AUTOCORRELATION ANALYSIS AND THE DESIGN OF EXPERIMENTS

BY
G. M. JENKINS AND J. CHANMUGAM

TECHNICAL REPORT NO. 2
AUGUST 2, 1960

PREPARED UNDER CONTRACT DA-04-200-ORD-996
FOR
OFFICE OF ORDNANCE RESEARCH

APPLIED MATHEMATICS AND STATISTICS LABORATORIES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AUTOCORRELATION ANALYSIS AND THE DESIGN OF EXPERIMENTS

by

G. M. JENKINS AND J. CHANGMUGAM*

TECHNICAL REPORT NO. 2
August 2, 1960

PREPARED UNDER CONTRACT DA-04-200-ORD-996
FOR
OFFICE OF ORDNANCE RESEARCH

* A portion of this work was performed under Office of Ordnance Research Contract No. DA 36-034-ORD-2297.
Princeton University, Princeton, N. J.

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

APPLIED MATHEMATICS AND STATISTICS LABORATORIES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
# TABLE OF CONTENTS

**SUMMARY**

**PART I: BASIC PRINCIPLES AND STATEMENT OF PROBLEM**

Section 1. The Role of Randomisation in Experimental Design: Systematic Designs 1

Section 2. An Industrial Problem in which the Errors are Correlated 4

Section 3. General Statement of Problem 8

Section 4. An Exact Form for the Variance of the Difference between Treatment Means 12

**PART II: DETAILED DERIVATIONS**

Section 5. Determination of the Optimum Block Size when the Autocorrelation Function has a Finite Cut-Off 15

Section 6. Determination of the Optimum Block Size when the Autocorrelation Function is of the Form (2,3) 22

Section 7. Likelihood Function Approach 29

Section 8. An Alternative Interpretation in Terms of the Spectrum and Signal to Noise Ratios 35

**PART III: APPLICATIONS**

Section 9. An Illustrative Example 44

Section 10. Some Further Considerations 49

**APPENDICES**

Appendix 1. Conditions on the Autocorrelations of Stationary Moving Average Processes 54

Appendix 2. Likelihood Ratio Criteria for the Modified Model 56

**ACKNOWLEDGEMENTS**

**REFERENCES** 65
SUMMARY

These are many examples in industry and in the sciences of situations where the uncontrolled variation in the observed responses is subject to autocorrelation when the independent variables are kept constant.

The present work is concerned with the investigation into the effect of this autocorrelation on the strategy of experimentation. One example, which is described in this report, is in the design of experiments for a chemical process where the yields from consecutive runs are negatively correlated. A simple physical model for this behavior has been constructed and this turns out to be an autoregressive process with coefficients which damp out quickly.

It has been shown that a considerable increase in efficiency results if a systematic design of experiments is used instead of randomising the order in which the experiments are performed. In the case of experiments conducted at two levels, systematic designs have been considered in which blocks of \( n \) consecutive experiments are conducted alternately at each level. If there are \( 2n \) blocks in all, then the total number of experiments \( N = 2bn \).

In the simplest situation, it is assumed that \( N \) is specified in advance from considerations of cost and the problem reduces to one of choosing the optimum value of \( n \). This approach may be replaced, for example, by one in which there is a specification on the variance of the estimated difference between the response at the two levels. In all cases, it turns out that the optimum value of \( n \) is determined by the autocorrelation structure of the process and that to a high degree of approximation, this choice can be made independently of the value of \( N \). Formulae for \( n_{opt} \) have been derived for various autocorrelation patterns.
By associating a frequency with the change from one level to the other, it has been shown that the problem may be regarded as one of detection of signal in noise and that asymptotically, choosing $n_{opt}$ is equivalent to minimising the spectral density.

These ideas can of course be extended to multi-factor experiments and also to the important case of regression problems with special reference to the estimation of the parameters of a response surface.

The report is divided into three main parts. Part I is concerned with a discussion of the basic principles involved and a statement of the problem. In Part II and the appendices, the necessary theory and formulae are developed in detail. Part II may be omitted by non-statisticians at first reading. The application of these results to a typical example and some of the practical implications are described in Part III.
PART I: BASIC PRINCIPLES AND STATEMENT OF PROBLEM

1. The Role of Randomisation in Experimental Design: Systematic Designs.

Some General Remarks on Randomisation

The introduction of the concept of randomisation in 1935 by R. A. Fisher [1] has had far reaching repercussions both on the application of statistical methods and in the development of statistical theory. A summary of the basic reasons for randomising the allocation of treatments to experimental units has been given recently by D. R. Cox [2] who distinguishes between three main motives:

1. To remove systematic errors (i.e. those not eliminated in a long series of experiments) in the estimated treatment differences.

2. To provide a valid estimate (i.e. one whose precision is known) of the error in the analysis of variance and also to justify this analysis under wide assumptions about the distribution of the uncontrolled variation.

3. As a device for concealment e.g. in situations where a knowledge of which treatments have been applied to various experimental units might conceivably influence the responses.

G. E. P. Box in a series of papers (see [3]) has extended the use of randomisation, introducing what he has termed the concept of angular randomisation. Box's approach may be summarised as follows:

1. Optimum designs are those for which the vectors of the design matrix \( (N \times k) \) (if there are \( N \) experiments and \( k \) quantitative factors) are mutually orthogonal and orthogonal to the vector consisting entirely of ones. This follows from a result due to R. L. Burman and J. P. Plackett [4] and leads to a number of designs possessing minimum variance but which differ by a rotation in the design space.
(2) The treatment vectors may be further orthogonalized with respect to systematic effects such as time and space trends.

(3) In order to avoid systematic biases due to other effects not known and also to justify the analysis under very wide distribution assumptions, the treatment vectors are chosen at random in the residual sub-space of dimension \( n - j \) where \( j \) is the number of vectors used to eliminate the trend. An interesting application of this technique has been given by G. E. P. Box and W. A. Hay [5]. It is to be noted that this approach leads to direct estimation of systematic effects under an assumed model for the form of the trend - if elimination of systematic effects is all that is required, then classical blocking techniques together with partial confounding within blocks may be used. In general, however, designs which are orthogonal with respect to unwanted systematic effects will maximise the efficiency of both trend and treatment effects.

**Randomisation and Autocorrelation**

Associated with the physical act of randomisation has grown up the notion of randomisation or permutation test (see e.g. O. Kempthorne [6] and E. Lehmann [7]). In these, the inference is based upon an internally generated reference set obtained by permuting the observations subject to various algebraic restrictions imposed by the combinatorial properties of the design. These tests have the property that they are insensitive (robust) with respect to fairly wide assumptions about the distribution of the uncontrolled variation.

It has been shown by G. E. P. Box and S. L. Andersen [8] that under certain conditions, these tests break down in the sense that if a significance level \( \alpha \) is derived from the randomisation distribution, then it is not correct that the null hypothesis is rejected when true on \( \alpha \% \) of occasions. This occurs when the joint probability density function of the random variables
associated with the sample values is not a symmetric function of these variables. Examples when this arises are:

(i) when the variables are non-identically distributed
(ii) when the variables are identically distributed but auto-correlated.

It is to be observed that if we can be sure that this joint density is a symmetric function, then we can apply the permutation test with confidence even if no randomisation has been conducted in practice. However, as pointed out repeatedly by Fisher, the physical act of randomisation will always result in a symmetrical distribution.

Hence in situations where the uncontrolled variation is autocorrelated, one way of ensuring that the permutation test is valid would be to randomise the allocation of treatments to experimental units and thereby effectively destroy the autocorrelation in the observations. The issue then arises as to whether this is an efficient procedure or not. The only alternative to randomisation is to employ a systematic design.

Systematic Designs:

This name has been given by D. R.Cox ([9],[10]) to designs in which it is expedient not to randomise at all except perhaps in the naming of the treatments. Cox advocates the use of such designs in the case of small experiments when there are smooth time or space trends present in the uncontrolled variation. The fact that the validity of the estimate of error is jeopardised by the absence of randomisation is justified by the fact that for small experiments the estimate of error is imprecise anyway whilst the use of randomisation is likely to yield an arrangement which may result in serious biases due to the presence of trend. There is very little comfort, therefore, in knowing that 'in the long run' it is advantageous to randomise if it might
ruin the only one (or few) small experiment which we are able to conduct.

In this report, an investigation is made of the properties of simple systematic designs which would be used in arbitrarily large experiments. It is shown that these designs result in a considerable increase in efficiency over randomisation if there is autocorrelation present. Since randomisation yields procedures which are essentially distribution free, one price which has to be paid in using systematic designs is that it is necessary to assume that the error distribution is in fact normal, or as is more akin to reality, that it may be described in terms of its mean and variance without too much discomfort.

2. An Industrial Problem in which the Errors are Autocorrelated

In this section, a description is given of an industrial problem which motivated the present study.

It is well known that in many production processes that even when an attempt is made to keep all the known independent variables constant, the uncontrolled variation from experiment to experiment is subject to low frequency drifts in the mean level (trends) and also to auto (serial) or time correlation. The latter may be present when there is no change in mean level and we now proceed to describe one such instance in which a simple and necessarily approximate physical explanation may be given for the observed correlation pattern of the responses of a production process. In this particular problem, the latter takes the form of negative autocorrelation between consecutive observations.

An attempt was made to determine the physical basis for the existence of this correlational pattern. It was evident from a study of the process that the dominant characteristics arose in that section of the process where the product was separated from the other constituents. In the manufacture of a
specific chemical compound, the entire contents of a run from a batch type reactor are transferred into a batch type distillation unit where the compound in question is separated. The yield of the distillate is the response of interest. The distillation unit consists essentially of a still, where the material is deposited initially for heating, and a vertical column along which "plates" or equilibrating surfaces are distributed axially. By maintaining a temperature gradient from top to bottom in the column, it is possible to separate out the distillate which boils below a given temperature. For purposes of a somewhat elementary description, the material fed into the distillation unit will be assumed to consist of a distillable portion and a non-distillable portion.

Thus, initially, the content \( C \) of the still is made up of an amount \( C - D \) of non-distillable material and a quantity \( D \) of distillable material. Distillation is not one hundred per cent efficient so that only a proportion \( k_1 \) of the distillable material \( D \) in the \( i \)th batch is distilled. At the end of the run, the distillation unit contains the remaining distillable material of amount \( D(1 - k_1) \) and the non-distillable material of amount \( (C - D) \). At the end of each distillation, a fraction \( (1 - \lambda_1) \) of the remaining contents of the unit is discharged as waste leaving behind an amount \( \lambda_1(1 - k_1)D \) of distillable material and \( \lambda_1(C - D) \) of non-distillable material. It is this fact which is responsible for the negative serial correlation in consecutive responses. A new batch of material from the \((i + 1)\)th run in the batch reactor is then transferred into the still for the beginning of the \((i + 1)\)th distillation run but the material which is retained in the distillation unit is ordinarily not removed, since it is neither economical nor desirable to wash out the column after each run. This means that the amount of each component present in the distillation unit at the beginning of the \((i + 1)\)th run,
assuming that an equal amount \( C \) is fed in at each stage, will be

\[
\begin{array}{c|c|c}
\text{Distillable} & \text{Non-distillable} \\
\hline
\text{Feed for the } (i+1)^{\text{th}} \text{ run} & D & C - D \\
\hline
\text{Retained from } i^{\text{th}} \text{ run} & \lambda_i (1 - k_i) D & \lambda_i (C - D) \\
\hline
\text{Total} & D(1 + \lambda_i (1 - k_i)) & (C - D)(1 + \lambda_i) \\
\end{array}
\]

It is easily verified that the contents of the distillation unit at various stages in the first two runs are as given in the following table*:

<table>
<thead>
<tr>
<th>Initial contents of Distillation unit</th>
<th>Yield (response) of distillate</th>
<th>Retained at the end of run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distillable</td>
<td>Non-distillable</td>
<td></td>
</tr>
<tr>
<td>Run 1</td>
<td>1</td>
<td>( k_1 )</td>
</tr>
<tr>
<td>Run 2 ( 1 + \lambda_1 (1 - k_1) )</td>
<td>( 1 + \lambda_1 )</td>
<td>( k_2 (1 + \lambda_1 (1 - k_1)) ) ( \lambda_2 ) (1 + ( \lambda_1 )) ( \lambda_2 (1 - k_1) (1 + \lambda_1 (1 - k_1)) )</td>
</tr>
</tbody>
</table>

[*These quantities are expressed as fractions of \( D \) in case of distillable material and of \( (C - D) \) in the case of non-distillable material.]

If \( Y_1 \) is the yield or response at the completion of the \( i^{\text{th}} \) run, then it is apparent that \( Y_2 \), the yield in the second run, is related to \( Y_1 \) by means of the expression

\[
Y_2 = D k_2 (1 + \lambda_1) - k_2 \lambda_1 Y_1
\]

and in general, for the response of the \( n^{\text{th}} \) run, it follows that

\[
Y_n = D k_n (1 + \lambda_1 \lambda_2 \cdots \lambda_{n-1}) - k_n \lambda_{n-1} Y_{n-1} - k_n \lambda_{n-1} \lambda_{n-2} Y_{n-2} \cdots \cdots \cdots - k_n \lambda_1 \lambda_2 \cdots \lambda_{n-1} Y_1
\]
If we now make the not unreasonable assumption that \( k_1 = k \) and \( \lambda_1 = \lambda \), it follows that

\[
Y_n = Dk(1 + \lambda^{n-1}) - k\lambda Y_{n-1} - k\lambda^2 Y_{n-2} \ldots \ldots k\lambda^{n-1} Y_1
\]

(2.1)

Clearly, terms such as \( k, \lambda, C, D \) etc. will be subject to random errors at each stage and these we amalgamate in one error term \( z_n \) so that the observed response \( y_n = Y_n + z_n \). For values of \( k \) and \( \lambda \) which are realistic in practice, the coefficients of the difference equation (2.1) will damp out fairly quickly so that it is reasonable to take a model which to first order approximation reads as

\[
y_n = Dk - k\lambda Y_{n-1} - k\lambda^2 Y_{n-2} \ldots \ldots k\lambda^m Y_{n-m} + z_n
\]

(2.2)

This model is termed an autoregressive scheme by statisticians; it has the desirable feature, however, that the coefficients damp out fairly quickly. To engineers the model corresponds to a representation of the uncontrolled variation as white noise (the \( z_n \) passed through a linear filter. Its autocorrelation function (abbreviated to a.c.f.) \( \rho_s = E \left\{ \left[ y_n - E(y_n) \right] \left[ y_{n+s} - E(y_{n+s}) \right] \right\} / \sigma^2 \) where \( \sigma^2 \) is variance of the observations, may be written in the form

\[
\rho_s = A_1 \pi_1 |s| + A_2 \pi_2 |s| + \ldots + A_m \pi_m |s|
\]

(2.3)

for \( s = 0, \pm 1, \pm 2, \ldots \), if the effects of the "transients" in (2.2) are assumed to have damped out. In (2.3) the \( \pi's \) are either real or else imaginary in conjugate pairs, all with moduli less than unity if the \( \{y_n\} \) series is to be stationary. Equation (2.3) corresponds to an a.c.f. of quite general nature.

In the example to be discussed in Section 9, it was found that the observed a.c.f. could be graduated to a high degree of accuracy by means of a sum of two damped exponentials. It may thus be concluded that as far as this
process is concerned, the response is dependent on the response of the two previous runs but that any dependence further back in time is sufficiently small to be neglected.

We now proceed to discuss the general question of designing experiments with a view to exploiting an autocorrelation pattern of the form (2.3).

3. General Statement of Problem

For simplicity it is assumed that only a single factor is varied and that it is varied between two levels in any one experiment. In the situation described in Section 2, experiments might be conducted at two temperature levels $T_1$ and $T_2$ of the reactor, all the other known independent variables being maintained at constant levels. This is sufficient to bring out the general principles involved; at a later stage it is proposed to extend these results to more general situations (for further discussion, see Section 10). It should be realised, however, that elaborate designs are not practical in large scale experimentation of this type and that very often, replicated $2^n$ or $3^n$ experiments which explore local behaviour, are all that is practical.

It is assumed that when conditions are maintained constant the uncontrolled variation $\{z_i\}$ constitute a stationary time series with known variance $\sigma^2$ and known autocorrelation function $\{\rho_s\}$, $s = 0, \pm 1, \pm 2 \ldots \ldots \ldots$. It is assumed further that at each set of experimental conditions, the statistical properties (assumed to be adequately described by the variance and a.c.f.) of the uncontrolled variation are the same. This is a reasonable practical assumption provided that the maximum separation of the levels is not too great. For the present it is assumed that $\sigma^2$ and $\{\rho_s\}$ are known; in Section 10, we discuss how this problem is modified when $\sigma^2$ and $\{\rho_s\}$ have to be estimated in advance from finite samples or realisations.
The problem to be considered may be regarded formally from one of the two following points of view.

(1) The size of the experiment, $N$, is fixed in advance from considerations of cost, disruption of normal plant working, etc. Given $N$, $\sigma^2$ and $\{\rho_s\}$, it is desired to estimate the difference between the treatment means with maximum precision. Instead of randomising the allocation of treatments to runs, a systematic design is employed in which $n$ consecutive runs are made alternately at each of the two levels. If the number of blocks of $n$ consecutive runs at each of the two levels is $b$, then $N = 2bn$. Given this arrangement, it is required to choose that value of $n$ which results in minimum variance of the estimated treatment differences. It is not claimed that designs in which the number of runs per block is a constant are optimum. However, as will be shown in later sections, little is lost in the way of efficiency by restricting ourselves to this class.

(2) Alternatively, the standard error of the difference in means may be specified in advance either from considerations of the width of the confidence interval required or else from the power of a test of significance (preferably the first of these since this is a problem of estimation of effects which are known to exist in advance from theoretical considerations). It will be shown subsequently that the important quantity in this problem is the ratio $\Omega = 2tb/N$, the optimum value of $\Omega$ being determined entirely by the $\{\rho_s\}$. In general, the standard error of the treatment differences will be a function of $\sigma^2$, $\{\rho_s\}$, $\Omega$ and $N$. Given $\{\rho_s\}$, $\Omega$ may be chosen so that this standard error is a minimum for any value of $N$. 
Given \( \sigma^2 \), \([\rho_s]\) and \( \Omega_{opt} \), then \( N \) may be chosen so that this minimal variance will be equal to a specified value. From now on we shall be concerned entirely with the choice of \( \Omega_{opt} \), leaving the decision as to whether to fix \( N \) in advance or to choose it to meet a specification on the standard error as a separate issue. It follows also, that double sample or sequential procedures could be devised for this problem without much difficulty.

The only other situation, as far as we are aware, in which considerations such as these have been used is some empirical work in agricultural trials by H. Fairfield Smith [11]. In order to show clearly the relation between this work and ours, we state a typical situation with which Fairfield Smith was concerned.

The \( n \) runs in a block in our example correspond to \( n \) plots of land on a block of land in which \( m \) different treatments are to be compared. In the simplest of situations, each treatment would occur an equal number of times, say \( r \), in each block so that \( n = rm \). As opposed to our example, treatments are allocated at random within blocks in order to avoid systematic biases in treatment comparisons. However, due to fertility gradients in the soil, neighboring plots give rise to correlated yields; within blocks the effect of this is eliminated by randomisation but it inflates the variation between blocks. If we denote the variance of a single observation due to variation between blocks by \( \sigma_B^2 \), then the variance of the difference between two treatments is given by the usual formula:

\[
\text{var}(\Delta) = \frac{2}{r \times \text{number of blocks}} (\sigma^2 + n\sigma_B^2). \tag{3.1}
\]

Fairfield Smith was able to show by analysis of the results from uniformity trials in which the crops were harvested in small units, that an empirical relation for

* We are grateful to Prof. G. E. P. Box and Mr. M. J. R. Healy for drawing our attention to this work.
\[ \sigma_B^2 = \frac{k}{n^c} \]  

(3.2)

where \(0 \leq c \leq 1\) is a measure of the positive correlation between contiguous plots. A consequence of this formula is that unless the relative costs of taking observations within and between blocks is known, \(\text{var}(\Delta)\) as given by (3.1) and (3.2) is minimised by taking \(n\) as small as possible. If the costs are known, then an expression for the optimum value of \(n\) will involve the ratio of these costs and the ratio of the variances (see e.g. R. L. Anderson and T. A. Bancroft [12], Chapter 22).

In our problem, formula (3.1) does not apply since there is no randomisation. However, we obtain analogous formulae in the form

\[ \text{var}(\Delta) = \frac{k}{N} \left( \sigma^2 + \frac{k_1(n)}{n} + \frac{k_2(n)}{N} \right) \]  

(3.3)

where \(k_1(n), k_2(n)\) depend on the autocorrelation structure of the residual variation. It is difficult to see from these results how any stationary correlation pattern could give rise to Fairfield Smith's law. It may well be that for the type of situations met within agricultural practice, the terms \(\frac{k_1(n)}{n} + \frac{k_2(n)}{N}\) are adequately approximated by (3.2). Further, it is necessary to remark that it is not true in general that positive correlation between contiguous plots will always give rise to minimum block size since in general it is necessary to look at the whole of the a.e.f.

It should also be emphasised at this stage that in most of the present work it is assumed that smooth time trends are absent. In many practical examples, there will be a further contribution to \(\text{var}(\Delta)\) as given by (3.3) due to the presence of trends. It is conceivable that if the effect of trends is ignored then the optimum design using the autocorrelation approach might be to conduct half of the experiments at one level and half at the other. It is clear that
this design will be very unrealistic if the uncontrolled variation is subject to unforeseen trends and serious biases might be introduced as a result. In general, the effect of trends will be to indicate smaller block sizes. Some preliminary results which will be useful in designing experiments to deal with trend and autocorrelation are given in Appendix 2.

Costs

It is assumed in the following analysis that the dominant costs are those associated with the conduct of a single run so that it is not supposed, for example, that an extra cost is introduced when a change from one level to the next is made. There would be no difficulty in general in modifying the results so as to take this into account when the total cost of the experiment is specified in advance. In view of this, optimum designs will be defined as those which minimise the variance of treatment differences since these will lead to a smaller number of experiments to attain a required level of precision. We do not consider the losses due to wrong decisions to be very relevant in technological applications of this calibre. In any case, these would be very difficult to assess quantitatively, but the fact that one is not interested in a cut decision between two or more alternatives since the whole process of experimentation is evolutionary in nature, is more important; the optimum conditions themselves change with respect to time. Furthermore, we do not wish to be faced with the situation where we accept one set of conditions on the basis of a few observations with a consequent large standard error in the estimate of treatment differences. For a discussion of the philosophy of costing in evolutionary experimentation, the reader is referred to the account given by G. E. P. Box [13].

4. An Exact Form for the Variance of the Difference between Treatment Means

If the actual observations are denoted by \( y_i \), \( i = 1, 2, \ldots, N \), the first stage is to find an expression for the variance of \( \bar{y}(1) - \bar{y}(2) \) where \( \bar{y}(k) \) is the mean yield or response at level \( k \). In more general situations,
calculation of an exact form for the variances becomes prohibitive and it is necessary to resort to approximation. One such method is discussed briefly in Section 7.

Let the mean yields from the \( n \) consecutive runs at one level be denoted by \( \bar{y}_j \), where

\[
\bar{y}_j = \frac{1}{n} \sum_{i=(j-1)n+1}^{jn} y_i, \quad j = 1, 2, \ldots, 2b
\]

so that the experimental design consists of \( 2b \) blocks of \( n \) runs, the blocks occurring alternately at each level. It then follows that

\[
\Delta = \bar{y}(1) - \bar{y}(2) = \frac{1}{b} \sum_{j=1}^{b} (\bar{y}_{2j-1} - \bar{y}_{2j})
\]

Let \( E(y_i) \) at level \( k \) be denoted by \( \mu_k \), so that we may write the \( y_i \) in terms of the uncontrolled variation in the form

\[
y_i = \mu_k + z_i, \quad (k = 1 \text{ or } 2)
\]

where

\[
E(z_i) = 0, \quad E(z_i^2) = \sigma^2, \quad E(z_i z_{i+s}) = \rho_s \sigma^2
\]

It may then be shown that

\[
E(\Delta) = \mu(1) - \mu(2), \quad \text{and}
\]

\[
\text{var}(\Delta) = \frac{1}{b^2} \left[ \sum_{j=1}^{b} \left\{ \text{cov} \left( \bar{z}_{2j-1} \bar{z}_{2j+1} \right) + \text{cov} \left( \bar{z}_{2j-1} \bar{z}_{2j+1} \right) \right\} \right]
\]

\[
- \text{cov} \left( \bar{z}_{2j} \bar{z}_{2j-1} \right) - \text{cov} \left( \bar{z}_{2j-1} \bar{z}_{2j+1} \right)
\]

\[
(4.3)
\]

It is clear that \( \text{var}(\Delta) \) may be written conveniently in terms of the auto-covariance function of the block means of the uncontrolled variation, viz.
\[ s_\ell = E \left\{ \frac{\bar{z}_j - E(\bar{z}_j)}{\bar{z}_{j+\ell} - E(\bar{z}_{j+\ell})} \right\} \] (4.4)

where \( \ell = 1, 2, \ldots, (2b - 1) \).

The calculation then proceeds in two stages:

(a) to express var(\( \Delta \)) as a function of the \( s_\ell \)

(b) to express \( s_\ell \) as a function of the \( \{\rho_i\} \) and \( \sigma^2 \).

Since the \( \{z_1\} \) process is assumed to be stationary, it follows that the \( \{z_j\} \) process will also be stationary and (4.3) may be written as

\[ \text{var}(\Delta) = V_1 - V_2 = \frac{2}{b^2} \left[ \sum_j \sum_{j'} \left\{ \text{cov}(\bar{z}_{2j-1}, \bar{z}_{2j'-1}) - \text{cov}(\bar{z}_{2j-1}, \bar{z}_{2j'}) \right\} \right] \] (4.5)

Further simplification yields

\[ V_1 = \frac{2}{b^2} \left\{ b\delta_0 + 2 \sum_{j=1}^{b-1} (b-j)s_{2j} \right\} \]

\[ V_2 = \frac{2}{b^2} \left\{ \sum_{j=1}^{b} (2b - 2j+1)s_{2j-1} \right\} \]

so that

\[ \text{var}(\Delta) = \frac{2}{b^2} \left\{ b\delta_0 + 2 \sum_{j=1}^{b-1} (b-j)(s_{2j} - s_{2j-1}) - \sum_{j=1}^{b} s_{2j-1} \right\} \] (4.6)

For future reference, we quote the terms in (4.6) up to \( \delta_4 \), viz.

\[ \text{var}(\Delta) = \frac{2}{b^2} \left\{ b\delta_0 + 2(b-1)(\delta_2 - \delta_1) + 2(b-2)(\delta_4 - \delta_3) - (\delta_4 + \delta_3) + \ldots \right\} \] (4.6a)

when \( b \geq 2 \).

It now remains to evaluate \( s_\ell \). For \( \ell = 0 \), this is just the variance of a mean of size \( n \) which was given by G. U. Yule [14] in the form

\[ \text{var}(\bar{z}_j) = \frac{\sigma^2}{n^2} \sum_{i=-n+1}^{n-1} (n \cdot |i|)\rho_i \] (4.7)
This may be extended quite easily to

\[ \delta^2 = \sigma^2 \frac{2}{n^2} \sum_{i=-n+1}^{n-1} (n - |i|) \rho_{|i|}^2 + 1 \]  \hspace{1cm} (4.8)

(4.5) and (4.8) together may thus be used to derive \( \text{var} (\Delta) \) for quite general autocorrelation functions.

In Part II we derive explicit expressions for \( \text{var} (\Delta) \) in certain special cases and then show how \( b \) may be chosen so as to minimize \( \text{var} (\Delta) \).

The application of these results is described in Part III.
5. Determination of the Optimum Block Size when the Autocorrelation Function has a Finite Cut-Off.

Case 1. \( \rho_1 \neq 0, \rho_s = 0 \ (s \geq 2) \).

For stationary series, this condition implies that \( |\rho_1| \leq \frac{1}{2} \). (See Appendix 1). Application of (4.4) then yields

\[
\delta_0 = \frac{\sigma^2}{n} \left\{ n + 2(n-1)\rho_1 \right\} \\
\delta_1 = \frac{\sigma^2}{n} \rho_1 \\
\delta_2 = O(\ell \geq 2).
\]

From (4.6a) we may write

\[
\text{var} (\Delta) = \frac{2}{b^2} \left\{ b\delta_0 - (2b-1)\delta_1 \right\}
\]

which simplifies to give

\[
\text{var} (\Delta) = \frac{8\sigma^2}{N} \left\{ \frac{N}{2} + \rho_1 (N - 4b + 1) \right\}
\]

(5.2)

It follows that there is a unique choice of \( b \) which minimizes \( \text{var} (\Delta) \), viz.

(i) if \( \rho_1 > 0 \), choose \( b \) as large as possible i.e. \( N/2 \).

(ii) if \( \rho_1 < 0 \), choose \( b \) as small as possible i.e. 1.

This conclusion has been mentioned by G. E. P. Box [15]. The optimum rule in this instance is thus seen to depend on the sign of the a.c.f. and not on its magnitude. It is to be observed, also, that it is possible to choose the optimum value of \( b \) for all values of \( N \). In most cases, the optimum choice of \( b \) will depend on the magnitudes as well as the signs of the autocorrelations and is only unique when \( N \) is sufficiently large.

If we now write

\[
\text{var} (\Delta) = \frac{4\sigma^2}{N} \phi(N) + O(\frac{1}{N^2})
\]
the relative efficiency of the systematic design compared to complete randomisation will be given for large \( N \) by the relation

\[
E_1 = 1/\phi_{\text{opt}}(N)
\]  

(5.3)

where \( \phi_{\text{opt}}(N) \) is the value of \( \phi(N) \) when the optimum value of \( \rho \) has been chosen. For Case 1, this becomes \( E_1 = 1/(1 + 2\rho_1) \) according as \( \rho_1 < 0 \) or \( \rho_1 > 0 \).

Since two treatments are to be compared, it is more likely that instead of complete randomisation, a design would be used in which the treatments are randomised within blocks of two. In this case, the relative efficiency of the randomised design is given by

\[
E_2 = \frac{1 - \rho_1}{\phi_{\text{opt}}(N)}
\]  

(5.4)

where \( \rho_1 \) is the first lag autocorrelation.

It may be seen by examination of \( E_1 \) or \( E_2 \) that arbitrarily large relative efficiencies may be obtained by using the systematic design. It is apparent, also, that \( E_1 > E_2 \) when \( \rho_1 > 0 \) and \( E_2 > E_1 \) when \( \rho_1 < 0 \).

Formulæ (5.3) and (5.4) hold quite generally; from now on we shall be concerned with the determination of \( \phi_{\text{opt}}(N) \) for various autocorrelation patterns.

Case 2: \( \rho_1 \neq 0 \neq \rho_2, \rho_s = 0 (s \geq 2) \)  

(5.5)

In Appendix 1, it will be shown that the \( (\rho_1, \rho_2) \) point must satisfy the following conditions in order that this be the a.c.f. of a valid stationary series:

\[
\begin{align*}
\rho_1 + \rho_2 &\geq -\frac{1}{2} \\
\rho_1 - \rho_2 &\geq -\frac{1}{2} \\
\rho_1^2 &\leq 4\rho_2(1 - 2\rho_2)
\end{align*}
\]  

(5.6)
This means that the stationarity region in the \((\rho_1, \rho_2)\) plane is bounded by two straight lines and the arc of an ellipse (see Figure 1). Application of (4.7) then yields for \(n \geq 2\) \((b \leq N/4)\)

\[
\begin{align*}
\delta_0 &= \frac{\sigma^2}{n} \left( n + 2(n - 1)\rho_1 + 2(n - 2)\rho_2 \right) \\
\delta_1 &= \frac{\sigma^2}{n} \left( \rho_1 + 2\rho_2 \right) \\
\delta_\ell &= 0, \quad (\ell \geq 2)
\end{align*}
\]  

(5.7)

In the case where \(n = 1\), it follows trivially that \(\delta_\ell = \sigma^2 \rho_\ell\).

Substitution in (4.6) then yields

\[
\text{var } (\Delta) = \begin{cases} 
\frac{8\sigma^2}{N^2} \left[ \frac{N}{2} + N(\rho_1 + \rho_2) - (4b - 1)(\rho_1 + 2\rho_2) \right], & n \geq 2, \ b \leq N/4 \\
\frac{8\sigma^2}{N^2} \left[ \frac{N}{2} - N(\rho_1 - \rho_2) + (\rho_1 - 2\rho_2) \right], & n = 1, \ b = N/2
\end{cases}
\]  

(5.8)

so that the possible values of \(b\) are \(b = 1\), \(b\) intermediate, \(b = N/4\), \(b = N/2\). It is apparent, therefore, that the optimum choice of \(b\) will now depend on the magnitudes of \(\rho_1\) and \(\rho_2\) as well as their signs. It is not enough, for example, to say that \(b\) must be large when \(\rho_1 + 2\rho_2 > 0\) because, depending on the term \(N(\rho_1 + \rho_2)\), it will pay to take \(b = N/4\) or \(b = N/2\). Writing \(\text{var } (\Delta)\) to \(O\left(\frac{1}{N}\right)\), we obtain

\[
\text{var } (\Delta) = \begin{cases} 
\frac{4\sigma^2}{N} \left( 1 + 2(\rho_1 + \rho_2) - \frac{8b}{N}(\rho_1 + 2\rho_2) \right), & n \geq 2, \ b \leq N/4 \\
\frac{4\sigma^2}{N} \left( 1 - 2(\rho_1 - \rho_2) \right), & n = 1, \ b = N/2
\end{cases}
\]  

(5.9)

The effect on \(\text{var } (\Delta)\) of terms of \(O\left(\frac{1}{N^2}\right)\) will be very small since
FIGURE 1 Stationarity region and choice of $b_{\text{opt}}$.

- $b_{\text{opt}}$
- $N/2$ ($n=1$)
- $N/4$ ($n=2$)
- $1$ ($n=N/2$)
\[ |\rho_1 + 2\rho_2| < 1 \] for example, so that in order to determine the optimum value of \( b \) it is sufficient to work with (5.9).

Suppose that \( \rho_1 + 2\rho_2 \leq 0 \). Then for \( n \geq 2 \), i.e. \( b \leq N/4 \), minimum variance is achieved by taking \( b \) as large as possible, i.e. \( N/4 \). In this case \( \text{var} (\Delta) = \frac{4\sigma^2}{N} (1 - 2\rho_2) + O\left(\frac{1}{N}\right) \). The latter will be less than the expression for \( \text{var} (\Delta) \), when \( b = N/2 \) provided that \( \rho_1 < 2\rho_2 \). This leads to the following rule:

(i) If \( \rho_1 + 2\rho_2 > 0 \), take \( b = N/4 \) or \( b = N/2 \) according as \( \rho_1 < 2\rho_2 \) or \( \rho_1 > 2\rho_2 \).

Now suppose that \( \rho_1 + 2\rho_2 < 0 \). This means that it is necessary to take \( b = 1 \) so that to \( 0(\frac{1}{N}) \) we obtain

\[ \text{var} (\Delta) = \frac{4\sigma^2}{N} (1 + 2(\rho_1 + \rho_2)) \]

and this will be less than \( \text{var} (\Delta) \) for \( b = N/2 \), provided that \( \rho_1 < 0 \). This leads to the following rule:

(ii) If \( \rho_1 + 2\rho_2 < 0 \), take \( b = 1 \) or \( b = N/2 \) according as \( \rho_1 < 0 \) or \( \rho_1 > 0 \).

It is to be observed that provided \( \rho_1 + 2\rho_2 < 0 \), the optimum choice of \( b \) is governed by the same rules as in the case where \( \rho_1 \neq 0 \), \( \rho_s = 0 \) (\( s \geq 2 \)). These results are summarised in Figure 1.
The terms of \( O\left( \frac{1}{N^2} \right) \) in (5.8) will have very little effect on the optimum choice of \( \theta \) as may be seen from the following considerations. In the case where \( \rho_1 + 2\rho_2 > 0 \) and \( \rho_1 < 2\rho_2 \), \( \text{var}(\Delta) \) is a minimum when \( \theta = N/2 \).

The effect of the term of \( O\left( \frac{1}{N^2} \right) \) will be to make this even smaller whereas the \( O\left( \frac{1}{N^2} \right) \) term will increase \( \text{var}(\Delta) \) when \( \theta = N/4 \) so that the optimum rule is unaffected. In the case where \( \rho_1 > 2\rho_2 \) the terms in \( O\left( \frac{1}{N^2} \right) \) will both be positive but of comparable magnitude and hence will tend to annul one another. Similar considerations apply in other cases.

**Case (3).** \( \rho_s \neq 0 \) \( (s = 1, 2, 3) \), \( \rho_s = 0 \) \( (s \geq 4) \). (5.10)

It has been found that the problem of finding the optimum regions for a time series which has an a.c.f. which is zero beyond a general lag \( k \) is a very complex matter. This has not been dealt with in its greatest generality in this report because the interesting a.c.f.'s from a practical point of view are those which can be expressed as a sum of damped exponentials and sine waves as given by (2,3). However, we give an explicit solution in the case where the a.c.f. is of the form (5.10).

In Appendix I it will be shown that (5.10) defines a valid stationary series provided that the autocorrelations satisfy the four conditions

\[
\begin{align*}
\rho_1 + \rho_2 + \rho_3 &> -\frac{1}{2} \\
\rho_1 - \rho_2 + \rho_3 &< \frac{1}{2} \\
\{27\rho_3^2 + 2\rho_2^3 - 9\rho_1\rho_2\rho_3 - 27\rho_2^2\rho_3^2\} + \phi \{6\rho_1\rho_3 - 18\rho_3^2 - 2\rho_2^2\} &> 0 \\
\phi &= \sqrt{\rho_2^2 - 3\rho_1\rho_3 + 9\rho_3^2}.
\end{align*}
\]

(5.11)

This means that the stationarity region is bounded by two planes and two surfaces in the 3-dimensional space of the \( \rho \)'s. The autocovariance function of the means may be determined using (4.7) and this is defined by a different functional form for different values of \( n \).
\[ s_0 = \frac{\sigma}{n} [n + 2(n - 1)p_1] \]
\[ s_1 = \frac{\sigma}{n} [p_1 + 2p_2 + 3p_3] \]
\[ s_2 = \frac{\sigma}{n} p_3 \]
\[ s_b = 0, b \geq 3 \]
\[ s_0 = \frac{\sigma}{n} [n + 2(n - 1)p_1 + 2(n - 2)p_2 + \ldots + 3p_3], n \geq 3, b \leq N/6 \]

Substitution in (4.6) then yields the following exact expressions for \( \text{var}(\Delta) \):

\[ \text{var}(\Delta) = \frac{4\sigma^2}{N^2} \left( 1 - 2(\rho_1 - \rho_2 + \rho_3) + \frac{2}{N}(\rho_1 - 2\rho_2 + 3\rho_3) \right), n = 1, b = N/2 \]

\[ \text{var}(\Delta) = \frac{4\sigma^2}{N^2} \left( 1 - 2\rho_2 + \frac{2}{N}(\rho_1 + 2\rho_2 - \rho_3) \right), n = 2, b = N/4 \]

\[ \text{var}(\Delta) = \frac{4\sigma^2}{N^2} \left( 1 + 2(\rho_1 + \rho_2 + \rho_3) - \frac{8(b-1)}{N} (\rho_1 + 2\rho_2 + 3\rho_3) \right), n \geq 3, b \leq N/6 \]

Formula (5.15) does not hold if \( b = 1 \) in addition to \( r = 1 \). Since this occurs when \( N = 2 \), this is not of any practical interest. The optimum rules for choosing \( b \) may now be formulated as follows:

1. If \( \rho_1 + 2\rho_2 + 3\rho_3 > 0 \), the choice is between \( b = \frac{N}{2} (n = 1) \), \( b = \frac{N}{4} (n = 2) \), and \( \frac{N}{6} (n = 3) \) depending on the magnitudes of \( 1 - 2(\rho_1 - \rho_2 + \rho_3) \), \( 1 - 2\rho_2 \) and \( 1 + \frac{2}{3}(\rho_1 - \frac{2}{3}\rho_2 + 2\rho_3) \).
(ii) If \( \rho_1 + 2\rho_2 + 3\rho_3 < 0 \), the choice is between \( b = \frac{N}{2} \), \( \frac{N}{4} \) and \( b = 1 \) depending on the magnitudes of \( 1 - 2(\rho_1 - \rho_2 + \rho_3) \), \( 1 - 2\rho_2 \) and \( 1 + 2(\rho_1 + \rho_2 + \rho_3) \).

In the general case where \( \rho_s \neq 0 \) (\( s = 1, 2, \ldots, k \)), \( \rho_s = 0 \) (\( s > k \)), the stationarity region is bounded by two hyper-planes and \( k - 1 \) hyper-surfaces in the \( k \)-dimensional space of the \( \rho \)'s. (See Appendix I). From previous considerations, it is apparent that this region will be divided into \( k + 1 \) regions corresponding to optimum block sizes of \( n = 1, 2, \ldots, k \) and \( n = N/2 \). The boundaries of the regions will be very difficult to obtain in general.

6. Determination of the Optimum Block Size when the Autocorrelation Function is of the Form \((2,3)\)

It is well known (c.f. M. G. Kendall [16]) that the a.c.f. of an auto-regressive scheme of order \( m \), i.e.

\[
(y_1 - \mu) = \alpha_1 (y_{1-1} - \mu) + \alpha_2 (y_{1-2} - \mu) + \cdots + \alpha_m (y_{1-m} - \mu) + z_1
\]

(6.1)

where \( E(y_1) = \mu \), satisfies a difference equation of the form

\[
\rho_s = \alpha_1 \rho_{s-1} + \alpha_2 \rho_{s-2} + \cdots + \alpha_m \rho_{s-m} \quad *
\]

This has a solution

\[
\rho_s = A_1 \pi_1 |s| + A_2 \pi_2 |s| + \cdots + A_m \pi_m |s|
\]

where the \( \pi \)'s are the roots of the characteristic equation

\[
x^m - \alpha_1 x^{m-1} - \alpha_2 x^{m-2} - \cdots - \alpha_m = 0
\]

(assumed to be all different) and the \( A \)'s are constants to be determined from the initial conditions. In order for the \( \{y_1\} \) series to be stationary, it is sufficient that \( |\pi_j| < 1, j = 1, 2, \ldots, m \).
Case 1: \( m = 1 \)

It follows that \( \bar{\alpha}_1 = \alpha_1 \), \( A = 1 \) so that \( \bar{\rho}_s = \alpha_1 \left| s \right| \) where \( |\alpha_1| < 1 \).

It is now possible to write down a single expression for \( \text{var} (\Delta) \), unlike the cases considered in Section 5. Summation of the finite series (4.7) and (4.8) yields

\[
\bar{\sigma}_0 = \frac{\sigma^2}{n} \left[ \frac{n(1 - \alpha_1^2) - 2\alpha_1(1 - \alpha_1^n)}{(1 - \alpha_1^2)^2} \right] \tag{6.2}
\]

\[
\bar{\sigma}_k = \frac{\sigma^2}{n} \frac{\alpha_1(\ell - 1)(n + 1)}{(1 - \alpha_1^2)^2}, \quad \ell \geq 2 \tag{6.3}
\]

These in turn may be used to sum the series (4.6) which results in the following exact expression for \( \text{var} (\Delta) \):

\[
\text{var} (\Delta) = \frac{4\sigma^2}{N(1 - \alpha_1)^2} \left[ 1 - \alpha_1^2 - \frac{8\bar{\sigma}}{N} \frac{1 - \alpha_1 \alpha_1^n}{(1 + \alpha_1^n)} - \frac{2\alpha_1}{N} \frac{(1 - \alpha_1^2)(1 - \alpha_1^n)}{(1 + \alpha_1^n)^2} (1 - \alpha_1^N) \right] \tag{6.4}
\]

This expression has the same broad features as formula (5.2), with \( \alpha_1 \) playing the same role as \( \bar{\rho}_1 \). However, it is a non-trivial matter to establish the optimum value of \( \bar{\rho}_1 \).

**THEOREM:** The rule which minimises \( \text{var} (\Delta) \), as given by (6.4), with respect to \( \bar{\rho} \) or \( N \) is the following:

(i) If \( \alpha_1 > 0 \), choose \( \bar{\rho} = N/2 \) (or \( n = 1 \)).

(ii) If \( \alpha_1 < 0 \), choose \( \bar{\rho} = 1 \) (or \( n = N/2 \)).

The rule is correct for all \( N \) if \( \alpha_1 < 0 \) but only to \( O(\frac{1}{N^2}) \) if \( \alpha_1 > 0 \).

**PROOF:**

Write \( \text{var} (\Delta) \) in the form

\[
\text{var} (\Delta) = \frac{4\sigma^2}{N(1 - \alpha_1)^2} \left[ 1 - \alpha_1^2 - \alpha_1 \bar{\sigma}(n) - \alpha_1 \eta(n)(1 - \alpha_1^N) \right] \tag{6.5}
\]
where
\[ \xi(n) = \frac{1}{n} \left( 1 - \alpha_1^n \right), \quad \eta(n) = \frac{2}{N} \left( \frac{1 - \alpha_1^n}{1 + \alpha_1^n} \right)^2 \]

The theorem follows as a consequence of the following results.

(i) \( \xi(n) \) is a maximum at \( n = 1 \) for all \( \alpha_1 \) such that \( |\alpha_1| < 1 \)

(ii) \( \eta(n) \) is a maximum at \( n = 1 \) only when \( \alpha_1 < 0 \).

Result (i) follows from the following inductions.

(a) If \( \alpha_1 > 0 \),
\[ \xi(n) > \xi(n + 1) \rightarrow \xi(n + 1) > \xi(n + 2) \]
Since \( \xi(1) > \xi(2) \), it follows that \( \xi(n) \) is monotonic decreasing.

(b) If \( \alpha_1 < 0 \),
\[ \xi(2n - 1) > \xi(2n + 1) \rightarrow \xi(2n + 1) > \xi(2n + 3) \]
\[ \xi(2n) > \xi(2n + 2) \rightarrow \xi(2n + 2) > \xi(2n + 4) \]
This establishes the monotonic nature of the sequences \( \xi(2n), \xi(2n - 1) \).

Result (i) then follows since \( \xi(1) > \xi(2) \).

The situation is much easier for \( \eta(n) \) when it may be established that

(a) \( \eta(2n) < \eta(2n + 2) \) for all \( \alpha_1 \)
(b) \( \eta(2n - 1) < \eta(2n + 1) \) if \( \alpha_1 > 0 \)
\( \eta(2n - 1) > \eta(2n + 1) \) if \( \alpha_1 < 0 \)

Hence, since \( \eta(1) < \eta(2) \), it follows that \( \eta(n) \) has a maximum at \( n = 1 \) when \( \alpha > 0 \). Since \( \eta(1) = \frac{2}{N} \left( \frac{1 - \alpha_1}{1 + \alpha_1} \right)^2 \) and \( \eta(\infty) = \frac{2}{N} \), it follows that \( \eta(n) \) has a maximum at \( n = 1 \) when \( \alpha_1 < 0 \).

This completes the proof.

Case 2: \( m = 2 \)

We now have \( \rho_s = A_1 \pi_1 |s| + A_2 \omega_2 |s| \). Using the initial conditions \( \rho_0 = 1, \rho_1 = \frac{\alpha_1}{1 - \alpha_2} \), it follows that
\[
\frac{\pi_1(1 - \pi_2^2)}{A_1(\pi_1 - \pi_2)(1 + \pi_1\pi_2)} = \frac{-\pi_2(1 - \pi_1^2)}{A_2(\pi_2 - \pi_1)(1 + \pi_1\pi_2)}
\]

The conditions for stationarity are most conveniently written in terms of the \(\alpha_1, \alpha_2\), viz. \(\alpha_1 + \alpha_2 < 1\), \(\alpha_1 - \alpha_2 > -1\), \(-1 < \alpha_2 < 1\).

Since \(\text{var}(\Delta)\) is linear in the \(\rho_s\)’s, it follows that we can write down immediately \(\text{var}(\Delta)\) in this case from (6.4). After some rearrangement, this becomes

\[
\text{var}(\Delta) = \frac{1 - \rho^2}{N} \left[ 1 + 2 \left( \frac{A_1\pi_1}{1 - \pi_1 \pi_2} + \frac{A_2\pi_2}{1 - \pi_1 \pi_2} \right) - \frac{8b}{N} \left( \frac{A_1\pi_1}{(1 - \pi_1^2)^2} \left( \frac{1 - \pi_1^2}{1 + \pi_1^2} \right) + \frac{A_2\pi_2}{(1 - \pi_2^2)^2} \left( \frac{1 - \pi_2^2}{1 + \pi_2^2} \right) \right) \right]
\]

\[+ O\left(\frac{1}{N}\right)\]  \(6.6\)

When \(b = 1\), it may be shown that

\[
\text{var}(\Delta) = \frac{1 - \rho^2}{N} \left( \frac{1 - \alpha_2}{(1 - \alpha_2)(1 - \alpha_1 - \alpha_2)} \right) + O\left(\frac{1}{N}\right)\]  \(6.7\)

The parallelism between this section and Section 4 now disappears since it will be shown that for \(m = 2\), values of \(b\) intermediate between \(b = 1\) and \(b = N/2\) will be possible optimum choices depending on the values of \(\pi_1, \pi_2\) or equivalently \(\alpha_1, \alpha_2\). In order to show this, it is sufficient to prove that \(\nu(n)\), the coefficient of \(bN\) in (6.6), changes sign as \(b\) increases from 1 to \(N/2\), i.e. as \(n\) decreases from \(N/2\) to 1. This follows from the fact that if \(\nu(n)\) were positive for \(b = 1\), increasing \(b\) indefinitely would not result in minimum variance if \(\nu(n)\) itself becomes negative. It is clear that in this case, therefore, some value of \(b\) intermediate between \(b = 1\) and \(b = N/2\) will be optimum. If we let \(\nu(n) = \frac{1}{(\pi_1 - \pi_2)(1 + \pi_1\pi_2)} f(\pi_1, \pi_2, n)\), then it is sufficient to show that \(f(\pi_1, \pi_2, n)\) changes sign. We state this in the form of a theorem.
THEOREM: For sufficiently large N, there exist values of \( \pi_1, \pi_2 \), namely values which are real and opposite in sign, which are such that a value of \( b \) intermediate between \( b = 1 \) and \( b = N/2 \) gives rise to a minimum value of \( \text{var}(\Delta) \) as given by (6,6).

PROOF:

More explicitly, we prove that \( f(\pi_1, \pi_2, 1) \) and \( f(\pi_1, \pi_2, \infty) \) have opposite signs for certain values of \( \pi_1, \pi_2 \).

Substitution for \( A_1, A_2 \) in \( f(\pi_1, \pi_2, n) \) leads to

\[
f(\pi_1, \pi_2, 1) = \frac{\pi_1^2(1-\pi_2^2)}{(1-\pi_1^2)} - \frac{\pi_2^2(1-\pi_1^2)}{(1-\pi_2^2)}
\]

\[
f(\pi_1, \pi_2, \infty) = \frac{\pi_1^2(1-\pi_2^2)}{(1-\pi_1^2)} - \frac{\pi_2^2(1-\pi_1^2)}{(1-\pi_2^2)}
\]

It is clear that \( f(\pi_1, \pi_2, 1) \geq 0 \) according as \( |\pi_1| \geq |\pi_2| \). For \( f(\pi_1, \pi_2, \infty) \), we distinguish between 4 cases.

Suppose that \( \pi_1, \pi_2 \) are real roots of the characteristic equation.

(i) If \( \pi_1, \pi_2 \) are both positive, then \( f(\pi_1, \pi_2, \infty) \geq 0 \) according as \( \pi_1 \geq \pi_2 \). This means that \( f(\pi_1, \pi_2, \infty) \) will always have the same sign as \( f(\pi_1, \pi_2, 1) \).

(ii) If \( \pi_1 = -\mu, \pi_2 = -\lambda \) where \( \lambda, \mu \) are both positive, then it is still true that \( f(\pi_1, \pi_2, \infty) \geq 0 \) according as \( \mu \gtrless \lambda \), i.e.

\[
|\pi_1| \gtrless |\pi_2|.
\]

(iii) If \( \pi_1 = -\mu, \pi_2 = \lambda \), where \( \lambda, \mu \) are both positive, then

\[
f(\pi_1, \pi_2, \infty) \geq 0 \quad \text{according as}
\]

\[
(\mu + \lambda)((\mu - \lambda)(1 + \lambda^2\mu^2) - 2\lambda\mu(1 - \lambda\mu)) \geq 0
\]

(6,8)
If \( \mu < \lambda \), this will always be negative but if \( \mu > \lambda \), then it could be positive or negative. For example, it is:

(a) positive when \( \mu = 0.75, \lambda = 0.25 \)

(b) negative when \( \mu = 0.5, \lambda = 0.4 \)

Hence if \( \mu > \lambda \), the relation (6.8) will define a region in the \((\pi_1, \pi_2)\) plane for which \( f(\pi_1, \pi_2, \omega) < 0 \). Since \( f(\pi_1, \pi_2, \omega) \) is always > 0 if \( \mu > \lambda \), it follows that this region is one for which an intermediate block size is optimum.

(iv) If \( \pi_1 = \mu, \pi_2 = -\lambda \), where \( \lambda, \mu \) are both positive, then by a similar argument it may be shown that if \( \mu < \lambda \), \( f(\pi_1, \pi_2, \omega) \) may be positive or negative.

This completes the proof.

The above theorem shows that for some \( \lambda_1, \lambda_2 \) such that \( \lambda_1 \lambda_2 < 0 \), an intermediate value of \( b \) will be optimum. This condition may also be written as \( \alpha_2 > 0 \) since \( \alpha_2 = -\lambda_1 \lambda_2 \).

For \( b = N/2 \), there is a contribution to \( \text{var}(\Delta) \) from the coefficient of \( b/N \) in (6.6). For large \( N \), this may be written as

\[
\delta = \frac{-4\alpha_1(1 + \alpha_2)}{(1 - \alpha_2^2) - \alpha_1^2}
\]

and if this is added to \( \text{var}(\Delta) \) when \( b = 1 \), it follows that when \( b = N/2 \),

\[
\text{var}(\Delta) = \frac{16 \delta^2}{N} \left( \frac{1 + \alpha_2^2}{1 - \alpha_2^2} \right) \left( \frac{1 - \alpha_1 - \alpha_2}{1 + \alpha_1 - \alpha_2} \right) + O \left( \frac{1}{N^2} \right)
\]

(6.10)

It also follows that \( \text{var}(\Delta) \mid b = N/2 \lesssim \text{var}(\Delta) \mid b = 1 \) according as \( \delta \gtrless 0 \) which is equivalent to writing \( \alpha_1 \gtrless 0 \) since \( 1 + \alpha_2 > 0 \) by virtue of the stationarity condition. It is to be noted that this is the same condition
as for the case \( m = 1 \) and is analogous to the result \( \rho_1 > 0 \) which was true for cases 1 and 2 of Section 4. A general result covering all these cases will be proved in Section 8.

It is theoretically possible to work out the regions in the \((\alpha_1, \alpha_2)\) plane which correspond to various optimum choices of \( b \) but in general these will be awkward functions of both the \( \pi \)'s and the \( \alpha \)'s. In practice it will be easier to calculate \( \text{var}(\Delta) \) as a function of \( b \) or \( n \) as given by formula (6.6) and to locate the minimum value of \( \text{var}(\Delta) \) numerically.

However, in this case \( m = 2 \), explicit expressions for the boundaries for large \( N \) have been obtained from the information variances. This will be discussed in Section 7.

Case 3: \( m \) general.

Using the fact that \( \text{var}(\Delta) \) is linear in the \( \{\rho_i\} \), it is possible to write down immediately an explicit expression for \( \text{var}(\Delta) \) directly from (6.4) in the case where the a.c.f. is of the form (2.3).

Thus

\[
\text{var}(\Delta) = \frac{4 \sigma^2}{N} \left\{ 1 + 2 \sum_{i=1}^{m} A_i \left[ \frac{\pi_1}{1-\pi_1} + \frac{4b}{N} \frac{\pi_1(1-\pi_1^n)}{(1-\pi_1)^2(1+\pi_1^n)} \right] - \frac{2}{N} \sum_{i=1}^{m} A_i \frac{\pi_1}{(1-\pi_1)^2} \frac{(1-\pi_1^n)^2}{(1+\pi_1^n)^2} (1-\pi_1^n) \right\} \tag{6.11}
\]

By substituting \( \frac{1}{2n} \) for \( \frac{b}{N} \), this may be plotted as a function of \( n \) and the minimum located numerically. Care is necessary in this case, however, since:

(i) When the \( \pi \)'s are negative, this function will oscillate locally.

(ii) If the oscillations are smoothed out, there will in fact be
several local minima so that careful tabulation of \( \text{var}(\Delta) \) will be necessary in order to pick out the true minimum.

The range of values of \( n \) which need be examined may be restricted since a good approximation to the value of \( n \) which minimises \( \text{var}(\Delta) \) may be obtained from a knowledge of the spectrum. This will be discussed in greater detail in Section 8.

7. Likelihood Function Approach

The likelihood function will be used in this section to derive results for the first and second order autoregressive schemes \((m = 1\) and \(2)\) which were considered in the previous section. Whilst this approach is not as exact as the preceding one, it gives a good indication of how well the information variances obtained from the likelihood function may be used as approximations to the exact variances. It is thus useful as a preliminary study to the extension of these results to more general designs in which more than two levels of several factors are considered. It will be shown in this section that not only do the information variances serve as good approximations to the variances derived in the previous section but also that they give much clearer indications of what are the important elements in the choice of the optimum value of \( n \) or \( b \).

A rather curious feature of the likelihood approach is that the final answer is not correct to \( O(\frac{1}{N}) \) for all \( n \). This is due to the fact that the information variance, which is correct to \( O(\frac{1}{N}) \) for fixed \( n \), is optimised further with respect to \( n \) which itself can be of \( O(N) \).

Case 1: \( m = 1 \)

In what follows we use circular likelihood functions i.e. it is assumed that the uncontrolled variation is such that \( y_{N+1} \equiv y_1 \). This is a convenient mathematical device which leads to a more symmetrical likelihood function and
leaves the information matrix unaltered to $O(\frac{1}{N})$. For a first order auto-
regressive scheme

$$z_i = \alpha_1 z_{i-1} + \epsilon_i \tag{7.1}$$

it is easily shown that the log-likelihood is given by

$$- \ln L(y | \alpha_1, \sigma^2) = \frac{N}{2} \ln 2\pi \sigma^2 - \ln (1 - \alpha_1^N)$$

$$+ \frac{1}{2\sigma^2} \{ (1 + \alpha_1^2) \sum_{i=1}^{N} z_i^2 - 2\alpha_1 \sum_{i=1}^{N} z_i z_{i+1} \} \tag{7.2}$$

where $\sigma^2 = \mathbb{E}(\epsilon_1^2)$.

If we now let $x_{j,k}$ be the response from the $k^{th}$ run in the $j^{th}$ block of size $n$, we may then rewrite (7.2) in terms of the observed responses in the form

$$- \ln L(x | \alpha_1, \sigma^2, \mu_1, \mu_2) = f_1(\alpha_1, \sigma^2, N)$$

$$+ \frac{(1+\alpha_1^2)}{2\sigma^2} \sum_{j=1}^{b} \sum_{k=1}^{n} \left\{(x_{2j-1,k} - \mu_1)^2 + (x_{2j+k} - \mu_2)^2 \right\}$$

$$- \frac{\alpha_1}{\sigma} \sum_{j=1}^{b} \sum_{k=1}^{n-1} \left\{(x_{2j-1,k} - \mu_1)(x_{2j-1,k+1} - \mu_1) + (x_{2j,k} - \mu_2)(x_{2j,k+1} - \mu_2) \right\}$$

$$- \frac{\alpha_1}{\sigma} \sum_{j=1}^{b} \left\{(x_{2j-1,n} - \mu_1)(x_{2j,1} - \mu_2) + (x_{2j,n} - \mu_2) \right\}$$

$$+ (x_{2j+1,1} - \mu_1)$$
Writing \( L' = \ln L \), it is easily verified that

\[
A = - \frac{\partial^2 L'}{\partial \mu_1^2} = \left\{(1 - \alpha_1)^2 n \beta + 2 n \alpha_2 \right\} / \sigma^2 = - \frac{\partial^2 L'}{\partial \mu_2^2}
\]

\[
B = - \frac{\partial^2 L'}{\partial \mu_1 \partial \mu_2} = - 2 n \alpha_1 / \sigma_2
\]

Consideration of the variance-covariance matrix of the maximum likelihood estimates \( \bar{x}_1, \bar{x}_2 \) of \( \mu_1, \mu_2 \) shows that

\[
\text{var} (\bar{x}_1 - \bar{x}_2) = \frac{\sigma^2}{A-B}
\]

Substituting \( \sigma^2 = \sigma^2 / (1 - \alpha_1^2) \), it may be shown after some simplification that

\[
\text{var} (\Delta) = \frac{k \sigma^2}{N} \left( \frac{1 + \alpha_1}{1 - \alpha_1} \right) \left\{ \frac{1}{1 + \frac{k}{n}} \frac{\alpha_1}{(1 - \alpha_1^2)} \right\}
\]

Comparison of (7.3) with (6.4) to \( O(\frac{1}{N}) \) shows that they agree exactly for \( n = 1, 2, \infty \). For intermediate values of \( n \) there is slight disagreement but this is small. It is to be noted also that (7.3) gives the same optimum rule as (6.4).

Case 2: \( m = 2 \).

Starting from the circular likelihood function

\[
- L' = f_2(\alpha_1, \alpha_2; \sigma^2, N) + \frac{1}{2\sigma^2} \left\{ (1 + \alpha_1^2 + \alpha_2^2) \sum_{i=1}^{N} z_i^2 - 2 \alpha_1 (1 - \alpha_2) \sum_{i=1}^{N} z_i z_{i+1} 
\]

\[
- 2 \alpha_2 \sum_{i=1}^{N} z_i z_{i+2} \right\}
\]

(7.4)
we may write for \( n \geq 2 \)

\[
L' = f_2(\alpha_1, \alpha_2, \sigma^2, N) + \frac{1 + \alpha_1^2 + \alpha_2^2}{2\sigma^2} \left\{ \begin{array}{c}
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,k} - \mu_1)^2 + (x_{2j,k} - \mu_2)^2 \\
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,k} - \mu_1)(x_{2j-1,k+1} - \mu_1) + (x_{2j,k} - \mu_2)(x_{2j,k+1} - \mu_2) \\
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,n-1} - \mu_1)(x_{2j-1,n-1} - \mu_2) + (x_{2j,n} - \mu_2)(x_{2j,n+1} - \mu_1) \\
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,n-1} - \mu_2)(x_{2j,n+1} - \mu_1) + (x_{2j,n} - \mu_2)(x_{2j,n+1} - \mu_2)
\end{array} \right\}
\]

\[
\frac{a_2}{\sigma^2} \left\{ \begin{array}{c}
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,k} - \mu_1)^2 + (x_{2j,k} - \mu_2)^2 \\
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,k} - \mu_1)(x_{2j-1,k+1} - \mu_1) + (x_{2j,k} - \mu_2)(x_{2j,k+1} - \mu_2) \\
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,n} - \mu_1)(x_{2j,n} - \mu_2) + (x_{2j,n} - \mu_2)(x_{2j,n+1} - \mu_1) \\
\sum_{j=1}^{b} \sum_{k=1}^{n-1} (x_{2j-1,n} - \mu_2)(x_{2j,n+1} - \mu_1) + (x_{2j,n} - \mu_2)(x_{2j,n+1} - \mu_2)
\end{array} \right\}
\]

As before we obtain

\[
- \frac{\partial^2 L'}{\partial \mu_1^2} = - \frac{\partial^2 L'}{\partial \mu_2^2} = bn(1 - \alpha_1 - \alpha_2)^2 + 2b \left\{ \alpha_1(1 - \alpha_2) + 2\alpha_2 \right\}
\]

\[
- \frac{\partial^2 L'}{\partial \mu_1 \partial \mu_2} = - 2b \left\{ \alpha_1(1 - \alpha_2) + 2\alpha_2 \right\}
\]

\[
\sigma_x^2 = \sigma^2 \left( \frac{1 + \alpha_2}{1 - \alpha_2} \right)^2 \left( 1 - \alpha_2^2 - \alpha_1^2 \right)
\]

\[
\text{var} (\Delta) = \frac{4\sigma_x^2}{N} \left( \frac{1 + \alpha_2}{1 - \alpha_2} \right)^2 \left( \frac{1 - \alpha_2 + \alpha_1}{1 - \alpha_2 - \alpha_1} \right) \frac{1}{\left( \frac{n}{n-1} \right)^2 \left( 1 - \frac{\alpha_1(1-\alpha_2) + 2\alpha_2}{\left( 1 - \alpha_1 - \alpha_2 \right)^2} \right)}
\]

For \( n = 1 \) the likelihood function is of a different form but it is easily verified that in this case, the information variance is given exactly by formula (6.10). As for the case \( m = 1 \), it may be verified that the informati
(33.

Variance agrees with the exact variance to \( O(\frac{1}{N^2}) \) for \( n = 1, 2 \) and \( \infty \).

Formula (7.6) is useful since it gives the following optimum rule based on the information variance:

(i) If \( f(\alpha_1, \alpha_2) = \alpha_1(1 - \alpha_2) + 2\alpha_2 > 0 \), take \( n = 2 \) or \( n = 1 \) according as formula (7.6) with \( n = 2 \) or formula (6.10) is the smallest.

(ii) If \( f(\alpha_1, \alpha_2) < 0 \), take \( n = \infty \) or \( n = 1 \) according as formula (7.6) with \( n = \infty \) is smaller or greater than (6.10). As far as a choice between \( n = 1 \) or \( n = \infty \) is concerned, this will give the same answer as the exact variance since to \( O(\frac{1}{N}) \) these quantities agree exactly. It follows that the optimum rules in this case will be identical.

The above rule is interesting since it only allows for a choice between three block sizes viz. \( n = 1, 2 \) and \( \infty \) (or \( n \) as large as possible) and is thus of exactly the same form as the rule for the second order moving average discussed in case 2 of Section 4. The exact variance, on the other hand, does allow for all block sizes. However, the information variance brings out the very important fact that for block sizes between .2 and \( \infty \), there is little reduction in variance. In Table 1 below we give for representative values of \( \alpha_1 \) and \( \alpha_2 \) in the region where an intermediate block size is optimum, a comparison of the optimum block sizes given by the exact and information variances together with the corresponding variances divided by \( \frac{4\sigma^2}{N} \) at these points. For the purposes of tabulation we define quantities \( \phi(n) \) and \( \psi(n) \) such that

\[
\text{var} (\Delta) = \frac{4\sigma^2}{N} \phi(n) + O(\frac{1}{N^2})
\]

(7.7)

for the exact variance and
\[ \text{var} (\Delta) = \frac{\lambda}{N} \psi(n) + O\left(\frac{1}{N^2}\right) \quad (7.8) \]

for the information variance. It follows that \( \phi(n) \), \( \psi(n) \) are measures of the efficiency of the design relative to one for which the errors are independent, or equivalently, relative to one where the dependence is destroyed by completely randomising the treatments (see Section 5).

| TABLE 1 |
| VARIANCES OF TREATMENT DIFFERENCES AS A FUNCTION OF BLOCK SIZE FOR THE CASE \( m = 2 \) |

<table>
<thead>
<tr>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
<th>( \phi(n) )</th>
<th>( \psi(n) )</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.4</td>
<td>0.778</td>
<td>0.307*</td>
<td>0.467</td>
</tr>
<tr>
<td>-0.318</td>
<td>0.186</td>
<td>3.326</td>
<td>0.542*</td>
<td>0.830</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.6</td>
<td>28.000</td>
<td>0.106*</td>
<td>3.226</td>
</tr>
<tr>
<td>0.1</td>
<td>0.8</td>
<td>3.000</td>
<td>0.083*</td>
<td>0.432</td>
</tr>
<tr>
<td>-0.667</td>
<td>0.25</td>
<td>28.333</td>
<td>0.098*</td>
<td>0.098</td>
</tr>
<tr>
<td>-0.47</td>
<td>0.143</td>
<td>4.571</td>
<td>0.449</td>
<td>0.472</td>
</tr>
</tbody>
</table>

[* denotes block size for minimum variance]

It is to be observed that the oscillations in the exact variance are smoothed out by the information variance but that in all cases they give the same value of \( n \) which minimises \( \text{var} (\Delta) \). In the fifth case, all block sizes greater
than two give the same variance; there are of course strong practical grounds
for keeping block sizes as small as possible so that \( n = 2 \) would be chosen.
Similar considerations apply even when the minimum variance is achieved by
using a large block size especially when the reduction in variance over say
\( n = 2 \) or 3 is not great as, for example, in the last case of Table 1.

8. An Alternative Interpretation in Terms of the Spectrum and Signal to Noise
Ratios.

In the previous sections it has been shown that the selection of an
optimum design, within the class of designs to which we have restricted our-
selves, is equivalent to a choice of optimum block size. To a high degree of
approximation this choice can be made independently of the size of the exper-
iment. We now proceed to give this another interpretation.

It becomes apparent that a choice of \( n = 1 \) corresponds to a rapid change
of treatments from one level to the other and this may be demonstrated pictorially
as follows:

\[
\begin{align*}
\text{Level 1:} & \quad x \quad x \quad x \quad x \quad x \\
\text{Level 2:} & \quad x \quad x \quad x \quad x 
\end{align*}
\]

If \( n \) is large, however, the treatments are changed slowly:

\[
\begin{align*}
\text{Level 1:} & \quad x \quad x \quad x \quad x \quad x \quad \cdots \quad x \\
\text{Level 2:} & \quad x \quad x \quad x \quad x \quad x \quad \cdots \quad x 
\end{align*}
\]

It becomes natural therefore to postulate a frequency of change of treatments
which we define as \( \Omega = \frac{2\pi \beta}{N} = \frac{\pi}{n} \). If \( n = 1 \), \( \Omega \) takes on its maximum value
of \( \pi \) and if \( n = N/2 \), \( \Omega = 2\pi/N \) which approaches zero frequency as \( N \to \infty \).

If the true difference in response at levels 1 and 2 is \( \beta \), then in
the absence of random errors, what would be observed is a signal of amplitude
\( \beta \) and frequency \( \Omega \). In actual experiments, this difference is masked by the
fact that superimposed on the signal is a random noise or time-series with
specified a.c.f. or equivalently a specified spectral density $g(\omega)$ where $\omega$
is a continuous frequency ranging from 0 to $\pi$ if the time-interval between
runs is regarded as the unit interval. The problem may thus be regarded as
one of detection of the signal in background noise and is typical of many
problems which arise in the optimisation of electrical systems subject to
random disturbances.

In problems of this nature, engineers use as an optimisation criterion,
a quantity known as the **signal-to-noise power ratio**. In our problem, the
signal-to-noise ratio is just $\beta^2/g(\omega)$ and since $\beta$ is a constant, this
quantity is maximised by choosing that value of $\omega$ which minimises $g(\omega)$.
It is to be emphasized, however, that there is nothing sacrosanct about the
various criteria based on signal-to-noise ratios apart from their obvious
intuitive appeal; they are, in fact, heavily dependent on certain assumptions,
notably the normal or gaussian nature of the noise. We now proceed to modify
the original problem slightly so as to give the signal-to-noise ratio a simple
statistical interpretation.

It is now assumed that the treatments are changed continuously from
level 1 to level 2 in the form of a cosine wave, thereby replacing the frequency
$\Omega$ which can only take values which are submultiples of $\pi$ by a continuous
frequency $\omega$ taking any value between 0 and $\pi$ radians per second. Our
model for the response then becomes

$$y_t = \alpha + \beta \cos \omega t + z_t$$

(3.1)

where $\alpha$ is a constant representing the theoretical response at a level inter-
mediately between levels 1 and 2 and $\{z_t\}$ is a sample from a discrete stationary
time-series defined at time points \( t = 1, 2, \ldots, N \) and possessing a theoretical spectrum \( g(\omega) \). The statistical problem is then one of testing the null hypothesis \( \beta = 0 \) or alternatively of making confidence intervals about \( \beta \).

Some word of explanation is necessary about the modified model (8.1). We are not assuming now that the production process is continuous since the \( z_t \) is still defined for discrete times \( t = 1, 2, \ldots, N \). This means that instantaneously (i.e., while the run is in progress) conditions are kept constant so that there are no transient responses as in the case of continuous time processes discussed by J. Chermuham and G. E. P. Box [17]. The model may be regarded, therefore, as one in which the conditions are changed in between runs and it has been adopted because cosine waves are much easier to deal with statistically than sampled square waves, as the previous analysis shows. There is of course no reason why the process variable should not be changed according to (8.1) in batch processes except possibly one of convenience.

If the \( \{z_t\} \) process is changed to a continuous process \( \{z(t)\} \), (8.1) corresponds exactly to the perturbed steady state response for small changes in the process variable discussed by Chermuham and Box. If the output \( y(t) \) is then sampled, the analysis of this section and Appendix 2 would apply exactly to this case.

It is shown in Appendix 2 that the likelihood ratio criterion for testing the null hypothesis \( \beta = 0 \) against alternatives \( |\beta| > 0 \) is approximated fairly closely by means of the criterion

\[
\Phi_N = \frac{\sum_{t=1}^{N} y_t \cos \omega t}{\sigma \sqrt{2\pi N g(\omega)}}
\]

(8.2)

and the null hypothesis rejected if \( \Phi_N > c \) where \( \Pr \{\Phi_N > c | \beta = 0\} = \alpha \).

The constant \( c \) may be determined from the fact that for large \( N \), under the
null hypotheses that \( \beta = 0 , \phi_N \) is normally distributed with zero mean and standard deviation unity.

Furthermore, it will be shown that for fixed \( \beta \), the asymptotic power function of this test is given by

\[
P = 1 - \Phi \left( c - \frac{\beta \sqrt{N}}{2\sigma \sqrt{2\pi g(\omega)}} \right)
\]

\( (8.3) \)

where \( \Phi(x) \) is the cumulative normal distribution function. Hence for fixed \( N \) and \( \beta \), \( P \) will be maximised by taking \( g(\omega) \) as small as possible. The power function, which is the probability of detecting a departure of size \( \beta \), if thus seen to be the statistical analogue of the detection probability in electrical problems. What is more important as far as the present problem is concerned is that in so far as (8.1) is an approximation to the original problem, an approximate solution to the problem of choosing the optimum block size is given by finding \( \omega_0 \) for which \( g(\omega) \) is a minimum and then solving for

\[
\omega_0 = \Omega_0 = \frac{\pi}{n}
\]

\( (8.4) \)

It is now proposed to compare this solution with the minimum variance solution in certain special cases.

Case (i). \( \rho_1 \neq 0 , \rho_s = 0 , (s \geq 2) \)

The spectral density \( g(\omega) \) is given by

\[
\pi g(\omega) = 1 + 2\rho_1 \cos \omega (0 \leq \omega \leq \pi)
\]

\( (8.5) \)

and it is easily verified that \( g(\omega) \) has a minimum at \( \omega = 0 \) if \( \rho_1 < 0 \) and at \( \omega = \pi \) if \( \rho_1 > 0 \). The approach of Section 5 gives for the same problem,

- if \( \rho_1 < 0 \), take \( n = N/2 \) so that \( \Omega = 2\pi/N \)
- if \( \rho_1 > 0 \), take \( n = 1 \) so that \( \Omega = \pi \).
It appears that for large $N$, the two approaches are identical and that the difference for finite $N$ is due to the discrete nature of the signal in the original problem.

Another link between the two approaches follows from (5.2) since this may be rewritten in the form

$$\text{var}(\Delta) = \begin{cases} \frac{4\sigma^2 \pi g(n)}{N} + O\left(\frac{1}{N^2}\right) & \text{if } n = 1 \\ \frac{4\sigma^2 \pi g(n)}{N} + O\left(\frac{1}{N^2}\right) & \text{if } n = N/2 \end{cases}$$

(8.6)

so that $n = 1$ or $n = N/2$ is chosen according as $g(n) \leq g(0)$, i.e., $\hat{\rho}_1 \geq 0$.

**Case (ii).** $\hat{\rho}_1 \neq 0 \neq \hat{\rho}_2$, $\hat{\rho}_3 = 0$ ($\hat{\delta} \geq 3$)

In this case the spectral density is given by

$$\pi g(\omega) = 1 + 2\hat{\rho}_1 \cos \omega + 2\hat{\rho}_2 \cos 2\omega$$

This has a minimum at $\omega_0$ where $\hat{\rho}_1 + 4\hat{\rho}_2 \cos \omega_0 = 0$, provided $\hat{\rho}_2 > 0$ and $|\hat{\rho}_1| < 4|\hat{\rho}_2|$. If $|\hat{\rho}_1| > 4|\hat{\rho}_2|$, then $g(\omega)$ has a minimum at $\omega = 0$ or $\omega = \pi$ according as $\hat{\rho}_1 < 0$ or $\hat{\rho}_1 > 0$.

It is clear that the spectral approach and minimum variance approach differ in two respects.

(i) The boundaries separating the choice of $\omega = 0$ or $\omega = \pi$ with that of an intermediate frequency (corresponding to a block size intermediate between $n = 1$ and $n = N/2$) differ in the two cases. In the spectral case, these are $\hat{\rho}_1 = \pm 4\hat{\rho}_2$, which is to be compared with $\hat{\rho}_1 = \pm 2\hat{\rho}_2$ for the minimum variance approach.

(ii) The spectral approach allows for the choice of all block sizes in the range $1 \leq n \leq N/2$ corresponding to the solution of
whereas the minimum variance approach leads to just 3 choices

viz. \( n = 1, 2 \) and \( N/2 \) (see Figure 1).

It is to be observed, however, that the boundary between the regions corresponding to the choice of \( n = 1 \) and \( n = N/2 \) is the same in both cases, viz. \( \rho_1 = 0 \).

**Case (iii).** \( \rho_s = A_1 \pi_1 |s| + A_2 \pi_2 |s| \cdot (|\pi_1| < 1) \) \hspace{1cm} (8.7)

In the case where \( \rho_1 = \pi_1 |s| \), the position is almost identical with case (i) so that this will not be discussed.

When the a.c.f. is of the form (8.7), it may be verified that the spectrum is given by

\[
\pi g(\omega) = g(\alpha_1, \alpha_2) \left\{ 1 + \alpha_1^2 + \alpha_2^2 - 2\alpha_1(1 - \alpha_2) \cos \omega - 2\alpha_2 \cos 2\omega \right\}^{-1}
\] \hspace{1cm} (8.8)

where \( g(\alpha_1, \alpha_2) = \frac{1 + \alpha_2}{1 - \alpha_2} \left\{ (1 - \alpha_2)^2 - \alpha_1^2 \right\} \).

The relation between the \( \pi \)'s and the \( \alpha \)'s has already been given in Section 6. It may be verified that this has a minimum at \( \omega_0 \) where

\[ 4\alpha_2 \cos \omega_0 + \alpha_1(1 - \alpha_2) = 0 \] provided \( |\alpha_1(1 - \alpha_2)| < 4|\alpha_2| \) and \( \alpha_2 < 0 \).

If \( |\alpha_1(1 - \alpha_2)| > 4\alpha_2 \) and \( \alpha_2 > 0 \) or else \( |\alpha_1(1 - \alpha_2)| > 4\alpha_2 \), then \( g(\omega) \) has a minimum at \( \omega = 0 \), or \( \pi \) according as \( \alpha_1 < 0 \) or \( \alpha_1 > 0 \).

It is not possible to compare these regions directly with those obtained from the minimum variance approach since the latter were too difficult to derive. However, they are in very close agreement with the corresponding boundaries obtained from the information variances in Section 7. In addition the boundaries separating the regions corresponding to \( n = 1 \) and \( n = N/2 \) are identical for the exact and information variances and for the spectral.
approach. This corresponds to a large percentage of the total stationarity region in the \((\alpha_1, \alpha_2)\) plane.

The region corresponding to an intermediate value of \(n\), i.e.

\[
\cos \frac{\pi}{n_0} = \frac{\alpha_1 (1 - \alpha_2)}{4 \alpha_2}
\]  

(8.9)
is thus small relative to the region corresponding to \(\omega = 0\) or \(\pi\); in order to compare the solutions given by the two approaches in this region, the values of \(n_0\) and the corresponding \(n_{\text{opt}}\) from the minimum variance approach are given in Table 2 for a range of \(\alpha_1, \alpha_2\) values.

As in Table 1, we use \(\phi(n)\) for tabulation and compare it with \(\pi \times \min g(\omega)\) \((0 \leq \omega \leq \pi)\).

**TABLE 2**

**COMPARISON OF OPTIMUM BLOCK SIZES OBTAINED FROM SPECTRAL AND MINIMUM VARIANCE APPROACHES**

<table>
<thead>
<tr>
<th>(\alpha_1)</th>
<th>(\alpha_2)</th>
<th>(n_{\text{opt}})</th>
<th>(\phi(n_{\text{opt}}))</th>
<th>(\pi \times \min g(\omega))</th>
<th>(n_0)</th>
<th>(\phi(n_0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.4</td>
<td>2</td>
<td>0.307</td>
<td>0.304</td>
<td>2</td>
<td>0.307</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.6</td>
<td>2</td>
<td>0.106</td>
<td>0.105</td>
<td>2</td>
<td>0.106</td>
</tr>
<tr>
<td>-0.667</td>
<td>0.25</td>
<td>2 (\rightarrow) (\infty)</td>
<td>0.098</td>
<td>0.087</td>
<td>3</td>
<td>0.098</td>
</tr>
<tr>
<td>-0.47</td>
<td>0.143</td>
<td>(\infty)</td>
<td>0.389</td>
<td>0.378</td>
<td>4</td>
<td>0.472</td>
</tr>
</tbody>
</table>

It appears that very little is lost by using \(n_0\) instead of \(n_{\text{opt}}\) except possibly in the last case. However, this raises a further interesting point; by restricting the class of designs to those in which the block sizes are equal, the systematic design which gives rise to the smallest variance might be excluded. It is conceivable, for example, that a design in which the block sizes are alternately two and one may in certain cases give rise to a variance which
is lower than that for which the block sizes have a constant size of two or one respectively. The resultant saving in this direction is likely to be small in view of the flatness of the minimum variance curve near \( n_{opt} \) so that it does not seem worth while to depart from a constant block size.

Some General Results for Choosing between \( n = 1 \) and \( n = N/2 \).

It has been observed that in all the special cases considered in Section 5 and 6, the minimum variance and spectral approaches give the same answer when the choice lies between \( n = 1 \) and \( n = N/2 \). This is due to the fact that

\[
\text{var} (\Delta) = \begin{cases} 
\frac{4\sigma^2 \pi g(\pi)}{N} + O\left(\frac{1}{N^2}\right) & \text{when } n = 1 \\
\frac{4\sigma^2 \pi g(0)}{N} + O\left(\frac{1}{N^2}\right) & \text{when } n = N/2
\end{cases} \tag{8.10}
\]

so that minimising \( \text{var} (\Delta) \) is equivalent to minimising \( g(\omega) \) in these cases. (8.10) follows from (4.7) and (4.8), the degree of accuracy of this formula depending on the rate at which \( \rho_s \to 0 \) as \( s \to \infty \). From formula (A2.19) of Appendix 2 it follows that if the treatment is varied sinusaicially, then

\[
\text{var} (\hat{\beta}) = \frac{2\sigma^2 \pi g(\omega)}{N} + O\left(\frac{1}{N^2}\right) \tag{8.11}
\]

so that the result is true for all \( \omega \). In the original problem, no relationship such as (8.11) seems to exist. It should be noted, however, that the region corresponding to a choice of either \( \omega = 0 \) or \( \omega = \pi \) occupies the greatest part of the stationary region so that it would seem worth while to have a general result to cover this case.

**Theorem:** If the choice of \( n_{opt} \) lies between \( n = 1 \) and \( n = N/2 \), then \( n = 1 \) or \( n = N/2 \) is chosen according as the sum of the odd autocorrelations
viz. \[ \sum_{s=1}^{\infty} \rho_{2s-1} \text{ is } < 0 \text{ or } > 0. \]

**PROOF:** By virtue of the previous discussion, \( n = 1 \) or \( n = N/2 \) is chosen according as \( g(\pi) < g(0) \) or \( g(\pi) > g(0) \) i.e. according as

\[
1 + 2 \sum_{s=1}^{\infty} \rho_{s} \leq 1 + 2 \sum_{s=1}^{\infty} (-1)^s \rho_{s}
\]

In view of the absolute convergence of the series \( \sum_{s=1}^{\infty} \rho_{s} \), this inequality may be rearranged to give

\[
\sum_{s=1}^{\infty} \rho_{2s-1} \leq 0. \tag{8.12}
\]

The results referring to the boundary separating the regions corresponding to \( n = 1 \) and \( n = N/2 \) which were derived for particular a.c.f.'s in Sections 5 and 6 may be verified to be special cases of formula (8.12).
PART III: APPLICATIONS

9. An Illustrative Example

The data used for illustrating the previous theoretical work (which has already been described in Section 2) was collected under conditions which must be regarded as ideal. By using identical feedstocks and through efficient process control, a uniformity trial was conducted (i.e., conditions were kept as nearly constant as possible) and the percentage yields in 70 consecutive runs obtained. These observations were used to estimate the autocorrelation structure of the uncontrolled variation so that optimum block size estimates could be obtained for further experiments.

From an examination of the data, there seemed little justification for transforming the observations before analysis. The sample serial (auto)-correlations $r_s$ are given as a function of the lags in Table 3 and plotted in Figure 2. The definition of $r_s$ used was $r_s = c_s/c_0$ where $c_s$ is the covariance between the first $N-s$ and the last $N-s$ terms of the series.

<table>
<thead>
<tr>
<th></th>
<th>$r_s$</th>
<th>$\hat{\rho}_s$</th>
<th>$v_s$</th>
<th></th>
<th>$r_s$</th>
<th>$\hat{\rho}_s$</th>
<th></th>
<th>$r_s$</th>
<th>$\hat{\rho}_s$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.390</td>
<td>-0.390</td>
<td>-0.390</td>
<td>6</td>
<td>-0.048</td>
<td>0.042</td>
<td>11</td>
<td>0.130</td>
<td>-0.004</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.310</td>
<td>0.310</td>
<td>0.186</td>
<td>7</td>
<td>0.040</td>
<td>-0.026</td>
<td>12</td>
<td>-0.082</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.175</td>
<td>-0.171</td>
<td>-0.005</td>
<td>8</td>
<td>-0.050</td>
<td>0.016</td>
<td>13</td>
<td>0.182</td>
<td>-0.002</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.074</td>
<td>0.112</td>
<td>-0.050</td>
<td>9</td>
<td>-0.006</td>
<td>-0.010</td>
<td>14</td>
<td>0.042</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.107</td>
<td>-0.068</td>
<td>-0.077</td>
<td>10</td>
<td>0.019</td>
<td>0.006</td>
<td>15</td>
<td>-0.010</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

In order to see how many terms of the autoregression given by equation (2.2) should be included to give an adequate fit to the observed a.c.f., the partial
a.c.f. $v_s$ was plotted as a function of lag as described by G. M. Jenkins [18]. The partial autocorrelation coefficient $v_s$ may be regarded as the ordinary partial correlation coefficient between the variables $y_t$ and $y_{t-s}$ when the effects of the intermediate variables have been eliminated. Another way of looking at this is that $v_s$ is the maximum likelihood estimator of $\alpha_s$ (subject to small end corrections) in an autoregressive scheme of order $s$. Referring to the autoregressive formula (6,1), $v_s$ is an estimate of $\alpha_s$ if $m = s$. It follows that since the $\alpha_s$ damp out as $s$ increases, the empirical partial a.c.f. $v_s$ should do likewise, subject to sampling fluctuations. In order to determine when to stop including terms in the autoregression, it is only necessary to observe that if the autoregression is of order $(s-1)$, $\alpha_s = 0$ so that $E(v_s) = 0$. If a few consecutive values of $v_s$ are small, the process may be terminated.

[The $v_s$ may be obtained as the coefficients of $\alpha_s$ in the following set of $s$ equations]

$$r_j = \alpha_1 r_{j-1} + \alpha_2 r_{j-2} + \ldots + \alpha_s r_{j-s}, \quad j = 1, 2, \ldots, s$$

with the provision that where $j < s$, $r_{j-s} = r_{s-j}$ in view of the symmetry of the autocorrelation function about the origin. For large $s$, the above formula is not too accurate and if a general plotting of the partial autocorrelation function $\{v_s\}$ for large $s$ is envisaged, it is best to calculate $v_s$ as follows:

For each value of $s$ introduce the series

$$X_1: x_1, x_2, \ldots, x_{n-s}$$
$$X_2: x_2, x_3, \ldots, x_{n-s+1}$$
$$\vdots$$
$$X_s: x_s, x_{s+1}, \ldots, x_n$$

Then calculate the ordinary partial correlation coefficient between $X_1$ and
$X_s$ when the effects of the intermediate variables have been eliminated.

Appropriate formulae for the calculations may be found in the book by M. G. Kendall [16].

The results in [18], together with those of H. E. Daniels [19] indicate that under the null hypothesis that the scheme is of order $s - 1$ against alternatives of order $s$, the significance of the lower order partial autocorrelations may be tested using the following approximate expressions for the mean and variance and taking $v_s$ to be normally distributed:

$$E(v_s) = \begin{cases} 
- \frac{1}{N} & \text{if } s \text{ is odd} \\
- \frac{2}{N} & \text{if } s \text{ is even}
\end{cases} \quad (9.1)$$

$$\text{var}(v_s) = \frac{1}{N-s}, \text{ all } s \quad (9.2)$$

For $N = 70$, this approximation will be more than adequate. The $v_s$ together with the approximate upper and lower $2.5\%$ significance limits are plotted in Figure 3. It is to be observed that it is only $v_1 \equiv r_1$ which is significant at the $5\%$ level but the fit of the a.c.f. based on a first order autoregressive scheme is not good. However, the second term is large and if a second order autoregressive scheme is fitted, there is very good agreement between the observed and fitted a.c.f.'s as may be seen from Table 3 and Figure 2. Like most statistical problems this is best regarded as a problem of estimation and not of hypothesis testing.

The fitted constants in formula (6.1) are given from the solution of the least squares equations viz.

$$\hat{\alpha}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} = -0.3176 \quad (9.3)$$

$$\hat{\alpha}_2 = \frac{r_2 - r_1^2}{1 - r_1^2} = 0.1864$$
FIGURE 3  Partial autocorrelation and $2\frac{1}{2}\%$ significance limits.
and these give rise to estimates of $\pi_1, \pi_2$ in the form

\begin{align*}
\hat{\pi}_1 &= 0.301 \\
\hat{\pi}_2 &= -0.619
\end{align*}

The fitted a.c.f. may then be generated from the recurrence relation

\[ \hat{\rho}_s = -0.3176 \hat{\rho}_{s-1} + 0.1864 \hat{\rho}_{s-2} \]  \hspace{1cm} (9.4)

It is to be noted that the second coefficient in (9.4) is positive whereas the theoretical model (2.2) predicted that all the autoregressive constants would be negative. It is suggested that one or some of the following reasons may be responsible for this effect.

(i) There may in fact be small trends in the mean in the data and these would normally inject positive correlations over and above the trend free situation.

(ii) Inertia in the equipment recording the response (percentage yield) would trend to induce positive correlation into the actual values recorded.

(iii) The a.c.f. given by (2.3) is the a.c.f. of the model (2.2) only if the effect of the transients in the solution have died out. Transients will be generated at the start of the experiment and also in this particular examples at weekly intervals due to the removal of some residues in the distillation column by a process known as "steaming".

Whatever the cause of the discrepancy, it is necessary to design experiments based on the observed a.c.f. and not on the predicted model. We shall now apply the theory of Sections 6, 7, 8 to the choice of optimum block size in the case of simple two level experiments.
In Table 4 we give various expressions for the variance of treatment difference as a function of block size for this particular example (as in previous tables, we quote the standardised variance i.e. the variance divided by the factor $4\sigma^2/N$). Also included in Table 4 are the fitted spectrum $\eta_0(\omega)$ corresponding to (8.8), evaluated at $\omega = \pi / n$ and an estimated spectrum based on a spectral analysis of the observed a.c.f. In the latter, the weight function used in the weighted Fourier transform of $r_s$ was $D_2^*(t)$ given by R. B. Blackman and J. W. Tukey [20], p. 14).

We may summarise information in Table 4 as follows:

1. The exact variances are higher than those to $O(1/N)$ but lead to the same choice of optimum block size viz. $n = 2$.

2. The first five quantities in Table 4 lead to a remarkably consistent estimate of the variance corresponding to the optimum block size $n = 2$. Compared with complete randomisation of the treatments, this corresponds to an estimated relative efficiency of 1.86.

3. The standardised variance corresponding to randomisation within blocks of size 2 is 1.4, representing a relative efficiency of the best systematic design of 2.6.

4. The choice of the Tukey-Blackman estimate of the spectral density was dictated by the availability of a computer programme. It is to be noticed that the fitted and estimated spectrum agree very well (they are plotted together in Figure 4) and this is a consequence of a reasonable choice of truncation point viz. $s = 8$, in the calculation of the spectrum. This was based on a visual examination of where the a.c.f. damped out.

5. From equation (8.9), the fitted spectrum gives rise to a minimum value at

$$\cos \frac{\pi}{n} = \frac{-\alpha_1 (1 - \alpha_2)}{4\alpha_2} = 0.388\pi$$
<table>
<thead>
<tr>
<th>( n \times \text{Estimated} ) Spectra</th>
<th>( n \times \text{Estimated} ) Spectra (cm⁻¹)</th>
<th>&lt; n \times \text{Estimated} ) Spectra (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.69</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>2.69</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>2.69</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>( \sigma = 79 )</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>( \sigma = 79 )</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>( \sigma = 79 )</td>
<td>0.63</td>
<td>0.63</td>
</tr>
</tbody>
</table>

| \( \sigma = 79 \)                       | 0.63                                           | 0.63                                           |
| \( \sigma = 79 \)                       | 0.63                                           | 0.63                                           |
| \( \sigma = 79 \)                       | 0.63                                           | 0.63                                           |

<table>
<thead>
<tr>
<th>Exact Value</th>
<th>Exact Value</th>
<th>Exact Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
</tr>
</tbody>
</table>

**Table 4**
and this corresponds to an optimum block size of 2.6. The estimated spectrum has a minimum at \( \omega = 3\pi/8 \) corresponding to \( n = 2.7 \). While both of these spectral estimates seem to point to \( n = 3 \) rather than \( n = 2 \), since the resulting saving is relatively small there are strong practical grounds for taking \( n = 2 \). In addition to this, the most important thing which emerges from this analysis is that blocks of size 1 must be avoided at all costs.

(6) The analysis clearly shows that in this particular example an increase in the efficiency results with larger negative serial correlations. This would suggest that the quantity \( \lambda \) in equation (2.2), the fraction of material retained in the distillation column and not discharged at the end of each run, should be made as large as possible. From the practical viewpoint because of the limited capacity of the distillation column and the possibility of flooding the column, this can be conveniently accomplished by recycling the "bottoms" (retained material) during distillation.

10. Some Further Considerations

In this section we discuss briefly some further questions which arise as a result of this investigation.

(i) In the previous analysis, it has been assumed that the variance \( \sigma^2 \) and autocorrelations \( \{\rho_n\} \) are known in the calculation of the optimum block size \( n_{\text{opt}} \). It is reasonable to ask how this procedure is likely to be affected by uncertainty in the knowledge of these parameters. In order to begin, it would be necessary to know the probability distribution of \( n_{\text{opt}} \) which would depend in a complicated way on the distribution of the estimates \( \sigma^2 \) and \( \{\beta_n\} \). Even if this calculation were possible, it would still be necessary to ensure that the size of the series, on which these initial estimates were based, was sufficiently
large for most of the mass of the distribution, \( p(n_{\text{opt}}) \), to be concentrated at the best estimate of \( n_{\text{opt}} \). If this were not the case, i.e. \( n_{\text{opt}} \) were subject to large uncertainties, it would be extremely difficult to assess the influence of this uncertainty on the optimum procedures. We are not aware of any statistical problem in which such an investigation has been made. An analogous situation arises in the theory of the design of optimum electrical systems subject to noise, which has been proposed by N. Wiener (\textit{etc.}[21]). Here, also, there is no provision for uncertainty in the statistical parameters, a knowledge of which is required for the calculation of the optimum design parameters.

In general, if \( n_{\text{opt}} \) is subject to large errors, then the actual procedure which is used, may be far from the best possible.

(ii) A problem of greater importance in the present discussion is concerned with the analysis of serially correlated observations. In the current context, this problem would arise if it were required to draw any conclusions on the basis of means \( \bar{y}(1) \) and \( \bar{y}(2) \) in an experiment of size \( N \) conducted at two levels using a previously determined block size \( n_{\text{opt}} \). In particular, it might be necessary to ask whether the difference \( \bar{y}(1) - \bar{y}(2) \) is sufficiently large to establish the existence of an effect in one direction and also to give confidence intervals for the true difference.

Two methods of approach seem possible.

(a) For a given block size, \( n_{\text{opt}} \), an approximate estimate of the variance of \( \bar{y}(1) - \bar{y}(2) \) may be obtained from the formula

\[
\text{var} (\bar{y}(1) - \bar{y}(2)) \sim \frac{\sigma^2}{N} g\left(\frac{\pi}{n_{\text{opt}}}\right)
\]

where \( \sigma^2 \) and \( g(\omega) \) are estimated from the previous analysis which was used to determine \( n_{\text{opt}} \).
As a rough approximation, we may then take

$$Z = \frac{\bar{y}(1) - \bar{y}(2)}{\sqrt{\frac{4\sigma^2\pi}{N} g(\frac{-\pi}{n_{\text{opt}}})}} \quad (10.2)$$

as being normally distributed with zero mean and standard deviation unity under the null hypothesis of no difference. This is subject to the criticism that $\sigma^2 g(\frac{-\pi}{n_{\text{opt}}})$ could be appreciably different for the series under analysis than for the preliminary estimate due to a change in the variance and autocorrelation structure. In view of this, it is better to work with an internal estimate of $\text{var} (\bar{y}(1) - \bar{y}(2))$.

(b) There are some difficulties which arise in the estimation of the $\{\rho_s\}$ from observations obtained at two levels. It would be necessary, first of all, to generate a primary series for analysis from the residuals $y_t - \bar{y}(k)$, $k$ being 1 or 2 depending on the level. This is likely to introduce sharp discontinuities in jumping from one level to the other and these might have serious effect on the estimation of the $\rho_s$. If the block size is not too small, say of the order of 5 or more, then a harmonic analysis could be conducted on each block and the spectrum obtained by averaging these harmonic analyses over frequency. This will mean the loss of a slight amount of information but would avoid the effect of discontinuities. The method would be equivalent to the use of the smoothed periodogram proposed by M. S. Bartlett (see [22]).

Alternatively, if it were felt that the residuals $y_t - \bar{y}(k)$ could be used, then the sample spectrum could be obtained from the serial (auto) correlations of this series by means of the relation

$$g_N(\omega) = \frac{1}{\pi} \left[ 1 + 2 \sum_{s=1}^{m-1} (1 - \frac{s}{n}) \lambda_s r_s \cos \omega s \right] \quad (10.3)$$
where \( m - 1 \) is called the truncation point of the a.c.f. and the \( \lambda_s \) are an appropriate set of weights for which a number of different forms have been proposed. The question then arises of judging the number of degrees of freedom associated with \( \sigma^2 \xi_N(\omega) \). This we can do by approximating its distribution by that of a \( \chi^2 \) distribution and calculating the equivalent number of degrees of freedom from the first two fitted moments using the relation

\[
f = \frac{2 \mathbb{E} \{ \sigma^2 \xi_N(\omega) \}}{\text{var} \{ \sigma^2 \xi_N(\omega) \}}. \tag{10.4}
\]

The value of \( f \) will depend on the choice of \( \lambda_s \) and in Table 5 below we give the appropriate number of degrees of freedom corresponding to various choices of the \( \lambda_s \).

### Table 5

**Equivalent Degrees of Freedom for Spectral Estimates**

<table>
<thead>
<tr>
<th>( \lambda_s^* )</th>
<th>Originator</th>
<th>Equivalent Degrees of Freedom**</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1, (0 \leq s \leq m - 1) )</td>
<td>Bartlett</td>
<td>0.5 n/m</td>
</tr>
<tr>
<td>( 1 - s/m, (0 \leq s \leq m - 1) )</td>
<td>Bartlett</td>
<td>1.5 n/m</td>
</tr>
<tr>
<td>( \frac{1}{2}(1 + \cos \frac{ns}{m}), (0 \leq s \leq m - 1) )</td>
<td>Tukey</td>
<td>1.33 n/m</td>
</tr>
<tr>
<td>( \frac{\sin sh}{sh}, (0 \leq s \leq m - 1) )</td>
<td>Daniell</td>
<td>nh/\pi (h = bandwidth)</td>
</tr>
<tr>
<td>( 1 - (s/m)^2, (0 \leq s \leq m - 1) )</td>
<td>Parzen</td>
<td>n/m</td>
</tr>
<tr>
<td>( 1 - 6(s/m)^2(1 - s/m), (0 \leq s \leq \frac{m}{2}) )</td>
<td>Parzen</td>
<td>1.85 n/m</td>
</tr>
<tr>
<td>( 1 - (s/m)^2, (m/2 &lt; s \leq m - 1) )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The sources of these weighting functions can be conveniently found in references [20] and [22].

** For the estimates at \( \omega = 0 \) and \( \pi \), the number of degrees of freedom will be equal to a half of these values.
As an approximation, we may then take

\[ t = \frac{\bar{y}(1) - \bar{y}(2)}{\sqrt{\frac{\sigma^2}{N} c_N \left( \frac{\pi}{n_{\text{opt}}} \right)}} \]  

(10.5)

as being approximately distributed as Student's "t" distribution with \( f \) degrees of freedom. It is apparent from Table 5 that \( f \) depends on \( m \), the number of lags taken in (10.3). Care must be taken so that \( m \) is not chosen too low so that whilst the estimate is seemingly accurate since it has a large number of degrees of freedom, it is a heavily biased estimate of \( g(\frac{\pi}{n_{\text{opt}}}) \), thus making the test of significance (10.5) insensitive.

This approach can be generalised quite easily when several means have to be compared.

(iii) In this report, we have been concerned entirely with experiments involving a single factor at two levels; this has been sufficient to bring out the general principles involved. In future work it is hoped to consider multi-factor experiments and also the design of regression experiments with special reference to the estimation of the parameters of a response surface.
APPENDIX 1

 Conditions on the Autocorrelations of a Stationary Moving Average Process.

In Section 5, the boundaries of the stationarity region were required for various finite moving average processes. H. O. Wold [23] has proved the following theorem which can be used in the derivation of these boundaries.

THEOREM:

A necessary and sufficient condition that $\rho_s (s = 1, 2, \ldots, m)$, $\rho_s = 0 \ (s > m)$ is the a.c.f. of a stationary finite moving average process

$$x_t = z_t + \beta_1 z_{t-1} + \cdots + \beta_m z_{t-m} \quad \text{(Al.1)}$$

where $\mathbb{E}(z_t z_{t+1}) = \delta_{tt'} \sigma^2$, is that the auxiliary polynomial

$$v(z) = v_0 z^m + v_1 z^{m-1} + \cdots + v_{m-1} z + v_m$$

has no real roots $z_0$ in the interval $-2 < z_0 < 2$. The auxiliary polynomial is obtained by making the substitution $z = x + x^{-1}$ in the generating function

$$\rho(x) = 1 + \rho_1 x + \rho_2 x^2 + \cdots + \rho_m x^m$$

This theorem may be used in certain cases to derive the critical boundaries but the calculations are difficult in the simplest of cases. An alternative theorem which lends itself more easily to calculation is the following.

THEOREM:

If $\rho_s (s = 1, 2, \ldots, m)$, $\rho_s = 0 (s > m)$, then this is the a.c.f. of a wide sense stationary process if the following $m + 1$ conditions are satisfied simultaneously.

(i) $\Sigma_{s=1}^{m} \rho_s > -\frac{1}{2}$

(ii) $\Sigma_{s=1}^{m} (-1)^s \rho_s > -\frac{1}{2}$

(iii) $\Sigma_{s=1}^{m} \rho_s \cos \omega_j s > -\frac{1}{2}$, $j = 1, 2, \ldots, m-1$
where the \( \omega_j \) are the roots of the equation

\[
\sum_{s=1}^{m} \frac{s \rho_s \sin \omega \sin \omega_j}{s \rho_s \sin \omega_j} = 0 \quad (\text{Al.2})
\]

**Proof**

We do not relate this a.c.f. to that of the spectrum of a finite moving average process which was Wold's procedure but make use of the fact that by the Herglotz theorem, the spectral density

\[
\rho_g(\omega) = 1 + 2 \rho_1 \cos \omega + 2 \rho_2 \cos 2\omega + \ldots + 2 \rho_m \cos m\omega, \quad (0 \leq \omega \leq \pi) \quad (\text{Al.3})
\]

must be a non-negative function for all values of the \( \rho_s \) if the series is to be stationary. Since all the derivatives of \( \rho_g(\omega) \) exist and are continuous in this interval, we may locate the extreme values of \( \rho_g(\omega) \) by differentiation. We proceed by finding the minimum values of \( \rho_g(\omega) \) and use the fact that the function will always be positive if it is positive at its minimum values.

It is clear that the stationary points of \( \rho_g(\omega) \) are at \( \omega = 0, \pi \) and at the \( m - 1 \) solutions of the equation

\[
\sum_{s=1}^{m} \frac{s \rho_s \sin \omega \sin \omega_j}{s \rho_s \sin \omega_j} = 0
\]

It is then easily shown by setting all the \( \rho_s \)'s except one equal to zero that these \( m + 1 \) roots will correspond for certain values to the absolute minimum of the function.

The statement of the theorem then follows.

**Special Cases**

\( m = 1 \): From (i) and (ii) we have immediately that \( |\rho_1| \leq \frac{1}{2} \)

\( m = 2 \): From (i) and (ii) we obtain

\[
\begin{align*}
\rho_1 + \rho_2 & > -\frac{1}{2} \\
\rho_1 - \rho_2 & > -\frac{1}{2}
\end{align*}
\]
whilst the solution of \((A1.2)\) viz. \(\rho_1 + 4\rho_2 \cos \omega_0 = 0\) when substituted in \((A1.3)\) gives the further condition
\[
\rho_1^2 < 4\rho_2(1 - 2\rho_2).
\]

\((A1.3)\)

\(m = 3\): From (i) and (ii) we again have
\[
\rho_1 + \rho_2 + \rho_3 > \frac{1}{2}
\]
\[
\rho_1 - \rho_2 + \rho_3 > \frac{3}{2}
\]

while the solutions to \((A1.2)\) are those of a quadratic equation
\[
\rho_1 - 3\rho_3 + 4\rho_2 \cos \omega_0 + 12\rho_2 \rho_0 \cos^2 \omega_0 = 0
\]
so that
\[
\cos \omega_0 = \frac{\rho_2 \pm \phi}{6\rho_3}
\]

where
\[
\phi = (\rho_2^2 - 3\rho_1\rho_3 + 9\rho_3^2)^{1/2}
\]

On substitution in \((A1.3)\), this gives
\[
(27\rho_3^2 + 2\rho_2^2 - 9\rho_1\rho_2\rho_3 - 27\rho_2^2\rho_3^2) \pm 2\phi(3\rho_1\rho_3 - 9\rho_3^2 + \rho_2^2) \geq 0
\]
which yields two further conditions.

---

**APPENDIX 2**

**Likelihood Ratio Criteria for the Modified Model**

In Section 8, it was found convenient to consider a modified model in the form
\[
x_t = A + \beta \cos \omega t + z_t \quad (A2.1)
\]

This approximation to the original model is useful since it leads directly to the interpretation of the optimum block size \(n_{opt}\) in terms of signal-to-noise ratios. We now make three assumptions about the residual variation all of which can be

---

The remainder of the document is not clearly visible in the provided image.
relaxed considerably without affecting the main results.

(a) The \( \{z_t\} \) process is generated by a mixed moving average auto-regressive process of the form

\[
z_t - \alpha_1 z_{t-1} - \cdots - \alpha_k z_{t-k} = \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_m \epsilon_{t-m}
\]

where \( m \leq k - 1 \) and \( \mathbb{E}(\epsilon_t \epsilon_t') = \sigma^2 \delta_{tt'} \).

It will be shown later that this assumption can be relaxed so that the \( \{z_t\} \) process is a general wide-sense stationary process with given a.c.f.

(b) The \( \{\epsilon_t\} \) process is normal and hence the \( \{z_t\} \) process also; again this assumption can be relaxed to a general distribution provided the process remains stationary.

(c) The \( \{\epsilon_t\} \) process is circular or periodic i.e. \( \epsilon_t = \epsilon_{t+Np} \) for all integers \( p \). This implies that the \( \{z_t\} \) process satisfies the circular condition as well.

Assumption (c) is the crucial assumption and if this holds, exact results may be obtained with assumptions (i) and (ii) in their most general form.

Models of the type (A2.1) have been considered by P. Whittle [24] and more general regression models by U. Grenander and M. Rosenblatt [25]. The latter use a method of approach in which they proceed by writing down the expressions for the variances of the best linear unbiased estimates of the regression parameters in terms of the covariance matrix of the \( z_t \). They then show by a rather complex limiting argument that these variances can be written in terms of the spectrum for large \( N \).

In this section we proceed by an entirely different method and write down exact expressions for the likelihood function in the circular case. These results will then be asymptotically true for large \( N \) in the non-circular case.

We now restrict ourselves to the model (A2.1) where without loss of generality we may take \( A \) to be zero. This is due to the fact that for circular processes \( \bar{x} \) is independently distributed of all other parameters and is also a
sufficient statistic for the estimation of $A$. With $A \neq 0$, the results are then identically the same as the following except that $\bar{x}_t$ is replaced by its deviation $(x_t - \bar{x})$ from the sample mean $\bar{x}$.

In the case of independence of the $z_t$, i.e., $\alpha_i = 0 = \beta_j$ (all $i$, $j$), it is easily seen that the log-likelihood function may be written in the form

$$L = \ln L = \frac{1}{2} \ln \frac{2\pi}{2\sigma_z^2} + \frac{1}{2} \left( \sum_{t=1}^{N} x_t^2 - 2N\beta \sum_{t=1}^{N} x_t \cos \omega t + \beta^2 N \right).$$

provided $\omega^2 = \frac{2\pi j}{N}$ for some $j = 1, 2, \ldots, N - 1$. \hfill (A2.3)

It then follows that the maximum likelihood estimator of $\beta$ is given by

$$\hat{\beta} = \frac{1}{N} \sum_{t=1}^{N} x_t \cos \omega t. \hfill (A2.3a)$$

and this is also the best linear unbiased estimate of $\beta$. In the case where the $z_t$ are not independent, it is necessary to derive the likelihood function by a rather indirect method.

From now on we assume circularity. The sample periodogram of the $z_t$ process is defined as

$$f_N(z) = \frac{1}{\pi N} \sum_{t=1}^{N} z_t e^{i \omega t} = \frac{1}{\pi N} g_N(z) g_N(z)^* \hfill (A2.4)$$

where

$$g_N(z) = \sum_{t=1}^{N} z_t e^{i \omega t}.$$

This may then be written in the form

$$f_N(z)(\omega) = \frac{1}{\pi} \sum_{s=0}^{N-1} c_s(z)e^{i \omega s} \hfill (A2.5)$$

provided $\omega = \frac{2\pi j}{N}$, $j = 0, 1, 2, \ldots, N - 1$, where $c_s$ is the sample circular serial covariance of lag $s$, namely

"
\[ c_s(z) = \frac{1}{N} \sum_{t=1}^{N} z_t z_{t+s} \]  

(A2.6)

This means that \( f_N(z)(\omega) \) may be written as a finite Fourier series. It is clear that \( c_s = c_{Np-s} = c_{s-Np} \) for all integral \( p \) so that (A2.5) could be written as a two-sided transform. In this case, however, it would be necessary to distinguish between \( N \) odd and \( N \) even so that the form (A2.5) is preferred.

The usual Fourier-Stieltjes inversion formula for (A2.5) then takes the form

\[ c_s(z) = \frac{\pi}{2N} \sum_{j=0}^{N-1} f_N(\omega_j) e^{-i\omega_j s} \]  

(A2.7)

The basic result in the following approach is an exact relationship connecting the sample periodograms of the \( \{z_t\} \) and \( \{e_t\} \) process as given by (A2.2) at the frequencies \( \omega_j \). This takes the form

\[ g_N(z)(\omega_j) = g_N(e)(\omega_j) |\psi(e^{i\omega_j})|^2 \]  

(A2.8)

where the transfer function \( \psi(z) \) is given by

\[ \psi(z) = \frac{1 + \beta_1 z + \beta_2 z^2 + \ldots + \beta_m z^m}{1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_k z^k} \]  

(A2.9)

(A2.8) is easily verified by multiplying both sides of (A2.2) by \( e^{i\omega j} \) and summing, using the fact that \( z_t \equiv z_{N+t} \) and \( e_t \equiv e_{N+t} \). It then follows from (A2.8) and (A2.4) that

\[ f_N(z)(\omega_j) = f_N(e)(\omega_j) |\psi(e^{i\omega_j})|^2 \]  

(A2.10)

Formula (A2.10) has the following consequences, using (A2.5) and taking expectations:

\[ \frac{\sigma_x^2}{\pi} \sum_{s=0}^{N-1} \rho_s(\omega) e^{i\omega s} = \frac{\sigma_e^2}{\pi} |\psi(e^{i\omega})|^2 \]  

(A2.11)

If we now define formally the spectrum of \( x_t \) (defined only at frequencies
\[ \omega_j = 2\pi j/N \] by
\[ g_x(\omega_j) = \frac{1}{\pi} \sum_{s=0}^{N-1} \rho_s(x)e^{i\omega_j s} \tag{A2.12} \]
then (A2.11) may be written as
\[ a_x^2 g_x(\omega_j) = \frac{\sigma^2}{\pi} |\psi(e^{i\omega_j})|^2 \tag{A2.13} \]
which is the usual relation for the spectral density of the output of a linear system for a white noise input.

If we now write \[ z_t = x_t - \beta \cos \omega t, \] in (A2.4), we obtain
\[ \mathbb{E}[N_N(x)(\omega_j)] = \mathbb{E}[N_N(x)(\omega_j)] - g_N(x)(\omega_j) c_N(-\omega_j) + |c_N(\omega_j)|^2 \tag{A2.14} \]
where
\[ c_N(\omega_j) = \sum_{t=1}^{N} \cos \omega e^{i\omega_j t} \tag{A2.15} \]
We may derive an expression for \[ c_0(\varepsilon) \] in terms of the observations by using (A2.7)
with \[ s = 0, \] (A2.10) and (A2.14). This simplifies very easily by observing that for fixed \[ \omega_j \] summation over \[ t \] leads to zero unless \[ \omega_j = \omega \] or \[ 2\pi - \omega \].
The final result is
\[ c_0(\varepsilon) = \frac{\pi}{2N} \sum_{j=0}^{N-1} \frac{f_N(x_j)(\omega_j)}{|\psi(e^{i\omega_j})|^2} - \frac{2\beta}{N} \sum_{j=1}^{N-1} \frac{\cos \omega e^{i\omega_j t}}{|\psi(e^{i\omega_j})|^2} + \frac{\beta^2}{2|\psi(e^{i\omega_j})|^2} \tag{A2.16} \]
If we let the Jacobian of the transformation from the \[ \varepsilon \)'s to the \[ x \)'s be \[ |J| \]
then the log likelihood is given by
\[ \ln L = \ln L - \ln N \ln \sigma + \ln |J| - \frac{N c_0(\varepsilon)}{2 \sigma_\varepsilon^2} \tag{A2.17} \]
It is easily verified that the maximum likelihood estimate of \( \hat{\beta} \) is exactly the same as in the case of independence, viz. (A2.3a). Application of classical least squares theory to \[ N c_0(\varepsilon) \] shows also that (A2.3a) is the best linear unbiased estimate and that its variance is given exactly by
The remarkable thing about this result is that the only ideas involved in its development are those of finite Fourier series. If we introduce formally the spectrum as given by (A2.12) then (A2.13) may be written alternately as

\[
\var \hat{\beta} = \frac{2\pi \sigma_x^2}{N} \mathcal{g}_x(\omega')
\]  

(A2.19)

We now proceed to derive the likelihood ratio criteria connected with various tests of the hypothesis \( \beta = 0 \) against alternatives \( |\beta| > 0 \).

(1) Assume that \( \sigma_x^2 \) (or equivalently \( \sigma_x^2 \)) and \( |\psi(1\omega')|^2 \) (or equivalently \( \mathcal{g}_x(\omega') \)) are known. Then the log of the likelihood ratio for the above hypothesis may be written as

\[
\lambda_1 = L(\hat{\beta}) - L(\beta = 0) = \frac{\beta}{\sigma_x^2} \sum x_t \cos \omega' t - \frac{N^2}{4\sigma_x^2} |\psi(1\omega')|^2
\]

so that we could use the test criterion \( \sum x_t \cos \omega' t \). This may be normalized using (A2.19) in the form

\[
\phi_N = \frac{1}{\sigma_x \sqrt{2\pi \mathcal{g}_x(\omega')}} \sum x_t \cos \omega' t
\]

(A2.20)

which is clearly normally distributed as a \( N(0,1) \) variable under the null hypothesis \( \beta = 0 \).

(11) If we modify the problem slightly so that we assume \( \sigma_x^2 \) unknown but \( |\psi(1\omega')|^2 \) known, then the Wilks likelihood ratio criterion for testing the composite hypothesis \( \beta = 0 \) may be written as

\[
\lambda_2 = \left[ 1 - \sum (x_t \cos \omega' t)^2 \right]^{N/2}
\]

and \( \lambda_2 \) is a \( \chi^2 \) variable with \( N \) degrees of freedom.
where \( N = \frac{\pi}{2N} \sum_{j=0}^{N-1} \frac{\psi(e^{i\omega})}{|\psi(e^{i\omega})|^2} \) is clearly \( \sigma_X^2 \).

In this case, therefore, we would take

\[
\phi_N' = \frac{\cos \omega t}{\sqrt{2\pi N} g_x(\omega)} e^{i\omega t}\sqrt{2\pi N} g_x(\omega) \sigma_X^2.
\]

This is shown in (A2.21) and (A2.24) as test criterion.

In the case where \( \sigma_X^2 \) and \( |\psi(\omega)|^2 \) are known it is a relatively trivial matter to show that the power function of a test of significance based on \( \phi_N' \) against departures of size \( |\beta| \) is given exactly by

\[
P = 1 - F\left(c = \frac{\frac{N}{2\sqrt{2\pi N} g_x(\omega)}}{2\sigma_X^2} \right)
\]

where \( F \) is the cumulative normal distribution and \( c \) is such that

\[
P \left( \phi_N' > c \mid \beta = 0 \right)
\]

is some prescribed value \( \alpha \). The interpretation of this quantity in terms of signal-to-noise ratios has already been made in Section 8.

**Extension to General Circular Processes**

A general representation for circular processes has been given recently by E. J. Hannan [26] who shows that a spectral theory can be developed which is a finite dimensional analogue of the spectral decomposition of unitary operators in Hilbert space which is fundamental in the theory of infinite processes.

Hannan shows that a general representation for circular processes may be made in the form

\[
z_t = \sum_{j=0}^{N-1} e^{i\omega_j t} \zeta_j
\]

and

\[
N = \sum_{j=0}^{N-1} e^{i\omega_j t} \zeta_j
\]

where \( \omega_j \) are the angular frequencies and \( \zeta_j \) are independent random variables with unit variance.
where the $e^{i\omega_j}$ ($\omega_j = \frac{2\pi j}{N}, \ j = 0, 1, 2, \ldots, N - 1$) are the latent roots of the matrix of shift operators which transform $z_t$ into $z_{t+1}$ and the $\epsilon_j$ are as in (A2.2). He also shows that the spectrum of the process consists of point masses $f_j$ concentrated at the frequencies $\omega_j$.

Starting from (A2.23), it may be shown that the basic equations (A2.8) and (A2.10) hold for general processes at frequencies $\omega_j = \frac{2\pi j}{N}$. The argument then proceeds exactly as before.

**Non-circular processes.**

The arguments proceed formally as for circular processes but in most instances the formulae are only true approximately. Thus if we define

$$c_s(z) = \frac{1}{N} \sum_{t=1}^{N-s} z_t z_{t+s}$$

then the analogous result to (A2.5) is

$$f_N(z)(\omega) = \frac{1}{\pi} \sum_{\omega = -N+1}^{N-1} c_s e^{i\omega s}$$  \hspace{1cm} (A2.24)

which is true for all $\omega$.

If we now define

$$F_N(z)(\omega) = \int_{-\pi}^{\omega} f_N(x)(y) \, dy$$  \hspace{1cm} (A2.25)

then the corresponding relationship to (A2.7) is

$$c_s(z) = \frac{1}{2} \int_{-\pi}^{\pi} dF_N(\omega) e^{-i\omega s}$$  \hspace{1cm} (A2.26)

However, relationship (A2.10) connecting the sample periodogram of the $\{z_t\}$ and $\{e_t\}$ process is no longer an identity and holds only if the end effects due to the finite length of the series are neglected. From then onwards, all the results which hold for circular processes are approximately true for large $N$ for non-circular processes.
General Harmonic Regression

If we now assume a regression model of the form

\[ x_t = \sum_{j=1}^{m} \gamma_j \cos \omega_j t + z_t \quad (t = 1, 2, \ldots, N) \tag{A2.27} \]

where \( m < N \) and the \( \omega_j = \frac{2\pi j}{N} \), then it is easily shown that

\[ c_0(\epsilon) = \frac{\pi}{2N} \sum_{j=0}^{N-1} \frac{f_N(x)(\omega_j)}{|\psi(e^{i\omega_j})|^2} - \frac{2}{N} \sum_{j=1}^{m} \gamma_j \frac{\sum_{t=1}^{N} x_t \cos \omega_j t}{|\psi(e^{i\omega_j})|^2} + \frac{1}{2} \sum_{j=1}^{m} \frac{\gamma_j^2}{|\psi(e^{i\omega_j})|^2} \]

\[ \tag{A2.28} \]

It then follows that the best linear unbiased estimate of \( \gamma_j \) is

\[ \hat{\gamma}_j = \frac{\pi}{N} \sum_{t=1}^{N} x_t \cos \omega_j t \]

and

\[ \text{cov}(\hat{\gamma}_j, \hat{\gamma}_k) = \frac{2 \sigma_x^2 g_x(\omega_j)}{N} \]

\[ \tag{A2.29} \]

These results may be extended very easily to other forms of regression, in particular to multivariate regressions which are important in the design of experiments in response surface situations. These will be considered elsewhere.

ACKNOWLEDGEMENTS

We wish to thank Professor George E. P. Box for his stimulating discussions which motivated this study and Professor John W. Tukey for his helpful suggestions and continued interest in the progress of the work. Our thanks are also due to Mrs. Ann Hillier of Stanford University for invaluable assistance in the computations.
REFERENCES


[18] G. M. Jenkins, Tests of hypotheses in the linear autoregressive model
I. Biometrika, Vol. 41, parts 3 and 4 (1954), pp. 405-419


