtained using the $\tilde{X}^*$ matrix. Estimates for s_6 , p_6 and t_6 are taken as $-\sum_{i=1}^5 \hat{s}_i$, $-\sum_{i=1}^5 \hat{p}_i$, $-\sum_{i=1}^5 \hat{t}_i$ respectively.

Appendix B

It was mentioned in section 3.5 that the Cochrane-Orcutt (1949) method of using the $(n - 1)$ first differenced variables cannot estimate an overall mean. Kadiyala (1966), page 16, considers the stationary autocorrelation model and the special case when $\tilde{X} = \underline{1}$. He finds that the Cochrane-Orcutt estimator has a limiting efficiency of zero as ρ approaches 1 when compared to the least squares estimator. The result is due to the fact that taking first differences of a column of ones leaves a column of zeros. Therefore the transformed model cannot estimate the constant term. We have seen from section 3.5 that the estimator $\tilde{\beta}(1)$ has the same variance as the minimum variance estimator as $\rho \rightarrow 1$ for the stationary model. The least squares estimator has this same limiting variance. From this one special case Kadiyala (1966) conjectures that it does not pay to make the Cochrane-Orcutt (1949) transformation when observations on the independent variables are 'close' for successive time periods. This conjecture certainly does not hold for the transformation suggested in section 3.5, even for the special case above. Fig. 6 gives an illustration of this fact for the case of polynomial regression.

REFERENCES

- ABRAHAMSE, A.P.J. and KOERTS, J. (1969). A comparison between the powers of the Durbin-Watson test and the power of the BLUS test.
J. Amer. Statist. Assoc., 64, 938-948.
- ABRAHAMSE, A.P.J. and LOUTER, A.S. (1971). On a new test for autocorrelation in least squares regression.
Biometrika, 58, 53-60.
- AITKEN, A.C. (1948). On the estimation of many statistical parameters.
Proc. Roy. Soc. Edinburgh, 62, 369.
- ANDERSON, T.W. (1948). On the theory of testing serial correlation.
Skand. Akt., 31, 88-116.
- ANDERSON, T.W. (1971). The statistical analysis of time series.
New York, Wiley.
- BARTH, W., MARTIN, R.S. and WILKINSON, J.H. (1967).
Calculation of the eigenvalues of a symmetric tridiagonal matrix by the method of bisection.
Numerische Mathematik, 9, 386-393.
- BELLMAN, R. (1960). Introduction to matrix analysis.
New York, McGraw-Hill.
- BERENBLUT, I.I. (1968). Experimental designs balanced for specified contrasts between residual effects.
Ph.D. thesis, University of London.
- BERENBLUT, I.I. and WEBB, G.I. (1972). First order designs in the presence of a time trend.
Paper presented at the European Meeting of Statisticians, Budapest.
- BERENBLUT, I.I. and WEBB, G.I. (1973). A new test for autocorrelated errors in the linear regression model.
J. Roy. Statist. Soc., (B), 35.
- BOX, G.E.P. (1950). Problems in the analysis of growth and wear curves.
Biometrics, 6, 362-389.

- BOX, G.E.P. (1952). Multifactor designs of first order.
Biometrika, 39, 49-57.
- BOX, G.E.P. (1960). Some general considerations in process optimization.
Trans. Amer. Soc. Mech. Eng., D, 82, 113-119.
- BOX, G.E.P. and GUTTMANN, I. (1966). Some aspects of randomization.
J. Roy. Statist. Soc. (B), 28, 543.
- BOX, G.E.P. and HUNTER, J.S. (1958). Experimental designs for the exploration and exploitation of response surfaces. From "Experimental Designs in Industry", edited by V. Chew, New York, Wiley.
- BOX, M.J. (1972). Contribution to the discussion of papers read by Wynn and Laycock.
J. Roy. Statist. Soc., (B), 34, 172.
- BOX, M.J. and DRAPER, N.R. (1971). Factorial designs, the $|X'X|$ criterion, and some related matters.
Technometrics, 13, 731-742.
- CHAMPERNOWNE, D.G. (1948). Sampling theory applied to autoregressive sequences.
J. Roy. Statist. Soc., (B), 10, 204-242.
- CHERNOFF, H. (1953). Locally optimal designs for estimating parameters.
Ann. Math. Stats., 24, 586-602.
- COCHRAN, W.G. and COX, G.M. (1957). Experimental designs.
New York, Wiley.
- COCHRANE, D. and ORCUTT, G.H. (1949). Application of least squares regression to relationships containing autocorrelated error terms.
J. Amer. Statist. Assoc., 44, 31-61.
- DE GROOT, M.H. (1970). Optimal statistical decisions.
New York, McGraw-Hill.

- DURBIN, J. (1970). An alternative to the bounds test for testing for serial correlation in least squares regression. *Econometrica*, 38, 422-429.
- DURBIN, J. and WATSON, G.S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, 37, 409-428.
- DURBIN, J. and WATSON, G.S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, 38, 159-178.
- DURBIN, J. and WATSON, G.S. (1971). Testing for serial correlation in least squares regression. III. *Biometrika*, 58, 1-19.
- FINNEY, D.J. (1956). Cross-over designs in bioassay. *Proc. Roy. Soc., (B)*, 145, 42-61.
- FISHER, R.A. and YATES, F. (1953). Statistical tables for biological, agricultural and medical research. (4th edition). Edinburgh, Oliver and Boyd.
- HANNAN, E.J. (1960). Time series analysis. London, Methuen.
- HANNAN, E.J. (1970). Multiple time series. New York, Wiley.
- HENSHAW, R. (1966). Testing single-equation least squares regression models for autocorrelated disturbances. *Econometrica*, 34, 646-660.
- HILDRETH, C. and LU, J.Y. (1960). Demand relations with autocorrelated disturbances. Technical Bulletin No. 276, Agricultural Experiment Station, Michigan State University.
- HOOS, S. and SHEAR, S.W. (1942). Relation between auction prices and supplies of California fresh Bartlett pears. *Hilgardia*, 14, No. 5, California Agr. Expt. Sta.
- IMHOF, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419-426.

- JEFFREYS, H. (1948). Theory of probability. (2nd edition)
Oxford, Clarendon Press.
- KADIYALA, K.R. (1966). Regression analysis with autocorrelated
disturbances.
Ph.D. thesis, University of Minnesota.
- KADIYALA, K.R. (1970). Testing for the independence of regression
disturbances.
Econometrica, 38, 97-117.
- KIEFER, J. (1959). Optimum experimental designs.
J. Roy. Statist. Soc., (B), 21, 272-319.
- KOERTS, J. and ABRAHAMSE, A.P.J. (1968). On the power of the
BLUS procedure.
J. Amer. Statist. Assoc., 63, 1227-1236.
- LEHMANN, E.L. (1959). Testing statistical hypotheses.
New York, Wiley.
- LEHMANN, E.L. and STEIN, C. (1948). Most powerful tests of
composite hypotheses. I. Normal distributions.
Ann. Math. Stats., 19, 495-516.
- MANDEL, J. (1964). The statistical analysis of experimental data.
New York, Wiley.
- VON NEUMANN, J. (1941). Distribution of the mean square successive
difference to the variance.
Ann. Math. Stats., 12, 367-395.
- VON NEUMANN, J., KENT, R.H., BELLINSON, H.R., and HART, B.I. (1941).
The mean square successive difference.
Ann. Math. Stats., 12, 153-162.
- OWEN, D.B. (1962). Handbook of statistical tables.
Massachusetts, Addison-Wesley.
- PAN JIE-JIAN (1968). Distribution of the non-circular serial
correlation coefficients.
Amer. Math. Soc. and Inst. Math. Statist.
Selected Transactions in Probability and Statistics,
7, 281-291.

- RAND (1955). A million random digits with 100,000 normal deviates.
The RAND Corporation. Illinois, Free Press.
- RAO, C.R. (1965a). Linear statistical inference and its applications.
New York, Wiley.
- RAO, C.R. (1965b). Least squares theory using an estimated dispersion matrix and its application to the measurement of signals.
Fifth Berkeley Symposium.
- STONE, R. (1948). The analysis of market demand.
Review of the International Statistical Institute.
- THEIL, H. and NAGAR, A.L. (1961). Testing the independence of regression disturbances.
J. Amer. Statist. Assoc., 56, 793-806.
- TOCHER, K.D. (1952). A note on the design problem.
Biometrika, 39, 189.
- WALD, A. (1943). On the efficient design of statistical investigations.
Ann. Math. Stats., 14, 134-140.
- WALLIS, K.F. (1972). Testing for fourth order autocorrelation in quarterly regression equations.
Econometrica, 40.
- WATSON, G.S. (1955). Serial correlation in regression analysis.
Biometrika, 42, 327-341.
- WATSON, G.S. (1967). Linear least squares regression.
Ann. Math. Stats., 38, 1679-1699.
- WILLIAMS, E.J. (1949). Experimental designs balanced for the estimation of residual effects of treatments.
Austral. J. Sci. Res., (A), 2, 149-168.
- WILLIAMS, R.M. (1952). Experimental designs for serially correlated observations.
Biometrika, 39, 151-167.
- ZELLNER, A. and TIAO, G.C. (1964). Bayesian analysis of the regression model with autocorrelated errors.
J. Amer. Statist. Assoc., 59, 763-778.